Coherent Optical Detection
with an Emphasis on Electronics

A Thesis
Submitted for the Degree
of
Doctor of Philosophy in Electrical and Electronic Engineering
in the
University of Canterbury
by
MARK KENNETH RAYMOND WALKER BE(Hons), MIEEE

University of Canterbury
1987
In the development of Optical Information Transmission Technology, considerable investigation of the loss mechanisms, in optic fibres, has been made. Subsequently drastic reductions in transmission path loss have been achieved. Somewhat less effort has been spent on methods of improving optical detection efficiencies.

By examining the signal-to-noise expressions associated with optical detection schemes, it is clear that by using coherent detection, output signal to noise ratios could be increased by about 3 dB over intensity detection systems. Moreover detection bandwidths could be increased using coherent detection methods.

Coherent detection permits the detection of optimum modulation schemes. By exploiting the optical properties of some materials, under the influence of external fields, FSK and PSK modulation can be introduced into a light beam. Using a local oscillator laser, which is added into the beam incident upon the detector, a mixer product, at the detector output, may be generated; either at an intermediate frequency or at the base-band.

To recover the information, the mixer product must be amplified. Amplifiers for such applications must produce constant gain over very wide bandwidths (over 8 decades). Moreover if they are to amplify digital signals they must be of linear phase. Such design requirements are difficult to meet, however by applying methods of equalisation, the designer may readily control the gain and phase of such amplifiers over the range of interest.

Concepts of broadband equalisation are verified by equalising a number of loads with frequency dependent power gains.
ACKNOWLEDGEMENTS

The author would like to express his gratitude to the New Zealand Post Office for their financial assistance and encouragement during the course of research and write up. He also wishes to thank members of the University technical staff who assisted in the construction of some equipment and the procurement of test instruments used.

The author also extends special thanks to Dr H.R. Siresena for many fruitful discussions on the theories of curve fitting and equalisation and to Dr I.D. Coope for his advice on the use and limitations of minimization routines.

Finally the author wishes to thank his project supervisor Dr P.T. Gough for his encouragement and interest in the familiar and unfamiliar fields of research.
Science is a very human form of knowledge. We are always at the brink of the known, we always feel forward for what is to be hoped. Every judgement in science stands on the edge of error, and is personal. Science is a tribute to what we can know although we are fallible.

- Jacob Bronowski (1908-1974) "The Ascent of Man", p.374
To my beloved wife Rusni

and my parents Lois and Alan.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter or Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td></td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>... References</td>
<td></td>
</tr>
<tr>
<td>2. OPTICAL COMMUNICATIONS, PAST, PRESENT AND FUTURE</td>
<td></td>
<td>2.1</td>
</tr>
<tr>
<td>2.1 Information Transmission Using Light - An Old Idea</td>
<td></td>
<td>2.3</td>
</tr>
<tr>
<td>2.2 Optical Information Transmission Today</td>
<td></td>
<td>2.6</td>
</tr>
<tr>
<td>2.3 The Structure of Optical Communications Systems and the Fibre Optic Equation</td>
<td></td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>... References</td>
<td>2.19</td>
</tr>
<tr>
<td>3. THE GENERATION OF COHERENT LIGHT AND ITS CHARACTERISTICS</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>3.1 The Generation of Coherent Light</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td>3.1.1 Spontaneous Light Emission From Semiconductors and Gases</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td>3.1.2 Stimulated Emission</td>
<td></td>
<td>3.10</td>
</tr>
<tr>
<td>3.1.3 Helium-Neon and Gallium-Arsenide Lasers</td>
<td></td>
<td>3.14</td>
</tr>
<tr>
<td>3.1.4 Cavity Modes - Frequency and Amplitude Stability</td>
<td></td>
<td>3.19</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS (Cont'd)

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>The Quantum Theory of Light</td>
<td>3.27</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Light and the Simple Harmonic Oscillator</td>
<td>3.28</td>
</tr>
<tr>
<td>3.2.2</td>
<td>The Principle of Complementarity - Photon Number - Phase Uncertainty</td>
<td>3.38</td>
</tr>
<tr>
<td>3.3</td>
<td>Propagation of Sound in Solids</td>
<td>3.43</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>3.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.</th>
<th>MODULATION OF LIGHT</th>
<th>4.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Optical Modulator Materials</td>
<td>4.3</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Electro-Optic Properties of Materials</td>
<td>4.3</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Magneto-Optic Modulator Materials</td>
<td>4.7</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Acousto-Optic Modulator Materials</td>
<td>4.10</td>
</tr>
<tr>
<td>4.2</td>
<td>Extra-Cavity Laser Modulation</td>
<td>4.21</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Electro-Optic and Magneto-Optic Modulation</td>
<td>4.22</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Acousto-Optic Modulation</td>
<td>4.30</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS (Cont'd)

4.3 Intra-Cavity Laser Modulation
- 4.3.1 Piezo-Electric Modulation
- 4.3.2 Electro-Optic, Magneto-Optic and Acousto-Optic Modulation
- 4.3.3 Energy Level Modulation
- 4.3.4 Power Supply Modulation

5. DETECTION OF MODULATED LIGHT

5.1 Intensity Detection
- 5.1.1 Detection Mechanisms
- 5.1.2 Quantum Efficiency
- 5.1.3 Optical Frequency Response of Optical Detectors
- 5.1.4 Electrical Frequency Response of Optical Detectors

5.2 Noise Processes in Dark Optical Detectors
- 5.2.1 Dark Current
- 5.2.2 Shot Noise
- 5.2.3 Thermal Noise

5.3 Light Dependent Noise Processes
- 5.3.1 Fluctuations in Light Intensity - Optical Granular Noise

References
TABLE OF CONTENTS (Cont'd)

5.3.2 Optical Granular Noise of an Ideal Photo-Receiver and its Association with Shot Noise 5.52

5.4 Detection of Amplitude Modulated and Amplitude-Shift-Keyed Signals 5.55
  5.4.1 Phases of Carrier Sidebands 5.56
  5.4.2 Envelope Detection of Amplitude Modulation 5.59
  5.4.3 The Photo-Detector as a Mixer 5.64

5.5 Detection of Modified-Amplitude-Modulated, Frequency-Shifted and Phase-Shifted Signals 5.72
  5.5.1 Detection of Single Sideband Suppressed Carrier Signals 5.74
  5.5.2 Detection of Frequency-Shift-Keyed Signals 5.75
  5.5.3 Detection of Phase-Shift-Keyed Signals 5.81

5.6 Improving the Signal-to-Noise Ratio of Detection 5.85
  5.6.1 Correlative Reduction of Detection Noise 5.86
  5.6.2 Reducing the Quantum Noise Limit - The Squeezed State 5.93

5.7 A Comparison of Homodyne and Heterodyne Detection of Optical Signals 5.100
6. ELECTRONIC DEVICES FOR MODULATION AND DETECTION

6.1 The General Requirements of Electronic Devices in Optical Applications 6.5

6.2 Detector Design
   6.2.1 Photo-Receiver Physical Design 6.11
   6.2.2 Stripline Design for the CD10 Silicon Photo-Diode 6.31

6.3 Mounting and Adjustment of the Acousto-Optic Modulator 6.44

6.4 A Temporary Photo-Detector
   6.4.1 Producing Mixer Products 6.51
   6.4.2 Frequency Sweep of Temporary Photo-Receiver 6.60

6.5 Increasing the Mixing Bandwidth by Reducing the Effects of Detection 6.64
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.6</td>
<td>Amplifier Design</td>
<td>6.69</td>
</tr>
<tr>
<td>6.6.1</td>
<td>Narrow-Band Amplifier Design using the Smith Chart</td>
<td>6.70</td>
</tr>
<tr>
<td>6.6.2</td>
<td>Broadband Amplifier Design using the Smith Chart</td>
<td>6.93</td>
</tr>
<tr>
<td>6.6.3</td>
<td>Broadband Hybrid Amplifier Design</td>
<td>6.111</td>
</tr>
<tr>
<td>6.7</td>
<td>Mixer Design</td>
<td>6.123</td>
</tr>
<tr>
<td>6.7.1</td>
<td>Differential Mixer Theory</td>
<td>6.125</td>
</tr>
<tr>
<td>6.7.2</td>
<td>Informal Differential Mixer Tests</td>
<td>6.127</td>
</tr>
<tr>
<td>6.7.3</td>
<td>A Current Stabilized Differential Mixer</td>
<td>6.129</td>
</tr>
<tr>
<td>6.7.4</td>
<td>A Full Multiplier using the Differential Mixer</td>
<td>6.140</td>
</tr>
<tr>
<td>6.8</td>
<td>Design of an Integrator for Long Term Correlation Applications</td>
<td>6.143</td>
</tr>
<tr>
<td>6.8.1</td>
<td>Digital Integrators versus Analogue Integrators</td>
<td>6.145</td>
</tr>
<tr>
<td>6.8.2</td>
<td>Stability of Analogue Integrators over Long Time Periods</td>
<td>6.154</td>
</tr>
<tr>
<td>6.8.4</td>
<td>Switching Techniques and the Two-Phase Clock</td>
<td>6.160</td>
</tr>
<tr>
<td>6.8.5</td>
<td>Adapting the Two Phase Clock to the Switched Capacitor Integrator</td>
<td>6.182</td>
</tr>
<tr>
<td>6.8.6</td>
<td>Switched Capacitor Integrator Construction</td>
<td>6.167</td>
</tr>
<tr>
<td>6.8.7</td>
<td>Testing the Switched Capacitor Integrator</td>
<td>6.170</td>
</tr>
<tr>
<td>6.8.8</td>
<td>Integrators Suitable for Data Application</td>
<td>6.171</td>
</tr>
</tbody>
</table>
7. BROADBAND CIRCUIT DESIGN AND THE TECHNIQUES OF EQUALIZATION 7.1

7.1 Curve Fitting 7.5

7.1.5 Rational Function Curve Fitting using Techniques of Non-Linear Minimization Subject to Non-Linear Constraints - The Summation Method 7.10

7.2 The Theory of Equalization 7.31

7.2.1 Background Theory of Youla Equalization 7.34
7.2.2 Youla Equalization 7.40
7.2.3 The Linear Diophantine Equation 7.54
7.2.4 Load Equalization using Linear Diophantine Equations 7.59
7.2.5 The Equalizer Back-Impedance 7.62
7.2.6 Imaginary Axis Zeros of Transmission 7.65
7.2.7 The Trivial Solution and the Bounding Polynomial 7.76
7.2.8 New Methods of Synthetic Division and Solutions to the Linear Diophantine Equation 7.80
TABLE OF CONTENTS (Cont'd)

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2.9</td>
<td>DEQDIO - A Program for Load Equalization</td>
<td>7.103</td>
</tr>
<tr>
<td>7.2.10</td>
<td>Realizing the Equalizer Back-Impedance</td>
<td>7.112</td>
</tr>
<tr>
<td>7.3</td>
<td>Load Synthesis</td>
<td>7.114</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Cauer Load Synthesis</td>
<td>7.116</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Brune Synthesis</td>
<td>7.122</td>
</tr>
<tr>
<td>7.3.3</td>
<td>Darlington Synthesis</td>
<td>7.133</td>
</tr>
<tr>
<td>7.3.4</td>
<td>DLREAL - A Load Synthesis Program</td>
<td>7.155</td>
</tr>
</tbody>
</table>

8. USING THE THREE PHASES OF EQUALIZATION 8.1

8.1 Curve Fit Examples 8.2

8.2 Equalization Examples 8.12

8.3 Load Impedances Realization Examples 8.18

9. FUTURE PROPOSALS FOR OPTICAL AND ELECTRONIC DESIGN AND EXPERIMENTATION

9.1 The Differential Optical Detector 9.3
TABLE OF CONTENTS (Cont'd)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.2 Reduced Carrier Optical Transmission Systems</td>
<td>9.9</td>
</tr>
<tr>
<td>9.3 Amplifier Design using Equalization and Load Realization Theory</td>
<td>9.13</td>
</tr>
<tr>
<td>9.4 Extension of Equalization Theory to Non-Positive Real Load Impedances</td>
<td>9.20</td>
</tr>
<tr>
<td>... References</td>
<td>9.24</td>
</tr>
<tr>
<td>10. CONCLUSION</td>
<td>10.1</td>
</tr>
<tr>
<td>APPENDICES</td>
<td></td>
</tr>
</tbody>
</table>
1.1 INTRODUCTION

During the early period of 1984 New Zealand's first fibre optic telecommunications link was commissioned between the switching centres at Wellington Central and Lower Hutt central. Its commissioning began a new period in the transmission of voice and data traffic in New Zealand. In three short years the New Zealand Post Office (NZPO) has installed over 1000 km of optic fibre. Each new system installed has rendered the technology of the last one obsolete as rapid advances in the technology continue.

This year (1987) the first 560 MB/s optic fibre transmission system is being installed, giving an increase in information capacity of 3.9 times over the existing 144 Mbps systems. Even with this speed increase the NZPO cannot satisfy the insatiable demand for long distance data and voice communications circuits.

The great demand for communications circuits in New Zealand is due, in part to a unique combination of Business and Social factors. Within designated toll free areas, NZPO telephone subscribers can make telephone calls free of charge. Each subscriber pays the NZPO a fixed telephone rental charge. Once paid he may make as many calls as desired within the free calling area without incurring extra charge. Charges in addition to the rental are made only if the subscriber makes calls between free calling areas using the toll network. For these reasons New Zealanders are the world's greatest telephone users per head of population and so New Zealand's telephone calling patterns are unique.
The business sector of New Zealand is also in a unique position and is, by far, the greatest user of long distance communications circuits. With the substantial decrease in the costs of computing technology, many firms have installed computing facilities for data processing applications. While it has fallen in price, computer technology is still relatively expensive in New Zealand. In most cases it is cheaper for a company to install a central computing facility and lease direct-circuit connections for terminals located vast distances away than to purchase additional computers for their other business premises.

In many other countries the cost of computer technology is significantly lower than the cost of direct-circuit connections and the relative demand for direct connection circuits is lower than in New Zealand.

The trends of direct circuit demand in New Zealand are quite different from the demand encountered overseas. With the increasing use of electronic data processing, the demand for leased direct connections continues to increase.

In order to meet the demand, the NZPO has continued to exploit new technology resources as they are developed.

The NZPO, like any service organisation, seeks to provide its services at the lowest possible cost to the consumer, yet it must acquire sufficient revenue to pay its employees and maintain a programme of development and installation of reducing the cost of procurement and installation.

In fibre optic communications systems, the regenerators placed at intervals along the transmission path, to regenerate a fresh signal,
represent a major cost of installation [1]. At each regenerator a photo­
receiver, digital detector and laser are required to receive, demodulate
and retransmit incoming information [1],[2],[3],[4],[5],[6],[7],[8],[9].
If the spacing between regenerators can be increased, the costs of
installing the system can be reduced, producing a corresponding reduction
in service costs to the consumer. Considerable research has already been
undertaken in the manufacture of optic fibres with very low loss. Large
increases in repeater spacings have occurred as a result [1],[2],[3],[4],
[5],[6],[7],[8],[9].

Somewhat less attention has been paid to improving the techniques of
modulation and detection. At present amplitude shift keying (ASK) is the
main form of modulation used commercially [1],[10],[11]. Yet it is well
known that amplitude modulation techniques do not give optimum modulation
and detection performance [10],[11]. By selecting other modulation forms,
improvement in detection efficiency can be produced and regenerator spacing
increased. Further improvements could also be made by reducing the noise
produced by the variety of noise sources present in optical communications
systems [10],[11],[12].

In the following chapters techniques of improving the performance of
optical communications systems are discussed together with the difficulties
encountered in trying to build practical equipment for the purpose. As a
result, new techniques of electronic equipment design have been developed
which should make the theoretical design of high quality signal processing
equipment possible.

In chapter 2, the past history of optical communications is reviewed
briefly. From this history it is clear that optical communication of
information is not a new idea [14]. The subsequent discussion then 
focusses on the present technologies of optical communication. The broad 
concepts of optical communication system design are introduced. These 
techniques are based on the allocation of the power budget required to 
achieve a specific receive signal-to-noise ratio (SNR) [1],[10],[11],[15]. 
The idealized power budget is summarised in the Fibre Optic Communications 
equation from which one can deduce the required regenerator spacings to 
achieve the SNR desired [1],[10],[11],[15].

The fibre optic equation also illustrated the areas in which 
improvement to the system can be made and their overall effect on the SNR 
or regenerator spacings at a fixed SNR.

The chapter ends in a discussion of the possible future development in 
fibre optic communications and draws parallel paths in the development of 
optical communications with the development of radio technology since the 
turn of the century [16].

Chapter 3 introduces the behaviour of light sources and the Laser in 
it's many forms. The characteristics of laser light are discussed in 
detail. The quantum theory of light is then introduced. A superficial 
understanding of the quantum theory of light is needed to explain the 
existance of quantum noise sources and the quantum limit of optical 
detection [1],[2],[17],[18],[19],[20],[21],[22],[23]. Recent advances in 
the quantum theory of light indicate a way in which quantum noise may be 
reduced below the values traditionally expected [23]. Using the theory of 
this chapter the theory of the squeezed state can be introduced in chapter 
5.
Techniques of modulating light are discussed in chapter 4. They fall into two categories, modulation of light inside the cavity of a laser [24] and modulation of light outside it. These techniques all involve the interaction of an externally applied field with the light inside a nonlinear medium [24],[25],[26].

Acousto optic (AO) techniques of light modulation [25],[26],[27],[28],[29],[30],[31],[32],[33],[34],[35],[36],[37] are discussed in some detail as a number of experiments were conducted using an AO modulator. The characteristics of beam deflection are analysed in detail, as these effect the bandwidth over which an AO modulator can be used.

In chapter 5 the mechanisms of detection are discussed. Starting with a detailed examination of the light detection process, the optical and electrical frequency responses are characterized [1],[17]. An integral equation of the detection process is thus produced. Thermally based noise processes [10],[11],[12] in the detector are discussed and quantized. The light dependent sources [1],[17],[18],[38],[39] of noise are then examined and the ways in which they contribute to the overall detection noise discussed.

Using the detector characterizations discussed, the detection of AM signals is reviewed and the process of mixing involved, emphasised. Using the mixing processes observed in AM detection, the general mixing characteristics of a photo detector are discussed.

Using optical mixing techniques [38],[39],[40],[41],[42],[43],[44],[45],[46],[47],[48],[49],[50],[51],[52],[53] of detecting modified forms of AM, such as single sideband suppressed carrier (SSB), and angle modulated
signals, are discussed. In each case expressions for the SNR's of detection are derived for each form of modulation.

Homodyne and Hetrodyne techniques of detection are compared and the significance of quantum mechanical phase uncertainty [19] discussed.

Techniques of improving the SNR of angle modulation detection are discussed. Emphasis is given to correlative methods of SNR reduction [54].

Finally the effects of beam misalignment on the conversion efficiency of a photo detector mixer is characterised [52].

To exploit the techniques of improving the SNR of photo detection, some electronic equipment is required. Chapter 6 reviews the techniques used to design suitable amplifiers, mixers and integrators. Finally the design of a photo receiver, for use with a 50 Ω standard impedance system, is discussed. The difficulties encountered with it are also revealed. While reasons for the poor photo receiver performance were sought, a temporary photo receiver was built for use in some early experiments.

Using the temporary photo detector a number of mixing experiments were conducted. The variation of deflection angle with modulation frequency of light emerging from the AO modulator enabled mixing to occur over a very small frequency range, without adjustment of the Mach Zehnder interferometer used. Methods were sought to reduce the frequency dependence of mixing and increase the bandwidth over which it could be performed.
Details and difficulties encountered in the design and construction of amplifiers, mixers and integrators are discussed. In order to improve the performance of these devices, methods of improving input, output and interstage coupling were examined, leading to the development of new techniques of equalization based on the theory of linear Diophantine equations, which is discussed in section 7.2.

Finally the construction and design details of a broadband, Hybrid amplifier based, gain module are discussed.

The development of a new theory of equalization was based on the Youla method of equalization to produce a constant power transfer through the amplifier with a linear phase response. To equalize the input and output impedance of an amplifier mathematical description of the amplifier input and output impedances are required. Generally the manufacturer presents the scattering parameters of the transistor at a number of spot frequencies, within its region of useful application. These are intended for use when designing a narrow band amplifier where a general description of the transfer frequency characteristics is not required. To obtain a more general description, a method of fitting curves, either to the manufacturers data, or to additional measurement data made by the designer, is developed. Application of existing methods to the characterization of measurement data was unsuccessful. Using techniques of non linear minimization, methods of fitting a rational function to measurements of impedance or admittance using non linear least squares regression, were investigated. Initially a product form of the general rational function was fitted to the measurement data. The results were satisfactory, however to apply equalization theory, the impedance characterization must be positive real (PR). It was not possible to enforce the PR constraints in
the product form of the rational function. By expressing the rational function as a sum of partial fractions, a PR rational function representation of the measurement data could be produced. Using the technique, a PR rational function curve fitting program was developed. Its structure and use is discussed in section 7.1.

Once a PR rational representation is found, the power transfer, through the amplifier must be modified, by a preconnected network, to give the desired response. The process is called equalization.

If the load is complicated, Youla's technique of equalization is difficult to apply, using the series expansion technique proposed in his paper. Using some of his key observations a new technique based on the theory of Linear diophantine equations (LDE)'s was developed. The salient points of Youla's theory are presented followed by the new LDE equalization technique. This technique is readily incorporated into a computer program. The structure and techniques used in computer based equalization are discussed in section 7.2.

The Youla and LDE techniques of equalization produce rational functions for the back impedance of the equalizer. In order to build the equalizer, the rational functions must be realized as reactive networks terminated in a resistor. Techniques of realizing rational functions as networks of real reactive and resistive components are discussed in section 7.1.

Using the three phases of equalization discussed above, a number of examples were tried. In chapter 8 these examples are given. Each equalizer and load combination is then simulated to verify that the
equalizer found, indeed, equalizes the load with the desired amplitude and phase response.

Much of the research discussed in chapter 2 to 8 has raised a number of problems and design techniques which have either not been investigated, or are still at an early stage of development. Areas in which development has been started, are introduced in chapter 9, which could form the basis for further research in these areas.

Techniques of reducing the effects of laser frequency instability are discussed, as an alternative to the use of expensive activity stablized lasers, to improve the SNR performance of optical communications systems.

The techniques of equalization were introduced to improve the coupling between source load and stages of an amplifier. This approach does not realize the full potential of equalization as an amplifier design tool. In section 9.7, new techniques of amplifier design using controlled mismatch are introduced, in which the theory of equalization is used to produce an optimum gain characteristic with any desired response. Unfortunately the impedances in amplifier design are sometimes non-PR, in which case the theory of equalization cannot be directly applied. Some theoretical considerations of non PR equalization are introduced and extensions to the computer based equalization procedure discussed.

In chapter 10 the research covered in chapters 2 to 9 is reviewed and a series of conclusions drawn, relating to the practical and commercial applications, of this research to optical systems now in exisstance and those it is proposed to build.
1.10

REFERENCES - CHAPTER 1

"Optoelectronics - An Introduction."

[2] FRODSHAM, D.G. and BAKER, D.J.

"1.55 \mu m Optical Transmission Experiments at 2 Gbits/s Using 51.5 km Dispersion Free Fibre."

"1 Gbit/s Transmission Experiment Over 101 km of Single-Mode Fibre Using a 1.55 \mu m Ridge Guide C3 Laser."

[5] OWEN, B.
"Pin - GaAs FET Optical Receiver with a Wide Dynamic Range."

[6] GOELL, J.E.
"Input Amplifier for Optical PCM Receivers."
[7] HULLETT, J.L. and MUOI, T.V.
"A Feedback Receive Amplifier for Optical Transmission Systems."

[8] CHERRY, E.M.
"Impedance Mismatch in Wide-Band Transistor Amplifier Design."
pp 131 - 132.

"The Design of Wide-Band Transistor Feedback Amplifiers.

[10] HAYKEN, S.
"Communications Systems"

"Introduction to Communication Systems."

[12] RICE, S.O.
"Mathematical Analysis of Random Noise."
Bell System Technical Journal, Vol. 23, July 1944, pp 282 - 333,

[13] HANBURY-BROWN, R.
"The Intensity Interferometer, Its Application to Astronomy."
Taylor and Francis Ltd. 1974.
1.12

[14] BRONOWSKI J.
   "The Ascent of Man"
   BBC - Angus and Robertson Pty Ltd (Aust.), 1973, 1976
   pp 179-181, 224-228, 247, 248-256

[15] GANDHI.
   "Microwave Engineering and Applications."

[16] DEMAW, D.
   "The Radio Amateurs Handbook."
   American Radio Relay League, Newington, C.T., USA 06111.
   pp 1.1 - 1.6.

[17] KINGSTON, R.H.
   "Detection of Optical and Infrared Radiation."

[18] CAVES, C.M.
   "Quantum Mechanical Noise in an Interferometer."

   "Limits on Electromagnetic Amplification Due to Complementarity."
   First International Conference on Quantum Electronics, High View, N.Y. 1939.

[20] GLAUBER, R.J.
   "Coherent and Incoherent States of the Radiation Field."
[21] GUDARSHAN, E.C.G.
   "Equivalence of Semiclassical and Quantum Mechanical Description
   of Statistical Light Beams."
   Physical Review Letters, Vol. 10, No. 7, 1 April 1963,
   pp 277 - 279.

[22] LOUDON, R.
   "The Quantum Theory of Light."

[23] WALLS, D.F.
   "Squeezed States of Light."

[24] HARVEY, A.F.
   "Coherent Light."

[25] HARVEY, A.F.
   "Coherent Light."

[26] MILEK, J.T. and NEUBERGER, M.
   "Linear Electro Optic Modulator Materials."
   IFI/Plenum, 1972.

[27] CHARLES, C. et TREFFE, F.
   "Application de l'Interaction Acousto-Optique Dans Les Solides a
   L'Analyse Ultra-Rapide D'un Spectre Lumineux."
[28] ALBRIGHT, R.J.
"Relationship of Doppler Ultrasonic Scattered Signal Characteristics to Flow and Beam Parameters."

[29] RIGROD, W.W. and KAMINOW, I.P.
"Wideband Microwave Light Modulation."

"Unified Approach to Ultrasonic Light Diffraction."

[31] KISHIOKA, K, and HASHIMOTO, M.
"Coupling of Guided Light Through a Periodic Multilayer."

"Interaction of Light and Microwave Sound."

[33] COHEN, M.G. and GORDON, E.I.
"Acoustic Scattering of Light in a Fabry-Perot Resonator."
[34] DIXON, R.W. and GORDON, E.I.

"Acoustic Light Modulators Using Optical Heterodyne Mixing."

[35] HANCE, H.V. and PARKS, J.K.


"Monolithic Surface Acoustic Wave (SAW) Charge Transfer Device and Its Applications."

[37] DAS, P., SCHUMER, D. and ESTRADA-VAZQUEZ.

Optical Communications Using Surface Acoustic Waves."

[38] DEGNAN, J.J. and KLEIN, B.J.

"Optical Antenna Gain."

[39] DEGNAN, J.J. and KLEIN, B.J.

"Optical Antenna Gain."

[40] OLIVER, B.M.

"Signal to Noise Ratios on Photo-Electric Mixing."
[41] TAKENAKA, T., TANAKA, K. and FUKUMITSU, O.
"Signal-to-Noise in Optical Heterodyne Detection for Gaussian Fields."

[42] NICHOLSON, G. and CAMPBELL, J.C.
"Performance Estimates for Heterodyne and Homodyne Optical Fibre Communications Systems."

[43] TARG, R.
"Optical Heterodyne Detection of Microwave Modulated Light."

[44] YOSHIMURA, T., HARA, K. and WAKABAYASHI, N.
"Modified-Triggered Photocounting and its Application to Optical Superhetrodyne Spectrum Analyses."

[45] CHURNSIDE, J.H. and McINTYRE, C.M.

[46] CHURNSIDE, J.H. and McINTYRE, C.M.
"Signal Current Probability Distribution for Optical Heterodyne Receivers in the Turbulent Atmosphere. 2 Experiment."
[47] TEICH, M.C. and YEN, R.Y.
"Three-Frequency Nonlinear Heterodyne Detection.
1. C.W. Radar and Analog Communications."

[48] TEICH, M.C. and YEN, R.Y.
"Three-Frequency Nonlinear Heterodyne Detection.
2. Digital Communications and Pulsed Radar."

[49] HEALEY, P.
"Fading in Heterodyne OTDR."

[50] PETTS, C.R., McCALL, M.W. and LAYCOCK, L.C.
"Two-Channel Heterodyne-Type Transmission-Type Experiment."

[51] FRIED, D.L.
"Optical Heterodyne Detection of an Atmospherically Distorted Signal Wave Front."

[52] COHEN, S.C.
"Heterodyne Detection: Phase Front Alignment, Beam Spot Size, and Detector Uniformity."
[53] FINK, D. and VODOPIA, S.N.


[54] HANBURY-BROWN, R.

Despite the recent interest in guided optics as a means of information transmission, the concept is itself a very old one, unlike the electrical and electromagnetic means which have been devised only over the last 200 years.

Recent interest is due to advances in optical technology and demand for greater capacity of communications circuits. Presently, optical communications systems offer mankind information transmission capacity that is practically limitless [1]. It is not possible to contemplate demand for transmission capacity that would exceed or even approach the theoretical bandwidth offered by light systems. However as man moves further into the information revolution, his need for transmission systems of growing capacity will increase.

In this chapter it is proposed to discuss the evolution of optical communications from the earliest experiments through the present technology and looking forward into its future. In section 2.1 some of the early applications of optical techniques are reviewed. They offer insights into the reasons for the way in which optical technology has developed.

In section 2.2 the advantages offered by optical transmission systems over electrical or electromagnetic systems are discussed in detail.

The general structure of optical communication systems is introduced in section 2.3. Using this structure the equation governing overall design of optical systems is discussed [2],[3],[4],[5],[6]. From the Optical
Fibre Communications equation one may decide where a system may be improved to reduce installation and operation costs.

Finally, in section 2.4 the future of optical communications is discussed. Some of the answers in the evolution of optical communications will come from a study of the evolution of electrical and electromagnetic means of communications, in particular the evolution of radio from the earliest experiments conducted by Heinrich Hertz to the modern techniques used today.
2.1 INFORMATION TRANSMISSION USING LIGHT - AN OLD IDEA

Much of man's early information was conveyed using light for essentially two reasons:

1. Sunlight was available during the day, and at night light could be readily generated by burning some sort of fuel.

2. Man himself could receive the optical information due to his sense of sight, no elaborate means of detecting light was needed.

In Europe beacons atop hills were used to convey information across countries using semaphore. An observer could read the information transmitted by one semaphore tower and relay it to the next. With towers scattered throughout a region, information could be transmitted from one place to another in the region.

More direct use could be made of the sun by reflecting the sun to a distant observer intermittently using some sort of code, an observer could receive the information by deciding the light fluctuations.

With the birth of electric telegraph and later the telephone, the ancient optical techniques of information transmission was rendered obsolete. However the inventor of the telephone, Alexander Graham Bell, did produce an optical version of his telephone [7]. A thin reflective diaphragm was mounted and aligned to reflect a light beam onto a vacuum tube photodiode which was connected to a telephone receiver. Vibrations due to the voice of a speaker caused the diaphragm to deflect the light
away from the photo detector, producing a fluctuating photo-current at the photodiode which stimulated bells telephone receiver reproducing the sound.

Although not practical this was the first optical speech transmission system, it showed that light could be used to convey information directly, and that light could be used with known electrical technology, without the use of human eyes.

The first commercial use of optical sound transmission was in talking movies. The sound was recorded by altering an aperture between a lamp and one edge of the unexposed film, an amplitude modulated version of the sound was recorded onto the film [8].

The sound is recovered during the process of projection by passing the sound track between a light source and a photo detector as it passes, the light incident upon the detector is amplitude modulated by the instantaneous aperture of the sound track. The photo detector produces a fluctuating photo-current, which is amplified and fed into a speaker. Thus the sound is recovered [5].

Few further commercial applications of optical communication were made until the realization of Einsteins theoretical proposal of light amplification by stimulated emission of radiation (LASER), the Laser [6].

The invention of the laser increased the interest in optical communications, however commercial systems using optical means of transmission were not considered seriously until the development of the optical fibre [2],[4],[5],[6].
As techniques of manufacture of lasers and optical fibres improved the cost of optical communications systems fell until they became commercially attractive as a means of information transmission. With this, interest rose and research gathered momentum until optical communication systems became more commercially attractive than coaxial and some terrestrial microwave systems.
2.2 OPTICAL INFORMATION TRANSMISSION TODAY

The decade of the 1980's has been described as the time of the information revolution. Development of very large scale integration (VLSI) has made it possible to build large bases of information. These may be readily accessed by a wide variety of users at low cost.

The availability of information from such information bases is dependent on the structure and coverage of the communications network to which they are connected.

The demand for readily accessible information has shifted the emphasis of demand for telecommunications circuits from a low growth rate audio network, to a rapidly growing digital network.

The changing emphasis on demand has put pressure on existing internal and international communications cables and on the electromagnetic spectrum. Communication administrations or companies cannot cope with the increasing demand for data communication circuits.

Coupled with the increasing demand for digital communication circuits is the increasing costs of copper cables which have made installation of coaxial transmission systems unattractive. Moreover, copper cables are fragile, their information carrying capacity is limited and they are lossey. Typical regenerator or repeater spacings on coaxial transmission systems, are of the order of 3 Km. Support plant is also expensive. The cables must be kept dry, thus they are pressurized. A compressor is required for each cable section which must be hermetically sealed. Seals are also required at each regenerator or repeater. To ensure that action
is prompt in the event of failure, alarms must be installed to monitor the pressure and internal humidity of the cable.

The marriage of laser and optic fibre technologies have led to systems in which less complicated plant is required. Regenerator spacings have been dramatically increased over those required for equivalent coaxial systems. The major cost of optic fibres themselves is one of technology and manufacturing. As use of optic fibres grows, these costs will fall. The actual raw material costs are very low, as most optic fibres are made of silicon which is the most abundant element in the earth's crust. Optic fibre costs have already fallen sufficiently to make optical transmission systems much more attractive than coaxial cable systems.

Presently optical systems pose little threat to terrestrial microwave systems unless information capacity is an overriding consideration. However, optical fibre transmission systems have begun to threaten the traditional role of the communication satellite as a bearer of information between large centres or nations. It is becoming more commercially attractive to sink optical fibre cables beneath oceans, than to launch satellites.

In low information capacity applications, optic fibre systems have not gained favour. The principal reasons being those of power supply and the complexity of terminal equipment. In telephony, for example, the telephone power supply is derived from the switching centre via the cable pair. To use optic fibres in this application, major changes in audio switching policy must be made. Common power feeds or use of domestic power supplies are needed to energize the telephone. Moreover the telephone itself will need digitization and coding hardware, presently available at the switching
2.8

centre. The cost of each telephone will therefore increase and its reliability may fall significantly. Similar problems are encountered with other terminal equipment.

Despite the many advantages that optical transmission systems offer, they are still very crude. The technology itself is still in its infancy [1],[2],[4],[5],[6].

The mode of transmission used in all commercial optical systems is Amplitude Shift Keying (ASK), which is a sub-optimum form of modulation.

The light sources are either Lasers or light emitting diodes (LED's). They emit broad bandwidths of light. The amplitude of light at a given frequency is unstable [2],[4],[9],[10],[11],[12],[13].

Relative to the optical carrier frequency, information rates are very low indeed, typically less than $30 \times 10^{-6}\%$ of the maximum speed possible theoretically; at 2 bits/Hz of bandwidth.

For this reason, fibre optic communication development, is at the same point as the development of radio 80 years ago, when radio transmissions were made using spark transmitters, built using large coils, leyden jars and spark gaps. The information was transmitted using morse code.

The emissions were very broad-band yet the information rate was very low, typically $1 \times 10^{-3}\%$ of the centre frequency of the emission.

Today radio has advanced a long way. Maximum use of bandwidth is made. Carrier frequencies have better than $0.025\%$ frequency stability.
Matched filter detection systems, maximizing detection SNR's signal to noise ratios, are used [14],[15]. Progressively higher frequencies have been used and new methods of generating radiation were found.

By reviewing the past 80 years of radio development, since the first experiments conducted by Heinrich Hertz, a view of the future of development of optical transmission systems can be gained.
2.3 THE STRUCTURE OF OPTICAL COMMUNICATIONS SYSTEMS AND THE FIBRE OPTIC EQUATION

The general structure of an optic fibre communications system is similar to any other electromagnetic communication system [15],[16],[17]. The major differences occur in the actual transducers used. Electromagnetic communications systems generally consist of 6 elements. Three in the transmitter and three in the receiver.

1. Carrier frequency generator.
Transmitter 2. Modulator.
3. Coupling the transmitter to the transmission medium.
4. The transmission medium itself.
Receiver 5. Coupling from the transmission medium to the receiver
6. Demodulator

The differences between microwave systems and fibre optic systems are found in the ways in which each element above is constructed, or behaves. In table 2.1 a comparison is made between the elements of a radio system and a fibre optic system.
Table 2.1 A Comparison of the Elements of a Typical Radio System and an Optical Transmission System

<table>
<thead>
<tr>
<th>System Element</th>
<th>Radio</th>
<th>Optical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier Generator</td>
<td>Quartz Crystal</td>
<td>Laser or LED</td>
</tr>
<tr>
<td></td>
<td>Oscillator and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amplifier Chain</td>
<td></td>
</tr>
<tr>
<td>Modulator</td>
<td>Electronic Mixer,</td>
<td>Cavity photon or</td>
</tr>
<tr>
<td></td>
<td>Multiplier or</td>
<td>power supply</td>
</tr>
<tr>
<td></td>
<td>Modulator</td>
<td></td>
</tr>
<tr>
<td>Coupling of transmitter to</td>
<td>Passive or Active</td>
<td>Lenses, interferometers,</td>
</tr>
<tr>
<td>medium</td>
<td>matching network,</td>
<td>Optical filters and</td>
</tr>
<tr>
<td></td>
<td>transmission line</td>
<td>gratings</td>
</tr>
<tr>
<td></td>
<td>and Antenna</td>
<td></td>
</tr>
<tr>
<td>Transmission Medium</td>
<td>Air, Waveside coaxial cable</td>
<td>Air, Optic fibre</td>
</tr>
<tr>
<td>Coupling from transmission</td>
<td>reactive components</td>
<td>Lenses, Gratings,</td>
</tr>
<tr>
<td>line</td>
<td>transformer, antennae</td>
<td>Optical filters and</td>
</tr>
<tr>
<td></td>
<td>transmission line,</td>
<td>polarizers</td>
</tr>
<tr>
<td></td>
<td>amplifiers</td>
<td></td>
</tr>
<tr>
<td>Demodulator</td>
<td>Square law,</td>
<td>Square Law</td>
</tr>
<tr>
<td></td>
<td>Discriminators,</td>
<td>Discriminators</td>
</tr>
<tr>
<td></td>
<td>and/or mixers</td>
<td>and mixers</td>
</tr>
</tbody>
</table>
A typical optical fibre system is shown, in block diagram form, in Fig 2.2(a). A typical microwave system is shown in Fig 2.2(b) for comparison.

Fig 2.2 Comparison of Optical Fibre Transmission System and a Digital Microwave Transmission System

(a) Typical Optic Fibre System

(b) Typical Microwave System
The techniques of optic fibre transmission system design are similar to those of radio system design. This is due to the similarity between the radio and optical fibre systems.

Optical fibre systems are designed on a power budget, in much the same way as a microwave system [2],[3],[4],[5],[6].

Firstly, a minimum receiver SNR figure is chosen to yield a bit error rate (BER) that is acceptable. Generally a figure of 1 error in $10^9$ bits or better is required for acceptable transmission of computer information.

The maximum transmit power available is dictated by two characteristics, one of the laser and the other of the optic fibre; to which it is connected [2],[3],[4],[5],[6]. The maximum laser output power is fixed by the technology available and the equipment cost. Moreover one can only inject a certain amount of optical power into the optic fibre before it acts as a non-linear transmission medium, in which further stimulated emission can occur [2],[3],[4],[5],[6]. Therefore the maximum amount of transmit power available is fixed.

As the modulated light travels down the fibre it is attenuated by a number of loss mechanisms in the fibre core and by the mismatch at joints in the fibre. The power arriving at the detector is given by the length of the fibre and the number of joints in the section [2],[3],[4],[5],[6].

One can define a system strength parameter which indicates the amount by which signal power at the receiver exceeds that required to obtain the minimum acceptable receiver SNR [2],[3],[4],[5],[6]. The system strength is chosen using an idealized transmission system in which there are no
joints. The system strength must be sufficient to ensure that, during installation when fibre joints are made, the power at the receiver is not reduced below the minimum SNR level, and as the system ages, and/or ground movements and other forms of degradation occur, the power at the receiver will not fall below the minimum level.

To ensure that some margin of system strength is left after installation, joints in the fibre must offer no more attenuation than a certain figure fixed by the system designer. If attenuation at a joint exceeds this figure it must be removed and the fibre rejoined. Some variation in the attenuation of individual fibre sections also occurs. The system strength must be sufficient to allow for such loss variations.

The parameters which effect the design of an optic fibre system can be summarised in the Optic Fibre Equation. Which is an equation of signal to noise of the optic fibre communications receiver [18],[19],[20],[21],[22],[23],[24],[25].

\[
\text{SNR}_r = \frac{2^{M_i} C_i C_0 C_e^2 \exp(-2\alpha_f L_f)}{N_i + N_f + N_0 + N_e}
\]  

(2.1)

Where \(P_i\) is the output power of the transmit laser or LED, 
\(M_i\) is the proportion of transmit power in the modulation, 
\(C_i\) is the efficiency of input coupling of the laser-fibre junction, 
\(C_0\) is the output coupling of the fibre-photo receiver coupling, 
\(C_e\) is the optical to electrical conversion efficiency of the photo receiver, it includes the quantum efficiency it is highly frequency dependent, 
\(\alpha_f\) is the attenuation of the optic fibre per unit length, 
\(L_f\) is the length of the optic fibre, 
\(N_i\) is the noise power due to the macro uncertainty of frequency and amplitude of the transmit laser,
2.15

\( N_f \) is the noise power of random scattering processes in the fibre core,

\( N_o \) is the optical noise at the receiver due to the quantum nature of light

and \( N_e \) is the electrical noise of the photo receiver.

The optical to electrical conversion efficiency of \( C_e \) is discussed in great detail in the signal to noise analyses of chapter 5. The macro uncertainty in amplitude and frequency of a laser source is discussed in chapter 3 together with the quantum noise of light.

Equation 2.1 is similar in nature to the system design equations used in sonar [26] and radar [27]. These equations illustrate which parameters, of the system performance, can be improved to increase the system SNR by the greatest amount. Obviously \( SNR_r \) is increased if terms in the numerator of (2.1) are increased and terms in the denominator are reduced.

By rearranging (2.1), an expression for the spacing between regenerators can be found,

\[
L_f = \frac{1}{2\alpha_f} \log_e \left[ \frac{P_{i}^2 M_{i}^2 C_{i}^2 C_{0}^2}{SNR_r (N_i^2 + N_f^2 + N_o + N_e)} \right]
\]

(2.2)

If the argument of the natural logarithm is large, adjustments to the parameters in the argument have a small effect on the SNR. However changes to the fibre attenuation \( \alpha_f \) have a marked effect. Moreover as \( \alpha_f \) becomes smaller and the argument of the logarithm becomes larger, smaller changes in \( \alpha_f \) produce large changes in the regenerator spacings \( L_f \).

Generally, the attenuation, of optical signals, by the fibre, over the length \( L_f \), tends to dominate the SNR, thus small reductions in \( \alpha_f \)
produce significant changes in the SNR. Considerable research is being undertaken in the fibre attenuation area and maximum regenerator spacings continue to increase.

Further increases in regenerator spacings can, however, also be achieved by, either reducing the noise terms in the denominator of the argument of the natural logarithm in (2.2), or by increasing the values of terms in the numerator.

Using amplitude shift keying and direct photo detection, the maximum value of $M_i$ is one third for 100% depth of modulation [28],[29]. By using a more efficient modulation scheme, such as phase shift keying, $M_i$ can be increased to unity giving an improvement in SNR, of 9.5 dB.

In commercial systems $C_i$, $C_0$ and $C_e$ are usually very close to unity, therefore the significance of any improvement in these values is small, and may cost more than the unimproved system.

Considerable improvement in $\text{SNR}_r$ may be gained if the transmit laser is stabilized reducing $N_i$ [2],[4],[10],[11],[12],[13], however this is also expensive.

Further improvements in the $\text{SNR}_r$ can be obtained by reducing the fibre channel noise power [14],[15] $N_f$, optical granulator noise power $N_0$ [19],[20] and electrical noise power at the output of the photo receiver. Methods of reducing the effects of these noise sources are discussed in chapter 5.
2.17

2.4 THE FUTURE OF OPTICAL COMMUNICATIONS

Many of the difficulties mentioned in sections 2.2 and 2.3 were encountered by the pioneers of radio. The first transmitters were crude and the emission bandwidths were very large and unstable. Little was known about transmission lines and antennae, therefore coupling of the transmitter to the antennae was poor. Transmitter power was also very low.

Techniques of detection were also crude, crystals of galena were used to form envelope detectors, later metallic salts were used to produce detector diodes. Information rates were also very low.

Little was known about the behaviour of transmission paths and attenuation mechanisms.

By experimentation, knowledge was gained which enabled improvement of transmission system reliability. The invention of the vacuum tube led to more precise control of frequency and power of the transmitter. Receive amplifiers could also be built to amplify the incoming signal.

More knowledge about the transmission of electromagnetic radiation through the atmosphere was gained and predictions, about attenuation of signals, could be made.

The characteristics of antennae and transmission lines were understood and coupling to receivers and transmitters improved. New methods of modulation were discovered, which improved the SNR of reception and made more efficient use of transmitter power.
Information rates were improved, from the morse code era, to the high speed digital microwave systems spanning the world today.

In the field of optical communications, man is at the spark transmitter stage. Methods of light generation modulation and detection must be improved, if the full capacity of optical communications is to be exploited. Inexpensive methods of building highly stable lasers must be found so that optimal modulation and detection techniques can be used in commercial applications [2],[4],[10],[11],[12],[13].

Still more must be understood about the nature of light itself, so that photon noise can be reduced [19],[20],[21],[22],[23],[24],[25].

Research into methods of reducing fibre core attenuation must continue.

Finally, the electronic equipment used in pre-modulation and post-demodulation stages must be improved. Presently bandwidths are small and phase characteristics uncontrolled [39],[40],[41],[42],[43],[44],[45],[46],[47],[48],[49],[50],[51],[52],[53]. New methods of sound theoretically based design must be found; for such equipment.
REFERENCES - CHAPTER 2

[1] COOPER, K.
"Clear Path Ahead for Unlimited Bandwidth."

"Optical Communications."

[3] BANCETTI, B. and TARTARA, G.
"Equalization and Quality Prediction in Digital Radio Systems."
GTE Telecommunicationi, Milan, Italy, pp 3 - 40.

"Optoelectronics - An Introduction."

"Contrasting Fibre-Optic-Component Design Requirements in Telecommunications, Analog and Load Data Communications Applications."

"System and Circuit Consideration for Integrated Industrial Fibre Optic Data Links."
"The Phototype and the Conversion of Radiant Energy."
Journal of the Society of Telegraph Engineers and of Electricians,
Vol. IX, pp 363.

[8] SNYDER, W.F.
"Acoustic Performance of 16 mm Sound Motion Picture Projectors."

[9] BROWN, R.
"Lasers - A Survey of Their Performance and Applications."

[10] BOYD, G.D. and GORDON, J.P.
"Confocal Resonator for Millimeter Through Optical Wavelength
Masers."
Laser Theory, F.S. BARNES Ed, IEEE Press, N.Y., 1972,
pp 115 - 134.

"Fields and Waves in Communications Electronics."

[12] BROWN, R.
"Lasers - A Survey of Their Performance and Applications."
Business Books Ltd., 1969, pp

[13] WILLIAMS, E.W. and HALL, R.
"Luminescence and the Light Emitting Diode."
[14] HAYKEN, S.
"Communications Systems."

[15] STREMLER, F.G.
"Introduction to Communication Systems."

[16] HAYKEN, S.
"Communications Systems."

[17] RODDY, D. and COOLEN, J.
"Electronic Communications."

[18] TAKENAKA, T., KAZUMASA, T. and FUKUMITSU
"Signal-to-noise ratio in Optical Hetrodyne Detection for Gaussian Fields."

[19] HANBURY BROWN, R.
"The Intensity Interferometer its Application to Astronomy."

[20] CAVES, C.M.
"Quantum Mechanical Noise in an Interferometer."

"Limits on Electromagnetic Amplification Due to Complementarity."
First international Conference on Quantum Electronics, High View,
N.Y., 1959.

[22] GLAUBER, R.J.

"Coherent and Incoherent States of the Radiation Field."

[23] SUDARSHAN, E.C.G.

"Equivalence of Semiclassical and Quantum Mechanical Descriptions
of Statistical Light Beams."
Physical Review Letters, Vol. 10, No. 7, 1 April 1963,
pp 277 - 279.

[24] STOLER, D.

"Equivalence Classes of Minimum Uncertainty Packets."

[25] LOUDEN, R.

"The Quantum Theory of Light."
Claredon Press, Oxford 1983, pp

[26] SKOLNIK, M.I.

"Introduction to Radar Systems."
McGraw-Hill Kogakusha Ltd, Tokyo, pp 3 - 5.

[27] KOCK, W.E.

"Radar, Sonar and Holography - An Introduction Academic Press,
N.Y., 1973."
[28] HAYKEN, S.
"Communications Systems."

[29] STREMLER, F.G.
"Introduction to Communication Systems."

[30] BECCIOLINI, B.
"Impedance Matching Networks Applied to R.F. Power Transection."
Motorola Application Note, AN-721, Motorola Inc.

[31] NORRIS, P.M.
"Using Linvill Techniques for R.F. Amplifiers."
Motorola Application Note, AN-166, Motorola Inc.

[32] KWOK, SCING-PIN
"UHF Broadband Amplifier Design."
Motorola Application Note, AN-406, Motorola Inc.

[33] FRATER, R.H.
"Accurate Wideband Multiplier-Square-Law Detector."
The review of scientific Instruments, Vol. 35, No. 7, July 1967,
pp 810 - 812.

[34] PERSONICK, S.D.
"Receiver Design for Optical Fibre Systems."
[35] PERSONICK, S.D.

"Receiver Design for Digital Fibre Optics Communications Systems (I & II)."


[36] HADLEY, M.

"Amplifier Gains 10dB over Nine Octaves."

Motorola Device Data, EB37, Motorola Inc., pp 8-62 - 8-63.

[37] MICROWAVE ASSOCIATES

"Transistor Designers Guide."


[38] HA, T.T.

"Solid-State Microwave Amplifier Design."

John Wiley and Sons Inc., 1981, pp

[39] HEJHALL, R.

"R.F. Small Signal Design Using 2-Port Parameters."

Motorola Application Note, AN-215, Motorola Inc.

[40] YOUNG, G.

"Broadband Network Design for U.H.F. Amplifiers."

Motorola Application Note, AN-704, Motorola Inc.

[41] NOLLESEN, D.L.


Motorola Application Note, AN-217, Motorola Inc.
[42] WEINERT, F.
"Scattering Parameters Speed Design of High-Frequency Transistor Circuits.

[43] GHAUSI, M.S.
"Principles and Design of Linear Active Circuits."

[44] GOELL, J.E.
"Input Amplifier for Optical PCM Receivers."
The Bell System Technical Journal, Vol. 53, No. 9, Nov. 1974,
pp 1771 - 1793.

"Transistor Circuit Design."

[46] CHERRY, E.M.
"An Engineering Approach to the Design of Transistor Feedback Amplifiers."

[47] HAMILTON, T.D.S.
Handbook of Linear Integrated Electronics for Research."

"The Design of Wide-Band Transistor Feedback Amplifiers."
[49] HULETT, J.L. and MUOI, R.V.
"A Feedback Receive Amplifier for Optical Transmission Systems."

[50] OHKAWA, N., YAMADA, J.I. and HAGIMOTO, K.
"Broadband GaAs FET Optical Front-End Circuit Up to 5.6
GHz."

[51] CHERRY, E.M.
"Impedance Mismatch in Wide-Band Transistor Amplifier Design."
IEEE TRANS. on Circuit Theory Vol. CT , No. , Feb. 1970,
pp 131 - 132.

[52] OWEN, B.
"Pin - GaAs FET Optical Receiver With a Wide Dynamic Range."

[53] LEHENY, R.F., NAHORY, R.E., POLLACK, N.A., BALLMAN, A.A., BEEBE, E.D.,
DeWINTER, J.C. and MARTIN, R.J.
"Integrated In0.53 Ga0.47 As p-1-n F.E.T. Photo
receiver."
3. THE GENERATION OF COHERENT LIGHT AND ITS CHARACTERISTICS

An understanding of the way in which coherent light beams are generated and their characteristics is important if the signal dependent noise, arriving at the detector, is to be described mathematically. Such descriptions are essential in the signal to noise analysis of a light transmission system.

Signal dependent noise is a consequence of two basic physical aspects; the first is noise due to the process of light generation and is attributable to the classical behaviour of amplification and resonance [1], [2],[3],[4],[5],[6],[7]. The second is due to the granular or quantum nature of light itself and is largely independent of the way in which it is generated [8],[9],[10],[11],[12],[13]. These aspects apply to all electromagnetic wave phenomena, however in radio systems they are insignificant compared with noises from other sources. Thus no analysis of quantum mechanical or generation noise and therefore signal dependent noise, is used in the design of radio systems [14].

In light systems however, if signal dependent noise is not used in system signal to noise ratio (SNR) calculations, serious discrepancies between theoretical predictions and experimental results occur. Moreover the behaviour of light systems becomes inconsistent with that of radio systems which intuitively should not occur since they encompass the same electromagnetic phenomena.

Failure to consider the quantum mechanical behaviour of light has led one author [1],[2],[15],[16] to the incorrect conclusion, that optical systems, employing homodyne detection techniques, experience a 3dB increase
in SNR over equivalent heterodyne systems. From a radio standpoint the idea is quite absurd [17],[18].

For these reasons the generation of coherent light and the accepted quantum mechanical description of coherent light is outlined in the following sections.

Initially the process of spontaneous emission of light from cold gases and solids will be discussed in 3.1.1 as a basis for the discussion of stimulated emission in gases and semiconductors given in section 3.1.2.

The construction of gaseous and semiconductor Lasers will be outlined in section 3.3.3. Finally section 3.3.4 will consider the general output characteristics of a laser.

The current quantum mechanical description of light is summarised in section 3.2.
3.1 THE GENERATION OF COHERENT LIGHT

In 1917 Einstein referred to a new process known as stimulated emission, which was introduced to fully explain the interaction of radiation and matter [7]. During the 1950's a full theoretical treatment of stimulated emission and the laser was published and on August 6 1960 the first report of a functional laser was revealed by T.H. Maiman [7].

With the development of the laser, man could generate coherent beams of light which could be used to carry information from one place to another.

The output of a laser is not however completely coherent and some uncertainty as to frequency or amplitude is present as a consequence of the process of laser action [1],[2],[3],[4],[5],[6],[7],[19],[20].

Therefore, it is intended to review the processes involved in light generation and the effects these have on laser light output. Initially the process of spontaneous light emission is outlined, to give a basis for an understanding of the process of stimulated emission, in gases and semiconductors. Practical details of laser conduction are given in section 3.1.3. The effects of internal processes and cavity size on laser output characteristics are also discussed.

3.1.1 Spontaneous Light Emission From Semiconductors and Gases

In 1913 Neils Bohr [21] proposed his quantization theory of electronic angular momentum about an atomic nucleus. Electrons, he said, could only assume a discrete number of orbits with discrete angular momenta.
and associated potential energy. The electron itself is kept in orbit by Coulombic force of attraction between the positively charged nucleus and negatively charged electron.

Between each permitted orbit lay regions in which the electron could not orbit. It could however move to a higher orbit if given precisely the right amount of energy, or downward to a lower orbit giving up a precise amount of energy. The energy absorbed or emitted is electromagnetic, with a frequency related to the energy difference between the two orbits given by Einstein's equation [21],

$$ E = h\nu $$

(3.1)

Where $E$ is the energy required to promote an electron from one orbit to another,

- $h$ is Planck's constant
- $\nu$ is the frequency of electromagnetic radiation.

Energy, $E$, is also the energy of one quantum of electromagnetic radiation called a photon [21]. The photon is the fundamental particle of electromagnetic radiation.

The subsequent theories have moved away from the Bohr orbital model to models containing regions of electronic probability separated by regions where the probability of finding an electron is low. Transition between these regions involves absorption or emission of radiation. The regions of high electron probability are known as energy levels. The differences between energy levels can be altered by external influence, such as either external electric or magnetic fields, or the proximity of other atoms and their electric or magnetic field strengths [22],[23],[24],[25].
In a gas, the atoms or indeed molecules are far apart [26], the effects of neighbouring atoms or molecules on each other is small therefore the predominant energy levels in gases are atomic or molecular. Fig. 3.1 shows an energy level diagram for the hydrogen atom [26].

![Energy Level Diagram for Atomic Hydrogen]

Using Bohr's postulates [20], which are based upon classical concepts of coulombic electronic attraction to the nucleus and momentum associated with the orbit by an electron about the nucleus, an expression for the energy of a particular energy level can be found [20]

\[
E = - \frac{mZ^2 e^4}{2n^2 \hbar^2}
\]  

(3.2)

Where

- \( E \) is the energy of the electron
- \( Z \) is the atomic number
- \( m \) is the electronic mass
- \( e \) is the electronic charge
- \( n \) is the energy level number
- \( \hbar \) is called "H-bar" and is the ratio
Using equation (3.1) and (3.2), the frequency of light emitted or absorbed by an electronic transition between two levels can be found.

Using Bohr's semiclassical theory, spontaneous emission of light from a gas can be described.

When an electrical discharge is passed through a gas collisions between atoms promote electrons from the ground state (N=1) to an excited state (N>1) [27]. After a brief period, the electrons return to the ground state emitting light with frequencies dependent on the energy transitions [20]. The process of energy promotion and decay occurs randomly with time and the gas is in motion, thus the phase of light is constantly changing and a broad band of frequencies is emitted. Emissions from individual atoms are said to be spontaneous [27].

Using these simple concepts, spontaneous emission of light from semiconductors may be described in a similar way [28],[29].

In crystals, the atoms are very close to each other so each atom is influenced by its neighbours. Electronic energy levels are determined not only by individual atoms but by groups of atoms. Electrons are tightly bound by the groups of atoms to which they belong and serve to bond the atoms together to form the crystal. As a result the distinct atomic energy levels are lost in semiconductor crystals and two broad energy levels are formed as shown in Fig. 3.2 [30],[31],[32].
The two energy levels are a function of the crystal as a whole [28]. Electrons involved in bonding are in the lower energy state, known as the valence band. At absolute zero all electrons are found in this band, the crystal is a perfect insulator [29],[30],[31]. At room temperature however, some electrons gain sufficient thermal energy to cross the energy gap into the conduction band where they are free to roam the crystal. An electron thus promoted leaves a hole where an electron should be. Neighbouring electrons in the valence band can fill the hole leaving a hole at another location, thus a hole can also roam the crystal in the valence band. If an electron meets a hole then it may combine with it giving up a photon of energy and return to the valence band [29],[30],[31]. The process of promotion and recombination is in equilibrium at room temperature [29],[30],[31].

The frequency of a photon emitted due to recombination is given by

$$\nu = \frac{E_C - E_V}{h} = \frac{E_g}{h}$$  \hspace{1cm} (3.4)

where $E_C$ is the conduction band energy
and $E_V$ is the valence band energy
and $E_g$ is the width of the energy gap,

$$E_g = E_C - E_V$$
Any photon thus emitted will be absorbed in the crystal as another electron-hole pair is created. For this reason no spontaneous light emission will occur due to thermal excitation alone [27],[28].

The controlled introduction of impurity atoms can be used to introduce charge carriers, either electrons in the conduction band or holes in the valence band [29],[30],[31]. Introduction of atoms of group 3 of the periodic table produces a p-type material in which the combination band is devoid of electrons and the valence band contains holes. The number of holes is proportional to the impurity density. If atoms from group 5 are introduced an n-type material is produced in which the valence band is completely full, and the conduction band is partially full, of electrons [29],[30],[31].

If a metallurgical junction between n-type and p-type semiconductors is formed a semiconductor diode is produced as shown in Fig 3.3.

![Fig 3.3 A Semiconductor Diode](image)

The combined energy level diagram is shown in Fig 3.4.
During formation of the junction, holes in the valence band tend to flow into the n-type material from the p-type, and electrons in the conduction band flow from the n-type material to the p-type, and a depletion region is formed [29],[30],[31].

If the PN junction is forward biased, electrons flow from the negative n-type region into the depletion and penetrate into the p-type region. Holes flow from the p-type region into the depletion region and on into the n-type region [29],[30],[31]. Both electrons and holes become minority charge carriers in their new regions, recombination between electrons and holes occur in the depletion region between minority and majority carriers. Thus for a suitable band gap photons of electromagnetic radiation are emitted from the crystal [27],[28].

The process of producing charge carriers in the depletion region and minority carriers outside it, is known as injection and forms the basis for Light Emitting Diode (LED) and Laser operation. If the number of injected charge carriers is small, radiative decay occurs spontaneously and incoherent light is emitted [27],[28].
In both gases and semiconductors spontaneous emission occurs when the transition energy between states is much larger than the thermal energy due to ambient conditions, that is

\[ h\nu \gg k_B T \]  (3.5)

Where \( k_B \) is Boltzmann's Constant and \( T \) is absolute temperature.

When (3.5) is satisfied, the probability that an electron will return to a lower energy level, from an excited state, is high [6].

3.1.2 Stimulated Emission

Using the knowledge of spontaneous emission processes in gases and semiconductors from section 3.1.1, an understanding of stimulated emission processes in gases and semiconductors is possible.

Stimulated emission is somewhat similar physically to semiconductor electronic avalanche or nuclear chain reaction.

Consider a group of electrons in an excited state that have not yet released their energy spontaneously and returned to the ground state [7]. Moreover suppose an electron promoted to the same state, much earlier, has reached the end of its lifetime in that state, and decays spontaneously emitting a photon of light. The photon encounters one of the excited electrons which immediately returns to the ground state emitting a second photon. This photon has exactly the same phase and frequency as the first.
When a photon is incident upon an atom in which the electrons are excited, the photon is not absorbed. The incident photon merely causes the return of the excited electron to a lower energy level. Thus two photons with the same frequency and phase have been generated [7].

The two photons encounter further excited electrons and each produce two more photons with the same frequency and phase, as the original photon. Thus 4 photons with exactly the same frequency and phase are generated. In this way each photon replicates itself until all the excited electrons have returned to the ground state. Hence the original photon has been amplified [7].

Under normal circumstances there are more electrons in the ground state than the excited state. The photons generated by the stimulated emission process are absorbed by those in the ground state. The rate of absorption will exceed the rate of stimulated emission [7]. Electrons in the excited state do not absorb radiation and so the rate of absorption is proportional to the number of electrons in the ground state. However, the rate of stimulated emission is proportional to the number of electrons in the excited state [7].

If the rate of absorption is reduced below the rate of stimulated emission it is possible to produce an overall amplification of a photon, since more photons of the same frequency and phase can be generated than absorbed [7]. To reduce the rate of absorption below that of stimulated emission, a greater number of photons must be in the excited state than the number in the ground state. This is known as a population inversion. When a population inversion is achieved Light Amplification by Stimulated Emission of Radiation (laser) is possible [7].
A population inversion is generated by pumping electrons from the ground state to the excited state. A number of pumping methods exist to produce a population inversion [6],[7].

The two of the greatest interest here are activated discharge techniques used in the Helium Neon gas laser [7], used widely in experiments to be subsequently discussed, and Injection pumping used in Laser diodes, which are widely used in telecommunication applications [7], [31]. Both techniques are discussed in the following section.

Having achieved a population inversion in the active material, it is possible to define a gain function $\alpha_L$, [32]

$$\alpha_L = N_2 - \left(\frac{g_2}{g_1}\right) N_1 \sigma_{12}$$

(3.6)

Where $N_1$ and $N_2$ are the populations of electrons in states 1 and 2 respectively, $g_1$ and $g_2$ are statistical weights for transition from level 1 to level 2 and level 2 to level 1. They reflect the probability of upward or downward electronic movement.

$\sigma_{12}$ is the single frequency absorption cross section for the particular transition.

The statistical weight $g_1$ and $g_2$ give the probabilities that an electron in state 2 will return to state 1 and an electron in state 1 will move to state 2 respectively. These are properties of the material only [32].

The absorption cross section $\sigma_{12}$ is a macro property of the active material and relates the chance of a photon in a plane wave interacting with an electron in the medium [32].
The threshold condition for stimulated emission gain occurs when $|\alpha| = 0$. Using this observation, an expression for the minimum pump power may be derived for specific lifetimes in the excited states [32]. The pump threshold power is reached when sufficient pump energy per unit time exceeds the energy radiated by spontaneous emission. Thus, assuming that no losses are present.

$$P_p = \frac{h V_{12} \gamma_{2/g, N_1}}{T}$$  \hspace{1cm} (3.7)

Where $P_p$ is the pump power

- $V_{12}$ is the frequency of expected radiation
- $T$ is the mean lifetime of a particular state.

To increase the power of stimulated emission produced, it is evident from (3.6) that the number of electrons in state 2 should be increased, which implies that the pump power should be increased. However stimulated emission power may also be increased by increasing the cross section of active material encountered $\sigma_{12}$. By placing mirrors around the active medium, stimulated emission photons will pass through the active material repeatedly [32]. Light amplification will continue and optical power will increase until the material is saturated and the rate of stimulated emission decay of electrons exceeds the rate at which they are pumped.

It is evident therefore, that the gain of a Laser is inversely dependent upon the optical power present in the material. As the output power rises, the gain tends to fall until saturation is reached, thus [2], [6]

$$\alpha(p) = \frac{\alpha(0)}{1 + SPI}$$  \hspace{1cm} (3.8)
Where $\alpha(p)$ is the gain as a function of power.

$\alpha(0)$ is the gain at zero output power.

$S$ is the saturation factor of the material and $P_0$ is the output power.

The general behaviour of both Gas and Semiconductor lasers is the same [32]. Specific differences in output behaviour and mechanisms do occur, but are of little interest in the present context.

### 3.1.3 Helium-Neon and Gallium Arsenide Lasers

The HeNe Laser belongs to the neutral atom class of gas Lasers [33]. As its name implies, the active medium of such lasers, is a mixture of the noble gases Helium and Neon. The mixture is enclosed in a quartz tube, with Helium at a pressure of 1 mmHg and Neon at a pressure of 0.1 mmHg [33].

The cavity around the gas mixture is formed, either by silvering the ends of the Laser tube as shown in Fig 3.5(a) [34], or by placing two external mirrors at the ends of the tube as shown in Fig 3.5(b) [32].
Fig 3.5 General Construction of the Helium Neon Laser

Laser action is dependent upon the presence of both gases. Helium forms part of the pumping process and is referred to as an activator. Fig 3.6 shows the energy distribution of the gas mixture [7].

Fig 3.6 The Energy Levels of Helium and Neon
It is obvious that a number of Helium energy levels correspond to those of Neon. If electrons in the Helium are excited to upper energy levels by an electrical discharge or Radio Frequency (RF) oscillator, they will collide with the Neon atoms in-elastically. The Helium atoms give up their energy to the neon atoms and the electrons of the neon atoms are promoted to a corresponding state, dependent on the amount of energy contained in the Helium atoms. The 2p energy level of Neon contains almost no electrons, therefore a population inversion between either the 2s and 2p energy levels, or the 3s and 2p energy levels occurs. An incident photon will cause stimulated emission as the excited neon electrons return to the 2p Laser terminating level [7].

If stimulated emission occurs between the 2s and 2p levels, the output light is contained in the infra red ($\nu = 260 \text{ THz}$) [7]. On the other hand, if stimulated emission occurs between the 3s and 2p levels, the output light is visible ($\nu = 474 \text{ THz}$) [7].

Electrons in the 2p level of Neon return rapidly to the 1s level and so the population inversion is maintained.

It is evident, from the discussion, that the Helium is instrumental in producing a population inversion in the Neon and is thus part of the pumping mechanism. For this reason the pumping technique is known as an activated discharge pump [7].

Light is emitted from the cavity formed by the mirrors if one mirror partially transmits light.
The physical structure of a Gallium Arsenide (GaAs) Laser is quite different from that of a Helium Neon Laser, however the purposes of some structures are the same [3].

A junction of n and p type GaAs is cleaved along the 110 crystal planes to form mirrors and enclose the junction in a cavity, the crystal is then sawn to shape, as shown in Fig 3.7 [3].

The GaAs P-N junction has a semiconductor energy level diagram similar to that of Fig 3.4. The energy level diagram for GaAs is shown in Fig 3.8 [3].
If the p-n junction is forward biased, electrons will flow into the depletion region from the n-type material and into the p-type material. Holes will flow from the p-type material into the n-type through the depletion region [3],[7]. Recombination of electrons and holes will occur in the depletion n-type and p-type regions [3],[7]. As a result of recombination, spontaneous emission of radiation, from the crystal, will occur. However in some cases, photons emitted from one combination will stimulate emission of others. In low current cases, however, emission from the crystal will be mainly spontaneous [3],[7].

If the current is increased beyond a threshold value, the population of electrons in the depletion region conduction band will exceed that of the valence band, which implies that the number of holes in the valence band exceeds the number of holes in the conduction band; a population inversion has occurred, thus light amplification is possible [3],[7]. However it is confined to a very narrow region close to the junction.
Power output levels of semiconductor lasers are very small typically 1μW to 2mW, although they exhibit very high efficiencies, typically 30% to 40%, which is 1 to 2 orders of magnitude better than neutral-atom gas Lasers [2],[32].

3.1.4 Cavity Modes - Frequency and Amplitude Stability.

Mirrors enclosing the Laser active material form a resonant cavity in which only certain modes of oscillation will be sustained [35],[36]. Each mode corresponds to an internal frequency of oscillation, which will be amplified by stimulated emission only if it corresponds with the stimulated emission frequency of the active medium. In practice the stimulated emission frequency is not confined to a single frequency. It occurs over a band of frequencies $\Delta \nu$ [32]. Any cavity mode, whose frequency is within this range, is amplified by stimulated emission.

Subsequent analysis will reveal that the frequencies of various modes are closely spaced; a number fall within the amplification bandwidth $\Delta \nu$. Consequently they are amplified and appear at the Laser output.

Specific analysis of various modes subtended in a cavity of arbitrary dimensions is a difficult mathematical problem. However insights into laser frequency output may be obtained from analysis of the confocal resonator, shown in Fig 3.9 [35],[36].
Using Huygens principle of secondary sources [37],[38], the field at point \( P(x,g) \), due to the sum of all points over the surface \( P'(x,y) \), can be found.

The free space Green's function for radiation, from an infinitely small radiator, is of the form [38].

\[
G(P_1) = \frac{\exp(jkr_{01})}{|r_{01}|}
\]  

(3.9)

Where \( r_{01} \) is a vector from the dipole to an arbitrary point \( p \),

Using Green's theorem the Helmholtz Kirchoff integral theorem is derived for a surface \( s \) [38].

\[
U(P_0) = \frac{1}{4} \int_S \frac{\partial U}{\partial n} \frac{\exp(jkr_{01})}{r_{01}} - U \frac{\partial}{\partial n} \frac{\exp(jkr_{01})}{r_{01}} \, ds
\]  

(3.10)

Where \( U \) represents the field, at a point due, to the Secondary Huygens sources over the surface \( S \).

\( n \) is the normal to the surface \( s \).

\( k \) is the wave vector \( k = \frac{2\pi}{\lambda} \)
If it is assumed that the field at $P'$ is linearly polarized, in the $y$ direction, it is given by [20],

$$E_{y'} = E_0 f_m(x') g_n(y') \quad (3.11)$$

Where $f_m(x')$ and $g_n(y')$ are the field variations over the aperture of the reflector.

The field at $P$ is given by [20],

$$E_y = \int_{S'} \frac{(1 + \cos \theta)}{4\pi \rho} e^{jk\rho} f_m(x') g_n(y') \, ds' \quad (3.12)$$

Where $E_0$ is the electric vector magnitude, at the source.

$\rho$ is the distance between $P$ and $P'$.

$\theta$ is the angle between the vector $PP'$ and the normal vector $n$.

It is assumed that the reflector is square of dimensions $2a$, $a$ is small compared with the reflector separation $b$, therefore $\theta$ is nearly zero. Moreover it is assumed that the active medium fills all space.

If it is required that the $x'y'$ field distribution reproduce itself over $xy$, to within a constant, the normal modes or eigen function of the resonator are found [20]. Thus

$$E_y = E_1 f_m(x) g_n(y) \quad (3.13)$$

Where $E_1 \sigma_m \sigma_n E_0$

$\sigma_m \sigma_n$ are proportionality factors and are generally complex, accounting for amplitude and phase changes.
and,

\[ \sigma m \sigma n \int f_m(x) g_n(y) = \int_{-a}^{a} \frac{jk}{2\pi \rho} \exp[jk\rho] f_m(x') g_n(y') \, dx'dy' \quad (3.14) \]

The paraxial approximation [39] to \( \rho \) is used in the phase term. It is assumed the \( p \approx b \) in all other terms, therefore, after some manipulation,

\[ \sigma m \sigma n F_m(X) G_n(Y) = \frac{\exp[-jkb]}{2\pi} \int_{\sqrt{c}}^{\sqrt{c}} F_m(X') \exp[(jXX')] \, dX' \quad (3.15) \]

\[ X \int_{\sqrt{c}}^{\sqrt{c}} G_n(Y') \exp[(jYY')] \, dY' \]

Where \( c = \frac{a^2}{b} = 2 \frac{a^2}{b} \)

\[ x = \frac{x\sqrt{c}}{a} \quad y = \frac{y\sqrt{c}}{a} \]

and \( F_m(X) = f_m(x) \quad G_n(Y) = J_n(Y) \)

It is possible to separate the variables in 3.15 such that \( F_m(X) \) may be written in terms of \( X \) varying quantities only [20].

\[ F_m(X) = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{c}}^{\sqrt{c}} F_m(X') \exp[(jXX')] \, dX' \quad (3.16) \]

Which is an homogeneous Fredholm equation of the second kind. It has solution [20],

\[ F_m(C, \eta) \propto S_m(C, \eta) \quad (3.17) \]
and \( \chi_m = \frac{2 \cdot c}{\pi} j^m R_{om}^{(1)} (c, l) \) \( m = 0, 1, 2, \ldots \) \hspace{1cm} (3.18)

Where \( S_{om} (c, \eta) \) and \( R_{om} (c, l) \) are the respective angular and radial wavefunctions in Prolate-Spherical co-ordinates, and \( \eta = x/c \).

The solution to the \( Y \) co-ordinate expression is similar. Thus the eigenfunctions of 3.15 are spheroidal wavefunctions \( S_{om} (c, X/a) \) \( S_{on}(c, y/a) \) \([20]\). The eigenfunctions are real, therefore the reflecting surfaces are of constant phase. The corresponding eigenvalues are

\[ \sigma_m \sigma_n = \chi_m \chi_n j \exp[(-jkb)] \] \hspace{1cm} (3.19)

For resonance, the round trip phase shift must be an integer multiple of \( 2\pi \). From (3.18) and (3.19) \([20]\),

\[ 2\pi q = 2 \left| \frac{\pi}{2} - kb + (m + n)\frac{\pi}{2} \right| \] \hspace{1cm} (3.20)

Where \( q \) is arbitrary.

Since \( k = \frac{2\pi}{\lambda} \), the condition for resonance is \([20]\),

\[ \frac{4b}{\lambda} = 2q + (1 + m + n) \] \hspace{1cm} (3.21)

The fractional loss for reflection due to diffraction is

\[ 0 = 1 - |\sigma_m \sigma_n|^2 = 1 - |\chi_m \chi_n|^2 \] \hspace{1cm} (3.22)

In most lasers light is emitted from a small aperture near the centre of the resonator \([20]\). Therefore the eigenfunctions of interest are those near the centre, thus

\[ \eta^2 \ll 1 \] \hspace{1cm} (3.23)
\[ F_m(X) \text{ can be approximated by,} \]

\[ F_m(X) \approx \frac{m+1}{\Gamma(m+1)} H_m(X) \exp \left( -\frac{X^2}{2} \right) \]  

(3.24)

Where \( \Gamma(\xi) \) is the gamma function of \( \xi \).

The mode shape is therefore approximately a Gaussian function times an Hermite polynomial [20].

Fields beyond the resonator can be found by again applying Huygen's principle [37],[38] across the outlet aperture subject to the Gauss-Hermite illumination.

For a mirror with transmission coefficient \( t \), the field at a point \( Z_0 \) outside the cavity is [20],

\[ E(x,y,Z_0) = \mu E_0 \sqrt{\frac{2}{1 + \xi^2}} \frac{m}{\Gamma(m+1)} \frac{n}{\Gamma(n+1)} \]

\[ \times \ H_m \left( X \sqrt{\frac{2}{1 + \xi^2}} \right) H_n \left( Y \sqrt{\frac{2}{1 + \xi^2}} \right) \exp \left[ -\frac{kw^2}{b(1 + \xi^2)} \right] \]

\[ \times \exp \left[ -j \left\{ k \left( \frac{b}{2} (1 + \xi) + \left( \frac{\xi}{1 + \xi^2} \right) \frac{w^2}{b} \right) - (1 + m + n)(\frac{\pi}{2} - \varphi) \right\} \right] \]  

(3.25)

Where \( W = x^2 + y^2 \)

\[ \xi = \frac{2Z_0}{b} \]

and \( \tan \) = \( \frac{1 - \xi}{1 + \xi} \)
The corresponding half power beam width is [20],

\[\omega_0 = 0.939 \sqrt{\frac{\lambda}{d}} \text{ radians}\]  

(3.26)

Using the condition for resonance (3.21), it is possible to define the frequency of separation between different orders of longitudinal mode, thus for fixed values of \(m\) and \(n\) [20],[36]

\[\Delta \omega = \frac{\pi c}{b}\]  

(3.27)

For stimulated emission to occur, the cavity dimensions must be chosen such that they produce modes with frequencies co-incident with the stimulated emission frequency.

In practice, the stimulated emission energy levels are broadened by processes in the active medium [5]. Thus each energy level may be represented by an absolute or centre energy term and an associated energy bandwidth. With reference to the ground state, the energy in level \(i\) lies between energies

\[-\frac{\Delta E_i}{2} \text{ and } \frac{\Delta E_i}{2}\]

\[E_i' = E_i \pm \frac{\Delta E_i}{2}\]  

(3.28)

The frequency range of light emitted due to transition between any two energy levels \(i\) and \(j\) is

\[\omega_{ji} = \frac{(\Delta E_j + \Delta E_i)}{n}\]  

(3.29)
It is therefore possible to define a frequency range, over which stimulated emission will occur, in terms of the energy level transition and a bandwidth parameter.

\[ \omega_{ji} - \frac{\omega_{ji}}{2} \leq \omega \leq \omega_{ji} + \frac{\omega_{ji}}{2} \]  

(3.30)

The frequency content of the Laser output is governed by the number of Laser cavity modes that lie within the range given by (3.30) [20],[35].

\[ \omega_{lc} = 2 \frac{\pi mc}{2bn} \]  

(3.31)

Where \( m \) lies within the interval

\[ \text{Int}\left[\frac{\omega_{ji} - \omega_{ji} \times bn}{2 \pi c}\right] \leq m \leq \text{Int}\left[\frac{\omega_{ji} + \omega_{ji} \times bn}{2 \pi c}\right] \]

\( n \) is the refractive index of the active material and \( \text{Int} \) is the integer truncation function.

Therefore, the output of a laser is generally characterized by the presence of a number of frequencies spaced by [20],[35],

\[ \Delta \omega = \frac{\pi c}{b} \]  

(3.32)
3.2 THE QUANTUM THEORY OF LIGHT

The concept of light dependent noise was introduced in section 2. It is a consequence of the internal processes of the Laser as discussed in section 3.1, and the physical characteristics of light as a consequence of its dual wave - particle nature.

Although laser action is itself a non classical phenomenon, the description of laser spatial and temporal behaviour was treated classically \[20\],\[35\]. Such classical treatment is reasonable because large numbers of photons are involved in the process of light amplification. Moreover, movement of the atoms and electrons occur at non relativistic speeds.

However, the various attenuation processes, mentioned in section 2, reduce the number of photons observed by the receive photo detector to sufficiently small values that quantum mechanical uncertainties become relevant \[8\],\[15\],\[16\]. These uncertainties manifest themselves as noise and must therefore be considered in any detector noise analysis. Ultimately, quantum mechanical noise processes fix the smallest detection noise attainable, assuming all other noise sources are eliminated. The minimum limit is known as the quantum noise limit, beyond which noise cannot be reduced.

Recently however, a number of authors \[40\],\[41\] have suggested that noise in optical signals may be reduced below the traditional quantum limit using the squeezed state concept, which is discussed in section 5.8.3.

In order to understand the traditional quantum limit to detection, and the possible implications of the squeezed state proposals, a review of the quantum mechanical description of light is required.
In the following section (3.1.1), Dirac's famous "bra" and ket" notation is introduced. While it is not necessary for an understanding of the quantum mechanical analysis of light, it is a useful shorthand. Use of Dirac's notation is also consistent with the notation used in the references [8],[9],[10],[11],[12],[42],[43].

Following the introduction of Dirac's notation, the quantum mechanical description of the harmonic oscillator and its relationship to optical processes is introduced [43].

3.2.1 Light and the Simple Harmonic Oscillator.

The quantum mechanical description of light is based upon the analogy between the radiation field and an ensemble of harmonic oscillators [8], [9],[10],[11],[12],[42],[43]. Using Maxwell's field equations [44] and the vector potential, the relationship between light and the harmonic oscillator is readily shown [43]. The details of the harmonic oscillator are well understood and may be used to predict the behaviour of light [45].

In time varying differential form, Maxwell's equations are [44],

\[ \nabla \times \mathbf{H} = \dot{\mathbf{E}} + \mathbf{J} \]  
(3.33)

\[ \nabla \times \mathbf{E} = -\mathbf{\mu} \frac{\partial \mathbf{H}}{\partial t} \]  
(3.34)

\[ \nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon} \]  
(3.35)

\[ \nabla \cdot \mathbf{H} = 0 \]  
(3.36)

Where \( \mathbf{H} \) is the magnetic field per unit length, \( \mathbf{E} \) is the electric vector, \( \mathbf{J} \) the current density, \( \rho \) the charge, \( \varepsilon \) the permittivity, and \( \mu \) the permeability of the medium of propagation. The dot above a variable denotes differentiation with respect to time.
\[ \nabla = \hat{i} \frac{\partial}{\partial x} + \hat{j} \frac{\partial}{\partial y} + \hat{k} \frac{\partial}{\partial z} \text{ in three dimensions} \]

\( \hat{i}, \hat{j}, \hat{k} \) are unit vectors of a co-ordinate system in 3 dimensions.

Let \( \mathbf{A} \) represent the vector potential such that

\[ \mu \mathbf{H} = \nabla \times \mathbf{A} \quad (3.37) \]

Substituting into (3.34) and using the identity \( \nabla \times \nabla \phi = 0 \), the most general expression for the electric field vector \( \mathbf{E} \) is obtained,

\[ \mathbf{E} = -\nabla \phi - \mathbf{A} \quad (3.38) \]

Where \( \phi \) is the scalar potential.

Substitution of (3.37) and (3.38) into (3.33) and (3.34) and using the identity \[44\], \( \nabla \times \nabla \times \mathbf{A} = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}, \] (3.33),

\[ \nabla^2 \mathbf{A} - \mu \varepsilon \mathbf{A} = -\mu \mathbf{J} + \mu \varepsilon \mathbf{\phi} + \nabla (\nabla \cdot \mathbf{A}) \quad (3.39) \]

and

\[ \nabla^2 \phi + \nabla \cdot \mathbf{A} = -\frac{\rho}{\varepsilon} \quad (3.40) \]

Using the Lorentz Gauge condition \[44\]

\[ \nabla \cdot \mathbf{A} = -\mu \varepsilon \phi \quad (3.41) \]

two partial differential equations, one in terms of vector potential \( \mathbf{A} \) the other in terms of scalar potential \( \phi \), may be derived \[44\].
\[ \nabla^2 A - \mu \varepsilon A = - \mu J \] (3.42)

and

\[ \nabla^2 \phi - \mu \varepsilon \phi = - \frac{\rho}{\varepsilon} \] (3.43)

If no electron currents are present within the Volume of interest, (3.42) becomes the vector-potential wave equation [44]

\[ \nabla^2 A - \frac{\dddot{A}}{c^2} = 0 \] (3.44)

where \( c = \frac{1}{\sqrt{\mu \varepsilon}} \).

Solution to (3.44) will have the form

\[ A = \sum_k A_k(t) \exp (j k \cdot r) + A_k^*(t) \exp (-j k \cdot r) \] (3.45)

Where the \( k \) are wave vectors and \( r \) is a position vector.

It is clear that (3.45) is a Fourier series. Each component of this series obeys the wave equation (3.1),

\[ k^2 \cdot A_k(t) + \frac{1}{c^2} \dddot{A}_k(t) = 0 \] (3.46)

If it is assumed that \( k \) and \( r \) are co-linear, a 1 dimensional harmonic oscillator equation is obtained.

\[ \frac{2A_k(t)}{t^2} + \omega_k^2 A_k(t) = 0 \] (3.47)

Where \( \omega_k = ck \).
The retardation terms of the vector potential are included in the exponential terms of (3.45). The spatial Fourier coefficients of (3.45) are dependent only upon time, therefore each wave vector \( \mathbf{k} \) is orthogonal to each corresponding vector potential Fourier coefficient [43],

\[
\mathbf{k} \cdot \mathbf{A}_k(t) = \mathbf{k} \cdot \mathbf{A}_k^*(t) = 0
\]

Moreover \( \mathbf{A}_k(t) \) is entirely independent of \( \mathbf{A}_k^*(t) \) [43].

The similarity between the optical radiation field and the simple harmonic oscillator is illustrated by the "mass on a spring" problem shown in Fig 3.10.

When no external force is applied the balance of forces during oscillation in one dimension is described classically by the equation [45],

\[
\frac{d^2 x(t)}{dt^2} - \omega^2 x(t) = 0
\]

Where \( \omega = \sqrt{\frac{k}{m}} \)
It is clear immediately that (3.49) has exactly the same form as (3.47). This similarity in form and the Fourier series of (3.45) indicate that the radiation field may be represented as an ensemble of harmonic oscillators [43].

Referring again to Fig 3.10, the potential energy of the spring is written in terms of the natural frequency $\omega$.

$$E_p = \frac{1}{2} m \omega^2 x^2 = 2m^2 \nu^2 x^2$$  \hspace{1cm} (3.50)

Using (3.50) the total steady state energy of the harmonic oscillator is given by Schrodinger's time independent wave equation [29],[45],[46].

$$-\frac{\hbar}{2m} \frac{\partial^2}{\partial x^2} (x) + 2\pi^2 m \nu^2 x^2 \psi(x) = E \psi(x)$$  \hspace{1cm} (3.51)

Which may be rewritten,

$$\frac{\partial^2}{\partial x^2} (x) + \left[ \frac{2mE}{\hbar} - \frac{2m \nu^2}{\hbar} x^2 \right] \psi(x) = 0$$  \hspace{1cm} (3.52)

Let $\frac{2\pi m \nu}{\hbar} = B = \frac{2mE}{\hbar^2}$

thus

$$\frac{\partial^2}{\partial x^2} (x) + (\beta - \alpha^2 x^2) \psi(x) = 0$$  \hspace{1cm} (3.53)

Making yet a further substitution. Let

$$\xi = \sqrt{\alpha} x$$

Then (3.53) becomes,

$$\frac{d^2}{d \xi^2} \psi + \left( \frac{\beta}{\alpha} - \xi^2 \right) \psi$$  \hspace{1cm} (3.54)
Which has a general solution of the form

\[ \psi = Ae^{\xi^2/2} + Be^{-\xi^2/2} \]  \hspace{1cm} (3.55)

However, the oscillations cannot grow in size since no amplification is present, thus \( A = 0 \). Moreover, the most general, "B", is a function of \( \xi \) also, thus,

\[ \psi(\xi) = H(\xi)e^{-\xi^2/2} \]  \hspace{1cm} (3.56)

Where \( H(\xi) \) is an arbitrary function of \( \xi \).

Using Eisberg's analysis [45], \( \frac{\beta}{\alpha} \) is restricted to values

\[ \frac{\beta}{\alpha} = 2n + 1 \]  \hspace{1cm} (3.57)

Substituting the definition of \( \alpha \) and \( \beta \) (3.52) into (3.57), it becomes clear that the total energy is given by [45],

\[ E = (n + \frac{1}{2}) \hbar \omega \]  \hspace{1cm} (3.58)

and the wavefunction is an exponential multiplied by an nth order Hermite polynomial [45],

\[ \psi(\xi) = e^{-\xi^2/2} H_n(\xi) \]  \hspace{1cm} (3.59)

Where \( H_n(\xi) \) is the nth order Hermite polynomial of \( \xi \).

According to the quantum theory, the squared magnitude is a probability that a particular state, \( n \), is occupied [46], therefore the sum of the squared magnitude over all states must be unity, in which case \( \psi(\xi) \) is normalized giving [45],
Using the following Hermite polynominal relationships [47],

\[
\frac{d}{d\zeta} H_n(\zeta) = 2n H_{n-1}(\zeta) \tag{3.61}
\]

and \(\zeta H_n(\zeta) = \frac{1}{2} H_{n+1}(\zeta) + n H_{n-1}(\zeta)\) \tag{3.62}

recursive relationships between wavefunctions may be derived \([45],[47]\),

\[
a^+ \psi_n(\zeta) = \sqrt{n+1} \psi_{n+1}(\zeta) \tag{3.63}
\]

and \(a \psi_n(\zeta) = \sqrt{n} \psi_{n-1}(\zeta)\) \tag{3.64}

Where \([45],[47]\)

\[
a = \sqrt{\frac{\alpha}{2}} X + \frac{1}{\sqrt{2}} \frac{\partial}{\partial X} = \frac{1}{\sqrt{2m\hbar}} + \frac{X}{\sqrt{2\hbar}} \tag{3.65}
\]

and \(a^+ = \sqrt{\frac{\alpha}{2}} X - \frac{1}{\sqrt{2}} \frac{\partial}{\partial X} = \frac{-1}{\sqrt{2m\hbar}} + \frac{X}{\sqrt{2\hbar}} \tag{3.66}\)

An is known as annihilation operator because it converts an \(n\)th state wavefunction to an \((n-1)\)th state, as in equation (3.63). Conversely \(a^n\) is known as a creation operator because it converts an \(n\)th order wavefunction to one of order \((n+1)\) as in equation (3.64).
Using (3.63) and (3.64), a composite operator $aa^+$ may be derived [45],[46]

$$a^+a\psi_n(\xi) = n\psi_n(\xi) \tag{3.67}$$

$n$ is interpreted as an eigen-value of the operator $a^+a$. Naturally, the operators $a, a^+$ may be generalized to matrices, and the wavefunctions generalized to vectors [43],[46].

Using the annihilation and creation operators, the wavefunctions and the energy expression (3.58), the characteristics of light may be discussed [10],[46].

Any coherent state of light is a linear combination of number states, as indicated by (3.45). Using the creation operator, each number state is written as a product of creation operators and the zero order, or vacuum, state [10]. Recall the relation (3.63),

$$\sqrt{n}\psi_n(x) = a^+\psi_{n-1}$$

then

$$11 = a^+_1|0\rangle \frac{|n\rangle}{1}$$

$$12 = a^+|1\rangle = a^+_1 a^+|0\rangle \frac{1}{\sqrt{2}} \frac{1}{\sqrt{1-\sqrt{2}}}$$

It is obvious that a recursion may be derived [10],

$$|nk\rangle_k = \frac{(a^+_k)^n_k}{(n_k!)^2} |0\rangle \frac{1}{1} \tag{3.68}$$

Where $n_k$ (0, 1, 2,...)

$k$ is the kth coherent state.
$|n_k\rangle_k$ is normalized thus [10],

$$<n_k|n_k\rangle_k = 1 \quad (3.69)$$

The complete field is written as [46],

$$|n_1\ n_2\ \ldots\rangle \quad (3.70)$$

by virtue of (3.45). However the different cavity modes are independent hence [46],

$$|n_1\ n_2\ \ldots\rangle = |n_1\rangle\ |n_2\rangle\ \ldots \quad (3.71)$$

Thus the application of any operator of the $k$th mode will effect only the photons of that mode [46],

$$a_k^n |n_1\ n_2\ n_3\ \ldots\rangle = (n_k + 1)^{1/2} |n_1\ n_2\ n_3\ \ldots\rangle \quad (3.72)$$

The eigen states of the modes obey a succession of relations [10],

$$a_k |\alpha_k\rangle = \alpha_k |\alpha_k\rangle \quad k = 0, 1 \ldots \quad (3.73)$$

Where $a_k$ is the eigenvalue of the $k$th mode and

$$|\alpha_k\rangle = \prod_k a_k |\alpha_k\rangle_k \quad (3.74)$$

The individual modes obey the relations [10],

$$a_k |\alpha_k\rangle_k = \alpha_k |\alpha_k\rangle_k \quad (3.75)$$
It is possible to find the eigenvalue $\alpha k$, in terms of the number state $n$. Dropping the subscript, it is intended to find $\alpha$ satisfying [10],

$$a \mid \alpha \rangle = \alpha \mid \alpha \rangle$$  \hspace{1cm} (3.76)

Using the Hermitian adjoint of (3.64), the recursion relation [10],

$$(n + 1)^{1/2} \langle n + 1 \mid > = \alpha \langle n \mid \alpha \rangle$$  \hspace{1cm} (3.77)

for the scalar product $\langle n \mid \alpha \rangle$ is derived. Using (3.68) the scalar product (3.77) may be expressed in terms of the vacuum number state [10],

$$\langle n \mid \alpha \rangle = \frac{\alpha^n}{(n!)^{1/2}} \langle 0 \mid \alpha \rangle$$  \hspace{1cm} (3.78)

The vector $\mid \alpha \rangle$ may be derived from the scalar product as the sum over $u$ [10],

$$\mid \alpha \rangle = \sum_u \mid n \rangle \langle n \mid \alpha \rangle$$

$$= \langle 0 \mid \alpha \rangle \sum_u \frac{n}{(n!)^{1/2}} \mid n \rangle$$  \hspace{1cm} (3.79)

The squared length of $\mid \alpha \rangle$ is

$$\langle \alpha \mid \alpha \rangle = \langle 0 \mid \alpha \rangle^2 \sum_u \frac{\mid \alpha \mid^2 n}{n!}$$

$$= \langle 0 \mid \alpha \rangle^2 e^{\mid \alpha \mid^2}$$  \hspace{1cm} (3.80)

The length of the vector must be normalized to unity, as it represents a probability [46],
Thus

\[
\langle 0 \mid \alpha \rangle = e^{-1/2 |\alpha|^2} \tag{3.81}
\]

Thus

\[
|\alpha\rangle = e^{-1/2 |\alpha|^2} \sum_n \frac{\alpha^n}{(n!)^{1/2}} |n\rangle \tag{3.82}
\]

The average occupation of the n-th state is given by [10],

\[
|\langle n \mid |\alpha\rangle|^2 = \frac{|\alpha|^2 n}{n!} e^{-|\alpha|^2} \tag{3.83}
\]

Which is a poisson distribution with mean value \(|\alpha|^2 [48],[49].

Using (3.78), (3.83) may be written in terms of the vacuum number state, \(|0\rangle\)

\[
\langle \alpha | = \exp[-\frac{1}{2} |\alpha|^2 + \alpha^* a] |0\rangle \tag{3.84}
\]

which is equivalent to [10],

\[
\langle \alpha | = \exp[\alpha^* a - \alpha a^*] |0\rangle \tag{3.85}
\]

Thus \(\exp(\alpha a + -\alpha^* a)\) is a displacement operator [10].

3.2.2 The Principle of Complementarity-Photon Number-Phase Uncertainty

Now referring again to the simple harmonic oscillator of Fig 3.10, the position and momentum may be described by the two sinusoidal equations [8].

\[
x = x_0 \cos(\omega t) + \frac{mV_0}{m\omega} \sin(\omega t) \tag{3.86}
\]
and \( mv = -mx_0 \sin t + mV_0 \cos \omega t \) \hspace{1cm} (3.87)

Where \( x_0 \) is the initial position at time \( t=0 \) and \( V_0 \) is the initial velocity.

Which may be written as

\[
q = q_0 \cos(t) + p_0 \sin(t)
\] (3.88)

and

\[
p = -q_0 \sin(t) + p_0 \cos(t)
\] (3.89)

Where the scales of distance and time are chosen such that the mass and frequency are unity i.e. \( m = \omega = 1 \) [8].

The mean square values of \( p \) and \( q \) may be written as [8],

\[
(\Delta q)^2 = \overline{q^2} - q^{-2}
\] (3.90)

and

\[
(\Delta p)^2 = \overline{p^2} - p^{-2}
\] (3.91)

Thus

\[
(\Delta q)^2 = (\Delta q_0)^2 \cos^2 t + (\Delta p_0)^2 \sin^2 t
\] (3.92)

and

\[
(\Delta p)^2 = (\Delta q_0)^2 \sin^2 t + (\Delta p_0)^2 \cos^2 t
\] (3.93)

since fluctuations in \( p \) and \( q \) are not correlated [8].
The only circumstances under which $\Delta p$ and $\Delta q$ remain constant are when $\Delta p_0 = \Delta q_0$ [8].

Equation (3.88) may be rewritten as a cosinusoid,

$$ q = q_0^2 + p_0^2 \cos [\omega t + \tan^{-1}(p_0/q_0)] $$  \hspace{1cm} (3.94)

Therefore the phase is $\phi = \tan^{-1}(p_0/q_0)$

For small values of initial uncertainty in $p_0$ and $q_0$, the variation in phase may be written using the total differential [50],

$$ \Delta \phi = \frac{q_0 \Delta p_0}{q_0^2 + p_0^2} - \frac{p_0 \Delta q_0}{q_0^2 + p_0^2} $$  \hspace{1cm} (3.95)

Again assuming that fluctuations in $p_0$ and $q_0$ are uncorrelated [8]

$$ (\Delta \phi)^2 = \frac{q_0^2 (\Delta p_0)^2 + p_0^2 (\Delta q_0)^2}{(q_0^2 + p_0^2)} $$  \hspace{1cm} (3.96)

Since $(\Delta p_0)^2 = (\Delta q_0)^2$, \hspace{1cm} (3.97)

$$ (\Delta \phi)^2 = \frac{(\Delta q_0)^2}{q_0^2 + p_0^2} = \frac{(\Delta q_0)^2}{2E} $$  \hspace{1cm} (3.98)

Where $E$ is the total energy of the oscillator.

For large numbers of quanta, the central limit theorem gives the quantum mechanical uncertainty [51],

.
\[ \Delta X \Delta p = \Delta q \Delta p = \frac{\hbar}{2} \]  

Moreover for large numbers of quanta

\[ E = (n + \frac{1}{2}) \hbar w \approx n\hbar w \]  

Using (3.97) and (3.99) \cite{8},

\[ (\Delta qo)^2 = \frac{\hbar}{2} \]  

and using (3.98) \cite{8},

\[ (\Delta \theta)^2 = \frac{\hbar}{4E} \]  

Substituting for \( E \)

\[ \Delta \theta = \frac{1}{2} \sqrt{\frac{\hbar}{E}} = \frac{1}{2\sqrt{n}} \]  

Similarly since \cite{8},

\[ 2E = qo^2 + po^2 \]  

The total differential assuming no condition between \( \Delta po \) and \( \Delta qo \), gives, \cite{51},

\[ (\Delta E)^2 = qo^2 (\Delta qo)^2 + po^2 (\Delta po)^2 \]

\[ (\Delta E)^2 = 2E (\Delta qo)^2 \]

\[ = Eh \]
Thus

$$(\Delta n)^2 = \frac{E}{\bar{n}}$$

and so \[8\],

$$\Delta n = \sqrt{\bar{n}}$$

Combining the two RMS uncertainties of photon number and phase

$$\Delta n \Delta \phi = \frac{\sqrt{\bar{n}}}{2\sqrt{\bar{n}}} = \frac{1}{2}$$

(3.105)

Serber and Townes \[8\] have shown that the phase and photon number uncertainty (3.99) applies to light generated by stimulated emission processes.
3.3 PROPAGATION OF SOUND IN SOLIDS

The theory of the harmonic oscillator, introduced for the description of light in section 3.2, is also applicable to the propagation of sound in a crystalline solid [30],[52].

Crystals are composed of regular arrays of atoms, which form internal planes, as shown in Fig 3.11 [30].

Fig 3.11 Internal Crystal Planes

Directions of atomic planes

Propagation of sound waves within a crystal involves translation of the atomic planes shown in Fig 3.11 as in Fig 3.12 [30].
Within the crystal, the locations of the planes are determined by balances between the internal ionic, covalent and van der Waals forces between the atoms [29],[30]. As a plane is translated from \(X_p\) to \(X_{p+s}\), the plane, as a whole, experiences a restoring force [30],

\[
F_s = \sum_{p=1}^{N_p} C_p (X_{s+p} - X_s)
\]  

(3.106)

Where \(C_p\) is the pth force constant, \(X_s\) is the initial plane position and \(X_{s+p}\) is the new plane position [30].

(3.106) is a result of the contributions of many atoms located throughout the crystal. It is a superposition of individual Hooks, or linear spring, laws [30].

Each atom in the plane has a specific mass which contributes to the mass of the plane. It is possible therefore, to define a force due to the acceleration of the plane.
\[ F_m = \sum_{q=1}^{N_q} M_q \frac{d^2 x_q}{dt^2} \]  

(3.107)

where \( M_q \) is the mass of the \( q \)th atom, \( \frac{d^2 x_q}{dt^2} \) in the acceleration of the plane, and \( F_m \) is the acceleration force of the plane.

The forces of acceleration and the spring must be equal and opposite, therefore,

\[ \frac{d^2 x}{dt^2} + \sum_{p=1}^{N_p} \frac{C_p}{N_q} x = 0 \]  

(3.108)

Equation (3.108) is an equation of simple harmonic motion of natural frequency

\[ \omega = \left[ \frac{N_q \sum_{q=1}^{N_q}}{\sum_{p=1}^{N_p} (C_p/M_q)} \right]^{1/2} \]  

(3.109)

Displacement of crystal planes, as a consequence of sound wave propagation, can therefore be described by an ensemble of Harmonic oscillators. Each oscillator describes an elastic mode.

The energy of a particular elastic mode is described by (3.58) where \( \omega_L \) is replace by \( \omega_s \) [45],

\[ E_s = (n + 1/2) \hbar \omega_s \]  

(3.110)

Where \( E_s \) is the soundwave energy, \( n \) in the quantum number of the mode, and \( \omega_s \) is the angular frequency of sound.
The minimum amount of energy that may be added to or removed from, an elastic mode is \( \Delta E_s \) [45],

\[
\Delta E_s = \hbar \omega_s \tag{3.111}
\]

Which represents one quantum of sound energy \( \Delta E_s \), known as a phonon, and is similar in nature to the photon [30],[45],[52].

When \( n = 0 \) in equation (3.110), the elastic mode energy is \( E_{so} \), and is called the zero point energy [45].

\[
E_{so} = \frac{\hbar \omega_s}{2} \tag{3.112}
\]

(3.112) represents the minimum energy state of the elastic mode.

When analysing real crystals, \( n \) is usually very large hence the quantum mechanical description of the elastic mode approaches the classical description. However the phonon concept is particularly useful when describing the interaction of light and sound [52].
REFERENCES - CHAPTER 3

[1] DEAN P.J.
"Optoelectronics - An Introduction."

[2] RATNER, A.M.
"Spectral, Spatial and Temporal Properties of Lasers."

[3] WILLIAMS, E.W. and HALL, R.
"Luminescence and the Light Emitting Diode."

"Coherent Light."


[7] BROWN, R.
"Lasers - A Survey of their Performance and Applications."

"Limits on Electromagnetic Application Due to Complementarity."
[9] CAVES, C.M.
"Quantum Mechanical Noise in an Interferometer."

[10] GLAUBER, R.J.
"Coherent and Incoherent States of the Radiation Field."

"Equivalence of Semiclassical and Quantum Mechanical Descriptions of Statistical Light Beams."

[12] STOLER, D.
"Equivalence Classes of Minimum Uncertainty Packets."

[13] LOUDON, R.
"The Quantum Theory of Light."

[14] BANCETTI, B and TARTARA, G.
"Equalization and Quality Prediction in Digital Radio Systems."
GTE Telecommunicazioni, Milan, Italy, pp 3 - 40.

[15] OLIVER, B.M.
"Signal-to-Noise Ratios in Photoelectric Mixing."
   "Comments on Noise in Photoelectric Mixing."

[17] HAYKIN, S.
   "Communications Systems." Hetrodyne and homodyne.

[18] STREMIER, F.G.
   "Introduction to Communication Systems."


[20] BOYD, G.D. and GORDON, J.P.
   "Confocal Multimode Resonator for Millimeter Through Optical
   Wavelength Masers."

[21] EISBERG, R.M.
   "Fundamentals of Modern Physics."

[22] EISBERG (as for [21]) p 80.

[23] EISBERG (as for [21]) pp 131 - 134.
"Physics of Solids."

[25] KITTEL, C.
"Introduction to Solid State Physics."

[26] HARVEY (as for [4]) pp 41 - 99.

[27] WILLIAMS & HALL (as for [3]) pp 1 - 38.

[28] WILLIAMS & HALL (as for [3]) pp 72 - 104.


[31] WILSON, J. and HAWKES, J.F.B.
"Optoelectronics - An Introduction."


[33] BROWN (as for [7]) pp 74 - 75.

[34] BROWN (as for [7]) pp 87 - 99.
3.51

"Fields and Waves in Communications Electronics."

[36] RAMO, WHINNERY and VAN DUZER. (as for [35]) pp 659 - 661.

[37] JORDAN, E.C. and BALMAIN, K.G.
"Electromagnetic Waves and Radiating Systems."

[38] GOODMAN, J.W.
"Introduction to Fourier Optics."


[40] WALLS, D.F.
"Squeezed States of Light."

[41] ROBINSON, A.L.
"Bell Labs Generates Squeezed Light."

[42] MARCUSE, D.
"Principles of Quantum Electronics"
[43] LOUDON (as for [13]) pp 120 - 129.

[44] JORDAN & BALMAIN (as for [37]) pp 314 - 315.


[46] YARIV, A.

"Quantum Electronics".
John Wiley & Sons INC., 1967, pp 1 - 34.

[47] ABRAMOWITZ, M. and SIESUN, I.A. (Editors)

"Handbook of Mathematical Tables."

[48] SPEIGEL, M.R.

"Schaum's Outline of Theory and Problems of Statistics in S.I. Units."

[49] KREYSZIG, E.

"Advanced Engineering Mathematics"

[50] ANTON, H.

"Calculus with Analytic Geometry."

[51] KREYSZIG (as for [49]) p 907.
[52] SAPRIEL, J.

"Acousto-Optics."

4. MODULATION OF LIGHT

The transmission of information using electromagnetic radiation is accomplished by changing the characteristics of the transmitted wave in sympathy with the input information.

Classically, an electromagnetic wave travelling through any medium is characterised by 4 parameters;
1. Amplitude
2. Frequency
3. Phase
4. Polarization

By varying any combination of the four parameters, the information may transmitted from one place to another. An electromagnetic carrier is thus modulated; the parameter varied describes the modulation type [1],[2].

In radio circuits the oscillations used to generate the electromagnetic wave eventually radiated from the antenna, are confined to electronic circuits. Using electronic components the oscillations are readily modulated prior to radiation from the antenna. The RF signal is thus directly modulated in the transmitter prior to radiation [1],[2],[3].

Modulation of optical signals is not possible using such direct means. Optical radiation is a consequence of electronic transitions within an atom [4]. The direct methods of modulation used in radio transmitters are not achievable at the atomic level. For this reason, optical radiation must be modulated by either indirectly varying the condition within the atom during emission [5], or by altering the radiation by some means, after emission [5].
In both cases materials which alter their characteristics under external influence must be employed.

When a laser is used as the source of optical radiation, emission conditions may be altered by changing the characteristics of the active material within the laser cavity [5], or the characteristics of the cavity itself [6]. Such modulation techniques are known as intra-cavity modulation techniques.

If the modulation is introduced by methods external to the optical cavity, the technique of modulation belongs to the Extra-cavity methods of modulation [6].

It is proposed to examine the characteristics of materials suitable for both intra-cavity and extra-cavity modulation in the following sections.

Specific techniques of intra cavity and extra cavity modulation will be introduced in sections 4.2 and 4.3. Particular emphasis is placed on acoustic-optic (AO) methods of modulation, [7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19] as AO methods were used extensively in experimentation.

Finally using a combination of theoretical predictions and experimental results, limitations on modulation bandwidth are examined in section 4.4. [7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19].
4.1 OPTICAL MODULATOR MATERIALS

Many materials in all three material phases exhibit optical properties that may be influenced externally [4],[8].

Materials which change their optical properties under the influences of electric or magnetic fields are of great interest as potential optical modulator materials [6],[9],[20]. Crystals in which light can interact with travelling sound waves are also of interest [7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19].

In the following 3 subsections the electro-optic (EO), magneto-optic (MO) and (AO) properties of material are discussed. Application of these properties to specific light modulators is detailed in sections 4.2 and 4.3. Production of angle modulation is emphasised as it is of greatest interest for coherent communications systems.

4.1.1 Electro-Optic Properties of Materials

As light propagates along an optical axis in the Z direction, as shown in Fig 4.1, the refractive index encountered is independent of the light polarization, in general [21].

Fig 4.1 Propogation of Light Through a Crystal
However when some materials are placed in an electric field, the refractive indices of the material change such that the refractive index for light polarized along the x axis differs from that for light polarized along the y axis. Therefore the speed of propagation of light becomes dependent, upon its polarization. One axis becomes the slow propagation axis, the other becomes the fast axis [21]. This effect is known as bi-refringence.

As a consequence of the bi-refringence of the material, it is possible to define a relative phase retardation of transmitted light, between the fast and slow axes [21]

\[
\Gamma = \frac{2\pi L (n_s - n_f)}{\lambda_L} = \frac{2\pi L \Delta n}{\lambda_L} \tag{4.1}
\]

Where \( \Gamma \) is the relative phase retardation, \( L \) is the effective transmission length through the crystal and \( n_s \) and \( n_f \) are the slow and fast axis indices of refraction respectively.

\[
\Delta n = n_s - n_f
\]

The dependence of the refractive index change on the applied electric field may be either linear or quadratic. Linear dependence of the refractive index change is due to the Pöckles effect [22], quadratic dependence is a consequence of the electro-optic Kerr effect [22].

The electro optic Kerr effect is described by the empirical relation [22].

\[
\Delta n = 2B \lambda_0 E_m^2 \tag{4.2}
\]

Where \( B \) is Kerr's Constant, \( \lambda_0 \) is the wavelength of light within the material and \( E_m \) is the magnitude of the electric field.
If the electric field is modulated, harmonics of the modulating signal will be generated as a consequence of (4.2) [22]. The introduction of such harmonics is generally undesirable, it can be avoided by superposing the modulation field on a large bias field [22], in which case,

$$E_m^2 \leq E_0 E_m < E_0^2$$

thus

$$\Delta n = 2B \lambda_0 (E_0^2 + 2E_0E_m)$$

(4.3)

where $E_0$ is the bias field.

The relative phase retardation $\Gamma$ between $x$ and $y$ polarization becomes

$$\Gamma = \frac{2\pi (\Delta n_0 + \Delta n_m)}{\lambda_0}$$

(4.4)

$$= \Gamma_0 + \Gamma_m$$

Where $\Gamma_0$ is the relative phase retardation due to the bias field and $\Gamma_m$ the relative phase retardation due to modulation.

Changes in optical signal due to modulation are primarily of interest, thus,

$$\Delta n_m = 4B \lambda_0 E_0 E_m$$

(4.5)

The Pöckles effect exhibits a linear dependence of refractive index change, on the electric field [22], thus

$$\Delta n_m = 2P \lambda_0 E_m$$

(4.6)

Where $P$ is a constant of proportionality.
The refractive index may be specified as a function of polarization angle, or as a function of polarization component along the x and y axes, using the refractive index ellipsoid [21].

The index ellipsoid for a specific field strength and direction is described by [21],

\[
\frac{x^2}{n_x^2} + \frac{y^2}{n_y^2} + \frac{z^2}{n_z^2} = 1
\]  \hspace{1cm} (4.7)

Where \( x, y \) and \( z \) are polarization direction cosines and \( n_x, n_y, n_z \) are the refractive indices for light polarized in the x, y and z directions respectively.

In general, the strength and direction of the electric field will effect the index ellipsoid to some extent in all directions, therefore in terms of the electric field [21],

\[
A_{11}X^2 + A_{22}Y^2 + A_{33}Z^2 + 2A_{23}YZ + 2A_{31}ZX + 2A_{12}XY = 1
\]  \hspace{1cm} (4.8)

Where \[ A_{11} - \frac{1}{N_x} = r_{11} \, Ex + r_{12} \, Ey + r_{13} \, Ez \]
\[ A_{22} - \frac{1}{N_y} = r_{21} \, Ex + r_{22} \, Ey + r_{23} \, Ez \]
\[ A_{33} - \frac{1}{N_z} = r_{31} \, Ex + r_{32} \, Ey + r_{33} \, Ez \]
\[ A_{23} = r_{41} \, Ex + r_{42} \, Ey + r_{43} \, Ez \]
\[ A_{31} = r_{51} \, Ex + r_{52} \, Ey + r_{53} \, Ez \]

and \[ A_{12} = r_{61} \, Ex + r_{62} \, Ey + r_{63} \, Ez \]
The $r_{ij}$ are proportionality constants relating a particular $i$th axis field component to a change in refractive index along the $j$th axis. $E_x$, $E_y$ and $E_z$ are the respective $X$, $Y$ and $Z$ axis components of the applied electric field.

The linearized Kerr effect (4.5) is similar [22], the electric field components become,

$$E_x = E_{ox} E_{mx} \quad E_y = E_{oy} E_{my} \quad E_z = E_{oz} E_{mz} \quad (4.9)$$

Equation 4.8 is generally summarized in matrix form, in which a $6 \times 6$ matrix $R$, with elements $r_{ij}$ is formed. Matrix $R$ is the EO matrix for a specific material, its values are fixed by the crystal characteristics alone [21].

Using the $R$ matrix for a specific material an EO modulator, exploiting induced bi-refringence, may be designed.

4.1.2 Magneto-Optic Modulator Materials

Magneto-Optic materials exhibit three optical effects that may be used to modulate a light beam.

The Cotton-Mouton or Voigt effect [22] is analogous to the EO Kerr effect. Thus, it is possible to define a relative phase delay between light polarized along the slow and fast axes,
\[ \frac{2 \pi L \Delta n}{\lambda} = C B_0^2 \]  

(4.10)

where \( C \) is the Cotton-Mouton Constant.

The analysis of the Cotton-Mouton phenomenon is similar to that of the electro-optic Kerr effect [22]. The magnetic vector \( \mathbf{B} \) replaces the electric vector in the relations of section 4.1.1.

Unfortunately the Cotton-Mouton effect is much smaller than the EO Kerr effect [22]. Large modulation powers are required to modulate the light beam, and therefore the Cotton Mouton effect is less attractive as a means of optical modulation than the electro-optic Kerr effect.

The strongest and most suitable effect for MO modulation is Faraday polarization rotation [22]. As a linearly polarized light beam travels through the material its polarization is rotated by \( \theta_F \) radians. Application of a magnetic field with a component in the direction of light propagation alters the angle of Faraday rotation, thus [22]

\[ \theta_F = V B_0 \cdot \hat{1} \]  

(4.11)

Where \( B_0 \) is the applied magnetic field, \( \hat{1} \) is a unit vector in the direction of light propagation and \( V \) is the Verdet constant.

Unfortunately the speed at which \( \theta_F \) can be changed is limited in some materials by an atomic relaxation process [22]. This restricts the maximum possible operating frequencies of MO modulators exploiting Faraday rotation. Some materials such as sodium vapour or Yttrium Iron Garnet
(YIG) will operate well into the visible spectrum with modulation frequencies of units of Gigahertz [22].

The MO Kerr effect may also be used to modulate light beams. By application of a suitable magnetic field, the state of polarization of light reflected from a metallic mirror may be altered. The actual polarization change is dependent not only on the strength and direction of the applied magnetic field, but also on the direction of light polarization at the surface of the mirror [22]. Fig. 4.2 shows three possible cases which occur [22].

Fig 4.2 The Three Cases of Magneto Optic Kerr Effect.

Mathematically, the magneto optic Kerr effect is described by the tensor relation [22].

\[
\begin{bmatrix}
E_{rx} \\
E_{ry}
\end{bmatrix}
= \begin{bmatrix}
P_{11} & P_{12} \\
P_{21} & P_{22}
\end{bmatrix}
\begin{bmatrix}
E_{ix} \\
E_{iy}
\end{bmatrix}
\]

(4.12)

Where \( P_{ij} \) are the Fresnel reflectance components and \( P_{ij} \) and \( i \neq j \) are due to MO effects.
To modulate the light beam, the magnetic field is varied causing $P_{12}$ and $P_{21}$ to change with time, whence the polarization of reflected light changes with time.

4.1.3 Acousto Optic Modulator Materials

EO and MO modulators exploit optical properties of materials which change under the influence of electric or magnetic fields. Such materials are particularly useful for the modulation of light, as will become evident in sections 4.2 and 4.3. These are however not the only properties of materials that may be exploited for the modulation of light.

Certain materials, such as Lead Molybdate, exhibit properties that may be used to modulate light when they are placed in a sonic field [10],[11],[12],[13],[14],[15],[16],[17],[18],[19],[23]. In such materials light and sound waves interact exchanging energy and momentum.

Sound may propagate through a material in essentially one of two ways, either through the bulk material or at the material surface [7]. With careful selection of the material, selection of suitable mounting of transducers and absorbers, the bulk of the sound energy may be concentrated near the material surface, in which case the sound travels as Surface Acoustic Waves (SAW) [7], or within the material as a planar bulk wave-front in which case the sound travels as a bulk wave within the whole of the crystal [7],[19].

Both Surface Acoustic waves and Bulk waves may be used to modulate incident light beams.
Surface Acoustic waves are launched by the transducer arrangement shown in Fig 4.3 [7],[19].

**Fig 4.3 Launching of Surface Acoustic Waves.**

The presence of the surface Acoustic wave produces surface ripples in the material much like waves on the surface of water. The depth of the penetration into the crystal is the order of the wavelength of visible light $\lambda$ [7],[19].

Within the region of the surface to a depth of $\lambda$ the refractive index of the material varies with the compressions and rarefactions produced by the sound wave [7],[19]. Fig 4.4 shows the nature of the effect of soundwaves on the surface of the material [7].
It is clear from Fig 4.4, that when an optical beam of width $d$ is incident upon the surface wavefront, light travelling along the ray paths, making up the beam, are incident upon the crystal at angles dependent on their position within the light beam and the position of the peaks and troughs of the surface ripple [7],[19].

Most of the light is neither reflected nor refracted and passes through in the simple 0 order direction given by Descartes laws, as shown in Fig 4.5 [5],[6].

\[ \sin \theta_{1r} = \sin \theta_0 \pm \frac{n_r \lambda_L}{n_0 \lambda_S} \]  

(4.13)

Where $\theta_{1r}$ is the angle of the first maximum, $\theta_0$ is the angle of incidence, $n_0$ is the refractive index of the first material encountered by the incident light ray and $n_r$ the refractive index of the light beam. $\lambda_L$ and $\lambda_S$ are the respective optical and sonic wavelengths.
The directions of the first order transmitted light are given by [7], [19],
\[
\sin(\theta_{1t}) = \sin(\theta_0) \pm \frac{n_t \lambda_l}{N_0 \lambda_s}
\] (4.14)

Where \( \theta_{1t} \) is the angle of transmitted maxima and \( n_t \) is the refractive index of the material in which the transmitted light beam propagates.

The proportion of light in the first reflected order, when the incident refractive index is unity, is given by [7],
\[
\eta_r = R^2 j^2 \frac{4\pi UN_0 \cos(\theta_0) \cos(\theta_{1r})}{\lambda} \frac{\cos(\theta_{1r})}{\cos(\theta_0)}
\] (4.15)

Where \( R \) is the surface reflectivity and is dependent upon \( \theta_0 \) according to Fresnel's Law [24], \( J \) is a Bessel function of first kind, and \( U \) is the amplitude of the surface wave.

The corresponding proportion of power diffracted into the first refracted order is [7],
\[
\eta_t = (1+R^2) j^2 \frac{2\pi UN_0}{\lambda} \left\{ \cos(\theta_0) - \cos(\theta_{1t}) \right\} \frac{\cos(\theta_{1t})}{n_1 \cos(\theta_0)}
\] (4.16)

The variations in refractive index in the active depth of the surface acoustic wave, have the same effect on light as the refractive index variations due to bulk sonic wave propagation; the details of which follows [7].

The optical properties of bulk acoustic waves and light interaction were used in much of the optical experimentation undertaken. Therefore the
theoretical aspects of light and sound interactions are examined in detail; drawing on the quantum mechanical basis introduced in chapter 3.

Consider a transparent crystal of refractive index $n$, in which a travelling sound wave has been launched. As the sound wave travels with a speed, $V_s$, and is of angular frequency, $\omega_s$, a light beam of angular frequency, $\omega_l$, and width, $d_l$, is directed at the crystal, at an angle $\theta$ to the horizontal, as shown in Fig 4.6 [7], [19].

The travelling bulk wave superimposes a variation in refractive index upon the natural refractive index of the material. The refractive index variation is described by [7],

$$\delta n = \delta n_0 \sin (\omega_s t - k_s \cdot r_s)$$

(4.17)

where $k_s$ is the sonic wave vector, $r_s$ is the direction of sound propagation, $\omega_s$ is the angular frequency of the sound wave and $t$ is time.

The travelling wave of refractive index variation forms a travelling phase-diffraction grating with spacing [7],
\[ \lambda_s = \frac{2 \pi V_s}{\omega_s} \]  

(4.18)

Where \( V_s \) is the propagation speed of sound in the crystal.

As mentioned in chapter 3 the sound wave is quantized into phonons [7],[8], which scatter photons of the incident light beam elastically, therefore energy and momentum are simultaneously conserved [7],[8]. If a photon encounters a large number of phonons it will absorb the energy of \( m_p \) photons, whence

\[ E_{1}' = E_1 + m_p E_s \]  

(4.19)

Where \( E_1 \) and \( E_1' \) are the energies of the photon before and after collision, respectively \( E_s \) is the energy of a phonon and \( m \) is the net number of phonons absorbed by the photon during its passage through the crystal.

The incident photon may also, either generate \( m_p \) new phonons by recoil processes [8], in which case \( m_p \) is negative, or it may pass through the medium with no net interactions [8], in which case \( m_p = 0 \).

The energies of (4.19) are related to the frequencies of optic and sonic radiation [7],

\[ \hbar \omega_{1}' = \hbar \omega_1 + m \hbar \omega_s \]  

(4.20)

from which the frequency of light, emerging from the crystal, is found [7],

\[ \omega_{1}' = \omega_1 + m \omega_s \]  

(4.21)
Moreover, the conservation of momentum requires that
\[
p_1 + m_p s = p_1' + m_p s
\]  
(4.22)
but the energy of $m_p$ photons is absorbed, therefore $p_s' = 0$.

The De Broglie relation, relates the photon momentum to the wave vector:
\[
\frac{2\pi}{\lambda} = k_l = \frac{2\pi p}{h} = \frac{p}{\hbar}
\]  
(4.23)

Using (4.23), the photon and phonon vectors are related by
\[
k_l + mk_s = k_l'
\]  
(4.24)
where $k_l$ and $k_l'$ are the respective incident and scattered wave vectors, of light, in the medium.

The vector diagram for the one photon $m$ phonon collision is shown in Fig. 4.7 [7].

**Fig 4.7 Vector Diagram of Photon Phonon Interactions.**
Using the vectors of Fig 4.7, an expression for the angle of the diffracted beam, from the horizontal, as a function of the input beam angle and number of phonons absorbed per photon, may be derived [7].

\[
\theta' = \tan^{-1}\left[\frac{m_k}{k_1 \cos \theta} + \tan \theta\right] = \tan^{-1}\left[\frac{m\lambda_1}{\lambda_S n \cos \theta} + \tan \theta\right] \tag{4.25}
\]

Where \( \theta \) and \( \theta' \) are the angles of incident and scattered rays respectively.

When the ratio \( \frac{\lambda_1}{n\lambda_S} \geq 1 \), a number of beams emerge from the crystal. The photon-phonon interactions are of the type described by Raman and Nath [7], [8],[9].

Each emerging beam is described by a particular order given by \( m_p \), which varies from \(-\infty \) to \( \infty \). The angle for each order is found from (4.24), and the frequency of each order is given by (4.21).

If the ratio \( \lambda_1/n\lambda_S \cos \theta \ll \tan \theta \) in equation (4.25), variations in \( m \) have little effect on the deflection angle, and \( \theta' \approx \theta \). In this case the output wave is said to be Bragg diffracted [7]. The vector triangle of Fig 4.7 is isocoles, the optical beam is specularly reflected from sonar planar wavefronts, thus,

\[
k_1 \sin \theta = k_1 \sin \theta = \frac{k_S}{2} = \frac{\pi}{\lambda_S} \tag{4.26}
\]

Therefore \( \theta \) becomes the bragg angle \( \theta_B \) [7].

\[
\theta_B = \sin^{-1}\left(\frac{\lambda_1}{2n\lambda_S}\right) = \sin^{-1}\left(\frac{\lambda\omega_S}{4\pi n\nu_S}\right) \tag{4.27}
\]
Which, for small angles, is,

\[ \theta_B = \frac{\lambda_1 \omega_s}{4\pi V_s n} \]  

(4.28)

At an incident angle \( \theta_B \), the proportion of incident optical power diffracted by Bragg waves is \( \eta_B \) [7], [11],

\[ \eta_B = \sin^2 \frac{\nu_B}{2} \]  

(4.29)

Where \( \nu_B = k_1 \delta n \omega L \) and \( L_s \) is the width of the acoustic beam.

At the Bragg angle \( \eta \) is a maximum all light rays emerging after crossing the acoustic beam are in phase; their effects positively interfere [7], [11].

If a light beam is aligned with the Bragg angle, at a particular sonic frequency \( \omega_{0s} \), the light beam is deflected in accordance with (4.28),

\[ \theta_B = \frac{\lambda_1 \omega_s}{4\pi V_s n} \]

If the acoustic frequency is raised to \( \omega_s + \Delta \omega_s \), without modifying the incident angle of the light beam, the Bragg angle change is given by [7], [11], [17].

\[ \Delta \theta_B = \frac{\lambda_1 \Delta \omega_s}{4\pi V_s n} \]  

(4.30)

The angle of incidence is now \( \Delta \theta_B \) smaller than the Bragg angle at \( \omega_s + \Delta \omega_s \), the angle of diffraction is greater than the Bragg angle at \( \omega_s + \Delta \omega_s \) by \( \Delta \theta_B \), as shown in Fig 4.8.
Using Fig 4.8, it is possible to define a path length difference between any two diffracted rays intersecting the acoustic wavefront, at points $L_1$ and $L_2$, about the centre ray [7],[11],[17],

$$Z = \frac{(L_2 - L_1)}{2} \cos (\theta_B - \Delta \theta_B) - \frac{(L_2 - L_1)}{2} \cos (\theta_B + \Delta \theta_B) \quad (4.31)$$

For a path length $\Delta Z < \lambda$, the path length difference stated as a phase delay is [7],[11],[17],

$$\varphi = \frac{\pi}{\lambda} (L_2 - L_1) \cos (\theta_B - \Delta \theta_B) - \frac{\pi}{\lambda} (L_2 - L_1) \cos (\theta_B + \Delta \theta_B) \quad (4.32)$$

By expanding the cosine terms of (4.32) and using (4.27) the phase delay becomes [7],[11],[17].

$$\varphi = \frac{(L_2 - L_1) \omega s \sin \Delta \theta_B}{2n V_s} \quad (4.33)$$

$$\approx \frac{(L_2 - L_1) \omega s \Delta \theta_B}{2n V_s}$$
Any two rays intersecting with the plane wave front and separated by \( L \) will destructively interfere giving an intensity minimum, when \( \varphi = \pi \) [7], [11], [17],

whence

\[
\Delta \theta_{B_{\text{min}}} = \frac{2\pi n v_s}{L \omega_s} \tag{4.34}
\]

Similarly a 3dB reduction in intensity occurs when \( \varphi = \pm \pi/2 \), hence

\[
\Delta \theta_{B_{-3dB}} = \frac{\pi n v_s}{L \omega_s} \tag{4.35}
\]

The corresponding -3dB frequency is found using (4.30) [7], [11], [17],

\[
\Delta \omega_{s_{-3dB}} = \frac{4\pi^2 n^2 v_s^2}{\lambda_1 L \omega_s} \tag{4.36}
\]

\( 2 |\Delta \omega_{s_{-3dB}}| \) yields the -3dB bandwidth.
4.2 EXTRA-CAVITY LASER MODULATION

In this section, it is proposed to discuss how the properties of materials, given in section 4.1, are used to modulate light, generated by a Laser source, when the modulator is placed at some distance away from the laser cavity.

In the following section (4.2.1), modulation using the magneto optic and electro optic properties of materials to produce both amplitude and angle modulated signals are discussed. Particular emphasis is given to the production of digital type angle modulation which is of greatest interest in modern commercial transmission applications.

Extra cavity AO modulation is examined in detail in section 4.2.2, as it was used extensively in experimentation. The AO modulator was chosen, for modulation experiments, because such modulators are simple to use and cheap to produce. Moreover with careful construction it should be possible to produce modulation over very wide bandwidths which may make AO modulators commercially attractive.

Using the AO modulator it is possible to produce Amplitude and Angle modulation, the details of which will be covered in section 4.2.2.

In addition to the bandwidth limitations discussed in section 4.1.3, the combination of beam alignment, angle of deflection and aperture size fix the modulation bandwidth of the system. These bandwidth limitations are discussed in section 4.2.2.
4.2.1 Electro-Optic and Magneto-Optic Modulation

By exploiting the field induced bi-refringence of electro optic and magneto optic materials, amplitude and angle modulation may be produced [4],[5],[6],[21].

Amplitude modulation is produced using a light value formed by placing the bi-refringent material between two polarizers as shown in Fig 4.9 [6].

![Figure 4.9 A Bi-refringent Amplitude Modulator](image)

The polarization axes, of the two polarizers, lie, either parallel or orthogonal, to each other and at 45° to the fast and slow axes of the bi-refringent crystal [21]. Light is incident upon the first polarizer and emerges at 45° to the x and y axes shown in Fig 4.8. The linearly polarized light may be separated into orthogonally polarized components lying along the fast and slow axes of the crystal [6]. Depending on the field strength and direction, one light component is delayed with respect to the other by a phase angle \( \varphi(E) \). The x and y components are written, [21]

\[
E_x = \frac{A}{\sqrt{2}} \cos (\omega t - \varphi_x(F)) \tag{4.37}
\]

\[
E_y = \frac{A}{\sqrt{2}} \cos (\omega t - \varphi_y(F)) \tag{4.38}
\]
Light emerging from the crystal is incident upon the second polarizer, a component of each signal $E_x$ and $E_y$ lies along the polarization axis of that polarizer, which is either parallel to the polarization axis of the first polarizer, as shown in Fig 4.10 (a), (b) or orthogonal to the polarization axis of the first polarizer, as shown in Fig 4.10 (a), (b).

![Fig 4.10 Light Emerging from Electro Optic Modulator](image)

(a) Parallel Polarizers
(b) Orthogonal Polarizers

The field lying along the polarization axis of the second polarizer, Fig 4.10 (a) is,

$$E_p = \frac{A}{2} \cos(\omega t - \varphi_x(E(t))) + \frac{A}{2} \cos(\omega t - \varphi_y(E(t)))$$

$$= A \cos[\varphi_x(E(t)) - \varphi_y(E(t))] \cos[\varphi_x(E(t)) + \varphi_y(E(t))] \cos\omega t$$

$$+ A \sin[\varphi_x(E(t)) - \varphi_y(E(t))] \cos[\varphi_x(E(t)) + \varphi_y(E(t))] \sin\omega t$$

The field lying along the polarization axis of the second polarizer, Fig 4.10 (b) is given by
\[ E_p = \frac{A}{2} \cos (\omega t - \phi_x(E(t))) + \frac{A}{2} \cos (\omega t - \phi_y(E(t))) \]  
\[ = A \cos[\phi_x(E(t)) - \phi_y(E(t))] \cos[\phi_x(E(t)) + \phi_y(E(t))] \cos \omega t \]
\[ + A \sin[\phi_x(E(t)) - \phi_y(E(t))] \cos[\phi_x(E(t)) + \phi_y(E(t))] \sin \omega t \]

A maximum of (4.39) occurs when \( \phi_x(E(t)) + \phi_y(E(t)) = 0 \) and a minimum at \( \phi_x(E(t)) + \phi_y(E(t)) = \pi \), similarly a minimum of (4.40) occurs when \( \phi_x(E(t)) - \phi_y(E(t)) = \pi \) and a maximum at \( \phi_x(E(t)) - \phi_y(E(t)) = 0 \). Amplitude modulation of light, using the light value is achieved by varying the amplitude of the electric field, between values giving the maximum and minimum amplitude phase delays \( \Delta \phi - \phi_x(E(t)) + \phi_y(E(t)) \), given above.

Using the same bi-refringence properties of crystals, light may be phase modulated. A simple phase modulator is constructed using a polarizer and bi-refringent crystal as shown in Fig 4.11 [6].

**Fig 4.11 Bi-refringent Phase Modulator**
The polarizer's direction of polarization is aligned with the slow axis of the crystal. By altering the transverse field, the phase of output light is changed by an amount given by (4.4) [6],

\[ \Delta \phi = \frac{2\pi l}{\lambda_0} (n_E - n_0) = \frac{2\pi l \Delta n_0}{\lambda_0} \]

Retardation as a consequence of the linearized EO Kerr effect is given by (4.5) [22], whence,

\[ \Delta \phi = 4\pi B_0 E_m(t) \quad (4.41) \]

Where \( E_m(t) \) is the time dependent modulation.

Similarly the EO Pöckles effect is summarized by (4.6) [22], whence

\[ \Delta \phi = 2\pi l \lambda_0 E_m(t) \quad (4.42) \]

Frequency modulation is readily produced by integrating the time varying phase thus [12],

\[ \Delta \omega = \int 4 B_0 E_m(t) dt = 4\pi B_0 \int E_m(t) dt \quad (4.43) \]

and the Pöckles effect,

\[ \Delta \omega = 2\pi l \int E_m(t) dt \quad (4.44) \]
Phase and frequency modulation is produced, using the MO Kerr effect, in a similar manner, the Kerr constant in (4.41) and (4.43) is replaced by the Cotton-Mouton constant "C", and the time varying electric field $E(t)$, by the time varying magnetic field $B(t)$. Modulation bandwidth and depth may be increased over that for 1 crystal by using several crystals [6].

To achieve modulation of frequencies above 100 MHz, travelling wave versions of the light value amplitude modulator and the phased modulator are constructed [6]. With careful matching, the electric vector, of the electromagnetic wave, may be launched, into a microwave cavity, with the incident light beam, and travel through a number of crystals, with the light signal, as shown in Fig 4.12 [6].

Naturally, the phase velocities of the light and microwaves must be equal [6].

MO Faraday rotation may also be exploited to modulate light. The angle of rotation is related to the field by (4.11)

$$\theta_F = V B(t) \cdot \hat{\imath}$$  \hspace{1cm} (4.45)

Where the line varying magnetic field $B(T)$ replaces $B_0$. 

\begin{center}
\textbf{Fig 4.12 Travelling Wave Microwave Modulation}
\end{center}
Using two polarizers, as shown in Fig 4.13, and the magneto optic crystal, it is possible to construct a light valve [22].

**Fig 4.13 Amplitude Modulation using Faraday Rotation**

The polarizer, polarization axes, are orthogonal to each other and no light passes through the valve, as the magnetic field is increased, the linearly polarized light is rotated through an angle, $\theta_F$, given by (4.45) [22]. The intensity of output light is thus varied with the magnetic field strength, and is described by

$$S(t) = E_1 \sin (\mathbf{V}\mathbf{B}(t) \cdot \mathbf{\hat{i}})$$ (4.46)

The modulation bandwidth is limited by the dipole relaxation period [22]. However if a large bias field is applied and the modulating field superimposed upon this, the relaxation periods may be reduced and modulation bandwidth increased. In this case the amplitude modulated signal is described by

$$S(t) = E_1 \sin [\mathbf{V}(\mathbf{B}(t) + \mathbf{B}_0) \cdot \mathbf{\hat{i}}]$$ (4.47)
Modulation depth is dependent upon the sensitivity, of the Faraday rotation, to the magnetic field strength. Maximum modulation depth is obtained when the angle of rotation varies with the magnetic field over 90°.

Faraday rotation can also be used to generate phase modulation. Two polarizers, a bi-refringent crystal and the magneto optic crystal are needed, to construct the phase modulator shown in Fig 4.14.

**Fig 4.14 Magneto Optic Phase Modulation using Faraday Rotation**

Light enters the first polarizer and is polarized at 45° to the fast and slow bi-refringent crystal axes. The bi-refringent crystal length is chosen to produce a relative phase delay of $\pi/2$ radians [21]. The light emerging from the bi-refringent crystal is circularly polarized. Faraday rotation of the magneto optic crystal produces a variable forward or reverse rotation of the circularly rotating field, either advancing or retarding, its rotation.

A second polarizer at 45° to the horizontal is placed after the magneto optic crystal. Light emerging from this polarizer is phase delayed and linearly polarized. (For proof see appendix 23). The light is thus phase modulated,
Where \( \varphi(t) \) is the time varying phase delay introduced by the magneto optic crystal.

Introducing the field variation

\[
E_1(t) = E_1 \cos(\omega t - \varphi(t))
\]  \hspace{1cm} (4.48)

Again, frequency modulation may be produced by integrating the phase with respect to time [1],[2], thus,

\[
E_1(t) = E_1 \cos(\omega t - V \int B(t) \cdot \hat{\nabla} dt)
\]  \hspace{1cm} (4.49)

Using the building blocks outlined, variations in modulator design may be produced, which improve the achievable frequency response and modulation depth.

Naturally, it is possible to produce travelling wave versions of the MO modulators discussed [6]. By selecting appropriate modes and matching guide phase velocities, light may be modulated at very high microwave frequencies.
4.2.2 **Acousto-Optic Modulation**

The interaction of light and sound is attractive as a modulation technique, because it is simple, requiring one crystal connected to a piezo-electric transducer and an absorber of sound waves. The crystal and piezo-electric transducer are cheaper than EO and MO crystals, used to build such modulators.

Alignment of the modulator is straightforward [23]; once aligned, a variety of modulation types may be produced without modifying the modulator.

Moreover AO devices are being used in high speed signal processing applications, hence the technology is already well known [7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19],[23]. For this reason acousto-optics will possibly become the most commercially attractive modulation technique, particularly for frequency and phase modulation.

Light may be amplitude modulated by amplitude modulating the sound wave travelling in the AO crystal. The crystal is aligned, such that the light encounters the plane sound waves at the Bragg angle, given by (4.28).

\[ \theta_B = \frac{\lambda L}{\frac{\omega_s}{\frac{4\pi V_s}{N}}} \]

The proportion of power deflected by the sound waves is given by (4.29), whence [7],[17],

\[ \eta = \sin^2 \left( \frac{K_1 K_{AO} A(t) L}{2} \right) \quad (4.51) \]
Where $K_{AO}$ is an acoustic coupling constant relating the change in refractive index to the amplitude of the electrical input signal and $A(t)$ is the modulating signal.

If a detector is placed at the Bragg angle, facing the modulator, as shown in Fig 4.15, the intensity of light measured by the detector is dependent upon the power of the sonic wave [27].

**Fig 4.15 Detection of Deflected Light**

When the sonic power is very low no light is deflected into the detector, the argument of (4.29) is zero. When the argument of (4.51) reaches $\pi/2$, all the incident optical power is deflected into the detector [7],[8].
As the modulation frequency of the amplitude modulated sonic wave is raised, the deflected beam shown in Fig 4.15 splits into 3 beams as shown in Fig 4.16 [7],[8].

Fig 4.16 Splitting of the Diffracted Optical Beam
Moreover, the undeflected beam also splits into 3 parts. Thus 6 independent light beams emerge from the AO modulator at differing angles. As the frequency is further increased, the angles between the outer two beams, of each 3 beam ensemble, increases, the central beam, in the zeroth and first order deflection angles, is unaffected. This beam is approximately twice as bright as the outer beams.

The AO modulator theory offers a useful explanation of the behaviour of the six beams. An amplitude modulated wave is composed of independent waves of differing frequency given by the Fourier transform of the AM wave. These are generally described as carrier, upper sideband and lower sideband.

An amplitude modulated wave is written, in the time domain, as, [1], [2]

\[ S(t) = A_1 [1 + \mu m(t)] \cos \omega_{sc} t \]  

(4.52)

Where \( m(t) \) is the modulating wave, which varies over the interval \([-1, 1]\), \( \mu \) is the modulation depth parameter, and \( \omega_{sc} \) is the sonic carrier angular frequency.

If \( m(t) \) is a cosinusoid, the Fourier transform reveals that there are 3 separate sonic signals travelling within the crystal, an upper sideband at \( \omega_{sc} + \omega_{sm} \), a lower sideband at \( \omega_{sc} - \omega_{sm} \) and a carrier at \( \omega_{sc} \). Each wave consists of phonons of energy given by (3.110) and momenta given by (4.23).

The photons of light entering the crystal interact with the phonons of the sound waves. As a result, six light beams are generated by single, and selected double phonon-photon interactions [7],[17].
The frequencies of photons emerging from the crystal are found using the conservation of energy \([7],[9],[17]\). The angles at which the beams emerge from the modulator are found from the conservation of momentum, between particles participating in collision \([7],[8],[9],[17]\).

For cosinusoidal modulation, of angular frequency \(\omega_{sm}\), light beams emerging from the AO modulator have frequencies given in Table 4.1.

Table 4.1 Frequencies of Modulator Optical Beams

<table>
<thead>
<tr>
<th>Diffraction Order (Refer to Fig 4.5)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Sideband (\text{LSB}_0)</td>
<td>(\omega_1 - \omega_{sm})</td>
</tr>
<tr>
<td>Carrier (\text{C0})</td>
<td>(\omega_1)</td>
</tr>
<tr>
<td>Upper Sideband (\text{USB}_0)</td>
<td>(\omega_1 + \omega_{sm})</td>
</tr>
<tr>
<td>Lower Sideband (\text{LSB}_1)</td>
<td>(\omega_1 + \omega_s + \omega_{sm})</td>
</tr>
<tr>
<td>Carrier (\text{C1})</td>
<td>(\omega_1 + \omega_s)</td>
</tr>
<tr>
<td>Upper Sideband (\text{USB}_1)</td>
<td>(\omega_1 + \omega_s + \omega_{sm})</td>
</tr>
</tbody>
</table>

The angles, at which the diffracted beams emerge from the modulator, are given by (4.25) with \(m_p = 1\) \([7],[8],[9],[17]\),

\[
\theta_1' = \tan^{-1} \left( \frac{C \omega_s}{nV_s \omega_1 \cos} + \tan \theta \right) \quad (4.53)
\]

where \(\theta_1'\) denotes the emergence angle of the first order beams, and \(\omega_s\) denotes the frequency of sound corresponding to either \(\text{LSB}_1\), \(\text{C1}\) or \(\text{USB}_1\).
Similarly, the angles of the zero order beam are given by [7],[8],[9],[17],

\[
\theta_0' = \tan^{-1} \left( \frac{c \omega_s}{nV_s \omega_1 \cos \theta} \right) - \tan \theta \tag{4.54}
\]

Where \( \theta_0' \) denotes the emergence angles of the zero order beams and \( \omega_s \) denotes the frequencies corresponding to either LSB_0, Co or USB_0.

For optimum operation the acousto optic crystal is aligned such that the incoming light is incident at the Bragg angle, which is given by (4.28) [7],

\[
\theta_B = \sin^{-1} \left( \frac{\lambda_1 \omega_s}{4 \nu_s n} \right)
\]

The first order optical carrier, C_1, emerges at the Bragg angle, the zero order carrier, C_0, emerges at the negative Bragg angle, -\( \theta_B \).

Light deflected, by the upper and lower sideband sonic signals, is not incident at the Bragg angle for the associated sideband frequencies. The incident light angle is constant, therefore at the modulation frequency rises, the intensity of the sideband falls by the mechanisms outlined in section 4.1.3 [7]. When the intensity of the sidebands falls by 3dB, the maximum modulation frequency has been reached. It is then possible to define a maximum modulation frequency, \( \omega_m \), using (4.36),

\[
\omega_{sm} = \frac{4\pi^2 n^2 \nu_s^2}{\lambda_1 \nu_s \omega_s} \tag{4.55}
\]
\( \omega_{sm} \) is a property of the AO crystal and the sonic carrier frequency, \( \omega_{sc} \) [7]. However geometric considerations of the detector position and aperture limit the maximum modulation frequencies to a much lower value.

When a detector of finite aperture is placed at a distance \( l_d \) from the modulator as shown in Fig 4.17.

![Fig 4.17 Detector Aperture Bandwidth Restriction](image)

The amplitude modulated signal is converted to a fluctuating electrical signal, provided that all 3 beams, of the deflected ensemble, enter the detector [27].

As the frequency is increased, the sidebands move away from the carrier until they meet the edges of the aperture. When the geometric centres, of the sideband beams, reach the edges of the aperture, the power of the detected amplitude fluctuations falls by 3dB. Thus it is clear that the detector aperture defines a maximum modulation frequency also.
For a given optical detector aperture width $\omega_D$, the maximum modulation frequency $\omega_{sm}$ is found by solving the quadratic equation derived in appendix 3, using (4.21), (4.22), (4.25)

$$(c^2 W \tan^2(\theta_B))\omega_m^2 + [2cnv_s \omega_1 (\tan^2 \theta_B + 1) \cos(\theta_B)] \omega_m$$

$$- 1 + \tan^2(\theta_B) wn^2 \omega_1^2 \sin^2(\theta_B) + (1 + \tan^2(\theta_B)) wn_v^2 \omega_c \omega_1 \sin(\theta_B)$$

$$ + wn^2 \omega_1^2 - c^2 \omega_1^2 \tan^2(\theta_B) = 0$$

(4.56)

Where $L$ is the distance between the modulator and detector.

Using Lenses, the significance of the aperture bandwidth limit may be reduced. This is discussed, in more detail, in section 5.9.3.

The analysis of aperture limited bandwidth was undertaken assuming that the detector aperture fixes the maximum modulation bandwidth, however the band limiting aperture may be a lens or the width of an optic fibre core.

It is also possible to exploit the angular separation of the optical carrier and sidebands to produce other forms of amplitude modulation, and ultimately to produce optical angle modulation. Using spatial filters, one can extract just one optical sideband and produce single sideband suppressed carrier light modulation.

The frequency range over which the sideband may be extracted, without the presence of significant carrier, fixes the minimum SSB modulation.
frequency. The problem is entirely similar to the problems encountered in filter generation of SSB radio signals. The maximum modulation frequency is fixed again by the aperture size of the spatial filter.

The minimum modulation frequency is determined by the sideband deflection angle, the width of the incident laser beam and its angular divergence. Fig 4.18 shows a diverging laser beam incident on plane wavefront.

The incident optical beam has a half power beam width of $\psi$ and is distance $l_d$ m away from the acousto optic modulator. After interaction, the diverging beam is deflected into diverging carrier and sideband beams. Each diverging beam is projected to a virtual focus for analysis purposes. It is considered that the carrier is suppressed if no light from within the 3dB beam width enters the detector, through the spatial filter.

Using the above definition for carrier suppression and analysis of the geometry of Fig 4.18, the minimum modulation frequency for which a sideband may be extracted is given by (see appendix 4).
\[
\frac{v}{\omega_m} = \frac{\pi}{2} - 2\pi - \cos^{-1}\left[ \frac{A}{A^2 + A^2 \cos^2(\alpha_1) + 2A \cos(\alpha_1) + 1} \right] \tan\left[ \theta_B - \frac{\psi}{2} \right]
\]

\[\omega_c\]

Where

\[A = \frac{(w_1 + w_2) \cos(\alpha_1) + \sqrt{(w_1 + w_2)^2 \cos(\alpha_1) + 1 + w_1^2}}{(w_1 + w_2)^2 + w_1^2}\]

and

\[\alpha'' = \tan^{-1}\left( \frac{c \omega_c}{n v_s \omega_1 \cos(\theta_B - \frac{\psi}{2})} + \tan(\theta_B - \frac{\psi}{2}) \right)\]

The upper modulation frequency is derived in a similar way to the derivation of the amplitude modulation cutoff frequency, as shown in Fig 4.19.
Again the geometry of Fig 4.19 is used to derive the maximum modulation frequency $\omega_{sm}$. If the upper sideband is to be extracted (appendix 5)

$$\hat{\omega}_{smUSB} = \frac{w^2 \omega_1^2 v_s^2 + w\omega_{sc}^2 c^2 + 2\omega_{sc} \omega \omega_{1} v_s \sin \theta}{\omega_{sc} \omega \omega_{1} v_s \cos \theta - w \omega_{sc}^2 c^2 - wcn \omega_1 v_s \sin \theta}$$ (4.58)

Where $w$ is the aperture width and $\theta$ is the incident angle, it is usually the Bragg angle, $\theta_B$.

If the lower sideband is extracted, the minimum modulation frequency $\omega_{sm}$ is

$$\hat{\omega}_{smLSB} = \frac{w^2 \omega_1^2 v_s^2 + w\omega_{sc}^2 c^2 + 2\omega_{sc} \omega \omega_{1} v_s \sin \theta}{\omega_{sc} \omega \omega_{1} v_s \cos \theta + w \omega_{sc}^2 c^2 + wcn \omega_1 v_s \sin \theta}$$ (4.59)

The minimum modulation frequency may be further limited by the spatial filter, in which case the modulation frequency range is found using a differential version of either (4.58) or (4.59).

Section 5.9.3 outlines ways in which lenses may be used to reduce the aperture width limitations, on the modulation frequency, in which case the ultimate frequency limit given by (4.55) [7],[11],[17] is reached.
4.3 INTRA-CAVITY LASER MODULATION

Methods of modulating light emerging from a laser cavity were discussed in the preceding section 4.2. These same methods together with a variety of others may also be used to modulate a laser by altering the conditions prevailing within the laser cavity, thereby altering the output optical amplitude or angle [6],[7],[9],[20],[28],[29].

Modulation techniques which act upon the optical cavity conditions are known as intra-cavity techniques. There are advantages and disadvantages in using them.

The advantages, in using intra cavity modulation techniques, are generally a consequence of the internal cavity gain, experienced by a light wave as it passes from one cavity reflector to another [28],[29]. For example, if a light valve is used to introduce a small amount of loss into the cavity, the single pass cavity gain can be used to effectively amplify the loss, causing large fluctuations in output power [28],[29].

The major disadvantages in using cavity modulation techniques, are the limitations placed upon the optical frequencies by the dimensions of the cavity [30]. It is not possible to alter the frequency, to any value inconsistent with (3.27), unless the cavity dimensions are modified.

With the above limitations in mind, it is proposed to examine a number of intra cavity modulation techniques which may prove useful for angle modulation of light [1],[2],[31], and may be easily incorporated into active frequency stabilization schemes [32].
Section 4.3.1 examines a piezo-electric method of altering the cavity dimensions [32], and therefore the permitted cavity frequencies. Whence frequency and phase modulation may be produced [32].

The EO, MO and AO techniques of modulation, discussed in section 4.2, will be reapplied to modulation of light within the optical cavity. Use of the materials is similar to extra cavity use, therefore the explanations of section 4.3.2 are brief.

Techniques of altering the energy level spacing involved in stimulated emission [33], and therefore the frequency of stimulated emission are examined in section 4.3.3. The Zeeman effect [32] of energy level splitting is of particular interest.

The modulation techniques outlined in this section are not exhaustive, they are however the simplest and possibly most attractive candidates for modulation of lasers in commercial transmission applications.

4.3.1 Piezo-Electric Modulation

By placing a piezo-electric crystal between the totally reflecting mirror, of a laser cavity, and its mount is shown in Fig 4.20, the cavity length may be altered by varying the electric field across the crystal [34].
Assuming that the piezo-electric strain is quadratically related to the electric field, the change in cavity length for a given electric field $E$ is [84],

$$\Delta l_p = A_{pz} l_p E^2$$

where $l_p$ is the length of the crystal, $\Delta l_p$ is the change in crystal length due to the electric field, and $A_{pz}$ is a constant relating the crystal strain to the magnitude of the electric field.

The optical output of the laser has frequency components spaced approximately by $\omega_{lc}$ (3.27) [30],

$$\Delta \omega_{lc} = \frac{\pi c}{bn}$$

where $n$ is the refractive index of the laser active medium, $c$ is the speed of light in vacuum and $b$ is the reflector spacing.
Around a stimulated emission frequency $\omega_1$, broadened by internal processes [31], to the range from

$$\omega_L - \frac{\Delta \omega}{2} \text{ to } \omega_L + \frac{\Delta \omega}{2}$$

Hence the laser output consists of frequencies, (3.31)

$$\omega_1 (m) = \frac{\pi mc}{bn}$$

where $m$ is an integer within the range

$$\text{Int} \left( \frac{(\omega - \Delta \omega)}{2\pi c} \right) \leq m \leq \text{Int} \left( \frac{(\omega + \Delta \omega)}{2\pi c} \right)$$

If the electric field across the piezo-electric crystal is modulated with a time varying signal, $E_m m(t)$, where $E_m$ is the signal amplitude and $|m(t)| \leq 1$, in the presence of a large bias field $E_0$, the cavity length will change in sympathy with the modulation,

$$\Delta b = l_p A_{pz} E_0 E_m m(t) \quad (4.61)$$

The output frequencies of the Laser will change to

$$\omega_1 (m,t) = \frac{\pi mc}{n (b + l_p AE_b E_m m(t))} \quad (4.62)$$

and $m$ will now lie within the range,

$$\text{Int} \left( \frac{n(\omega_L - \omega_1)}{2\pi c} \right) \leq m \leq \text{Int} \left( \frac{n(\omega_L + \omega_1)}{2\pi c} \right)$$

From (4.62) it is clear that, for a sufficiently large amplitude $E_m$, one frequency of oscillation can be quenched, at one frequency limit of the ensemble, and another generated at the other.
Now if a secondary cavity, or etalon, is inserted within the first as shown in Fig 4.21 [30],[35], one member of the frequency ensemble may be selected, if it satisfies the boundary conditions of the secondary cavity [30], that is

\[ \omega_{le} = \frac{2\pi m_e}{l_e} \]  

(4.63)

Where \( l_e \) is the effective length of the second cavity and \( m_e \) is the number of wavelengths within the cavity.
As the main cavity dimensions change, the internal light frequency changes. The associated mode no-longer satisfies the boundary conditions of the second cavity or etalon, regeneration ceases and the light is extinguished. Using these characteristics, the light is amplitude modulated.

If the dimensions of the etalon are suitably chosen; by varying the main cavity dimensions modes at the upper and lower edges of the stimulated emission bandwidth may regenerated and extinguished alternately. Thus Frequency Shift keying may be generated.

By changing the cavity length appropriately, the optical output can also be stabilized to one narrow Gaussian frequency mode [35].

4.3.2 Electro-Optic, Magneto-Optic and Acousto-Optic Modulation

Using the EO, MO and AO properties of materials, within the optical cavity, lasers may readily be modulated [28],[29]. Experiments have been conducted using such techniques to produce mainly amplitude modulated signals [5]. However other forms of modulation may be produced by these methods.

Using electro optic and magneto optic light valves, amplitude modulation may be produced by placing the valve between the active medium and a mirror, forming the laser cavity as shown in Fig 4.22 [5].
Regenerative oscillations within the cavity can be either quenched or re-enforced, depending upon the applied magnetic or electric field. Thus the laser is amplitude modulated. When inside the cavity the light valve need only provide attenuation greater than or equal to the single pass amplification of the active medium, to extinguish the light; the modulating signal is amplified by the multi-pass cavity gain [28],[29]. Unfortunately such modulation techniques are confined to low frequencies, as a finite time is required for regenerative oscillations to build up [5].

If however, the light valve is placed between the active medium and the partially transmitting mirror, (see Fig 4.22) the laser may be amplitude modulated at very high frequencies. If these frequencies are sufficiently high, the regenerative build up and decay of light power will not follow the changing transmission of the light valve. The light valve
will effect only the coupling of light from the cavity to the outside, through the partially transmitting mirror [5]. Therefore the laser may be amplitude modulated at very high frequencies using the light valve.

Amplitude modulation may also be produced by exploiting the MO Faraday effect. The magnetic field produces changes in light polarization, or senses of polarization rotation, of light reflected from a cavity mirror, to cancel the light incident on that mirror [5].

Light may be frequency modulated using crystal bi-refringence, by exploiting the adjustable phase retardation of the slow axis [5],[6]. As the modulating field increases, the refractive index, and hence the speed of light falls. Consequently the wavelength of light within the crystal increases; hence the cavity electrical length increases. The corresponding output frequencies fall. The behaviour of such a modulator is very similar to piezo-electric modulator discussed in section 4.3.1.

Using an AO modulator, both amplitude and angle modulation may be produced. Amplitude modulation is produced by altering the amplitude of the sonic wave within the crystal thereby altering the amount of power deflected out of the cavity, thus preventing optical regeneration [5]. Angle modulation is produced by using the frequency shifting capabilities of the acousto optic interaction, as discussed in section 4.2.2.

The frequency spacing of adjacent optical cavity modes tends to limit the possible modulation frequencies to only those permitted by the cavity dimensions [30].
4.3.3 Energy Level Modulation

A laser may also be modulated by altering the difference between energy levels involved in stimulated emission using an external field [32]. The Zeeman effect may be used to alter the energy level spacing [32]. When influenced by a magnetic field, the energy levels of a material split into several closely spaced levels, as shown in Fig 4.23 [32]. Such Zeeman splits are a consequence of electron spin and its associated magnetic dipole moment [32],[36]. If the spin magnetic dipole moment is aligned with the applied magnetic field, the energy of the electron is minimal [36]. The electron energy rises as the electron spin dipole moment becomes more misaligned with the magnetic field.

![Fig 4.23 Zeeman Energy Level Splitting](image)

The change in energy as a result of the Zeeman effect is summarized [36], by

$$\Delta E = \mu \cdot H$$  \hspace{1cm} (4.64)
Where $\mu$ is the net magnetic moment of the electron spin, $H$ is the magnetic field intensity and $\Delta E$ is the change in energy caused by the applied magnetic field.

Therefore an energy level with $m$ different dipole moments $\mu_i$ splits into $m$ separate energy levels under the influence of an applied electric field. Thus

$$\Delta E_i = \mu_0 \mu_r \mu_i \cdot B \quad i = 1, 2, \ldots m \quad (4.65)$$

Where $B$ is the magnetic field vector, $\mu_0$ is the permeability of free space and $\mu_r$ is the relative permeability of the active medium.

The change in the difference between energy levels is given by

$$\Delta E = \mu_0 \mu_r B \cdot (\mu_2 - \mu_1) \quad (4.66)$$

It is obvious that a change in stimulated emission frequency induced by the magnetic field, will only occur if $\mu_2 \neq \mu_1$. That is, that the magnetic dipole moments are not equal in magnitude and/or orientation.

Using (3.4), the angular frequency for the stimulated emission is defined [28],[36],

$$\omega_1 = \frac{\mu_0 \mu_r B \cdot (\mu_2 - \mu_1)}{\hbar} + \omega_1$$
Where $\omega_1$ is the angular frequency of the original energy level transition and $\omega'_1$ is the new stimulated emission frequency.

If a number of levels are involved in stimulated emission then the emission frequencies are summed as a Fourier series.

When the Zeeman induced emission frequency change is given by,

$$\Delta \omega_1 = \frac{\pi c}{2nb}$$

(4.67)

the cavity will not sustain oscillation, the mode will be quenched. Using this technique it is possible to amplitude modulate the light using the Zeeman effect.

The magnetic vector magnitude required for amplitude modulation is given by

$$|B| = \frac{\pi\chi h}{2nb} \frac{\mu_\mu_2 - \mu_1}{\mu_2 - \mu_1}$$

(4.68)

If the Zeeman split generates a frequency change of

$$\Delta \omega = \frac{\pi c}{nb}$$

oscillation at the new frequency will be sustained by the optical cavity, thus a frequency modulated signal can be generated.

Using the secondary cavity etalon the laser can be forced to emit only one of two frequencies, dependent on the magnitude of the magnetic field,
hence Frequency Shift keyed and Phase Shift keyed signals can be generated.

The maximum frequency of modulation is determined by the period of magnetic dipole relaxation, which is a characteristic of the active medium [32]. It may, however, be possible to overcome such modulation frequency limitations using travelling wave techniques [5].

In some cases it may be necessary to alter the pump frequency to sustain the population inversion at the new energy levels produced by the Zeeman split. Moreover, if an activator is used to generate the population inversion [37], the change in activator energy levels must be the same as those of the active medium under the influence of the applied magnetic field.

4.3.4 Power-Supply Modulation

Power supply modulation is the most commonly used form of light modulation in commercial optical transmission systems [38].

The power supply is turned on and off in sympathy with the digital information and the laser is amplitude modulated. Maximum modulation frequencies are fixed by the time taken for light to build up to saturation and decay to nothing [38].

Power supply modulation is a simple amplitude modulation technique, however each time the laser is keyed, sharp transients occur and frequencies well outside the operating bandwidth are generated. The
transients are particularly significant if it is intended to wavelength division multiplex more than one signal into one fibre [38]. The same problem occurs in radio transmission [39].

Moreover, during the build up of oscillation, the frequency tends to change until saturation is reached, the laser output tends to "chirp". Demodulating such pulses is difficult [39].

For these reasons, power supply amplitude modulation, while simple, uses more bandwidth than necessary and gives a worse signal to noise performance than other amplitude modulation systems [1],[21].

Power supply modulation cannot be used to generate angle modulation, and is of little interest in this dissertation.
REFERENCES - CHAPTER 4

[1] HAYKEN
"Communications Systems."
John Wiley and Sons, Inc. 1978, pp 244 - 377, 530 - 521.

[2] STREMNER, F.G.
"Introduction to Communication Systems."

[3] BANCETTI, B. and TARTARA, G.
"Equalization and Quality Prediction in Digital Radio Systems."
GTE Telecommunicazioni, Milan, Italy, pp 3 - 40.

"Coherent Light."


[7] SAPRIEL, J.
"Acousto-Optics."

[8] KITTEL, C.
"Introduction to Solid State Physics."
[9] MARCUSE, D.

"Light Transmission Optics."

[10] ALBRIGHT, R.J.

"Relationship of Doppler Ultrasonic Scattered Signal
Characteristics to Flow and Beam Parameters."
Journal of the Acoustical Society of America, Vol. 59, No. 4,
April 1976. pp 786 - 789.


"Wideband Microwave Light Modulation."


"Unified Approach to Ultrasonic Light Diffraction."
IEEE Trans. On Sonics and Ultrasonics, Vol. SU-14, No. 3, July

[13] KISHIOKA, K., and HASHIMOTO, M.

"Coupling of Guided Light Through a Periodic Multilayer."


"Interaction of Light and Microwave Sound."
pp 1604 - 1610.

"Acoustic Scattering of Light in a Fabry-Perot Resonator."

[16] DIXON, R.W. and GORDON, E.I.

"Acoustic Light Modulators Using Optical Heterodyne Mixing."

[17] HANCE, H.V. and PARKS, J.K.


[18] PAPANICOLAOU, N.A., LIN, H.C. and BENZ, H.F.

"Monolithic Surface Acoustic Wave (SAW) Charge Transfer Device and Its Applications."

[19] DAS, P., SCHUMER, D. and ESTRADA-VAZQUEZ.

Optical Communications Using Surface Acoustic Waves."


"Optoelectronics - An Introduction."

[21] MILEK, J.T. and NEUBERGER, M.

"Linear Electro-optic Modulator Materials."
IFI/Plenum, 1972.

[22] HARVEY (as for [4]) pp 160 - 178.
[23] CHARLES, C. et TREFFE.  

[24] GOODMAN, J.W.  
"Introduction to Fourier Optics."  

"Physics of Solids."  

[26] ISOMET CORPORATION  
"Instruction Manual Acousto Optic Modulator Driver Series 230 (Analogue Modulation)"  
Isomet Corp., P.O. Box 1634, 5263 Port Royal Rd., Springfield, VA 22151.

[27] KINGSTON, R.H.  
"Detection of Optical and Infrared Radiation."  

[28] HARVEY (as for [4]) pp 100 - 102, 319 - 320.

[29] RAINER, A.M.  
"Spectral, Spatial and Temporal Properties of Lasers."  
   "Fields and Waves in Communication Electronics."

[31] HARVEY (as for [4]) pp 591 - 592.

[32] HARVEY (as for [4]) pp 681, 793.

[33] EISBERG, R.M.
   "Fundamentals of Modern Physics."

[34] WERT and THOMPSON (as for [25]) pp 403 - 405.

[35] HARVEY (as for [4]) pp 702 - 705; 781 - 796.

[36] EISBERG (as for [33]) pp 443 - 449.

[37] BROWN, R.
   "Lasers - A Survey of Their Performance and Application."

[38] NILSON, J. and HAWKES, J.F.B.
   "Optoelectronics - As Introduction."

[39] DEMAW, D.
   "The Radio Amateur's Handbook"
   America Radio Relay League, Newington, CT, USA, 06111, pp 6.1 - 7.27.
5. DETECTION OF MODULATED LIGHT

Detection of light is the process of converting optical energy to electrical energy.

A variety of light detectors; sensitive to different parts of the visible and infra-red spectra exist. They have differing efficiencies and noise behaviour, and are built in a variety of ways. Photo-detectors do however, rely on one basic mechanism of interaction between incident photons and electrons, of a material.

In a perfect photo detector, each incident photon liberates an electron, which contributes to current generation or conduction within a detector [1],[2],[3],[4]. Therefore photo detectors conduct or produce currents proportional to the photon flux incident upon the surface of the detector [1],[2],[3],[4].

At a particular frequency the photon flux is proportional to the power of the incident light beam. Therefore the current conduction or generation within the detector is proportional to the incident power [1],[2],[3],[4].

Unfortunately not every photon liberates a photo-electron therefore photo-detectors have conversion efficiencies of less than unity [1],[2],[3],[4]. Moreover, detectors produce electrical noises due both to light dependent and heat dependent mechanisms [5],[6],[7],[8].

The purpose of this chapter is to characterize the optical detector and its noise sources. From such characterizations, expressions for signal to noise ratios may be derived for a variety of optical modulation schemes.
It is then possible to select modulation schemes which are suitable for detection with optimum signal to noise ratios. Moreover techniques for improving the detection signal to noise ratios may become evident.

The details of optical detection processes are discussed in section 5.1, where optical efficiency, optical and electrical frequency responses are examined and transfer functions are derived.

The noises introduced by darkened optical detectors are examined in section 5.2. These noise sources are a consequence of thermal charge carrier generation, and are a characteristic of the detector and its thermal environment only [5],[6].

Section 5.3 discusses the optically based sources of noise which have no equivalent in other electronic systems. Such noises are due to the corpuscular nature of light [7],[8],[9],[10],[11],[12],[13],[14], the uncertainty in arrival times of photons at the detector and the stochastic nature of photon-electron interaction [7],[8]. Such noise sources are present in radio signals, however the frequencies of radio transmissions are much lower than those frequencies of visible and infra red light. Consequently for a given transmitter power the number of quanta is much larger in radio waves than it is a light wave [7],[8]. Therefore classical treatment of radio waves introduces little error in signal to noise expressions. In optical systems the error, introduced by classical treatment, cannot be ignored [7],[8],[9],[10],[12].

Using the mathematical characterization of optical detection processes and noise sources, it is possible to discuss detection of modulated signals. In section 5.4 detection of amplitude modulated and amplitude
shift keyed signals is discussed. From the discussion, it becomes clear that the detection of amplitude modulated signals is an optical mixing process [13],[15],[16],[17],[18],[19],[20],[21],[22],[23],[24],[25],[26], [27],[28],[29],[30]. The optical detector is therefore a mixer of optical signals.

By exploiting the mixing characteristic of the optical detector, it is possible to detect other forms of amplitude and angle modulation by Hetrodyne methods, where a local oscillator is optically mixed with the incoming signal. The characteristics of such techniques are examined in section 5.5 and signal to noise expressions are derived.

Hetrodyne and Homodyne methods of detection are compared in section 5.6, where it is shown that the two methods are equivalent and produce the same output signal to noise ratios for similar equipment configurations [31],[32]. Homodyne systems cannot produce a 3dB improvement over equivalent hetrodyne systems as claimed by some authors [31],[32].

In section 5.7, methods of improving the signal to noise ratios of the modulation types discussed in sections 5.4 and 5.5 are examined. It is intended that optical detection should reach the quantum limit [4],[7], [15],[16],[17].

New theories of the squeezed states of light are also introduced. They challenge the traditional ideas of the quantum limit. Using special techniques it may be possible to reduce SNR's below the traditional limit of detection [33].
Finally, the alignment of beams, to be mixed by the optical detector and the effects of misalignment on the detector frequency responses, is discussed in section 5.8 [27].

Misalignment may be overcome by the methods discussed in section 5.9.
5.1 INTENSITY DETECTION

In 1906 Einstein characterized the photo electric effect [34],[35]. He observed that the effect occurred when a characteristic value of radiation frequency, for the material under test, was exceeded. Electrons were also emitted with specific energies. From these observations he concluded that the emitted electron energy \( E_e \) is given by [4],[13],[34]

\[
E_e = h
\nu_1 - WF = \pi \omega_1
\]  

(5.1)

Where \( h \) is Planck's constant, \( \nu_1 \) is the frequency of radiation and \( WF \) is the work function of the material. \( \pi = h/2\pi \) and \( \omega_1 \) is the angular frequency of radiation.

Therefore, provided that the radiation frequency \( \nu_1 > WF/h \), photons will be emitted from the surface of the material [4],[13],[34]. Einstein's observations were made for metals, and the work function value is determined by the surface binding characteristics, of metals, in evacuated envelopes [4],[13],[34]. Electrons thus liberated from the metal surface are available for conduction. Currents may flow inside the evacuated container between a photo cathode and positivity charged anode [4].

The photo electric effect is not confined to metals. It is also observed in semiconductors [4],[13]. Section 3.1 discussed the two band structure of semiconductor materials. An electron of sufficient energy may cross the forbidden energy zone, from the valence band to the conduction band, generating an electron-hole pair. These are available for conduction [4]. Again, the electron energy is given by (5.1), however the work function \( WF \) is replaced by the energy width of the forbidden zone [4], that is
\[ E_e = \eta \omega_1 - E_g \]  

(5.2)

The optical radiation must be of a frequency \( \omega_1 > \frac{E_g}{\eta} \) for optical generation of charge carriers to occur.

It is therefore reasonable to conclude that detection mechanisms, for optical radiation in vacuum tube or semiconductor devices, are essentially the same.

In section 5.1.1, the general types of detector and associated detection mechanisms are examined. It will become evident that subtle differences between detectors occur, which fix their sensitivities and frequency responses. Particular emphasis is given to the semiconductor photo-diode, which is the most commonly used detection device, in commercial applications; it was also used extensively for experimentation.

It is, however, instructive to discuss the characteristics of the vacuum-tube photo diode, as it is a simple device and gives insights into the general behaviour of photo detectors [4].

The average number of electrons emitted per photon gives a measure of the detector quantum efficiency [4],[5],[6],[7],[8]. Mechanisms fixing the quantum efficiency, of SPD's, are discussed in section 5.1.2. However, the detector quantum efficiency varies with optical radiation frequency and the frequency of optical intensity fluctuations [4]. The nature of the optical frequency variations is covered in section 5.1.3 and the electrical response to fluctuations, in intensity, are discussed in section 5.1.4.
Using the optical and electrical frequency behaviour, introduced in the two preceding sections, an expression for the detector transfer function is derived in section 5.1.5.

5.1.1 Detection Mechanisms

A variety of radiation detectors exist. Each has merits for detection of radiation of particular wavelengths, in particular applications. However, all methods currently in use are based upon photon electron interaction, as described by equation (5.1) [4],[13],[14].

It is instructive to discuss the vacuum-tube-photo-diode, initially because its operation is simple and it provides insights into the behaviour of the much more complicated semiconductor diode, on which the subsequent analysis will concentrate. However to gain an overview of detectors that may be used in various applications, all common photo detector types will be introduced.

The vacuum-tube-photo-diode consists of a metallic photo-cathode and an anode, in an evacuated envelope as shown in Fig 5.1 [4].
A power supply is connected between the cathode and the anode, such that the anode is positively charged with respect to the cathode. Light incident upon the cathode releases electrons, with energies given by (5.1).

$$E_e = \hbar \omega - WF$$

Electrons flow from the cathode to the anode and into the external circuit.
After its release from the cathode, the electron is accelerated until it reaches the anode. Assuming that the electrodes are flat plates, the electron receives an acceleration of,

\[ a = \frac{q_e v}{m_e d} \]  

(5.3)

Where \( a \) is the acceleration, \( e \) the electronic charge, \( m_e \) is the electronic mass, \( v \) and \( d \) are the respective voltage and distance between the anode and cathode.

Using simple calculus, the time, \( \tau \), taken for an electron to travel the cathode and anode is found,

\[ \tau = d \sqrt{\frac{2m}{q_e v}} \]  

(5.4)

At time \( t = 0 \), an electron has been freed from the surface of the metal; it is assumed that the initial velocity is zero. The electron is then accelerated by the electric field where it gains a velocity of \( u \) \n
\[ u = \frac{q_e v \tau}{m_e d} \]  

(5.5)

The electron it collides with the anode and is stopped. The current over the acceleration period is proportional to the velocity of the electron, which is linearly increasing, therefore,

\[ i(t) = k_e q_e t \]  

(5.6)

Where \( i(t) \) is the electron current and \( k_e \) is a constant of proportionality.

The net charge, that has passed from cathode to anode, is therefore
\[ \int_0^T i(t) \, dt = q_e \]  \hspace{1cm} (5.7)

Using (5.6), the constant of proportionality \( k_e \) is found [1],[2],[4],[12],[13],[36]
\[ k_e = \frac{2}{\tau^2} \]  \hspace{1cm} (5.8)
and the current is,
\[ i(t) = \frac{2 \, q_e \, t}{\tau^2} \]  \hspace{1cm} (5.9)

Equation (5.9) therefore gives the impulse response of the detector.

The detector frequency response is readily found by taking the Fourier transform of (5.9), whence
\[ i(\omega) = \int_{-\infty}^{\infty} i(t) \exp[j\omega t] \, dt \]
\[ = \frac{2 \, q_e}{\omega^2 \, \tau^2} \left\{ (1 + j\omega \tau) \exp[j\omega \tau] - 1 \right\} \]  \hspace{1cm} (5.10)

Implicit in (5.10) is the assumption that the optical radiation is of sufficiently high frequency to liberate a photo electron. "\( \omega \)" in (5.10) is the light intensity fluctuation frequency, therefore (5.10) defines an electrical bandwidth of \( \frac{1}{\tau} \) rad/s.

Current conduction is naturally, not confined to single electron transitions, but to the number of transitions per unit time given by the number of quanta arriving at the detector per unit time [4], thus
\[ i(\omega) = \left( \frac{p_0}{\hbar \omega_1} \right) \left( \frac{2 \, q_e}{\omega_m^2 \, \tau^2} \right) \left( 1 + j\omega \, \tau \right) \exp[j\omega \, \tau] \]  \hspace{1cm} (5.11)
Where $\omega_n$ in (5.10) has been replaced with $\omega_m$, the optical intensity fluctuation frequency.

It is clear therefore, that the transit time, within an optical detector, is a major limitation on the amplitude modulation frequency passed by the detector [4].

In subsequent analysis, it will become evident that the transit time limits the electrical frequency response of SPD detectors also [4].

When the detector is to operate in low light level situations, it may be desirable to use a vacuum-tube-photo-detector that multiplies the current beyond the 1 photo electron per photon limit, of the photo-diode, discussed [4].

By positioning a number of dynodes, within the envelope of the photo diode as shown in Fig 5.2, a photo-multiplier tube may be produced [1],[4], [14].
When a photo-electron is emitted from the cathode it strikes a dynode as it accelerates towards the anode. A second electron is emitted from the dynode and is accelerated towards the anode. The two electrons strike other dynodes liberating further secondary electrons. In this way the original photo current is multiplied [1],[4],[14].

The photo multiplier is much more sensitive to incident light, but is noisier and its transit times are much larger, than the photo-diode, thus the photo-multiplier tube has a much smaller electrical bandwidth.
Semiconductor devices exist which exhibit similar characteristics to the vacuum tube devices discussed. There are, however, devices which have no equivalents in the vacuum tube realm.

Again, recall the double banded structure, of intrinsic semiconductor crystals, discussed in chapter 3. Photons of incident light with energy greater than the energy gap $E_g$ will promote electrons from the valence band. Such hole-electron pairs are available for conduction within the crystal. Therefore, the intrinsic semiconductor crystal will act as a detector of optical radiation [4]. As in the vacuum-tube-photo-diode case, the conductance is proportional to the incident light power. A full analysis of the electric frequency dependence, of intrinsic photo conductive detectors, is given in Kingston [4].

Photo conductive detectors may also be made out of extrinsic or doped semiconductors [4]. The dopants modify the energy level diagrams of the semiconductors as shown in Fig 5.3 (a) and (b).
A photon is absorbed at the impurity level producing either a free electron in the N-type material, or a free hole in the P-type material. Therefore the cutoff wavelength (5.2) is determined by the ionization energy of the dopant [4]. Thus conduction, proportional to the optical intensity, occurs if

$$\omega > \frac{E_i}{n}$$  \hspace{1cm} (5.12)
Where $E_i$ is the impurity ionization energy.


The device of greatest interest, in this dissertation, is the semiconductor-photo-diode, (SPD), which may be used in two modes, either photo-conductive mode, or photo-voltaic mode [4],[37].

When the P-N metallurgical junction is formed, electrons from the N-type material move across the junction combining with holes in the P-type region. Similarly holes move from the P-type region into the N-type region and combine with electrons [4]. A depletion region is formed as mentioned in chapter 3. The P-type region becomes negatively charged and the N-type region positively charged. The Fermi levels, of the two crystals, move until a constant Fermi level between the two crystal types achieved, as shown in Fig 5.4.

![Energy Levels of P-N Junction](image)

The depletion region is small, therefore in the region outside the depletion region there is no appreciable electric field, due to the charge movement [4].
In the photo conductive mode, the diode is reverse biased, and the depletion region broadens. However, again, outside the depletion region the electric field is negligible, therefore the voltage of the reverse bias is found over the width of the depletion region only [4]. The energy levels are moved further apart as shown in Fig 5.5.

**Fig 5.5 Energy Levels of Reverse-Biased P-N Junction**

![Energy Levels Diagram](image)

Photons, of sufficient energy, incident upon the crystal, separate electrons and holes. Outside the depletion region, they experience no net electric field. However a concentration gradient does occur, which produces a diffusion of carriers away from the site of separation [4]. Some minority charge carriers will drift towards the depletion region. Some of these minority carriers will be lost by recombination with majority carriers.

The minority carriers reaching the edge of the depletion region will be swept across the depletion region into the crystal on the other side of the junction [4]. Those charge carriers crossing the junction represent a conduction current, the magnitude of which is proportional to the optical intensity. Therefore, when reverse biased, the P-N SPD junction diode may
be used, as a photo-conductive optical detector, in a similar way to the vacuum tube photo diode.

The SPD can be operated in photo voltaic mode by connecting the unbiased diode to a high impedance external circuit [4],[37].

Photons of incident light create electron-hole pairs, which may diffuse towards the depletion region. Some will recombine before the junction is reached [4]. Minority carriers, which reach the depletion region, are swept from one side of the depletion region to the other side by the junction-potential electric field [4]. They emerge as majority carriers on the opposite side of the junction. Such charge carriers are free for conduction, in the external circuit, under the influence of the junction potential; in much the same way as current is conducted in the collector-emitter circuit of a bi-polar transistor.

Most SPD's will operate in either photo-conductive or photo-voltaic mode, however doping levels and physical dimensions are chosen to enhance the behaviour of a SPD in one mode only [37].

The anode current voltage curves for a typical SPD are shown in Fig 5.6 [37].
In applications where great optical sensitivity is required, an avalanche photo-diode may be used. It is analogous to the vacuum tube photo multiplier.

An avalanche photo diode is used in the photo conductive mode. It is reverse biased just below avalanche breakdown. Incident photos produce charge carriers which diffuse to the depletion region. When the depletion region is reached the minority carriers are accelerated at a very high rate. They gain sufficient energy, that on collision with other atoms,
they liberate new charge carriers, which are also accelerated rapidly. These too, collide with atoms liberating still more charge carriers and a diode avalanche occurs. Currents much larger than those due to single photon-electron interactions are generated, therefore a current gain is achieved [4].

The time taken for the avalanche to build up restricts avalanche photo-diodes to applications in which light intensity fluctuation frequencies are low [4].

Details of optical and electrical bandwidth restrictions, of SPD's, are discussed, in detail, in sections 5.1.3 and 5.1.4 respectively.

5.1.2 Quantum Efficiency

In the preceding analysis it is clear that, irrespective of the detection mechanisms used, the current is given by the product of the photon flux and the electronic charge [4],[36], if device ideality is assumed [4],[36]. Thus

\[ i = \frac{Pq_e}{\hbar \omega_i} \]  

(5.13)

Where \( P \) is the incident optical power and \( q_e \) is the electronic charge.

However, in reality, not all photons, incident upon the crystal, generate charge carriers. In fact, with the exception of the photo
multiplier tube and avalanche photo diode, the number of charge carriers generated is less than the number of incident photons \([4],[36]\). By taking a large number of observations it is possible to define the average number of charge carriers generated per photon. This figure is the quantum efficiency of the detector. The detector current is therefore,

\[
i = \frac{\eta p q_e}{n \omega l}\]

(5.14)

Where \(\eta\) is the quantum efficiency.

The quantum efficiency is a characteristic of the detector itself. It varies with both optical frequency and optical-intensity-fluctuation frequency.

The frequency analysis of the vacuum tube photo diode, of section 5.1.1, may be viewed as the frequency dependent part of the quantum efficiency \([4]\). In this case the quantum efficiency is dependent upon the frequency dependent transmission of the glass envelope, and the surface of the photo cathode.

Similarly, the quantum efficiency of an SPD is dependent on the transmission of the crystal itself, and also upon the diffusion life-time of the liberated charge carriers. The quantum efficiency is also dependent upon the orientation of the crystal in the optical field \([4]\).

From simple analyses of the behaviour of minority carriers within a SPD, an expression for the quantum efficiency may be derived. Assuming that the SPD is in electrical and thermal equalibrium, any minority carriers generated in the crystal will obey the diffusion equations \([4]\),
\[
\frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} - \frac{p}{\tau} \tag{5.15}
\]
\[
\frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} - \frac{n}{\tau} \tag{5.16}
\]

Where \( D_p \) and \( D_n \) are the hole and electron diffusion constants respectively, \( p \) and \( n \) are the respective hole and electron densities, above their equilibrium values, \( x \) is the distance travelled by diffusion and \( \tau \) is the carrier lifetime.

Equations (5.15) and (5.16) represent the rates of change of minority carrier densities. The second terms, in each equation, represent the rate of change of concentration gradient, with distance and the last terms represent the recombination rates [4].

Now consider the P-N junction shown in Fig 5.7.

![Fig 5.7 P-N Junction](image-url)
Setting the junction at the origin of coordinates, it is assumed that an optical signal, of constant intensity, is generating an excess charge density of $P_0$ across the full cross section, of the P-type region, at a distance $-d$ from the origin. Moreover, the excess carrier density, at the junction, should be zero since carriers, from the opposite side, will cancel any minority carriers at $X=0$.

By solving 5.14 with $P = 0$, the current due to an excess charge carrier density at $X=-d$ may be found [4],

$$\frac{D}{p} \frac{d^2 p}{dx^2} - \frac{p}{\tau} = 0 \quad (5.17)$$

The solution to (5.17), which also satisfies the boundary conditions above, is

$$p = P_0 \exp \left[ -\frac{X+d}{\sqrt{\tau \frac{D}{p}}} \right] + \exp \left[ \frac{(X-d)}{\sqrt{\tau \frac{D}{p}}} \right] \quad (5.18)$$

The carrier flow is given by the diffusion equations,

$$J_p = -q_e D_p \frac{dp}{dx} \quad (5.19)$$

and,

$$J_n = q_e D_n \frac{dn}{dx} \quad (5.20)$$

Using (5.18) and (5.19), the current collection efficiency may be found,

$$\eta_i = \exp \left[ -\frac{d}{\sqrt{\tau \frac{D}{p}}} \right] \quad (5.21)$$

Consider now a SPD comprising a transparent electrode and an epitaxially deposited N-type layer on a P-type substrate as shown on Fig 5.8 [4].
A proportion of the light, incident upon the surface of the SPD, is reflected according to

\[
\frac{P_r}{P_i} = R_0
\]  

(5.22)

Where \( P_i \) and \( P_r \) are the incident and reflected powers respectively. \( R_0 \) is the power reflection coefficient.
The remaining optical power is transmitted into the crystal, as described by [4][38],

\[
\frac{p_t}{p_0} = 1 - R_0 = T_0 \tag{5.23}
\]

Where \( p_t \) is the transmitted optical power.

As electrons are promoted from the valence band to the conduction band, the light is absorbed. Therefore the intensity of light at any distance \( x \), from the surface of the crystal, is given by [4][39],

\[
p(x) = p_0 \exp[-\alpha x] \tag{5.24}
\]

\[
\Delta p(x) = p_0 \left( \exp[-\alpha x_2] - \exp[-\alpha x_1] \right)
\]

\[
= p_0 \exp[-\alpha x_1] \left( 1 - \exp[-\alpha(x_2 - x_1)] \right) \tag{5.25}
\]

but \( x = x_1 - x_2 \) defines a displacement, and \( p_0 \exp[-\alpha x] \) defines the amount of power arriving at the section of interest. Therefore, the amount of input power absorbed in displacement \( x \) is,

\[
\frac{p(x)}{p_0 \exp[-\alpha x]} = (1 - \exp[-\alpha \Delta x]) \tag{5.26}
\]

and the amount of input power absorbed, per unit displacement, is the differential of (5.26), thus

\[
\alpha(x) = \frac{d}{dx} \left( 1 - \exp[-\alpha \Delta x] \right) = \alpha \exp[-\alpha \Delta x] \tag{5.27}
\]

\[
\alpha(x) = \alpha \exp[-\alpha x] \quad \text{for} \ x_1 = 0 \tag{5.28}
\]

The number of charge carriers generated per unit length is equal to the number of photons absorbed per unit length, thus,
The corresponding current is

$$i(\omega_1, x) = \frac{Pq_e \alpha \exp[-\alpha x]}{\hbar \omega_1} \quad (5.30)$$

If the N-type layer of the SPD shown in Fig 5.9, is small, little appreciable absorption occurs within it [4]. Most of the optical power passes into the P-type substrate, where the bulk absorption occurs. At any distance, $x$, from the junction, the current generated by a sheet of light is [4]

$$i(\omega_1, x) = \frac{Pq_e \alpha \exp[-\alpha x] \exp\left(-\frac{x}{\sqrt{D_n \tau_n}}\right)}{\hbar \omega_1} \quad (5.31)$$

If the N-type layer thickness, $l_n$, is not sufficiently small to ignore, the current contribution generated in that layer is

$$i_n(x) = \left(\frac{Pq_e}{\hbar \omega_1}\right) (1 - R) \alpha_n \exp[-\alpha_n (x - l_n)] \exp\left(-\frac{x - l_n}{\sqrt{D_n \tau_n}}\right) \quad \text{for } x \leq l_n \quad (5.32)$$

Where $i_n$ is the current in the N-type region at a distance $+x$ from the junction, $\alpha_n$ is the N-type material absorption coefficient, $l_n$ is the N-type layer width, $D_n$ and $\tau_n$ are the respective carrier diffusion constant and carrier lifetime in the N-type material.

After attenuation through the thickness of the N-type region, the light is absorbed in the P-type substrate,
If the SPD is built from a P-type epitaxy on an N-type substrate, the p and n subscripts of equation (5.31) - (5.33) are interchanged.

The performance of the MRD 500, used in experimentation, is readily described by (5.31), therefore subsequent analysis will use this equation. Devices described by (5.32) and (5.33) are analysed in a similar way.

Using a non idealized version of (5.31), the total current for the SPD of Fig 5.9 is

\[ i(x) = \frac{Pq_e (1 - R) \exp[-\alpha n x]}{\hbar \omega} \exp\left[-\frac{-x}{\sqrt{D_p \tau_p}}\right] \]  

(5.33)

for \( x > 0 \)

The product, \((1-R)\frac{\alpha_p}{\alpha_p + \frac{1}{\sqrt{D_p \tau_p}}}\), may be viewed as a frequency varying quantum efficiency, which is dependent on the surface transmissivity of the SPD, and the diffusion length \(\sqrt{D_p \tau_p}\). When the absorption coefficient is large, the quantum efficiency approaches \((1-R)\), the optical transmittance of the surface [4].
5.1.3 Optical Frequency Response of Optical Detectors

The optical frequency response of an optical detector, may be derived from the quantum efficiency (5.35) [4],

\[ \eta = (1-R) \left( \frac{\alpha_p}{\alpha_p + \frac{1}{\sqrt{D_p \tau_p}}} \right) \]

\( \alpha_p \) and \( \tau_p \) are properties of the extrinsic P-type material alone. The absorption coefficient \( \alpha_p \) and the reflectivity \( R \) are both frequency dependent and, therefore, determine the optical frequency response of the detector.

Absorption of radiation by the semiconductor involves promotion of electrons from the valence band to the conduction band. One of two mechanisms may be involved in the promotion, depending on the particular crystal in which absorption is occurring [40]. The transition may be, either direct, involving the interaction of 1 photon and 1 electron [40], or indirect, in which case a phonon is either absorbed or emitted, during transition [40]. At room temperature, direct or indirect absorption can be related to the optical frequency by [40],

\[ \alpha = \begin{cases} (A \hbar |\omega| - E_g)^n & \text{for } |\omega| > E_g \\ 0 & \text{for } |\omega| > E_g \end{cases} \] (5.36)

Where \( A \) and \( n \) are constants of the crystal.

The values of \( A \) and \( n \) are fixed by the method of electron transition. For direct transitions \( n = 1/2 \), for indirect transitions \( n = 2 \) [40].
The frequency dependence of the power reflection coefficient, $R$, is more complicated.

For normal incidence, the reflection coefficient of the electric field is \[ r(\omega_1) = \frac{n(\omega_1) + jk(\omega_1) - 1}{n(\omega_1) + jk(\omega_1) + 1} \quad (5.37) \]

Where $n(\omega_1)$ is the frequency dependent index of refraction, and $k(\omega_1)$ is the frequency dependent extinction coefficient.

The power at the detector surface is proportional to the squared magnitude of the electric field thus \[ R = r(\omega_1)r^*(\omega_1) = \frac{(n(\omega_1) - 1)^2 + k^2(\omega_1)}{(n(\omega_1) + 1)^2 + k^2(\omega_1)} \quad (5.38) \]

The extinction coefficient is related to the absorption coefficient by \[ k(\omega_1) = \frac{c\alpha}{2\omega_1} \quad (5.39) \]

Where $c$ is the velocity of light in a vacuum.

Substituting (5.36) into (5.39), the extinction coefficient becomes \[ k(\omega_1) = \frac{cA(h|\omega_1| - E_g)^n}{2\omega_1} \quad (5.40) \]

Using the kramers-kronig relations \[ [13],[40] \], the refractive index may be written in terms of the absorption coefficient
\[ n(\omega) = \frac{4\pi \sigma}{\alpha c} \] (5.41)

Where \( \sigma \) is the material conductivity.

Again substituting for \( \alpha \) using (5.36),

\[ n(\omega) = \frac{4\pi \sigma}{cA(\hbar |\omega| - E_g)^n} \] (5.42)

Substituting (5.40) and (5.42) into (5.38), the reflection coefficient becomes

\[
R = \frac{(8\pi \omega - 2\omega_c cA(\hbar |\omega| - E_g)^n)^2 + c^2 A^2(\hbar |\omega| - E_g)^{2n}}{(8\pi \omega + 2\omega_c cA(\hbar |\omega| - E_g)^n)^2 + c^2 A^2(\hbar |\omega| - E_g)^{2n}}
\] (5.43)

Substituting (5.27) and (5.34) into (5.37), the frequency dependent quantum efficiency is obtained, after some rearrangement,

\[
\eta(\omega) = \left\{ \frac{\sqrt{32\pi \sigma c A}}{2\omega_c cA(\hbar |\omega| - E_g)^n} \right\} \left\{ \frac{\sqrt{8\pi \omega + 2\omega_c cA(\hbar |\omega| - E_g)^n + c^2 A^2(\hbar |\omega| - E_g)^{2n}}} \right\}
\] (5.44)

\[
\times \frac{\sqrt{D_p \tau_p)}A(\hbar |\omega| - E_g)^n}{\sqrt{D_p \tau_p)}A(\hbar |\omega| - E_g)^n + 1}
\]

The photo detector direct current, \( i_{dc} \), is given by (5.13),

\[ i_{dc} = \frac{p q e}{\hbar \omega} \]

The direct current dependence on optical frequency is therefore
Unfortunately (5.45) still assumes that every photon produces a photo-electron within the crystal. However a small proportion will pass through the crystal without any interaction with electrons, therefore a fixed ratio of the incident photons will be absorbed, generating the current given by (5.46), thus

\[
\begin{align*}
\text{i}_{dc} &= \frac{Pq}{h |\omega|} \left( \frac{\sqrt{32\pi\sigma c^2 \omega_1^2 (h |\omega_1| - E_g)^n}}{[8\pi\sigma_\omega_1 + 2\omega_1 cA(h |\omega_1| - E_g)^{n+1} + c^2 A^2 (h |\omega_1| - E_g)^2 n]} \right) \\
&\times \left\{ \frac{(\sqrt{D_p T_p}) A h |\omega_1| - E_g)^n}{(\sqrt{D_p T_p}) A (h |\omega_1| - E_g)^n + 1} \right\}
\end{align*}
\]

Where \( \eta \) is the maximum quantum efficiency.

Thus far the analysis assumes that the optical power is present at one frequency only, and that it is constant, producing a direct current in the detector.

In general, the power arriving at the detector is a function of time \( P(t) \). The instantaneous power, \( P(t) \), is related to the squared magnitude of the electric field, by the atmospheric conductance [41],

\[
P(t) = G_a A_d E^*(t)E(t)
\]

Where \( E(t) \) is the time varying electric field, \( G_a = \sqrt{e/\mu} \) and \( A_d \) is the effective area of the detector.
The filtering effects, of the frequency dependent quantum efficiency, can be introduced as filter functions of the electric vector, thus the time dependent current is,

$$i(t) = G_a A_d [E(t) \circ h_1(t)][E(t) \circ h_1(t)]^*$$  \hspace{1cm} (5.48)

Where $i(t)$ is the time varying current and $h_1(t)$ is the electric vector impulse response. ($\circ$ denotes convolution.)

The Fourier transform of $i(t)$ gives the frequency dependent current [42],

$$i(\omega) = G_a A_d E^*(\omega) H^*(\omega) \otimes E(\omega) H(\omega)$$  \hspace{1cm} (5.49)

Where $E(\omega)$ is the frequency dependent electric field, and $H(\omega)$ is the electric field transfer function. ($\otimes$ denotes correlation.)

(5.49) may be written in its expanded form [42]

$$i(\omega) = G_a A_d \int_{-\infty}^{\infty} E^*(a) H_1^*(u) E(u+\omega) H_1(u+\omega) du$$  \hspace{1cm} (5.50)

$u$ is a dummy variable of integration.

If the detector is illuminated an analytic signal of frequency $\omega_1$, the Fourier transform of the signal is [42],

$$E(u) = 2\pi \delta(\omega-\omega_1)$$  \hspace{1cm} (5.51)

therefore
\[ i(\omega) = 4\pi^2 E^2 g_a d \int_{-\infty}^{\infty} H_1^*(u) H_1(u+\omega) \delta(u+\omega) s(u+\omega-\omega_1) du \quad (5.52) \]

\[ i(\omega) = 2\pi E^2 g_a d H_1^*(\omega_1) H_1(\omega_1 + \omega) \delta(\omega) \quad (5.53) \]

Which is a direct current of magnitude,

\[ i_{dc} = E^2 g_a d H_1^*(\omega_1) H_1(\omega_1) \quad (5.54) \]

which is equal to the magnitude of the direct current \( i_{dc} \) in (5.45), for a power of

\[ P = E^2 g_a d \quad (5.55) \]

Therefore, by spectrally factorizing (5.45) the optical electric field response can be found [43]. Unfortunately (5.45) is applicable only to direct currents, because it was assumed that

\[ \frac{\partial P}{\partial t} = 0 \]

in equation (5.15). To find the intensity-fluctuation-dependent quantum efficiency, equation (5.15) must be solved for a time varying excess carrier density, \( p \).

### 5.1.4 Electrical Frequency Response of Optical Detectors

The current collection efficiency of an optical detector can be found by solving partial differential equation (5.15)

\[ \frac{\partial P}{\partial t} = D_p \frac{\partial^2 P}{\partial x^2} - \frac{P}{\tau} \]
Subject to the boundary conditions of section 5.1.3. Equation (5.15) is readily solved by separation of variables.

\[ p = p_0 \left\{ \exp \left[ -\frac{(x+d)}{L} \right] + \exp \left[ \frac{(x-d)}{L} \right] \right\} \quad (5.56) \]

Where \( L \) is the complex diffusion length,

\[ L = \sqrt{\frac{D_p}{j \omega + \frac{1}{\tau_p}}} = \sqrt{\frac{D_p \tau_p}{j \omega + \frac{1}{\tau_p}}} \quad (5.57) \]

and \( \omega \) is the angular frequency of excess carrier density fluctuations.

In the direct current case, replacement of \( \sqrt{D_p \tau_p} \) by the complex diffusion length \( L \) will produce the same equation as (5.45), because \( \omega = 0 \). However, it is desired to generalize the electric vector, of the current transfer function, to alternating currents produced by optical intensity fluctuations.

Consider a planar wavefront of light, of electric field \( E_0(\omega_1) \), incident upon the SPD. Some of the incident will be reflected according to equation (5.37).

\[ r(\omega_1) = n(\omega_1) + jk(\omega_1) - 1 \]

The electric field transmitted, per unit incident electric field, is found by spectrally factorizing the power transmission coefficient \( t(\omega_1) \) which is related to \( r(\omega_1) \) through the energy conservation relations [43].

\[ t^*(\omega_1) t(\omega_1) = 1 - r^*(\omega_1) r(\omega_1) \quad (5.58) \]
$n(\omega_l)$ and $k(\omega_l)$ are real, thus,

$$t(\omega_l) = \frac{\sqrt{2} \sqrt{n(\omega_l)}}{(n(\omega_l) + 1) - jk(\omega_l)} \quad (5.59)$$

Substituting for $n(\omega_l)$ and $k(\omega_l)$ using (5.40) and (5.42),

$$t(\omega_l) = \frac{j\omega_l \sqrt{32\pi \sigma c A(\pi |\omega_l| - E_g)^n}}{j\omega_l [8\pi \sigma + 2c(\pi |\omega_l| - E_g)^n] + cA(\pi |\omega_l| - E_g)^2} \quad (5.60)$$

Let,

$$E_t(\omega_l) = E_t(\omega_l) t(\omega_l) \quad (5.61)$$

$E_t(\omega_l)$ is the frequency dependent electric field within the crystal, near the surface. At any distance $x$ from the surface of the crystal, the electric field is retarded and attenuated according to [39],

$$E_t(\omega_l, x) = E_t(\omega_l) t(\omega_l) \exp(-j2\pi x n(\omega_l)) \exp(-\alpha(\omega_l) x) \quad (5.62)$$

The Fourier transform of the instantaneous power, at $x$, is given by the correlation [42].

$$P(\omega, x) = \int_{-\infty}^{\infty} G_A dE_t(u, x) E_t(u + \omega, x) du$$

$$= G_A d \int_{-\infty}^{\infty} E_t(u) E_t(u + \omega) \exp\left[\frac{j2\pi n(u)}{c}\right] \exp\left[-\frac{j2\pi n(u + \omega)}{c}\right] \exp\left[-\frac{\alpha(u)x}{2}\right] \exp\left[-\frac{\alpha(u+\omega)x}{2}\right] du \quad (5.63)$$

The difference between the input power $p(\omega, 0)$ and the power found at point $x$, $p(\omega, x)$ is equal to the amount of power absorbed by the crystal, thus,
\[
\begin{align*}
p_{ab}(\omega, x) &= \int_{-\infty}^{\infty} G_{a} A_{d} E_{t0}^*(u) E_{t0}(u+\omega) du \\
&= A_{d} G_{a} \int_{-\infty}^{\infty} E_{t0}^*(u) E_{t0}(u+\omega) \exp \left( \frac{j2\pi(u)(x)}{c} \right) \exp \left( \frac{-j2\pi(u+\omega)x}{c} \right) \exp \left( -\alpha(u)x \right) \exp \left( -\alpha(u+\omega)x \right) du.
\end{align*}
\]

(5.64)

The power absorbed per unit length is the differential of \(p_{ab}(\omega, x)\) with respect to \(x\). The second term of (5.64) is an integral with respect to the dummy frequency variable \(u\), therefore the argument of the integral may be differentiated with respect to \(x\), thus

\[
\begin{align*}
\frac{dP_{ab}(\omega, x)}{dx} &= \int_{-\infty}^{\infty} A_{d} G_{a} E_{t0}^*(u) E_{t0}(u+\omega) \left( \frac{\alpha(u)+\alpha(u+\omega)}{2} \right) + \frac{j2\pi}{c} \left\{ n(u,\omega)-n(u) \right\} \exp \left( \frac{j2\pi(u)(x)}{c} \right) \exp \left( \frac{-j2\pi(u+\omega)x}{c} \right) du \left( \frac{\alpha(u)+\alpha(u+\omega)}{2} \right) + \frac{j2\pi}{c} \left\{ n(u,\omega)-n(u) \right\} \right) \right) \right) \right) \right)
\end{align*}
\]

(5.65)

The photon flux \(\varphi_{p}\), for a given single frequency optical signal of power \(p\), is

\[
\varphi_{p} = \frac{p}{\hbar \omega_1}
\]

(5.66)

For a single frequency optical signal, the power is given by

\[
P = G_{a} A_{d} E^*(\omega_1) E(\omega_1)
\]

(5.67)

Therefore the Fourier transform of the instantaneous photon flux may be written as the convolution [42],

\[
\varphi_{p} = \frac{G_{a} A_{d}}{\hbar} \int_{-\infty}^{\infty} \frac{E^*(u) E(u+\omega)}{\sqrt{u} \sqrt{u+\omega}} du
\]

(5.68)

and so the number of photons absorbed per unit length of crystal is given by
The current, injected at $x$, is given by the product,

$$i_i(W,x) = q e \Phi_p(W,x) dx$$

and the diode current, for an optical signal at $x$, is related to the injection current by the collection efficiency, $\eta_i(\omega)$ (5.21), where

$$\eta_i(\omega) = \exp \left[ -\frac{x}{L} \right]$$

The frequency dependent current at any point $x$ is thus

$$i(\omega,x) = q e \Phi_p(\omega,x) \int_{-\infty}^{\infty} E_t^*(u) E_{t_0}^*(u+\omega) \left( \frac{\alpha(u) + \alpha(u+\omega)}{2} + \frac{j2\pi}{c} \left( n(u) - n(u+\omega) \right) \right) x \exp \left[ \left( \frac{\alpha(u) + \alpha(u+\omega)}{2} + \frac{j2\pi}{c} \left( n(u) - n(u+\omega) \right) \right) x \right] du \tag{5.72}$$

Substituting $\alpha(\omega)$ (5.36), $K(\omega)$ using (5.40), $n(\omega)$ using (5.42), $E_{t_0}(\omega)$ using (5.60) and (5.61), and $L$ using (5.57),
\[
\begin{align*}
\xi(\omega, x) &= \frac{q_e G_a A_d}{\hbar} \exp \left[ -x \left( \frac{j \omega \tau_n + 1}{D_n \tau_n} \right) \right] \int_{-\infty}^{\infty} \frac{E_{i}^{*}(u) E_{i}(u+\omega)}{\sqrt{u} \sqrt{u+\omega}} \\
& \quad \frac{-ju\sqrt{32\pi \sigma cA(\pi \mid u \mid -E_{g})^{n}}}{-ju[8\pi \sigma + 2cA(\pi \mid u \mid -E_{g})^{n} + c^2 A^2(\pi \mid u \mid -E_{g})^{2n}]} \\
& \quad \frac{-j(u+\omega)\sqrt{32\pi \sigma cA(\pi \mid u+\omega \mid -E_{g})^{n}}}{j(u+\omega)[8\pi \sigma + 2cA(\pi \mid u+\omega \mid -E_{g})^{n} + c^2 A^2(\pi \mid u+\omega \mid -E_{g})^{2n}]} \\
& \quad \left\{ \frac{A(\pi \mid u \mid -E_{g})^{n} + A(\pi \mid u+\omega \mid -E_{g})^{n}}{2} \right\} + \frac{j8\pi^2 \sigma}{c^2 A} \left\{ \frac{(\pi \mid u \mid -E_{g})^{n} - (\pi \mid u+\omega \mid -E_{g})^{n}}{(\pi \mid u \mid -E_{g})^{n} - (\pi \mid u+\omega \mid -E_{g})^{n}} \right\}
\right)^x \exp \left[ \left( \frac{A(\pi \mid u \mid -E_{g})^{n} + A(\pi \mid u+\omega \mid -E_{g})^{n}}{2} \right) + \frac{j8\pi^2 \sigma}{c^2 A} \left\{ \frac{(\pi \mid u \mid -E_{g})^{n} - (\pi \mid u+\omega \mid -E_{g})^{n}}{(\pi \mid u \mid -E_{g})^{n} - (\pi \mid u+\omega \mid -E_{g})^{n}} \right\} \right] \ \text{du} \\
& \quad \text{Let,} \quad H(u) = \sqrt{\frac{q_e}{\hbar}} \left( \frac{1}{\sqrt{u}} \right) \frac{ju\sqrt{32\pi \sigma cA(\pi \mid u \mid -E_{g})^{n}}}{ju[8\pi \sigma + 2cA(\pi \mid u \mid -E_{g})^{n} + c^2 A^2(\pi \mid u \mid -E_{g})^{n}]} \\
& \quad F_1(u, [u+\omega]) = \frac{A(\pi \mid u \mid -E_{g})^{n} + A(\pi \mid u+\omega \mid -E_{g})^{n}}{2} + \frac{j8\pi^2 \sigma}{c^2 A} \left\{ \frac{(\pi \mid u \mid -E_{g})^{n} - (\pi \mid u+\omega \mid -E_{g})^{n}}{(\pi \mid u \mid -E_{g})^{n} - (\pi \mid u+\omega \mid -E_{g})^{n}} \right\} \\
& \quad G_1(\omega, x) = \exp \left[ -x \left( \frac{j \omega \tau_n + 1}{D_n \tau_n} \right) \right]^{1/2}
\end{align*}
\]
and
\[ s(u) = E_i(u) \]  
(5.77)

thus
\[
i(\omega, x) = G_a A_d G_1(\omega, x) \int_{-\infty}^{\infty} H^*(u) H(u+\omega) F_1(u, [u+\omega]) \exp[-F(u, [u+\omega])x]S^*(u)S(u+\omega)du
\]
(5.78)

\( G_1(\omega, x) \) may be viewed as an electrical filter. It is the bandwidth limit, fixed by the carrier lifetime, and is similar to the vacuum tube transit-time frequency limit [4].

\( G_1(\omega, x) \) is not the only limit in the SPD electrical frequency response however. The diode depletion region forms a capacitor, within the diode, which interacts with the ohmic resistance of the semiconductor, outside the depletion region, and the metallic electrode contacts [4]. The depletion-region-capacitance usually limits the SPD bandwidth further than the carrier lifetime [4].

Kingston [4] derives an expression for the junction capacitance per unit area, in terms of the SPD properties and bias voltage,

\[
\frac{C}{A_d} = \sqrt{\frac{\varepsilon_d q_e}{2(V_0-V_B)\left[\frac{1}{N_A} + \frac{1}{N_B}\right]}}
\]
(5.79)

Where \( A_d \) is the detector junction cross sectional area, \( \varepsilon_d \) is the depletion permittivity, \( V_B \) is the forward bias voltage, \( V_0 \) is the junction potential, \( N_A \) and \( N_B \) are the acceptor and donor impurity densities, respectively.
The junction capacitance is connected to the device terminals via the semiconductor and lead resistances. The leads are in turn connected to a resistive load, which is usually the input impedance of an amplifier, it is denoted $R_L$.

Let $R_n$ and $R_p$ denote the resistance of the N-type and P-type regions, between the edge of the depletion region and the terminals, respectively. In which case, the equivalent circuit giving the junction capacitance frequency limit is given by Fig 5.9 [40].

Fig 5.9 Equivalent Circuit for Junction Capacitance Frequency Limit
The voltage across the load resistor is given by,

\[ V_L(\omega) = \frac{j\omega[R_n + R_p + R_L]}{j\omega(R_n + R_p + R_L) + 1} i_L(\omega) \]  

(5.80)

and the load current, \(i_L(\omega)\), by

\[ i_L(\omega) = \frac{j\omega(R_n + R_p) + 1}{j\omega(R_n + R_p + R_L) + 1} \]

(5.81)

The junction capacitance filter-function is therefore,

\[ G_2(\omega) = \frac{j\omega(R_n + R_p) + 1}{j\omega(R_n + R_p + R_L) + 1} \]

(5.82)

The total electrical filtering function is \(G(\omega, x)\), where,

\[ G(\omega, x) = G_1(\omega, x)G_2(\omega) \]

(5.83)

The total time varying current, generated by a time varying optical signal, at a distance \(x\) from the junction, is

\[ i(\omega, x) = G_A d G(\omega, x) \int_{-\infty}^{\infty} H^*(u)H(u+\omega)F_1(u, [u+\omega])exp[-F(u, [u+\omega])x]S^*(u)S(u+\omega)du \]

(5.84)

The total SPD current is the sum of each individual current contributions, in the region of absorption, thus,

\[ i(\omega) = G_A d \int_{0}^{\infty} G(\omega, x) \int_{-\infty}^{\infty} H^*(u)H(u+\omega)F_1(u, [u+\omega])exp[-F_1(u, [u+\omega])x]S^*(u)S(u+\omega)du dx \]

(5.85)

\[ = G_A d G_2(\omega) \int_{-\infty}^{\infty} H^*(u)H(u+\omega) \frac{F_1(u, [u+\omega]) \sqrt{\frac{\sigma_n}{\tau_n}}} {F_1(u, [u+\omega]) \sqrt{\frac{\sigma_n}{\tau_n}} + \frac{1}{\sqrt{j\omega\tau_n + 1}}} S(u)S(u+\omega) du dx \]

(5.86)
Let

\[ F(u, \omega, [u+\omega]) = \frac{F_1(u, [u+\omega]) \sqrt{D_n \tau_n}}{F_1(u, [u+\omega]) \sqrt{D_n \tau_n} + \sqrt{3\omega \tau_n + 1}} \]  

\[ = \left( \sqrt{D_n \tau_n} \right) \left\{ c^2 A^2(\tau | u | -E_g)^n(\tau | u+\omega | -E_g)^n \right\} \]

\[ + j16\pi^2 \alpha \left[ (\tau | u+\omega | -E_g)^n (\tau | u | -E_g)^n \right] \]

\[ \phi_n \left( \tau_n \right) \left( \sqrt{D_n \tau_n} \right) \left\{ c^2 A^2(\tau | u | -E_g)^n(\tau | u+\omega | -E_g)^n \right\} \]

\[ + j16\pi^2 \alpha \left[ (\tau | u+\omega | -E_g)^n (\tau | u | -E_g)^n \right] + 2c^2 A(\tau | u+\omega | -E_g)^n(\tau | u+\omega | -E_g)^n \sqrt{3\omega \tau_n + 1} \]

thus

\[ i(\omega) = G_2(\omega) \int_{-\infty}^{\infty} H^* (u) H(u+\omega) F(u, \omega, [u+\omega]) \Phi(u) S(u) S(u+\omega) du \]
5.2 NOISE PROCESSES IN DARK OPTICAL DETECTORS

SPD's are usually operated at room temperature, where a certain number of electrons can gain sufficient energy to cross the energy gap, producing two charge carriers which are available for conduction [1],[2],[3],[4],[35],[39]. Statistically thermal charge carrier generation is random, therefore random current fluctuations occur due to the thermal charge carrier generation process [1],[2],[3],[4],[35],[39]. Uncertainty associated with these current fluctuations produces light intensity independent noise at the detector output terminals, and corresponding bit errors in digital systems [5],[6].

Section 5.2.1 reviews the thermal charge carrier generation process giving rise to a mean square leakage or dark current. Associated with this current, is a noise which is known as noise. It is similar in nature to the well known noise processes of many vacuum tube and semiconductor devices [4],[6],[44],[45]. Dark current associated noise is reviewed in section 5.2.2.

Associated with the bonding of wires to the crystal, of the SPD, is a resistance. This is also a source of noise due to the random nature of phonon-electron collisions; it is known as Johnson noise [4],[6],[44],[45].

Johnson noise is present in any realistic circuit. It is dependent upon resistance and absolute temperature of the circuit [4],[6],[44],[45]. Johnson noise is present in SPD's and is therefore reviewed in section 5.2.3.
5.2.1 Dark Current

Thermal Generation of charge carriers, within the SPD, produces a leakage current which is present even if the detector is darkened [4],[45]. Associated with this current is a noise that appears at the output of the detector.

Recall differential equation (5.17), which associates the rate of change of diffusion gradient with distance away from the SPD junction.

\[
D_p \frac{d^2 p}{dx^2} - \frac{p}{\tau} = 0
\]

In this case \( p \) is the excess charge carriers density in the N-type region and has the general solution,

\[
p = c_1 \exp \left[ \frac{x}{\sqrt{D_p \tau_p}} \right] + c_2 \exp \left[ -\frac{x}{\sqrt{D_p \tau_p}} \right] \tag{5.90}
\]

The total charge carrier density is therefore the sum of the excess and equilibrium charge carrier densities [4],

\[
\hat{p} = p_0 + c_1 \exp \left[ \frac{x}{\sqrt{D_p \tau_p}} \right] + c_2 \exp \left[ -\frac{x}{\sqrt{D_p \tau_p}} \right] \tag{5.91}
\]

Similarly the total charge carrier density in the P-type region is,

\[
\hat{n} = n_0 + c_3 \exp \left[ \frac{x}{\sqrt{D_n \tau_n}} \right] + c_4 \exp \left[ -\frac{x}{\sqrt{D_n \tau_n}} \right] \tag{5.92}
\]

The boundary conditions for the SPD, shown in Fig 5.9, determine the values of the coefficients, \( c_1, c_2, c_3, c_4 \). At the junction, that is where \( x=0 \), the minority carrier densities are given by

\[
n = n_0 \exp \left[ \frac{qE}{kB} \right] \tag{5.93}
\]
and

\[ p = p_0 \exp\left(\frac{qV_B}{KT}\right) \tag{5.94} \]

Where \( V_B \) is the applied bias voltage, \( K \) is Boltzman's constant, \( T \) is absolute temperature, \( n_0 \) and \( p_0 \) are the zero bias equilibrium charge carrier densities in the \( P \) and \( N \)-type regions respectively.

At the edges of the \( N \)-type and \( P \)-type layers, the charge carrier diffusion gradients must be zero, as the metallic contacts for the electrodes are connected at this point [4].

Let \( l_n \) and \( l_p \) represent the thickness of the \( N \)-type and \( P \)-type regions respectively, in which case the boundary conditions give the minority carrier densities, in the \( N \)-type and \( P \)-type regions, as,

\[ \hat{p} = p_0 + p_0 \left( \exp\left[\frac{qV_B}{KT}\right] - 1 \right) \left( \frac{\exp\left[\frac{x}{l_0}\right] + \exp\left[-\frac{2l_n}{l_0}\right]}{1 + \exp\left[-\frac{2l_n}{l_0}\right]} \right) \tag{5.95} \]

and

\[ \hat{n} = n_0 + n_0 \left( \exp\left[\frac{qV_B}{KT}\right] - 1 \right) \left( \frac{\exp\left[\frac{x}{l_0}\right] + \exp\left[-\frac{2l_p}{l_0}\right]}{1 + \exp\left[-\frac{2l_p}{l_0}\right]} \right) \tag{5.96} \]

However, the thickness of the \( P \)-type region of Fig 5.8 is many diffusion lengths and is, in fact, infinite for all practical purposes, therefore (5.96) may be rewritten as [4],

\[ \hat{n} = n_0 + n_0 \left( \exp\left[\frac{qV_B}{KT}\right] - 1 \right) \left( \exp\left[-\frac{x}{l_0}\right] \right) \tag{5.97} \]
Using the diffusion equations, (5.15) and (5.16), currents, due to the minority carriers in each region, may be found [4]. From which an expression for the total current may be derived,

\[ i = A_d \left( \frac{q_D p_0}{l_0} \left( \frac{\exp \left( \frac{-2l_n}{l_0} \right) - 1}{\exp \left( \frac{-2l_n}{l_0} \right) + 1} \right) \frac{q_D n_0}{l_0} \right) \times \left( \exp \left[ \frac{q e V_B}{kT} \right] - 1 \right) \]  

(5.98)

For large reverse bias voltages, the total current is independent of applied voltage; becoming the device dark current \( i_D \) [4].

\[ i_D = A_d \left( \frac{q_D p_0}{l_0} \left( \frac{\exp \left( \frac{-2l_n}{l_0} \right) - 1}{\exp \left( \frac{-2l_n}{l_0} \right) + 1} \right) \frac{q_D n_0}{l_0} \right) \]  

(5.99)

The total current may be written in terms of the dark current \( i_D \), which flows in the opposite direction to a forward bias current, thus,

\[ i = i_D - i_D \exp \left[ \frac{q e V_B}{kT} \right] \]  

(5.100)

5.2.2. Shot Noise

Associated with the dark current, of an SPD, is a shot noise, which is due to the uncertainty of electron arrival times at the junction of the SPD. Shot noise obeys Poisson statistics and is observed as a flat power spectrum up to frequencies approaching the reciprocal of the carrier lifetime [6],[14], that is,
\[ \omega_{SN0} \sim \frac{2\pi}{T_p} \quad \text{or} \quad \omega_{SN0} \sim \frac{2\pi}{T_n} \]

Where \( \omega_{SN0} \) is the shot noise spectral density cutoff frequency.

It is also possible to view the spectral density as a filtered spectral density which stretches from \(-\infty\) to \(+\infty\). The power spectrum below \( \omega_{SN0} \) is given by

\[ n_s(\omega) = 2q_e |i_{dc}| \quad (5.101) \]

Where \( i_{dc} \) is the average, or direct, current.

The shot noise corresponding to a total current flow of \( i \), given by (5.100) is [4],[6],

\[ n_s(\omega) = q_e \left( |i_o| + |i_o| \exp \left[ \frac{q_e V_B}{kT} \right] \right) \quad (5.102) \]

The conductance at the bias voltage \( V_B \) is found by differentiating (5.93) thus

\[ G_D = \frac{d}{dV_B} = q_e \frac{\exp \left[ \frac{q_e V_B}{kT} \right]}{kT} \quad (5.103) \]

At zero bias [4],

\[ G_D^0 = \frac{q_e}{kT} \quad (5.104) \]

and the zero bias shot noise is [4],[6],

\[ n_s(\omega) - 2q_e |i_D| \quad (5.105) \]

Substituting for \( q_e \) in (5.105), using (5.104), gives [4],

\[ n_s(\omega) = 2G_D^0 kT \quad (5.106) \]

which is the same as a double sided spectral density due to Johnson noise [5],[6].
The mean square current is found by integrating the filtered spectral density, thus [6]

\[ i_{SN}^2 = \int_{-\infty}^{\infty} n_s(\omega) G_g(\omega) |^2 d\omega \]  

(5.107)

Where \( G_g(\omega) \) is a general filter function.

Therefore, for an electrical filter function \( G_2(\omega) \), the shot noise power is [4],[6]

\[
\begin{align*}
    i_{SN}^2 &= q_e \left| i_d \right|^2 \left| \exp\left[ \frac{q_e V_B}{KT} \right] \right|^2 \\
    &= q_e \left| i_d \right|^2 \left( \exp\left[ \frac{q_e V_B}{KT} \right] \right)^2 \\
    &= q_e \left| i_d \right|^2 \left( \exp\left[ \frac{q_e V_B}{KT} \right] \right)^2 \frac{\omega^2 r_D + 1}{\omega^2 (r_D + r_L) + 1} d\omega \left( \frac{\omega^2 r_D + 1}{\omega^2 (r_D + r_L) + 1} \right)
\end{align*}
\]

(5.108)

Which is unfortunately infinite. Generally however the designer will limit the frequency range to the -3dB points of the electrical filter, or integrate (5.108) over the bandwidth of a post detector electrical filter.

5.2.3 Thermal Noise

In addition to the shot noise derived within the actual semiconductor, noise is generated in the contact resistance and lead resistance. Stremmler [6] gives an expression for the power spectral density of the thermal noise,

\[ n_{TH}(\omega) = \frac{2\pi \omega}{\exp\left[ \frac{\pi \omega}{KT} \right] - 1} \]

(5.109)

in watts per Hertz per ohm.
When
\[ |\omega| \ll \frac{KT}{\hbar} = 38.32 \text{ THz} \] (5.110)

\( n_{TH}(\omega) \) may be approximated by,
\[ n_{TH}(\omega) \approx 2KT \] (5.111)

Which is valid for a double sided spectrum. Again \( n_{TH}(\omega) \) is filtered by the electrical filter function as in the shot noise case, thus
\begin{equation}
I_{JN}^2(\omega) = 2KG_c \int_{-\infty}^{\infty} |G_2(\omega)|^2 d\omega \\
= 2KG_c H_{JN}
\end{equation} (5.112)

Generally, the thermal or Johnson noise contribution is small compared with the shot noise sources, discussed in the preceding section [6].
5.3 LIGHT DEPENDENT NOISE PROCESSES

In addition to the thermal noise sources of the darkened optical detector, there are noises associated with the presence of light, at the detector [4],[7],[8],[9],[10],[11],[12],[36]. In many treatments of optical detection, these are called shot noise [4],[7],[8],[9],[10],[11], [12],[36]. This nomenclature is suitable for idealized detectors, which have rectangular bandpass and zero phase shift, as the subsequent analysis will show. However when the detector is non ideal, the statistics associated with optical shot noise will differ from those of electrical shot noise. Therefore, in order to avoid confusion, the optical shot noise is referred to as Optical Granular Noise (OGN).

Optical granular noise (OGN) will be analysed in section 5.3.1 using the characteristics of light derived in chapter 3. The reasons for its more common name, shot noise, will be discussed in more detail in section 5.3.2, and the nomenclature used in section 5.3.1 justified.

5.3.1 Fluctuations in Light Intensity - Optical Granular Noise

In section 3.2.1, expressions for photon number and phase uncertainty were derived. Uncertainties in the number of photons arriving at the detector, in a given time, produce uncertainties in the output current, which are experienced as noise.
If $n_p$ photons arrive at a detector, the root mean square uncertainty in that number is given by $\Delta n_p$ [8],

$$\Delta n_p = \sqrt{n_p}$$

Therefore, for a given photon flux,

$$\Phi_p = \frac{p}{\hbar \omega}$$

(5.113)

the root mean square uncertainty in the photon flux is,

$$\Delta \Phi_p = \sqrt{\frac{p}{\hbar \omega}}$$

(5.114)

Which may be written as an uncertainty in power,

$$\Delta P = \sqrt{\frac{\hbar \omega}{p}}$$

(5.115)

Optical power may be written in terms of the radiation electric field, and the conductance of free space [41],

$$\Delta E^* \Delta G_a A_d = \sqrt{\frac{\hbar \omega}{G_a A_d}}$$

(5.116)

It is therefore possible to associate an uncertainty in photon number, with the magnitude of an electric field

$$\Delta E = \left(\frac{\hbar \omega E^2}{G_a A_d}\right)^{1/4}$$

(5.117)

Which represent a spectral density of uncertainty in volts/m$\sqrt{\text{Hz}}$. 

The detector input signal, is,

\[ G_{\text{in}} = \sqrt{E_{\text{det}}^2} = \sqrt{G \omega_1 E^2} \]  

(5.118)

The corresponding current distribution, over the correlation width of the detector, is,

\[ i(W'X) = \int_{-\infty}^{\infty} H(u)H(u+\omega)F_1(u,[u+\omega])\exp[-F_1(u,[u+\omega])du \]  

(5.119)

The corresponding power is then,

\[ i^2(\omega,x) = R_1 \omega_1 E^2 G A_d \int_{-\infty}^{\infty} H(u)H(u+\omega)F_1(u,[u+\omega])\exp[-F_1(u,[u+\omega])du \]  

(5.120)

Therefore, at any distance \( x \), the power spectral density is found by differentiating (5.120) [36],[44],

\[ n_{\text{GN}}(\omega,x) = R_1 \omega_1 E^2 G A_d \left\{ \frac{1}{\omega_0} \int_{-\infty}^{\infty} H(u)H(u+\omega)F_1(u,[u+\omega])\exp[-F_1(u,[u+\omega])du \right\} \times \left\{ \frac{1}{\omega_0} \int_{-\infty}^{\infty} H(u)H(u+\omega)F_1^*(u,[u+\omega])\exp[-F_1^*(u,[u+\omega])du \right\} \]  

(5.121)

Including the electrical filter function and the integrating over the spectrum.
The integral expression of (5.122) is, however, a constant and is evaluated only once, therefore the noise power $i^2(\omega) R_L$ may be written as,

$$i^2(\omega) R_L = R_L \hbar \omega_1 E^2 G_{\alpha d} A_d$$

If a number of optical signals are present, the noise power is the sum of the individual noise powers due to each signal.

5.3.2 Optical Granular Noise of an Ideal Photo-Receiver and its Association with Shot Noise

Consider now, a somewhat more ideal photodiode which has a rectangular optical bandwidth centred at $\omega_1$ and extending from $-\omega_1/2$ to $\omega_1/2$ [36]. Incident upon this detector is a light beam of electric field [41],

$$E = \sqrt{\frac{\hbar \omega_1}{G_{\alpha d}}}$$
The uncertainty in electric field is,

\[ \Delta E = \left( \frac{Pn\omega_1}{G\lambda d} \right)^{1/4} \]

\[ i(\omega) = \left( \frac{Pn\omega_1}{\hbar\omega_1} \right)^{1/2} q_e \left( \Delta \omega_1 + \omega_e \right) \]

and where \( \omega_e \) is electrical frequency.

\[ i^2(\omega_e) = \frac{Pq_e^2}{\hbar\omega_1} (\Delta \omega_1 + \omega_e)^2 \]  

The power spectral density is found by differentiating \( i^2(\omega_e) \) with respect to \( \omega_e \) [36],[44],

\[ \eta(\omega_e) = 2Pq_e^2 \left( \frac{\Delta \omega_1 + \omega_e}{\hbar\omega_1} \right) \]

but \( \frac{Pq_e \Delta \omega_1}{\hbar\omega_1} \) is the expression for the direct current due to a signal occupying a bandwidth of \( \Delta \omega_1 \). If the electrical filter is of bandwidth \( \Delta \omega_e \),

\[ i^2 = 2q_e i_{dc} \Delta \omega_e \]

This expression is the same as the expression for shot noise internally generated within the photo detector, for an optically generated direct current [6],

\[ i_{dc} = \frac{Pq_e}{\hbar\omega_1} \Delta \omega_1 \]

However the filter function of a real SPD colours the white OGN spectral density and produces a non white spectral density near the DC.
For detectors with small electrical bandwidths, the error introduced by the optical filter is small, however the effects in detectors with larger bandwidths cannot be ignored.
5.4 DETECTION OF AMPLITUDE MODULATED AND AMPLITUDE-SHIFT-KEYED SIGNALS

Amplitude modulated, and amplitude shift keyed signals, are currently the most common methods of signalling in commercial optical fibre systems [1],[2]. Detection of such signals is simple, requiring only one photodiode. Variations in intensity are translated directly into variations in detector output current. These are then amplified to a level where a logical 1 or 0 decision can be made [1],[2],[5],[6]. Any encryption and redundancy, present, is removed and the original data recovered for transmission into a local area network.

It is instructive to review amplitude detection processes. Their analysis provides insights into the detection of other types of modulation. The Amplitude Modulation (AM) detection process also illustrates the process of optical mixing, which takes place in the photo detector.

Initially the phase relationships between an AM carrier and sidebands is examined in section 5.4.1. Changes in phase relationships can reduce the maximum signal power that may be detected by the photo detector.

In section 5.4.2, the relationship between carrier levels and sideband power is examined and an expression for the maximum SNR derived.

Finally, in section 5.4.3, an alternative view of the photo detector, as a mixer, is discussed, together with ways of improving the detection SNR of an AM signal.
5.4.1 Phases of Carrier and Sidebands

An amplitude modulated signal may be described by the time domain
equation [6],

\[ S_{AM}(t) = A_c (1 + \mu m(t)) \cos \omega_c t \quad |m(t)| < 1 \]
\[ |\mu| < 1 \]

(5.129)

Where \( \omega_c \) is the angular frequency of the carrier, \( m(t) \) is the time
varying modulation, and \( \mu \) is the modulation depth.

In the frequency domain, an AM signal appears as a carrier and two
sidebands, as shown in Fig 5.10 [6].

**Fig 5.10 A Typical Amplitude Modulated Spectrum**

The lower sideband lies between \( \omega_c - \omega_m \) and \( \omega_c \), where \( \omega_m \) is
the maximum modulation frequency. The carrier is located at \( \omega_c \) and the
upper sideband lies between \( \omega_c \) and \( \omega_c + \omega_m \). For a double sided
spectrum the positive frequency axis is mirrored in the D.C. axis as shown
in Fig 5.10. The receiver is designed to accommodate the full AM spectrum,
its filters -3dB points are set at \( -\omega_m + \omega_c \) and \( \omega_m + \omega_c \). Signals at
these points are reduced by 3dB and will thus give the worst case signal to
noise ratio. Thus, the analysis of AM systems will be undertaken using
sinusoidal modulation at \( \omega_m \).
Amplitude shift keyed signals are similar. The maximum modulation frequency of a digital system occurs when an alternate 1-0 sequence is transmitted. This is a square wave. It produces a spectrum as shown in Fig 5.11.

The harmonics, of the square modulating wave, are disposed around the carrier frequency, however the essential information is contained between $\omega_m + \omega_c$ and $\omega_m + \omega_c$. For analysis purposes it is assumed that only the carrier and two sidebands at $\omega_m + \omega_c$ are of interest. Such restrictions produce a sinusoidally amplitude modulated wave, therefore the analyses of ASK and AM are the same.

The cosinuoidal modulation at $m$ is written as

$$S_{\text{ASK}} = A_c (1 + \mu \cos \omega_m t) \cos \omega_c t$$

$$= A_c \cos \omega_c t + \frac{A_c \mu \cos (\omega_c - \omega_m) t}{2} - \frac{A_c \mu \cos (\omega_c + \omega_m) t}{2}$$

All information is contained in the sidebands, thus for a total transmitted power $P_T$, the proportion of $P_T$ containing the information is [5],[6],
5.58

\[ P_I = \frac{\mu^2}{\mu^2 + 2} \]  

(5.131)

Where \( P_I \) is the information power.

Even if the carrier is modulated to a depth of 100%, the maximum information power is one third of the total transmitted power. Therefore, for a given channel to noise ratio the maximum SNR obtainable is 1/3 of that figure [5],[6]. It is obvious that this method of detection is non optimum. Subsequent sections will discuss the ways in which detection SNR may be improved.

The SNR of detected AM signals may be further degraded by changes in amplitude and phase between the sidebands. In chapter 4 it was shown that the AO modulator deflects each sideband, and the carrier, through different angles, in which case each may travel through a different path, receiving differing phase delays and amounts of attenuation. Introducing arbitrary phase delays and amounts of attenuation into an amplitude modulated signal.

\[ E(t) = E_c \cos c t + \alpha_1 E_c \frac{\mu \cos[(\omega_c - \omega_m) t - (\omega_c - \omega_m) T_1]}{2} \]

\[ + \alpha_2 E_c \frac{\mu \cos[(\omega_c + \omega_m) t - (\omega_c - \omega_m) T_2]}{2} \]

(5.132)

By rearranging (5.132), the phase delay may be introduced into the amplitude of an in-phase and quadrature modulation signal, thus

\[ E_m(t) = \frac{E_c^2 G_a A_d \mu}{2} (\alpha_1 \cos(\omega_c - \omega_m) T_1 + \alpha_2 \cos(\omega_c + \omega_m) T_2) \]

\[ \times \cos \omega_m t \]

\[ - (\alpha_1 \sin(\omega_c - \omega_m) T_1 - \alpha_2 \sin(\omega_c + \omega_m) T_2) \sin \omega_m t \]

(5.133)
In terms of amplitude and phase

\[ E_m(t) = E_c G_A d \mu \sqrt{ \frac{\alpha_2^2 + 2\alpha_1 \alpha_2 \cos[\omega_c (T_1 + T_2) + \omega_m (T_2 - T_1)] \cos(\omega_m t + \theta)}{2}} \]

where

\[ \cos \theta = \frac{\alpha_1 \cos(\omega_c \omega_m T_1 + \omega_c \omega_m T_2) + \alpha_2 \cos(\omega_c + \omega_m) T_2}{\alpha_2^2 + 2\alpha_1 \alpha_2 \cos[\omega_c (T_1 + T_2) + \omega_m (T_2 + T_1)]} \]

(5.134)

If \( \omega_c (T_1 + T_2) + \omega_m (T_2 - T_1) = \pi \), the amplitude of recovered modulation falls to 0.

From the above analysis, it is clear that the phase relationship of each sideband, to the carrier, is of greatest importance to the modulation recovery. When undisturbed, the sidebands mix with the carrier, in an envelope detector, producing two separate modulation signals which are added in phase to each other. When the two differ in phase, by different amounts from the carrier, the two modulation signals are added at a finite phase angle.

5.4.2 Envelope Detection of Amplitude Modulation

The detection of amplitude modulated optical signals is a simple process in practice. In the time domain, the AM wave, incident upon the SPD, produces a current which varies in sympathy with the variation in amplitude of the optical carrier [2],[4],[36]. Ideally, the detector output is independent of the modulation frequency. However in practice, the optical detector performs a number of filtering functions, as indicated by equation (5.89).
\[ i(\omega) = G_2(\omega) \int_{-\infty}^{\infty} H(u) H(u+\omega) F(u, \omega, [u+\omega]) S^*(u) S(u+\omega) du \]

The frequency dependence, of output current, is easier to analyse in the frequency domain, than in the time domain. Therefore, all subsequent analysis will be discussed in frequency domain.

An undistorted AM signal may be represented in the time domain by (5.129),

\[ S_{AM}(t) = A_c \cos \omega_c t + A_c \mu \cos(\omega_c - \omega_m) t + A_c \mu \cos(\omega_c + \omega_m) t \]

The Fourier transform of \( S_{AM}(t) \) is given by [6],[42]

\[
S_{AM}(\omega) = A_c \frac{\pi}{\omega_c + \omega_m} + A_c \frac{\pi}{\omega - \omega_c} + A_c \frac{\pi}{\omega + \omega_m} + A_c \frac{\pi}{\omega - \omega_c - \omega_m} + A_c \frac{\pi}{\omega_c + \omega_m} + A_c \frac{\pi}{\omega_c - \omega_m} \]

(5.135)

The product \( s(u) s(u+\omega) \) generates 36 terms, however only a small number of these will pass through \( G_2(\omega) \). Moreover, the power in the AM distortion terms is of no interest. Therefore the product terms of interest are,

\[
S^*_m(\omega) (u+\omega) = \frac{A_c^2 \pi^2 \mu}{2} \delta(u+\omega_c) \delta(u+\omega_c + \omega_m) \\
S^*_m(\omega) (u+\omega) = \frac{A_c^2 \pi^2 \mu}{2} \delta(u+\omega_c) \delta(u+\omega_c - \omega_m)
\]
\[ \frac{A_c^2 \pi^2 \mu \delta(u+\omega_c) \delta(u-\omega_m)}{2} \]

\[ \frac{A_c^2 \pi^2 \mu \delta(u+\omega_c) \delta(u+\omega_c + \omega_m)}{2} \]

\[ \frac{A_c^2 \pi^2 \mu \delta(u+\omega_c) \delta(u+\omega_c - \omega_m)}{2} \]

\[ \frac{A_c^2 \pi^2 \mu \delta(u-\omega_c) \delta(u+\omega_c + \omega_m)}{2} \]

\[ \frac{A_c^2 \pi^2 \mu \delta(u+\omega_c) \delta(u+\omega_c - \omega_m)}{2} \]

\[ \frac{A_c^2 \pi^2 \mu \delta(u-\omega_c) \delta(u-\omega_c + \omega_m)}{2} \]  

(5.136)

Each term in the series may be integrated individually, the autocorrelation of the first term becomes

\[ i(\omega) = \frac{A_c^2 \pi \mu G_2(\omega)}{2} \int_{-\infty}^{\infty} H(u) H(u+\omega) F(u, \omega, [u+\omega]) \delta(u-\omega_c) \delta(u+\omega_c + \omega_m) du \]  

(5.137)

\[ = \frac{A_c^2 \pi \mu G_2(\omega)}{4} \left[ H(\omega-\omega_c) H(\omega+\omega_c) F(\omega_c, \omega, [\omega+\omega_c]) + H(\omega+\omega_c) H(\omega-\omega_c) F(\omega_c, \omega, [\omega+\omega_c]) \right] \]

The total detector current at the modulation frequencies \( \pm \omega_m \) is

\[ i(\omega) = \frac{A_c^2 \pi \mu G_2(\omega)}{4} \left[ H(\omega-\omega_c) H(\omega+\omega_c) F(\omega_c, \omega, [\omega+\omega_c]) + H(\omega+\omega_c) H(\omega-\omega_c) F(\omega_c, \omega, [\omega+\omega_c]) \right] \nonumber \]

\[ \left[ \delta(\omega+\omega_m) + \delta(\omega-\omega_m) \right] \]  

(5.138)
The signal power is given by,

\[ i^2(\omega)R_L = \left( \frac{\mu_A R_L}{4} \right) G_2(\omega_m)G_2^*(\omega_m) \]  \hspace{1cm} (5.139)

\[
\begin{align*}
&H^*(-\omega_c)H(-\omega_c)H^*(\omega_m+\omega_c)H(\omega_m-\omega_c)F(-\omega_c,\omega_m,[\omega_m+\omega_c])F^*(-\omega_c,\omega_m,[\omega_m+\omega_c]) \\
&+H^*(-\omega_c)H(\omega_c)H^*(\omega_m-\omega_c)H(\omega_m+\omega_c)F(-\omega_c,\omega_m,[\omega_m-\omega_c])F^*(-\omega_c,\omega_m,[\omega_m+\omega_c]) \\
&+H^*(-\omega_c)H(\omega_m-\omega_c)H^*(\omega_m+\omega_c)F(-\omega_c,\omega_m,[\omega_m+\omega_c])F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H^*(-\omega_c)H^*(\omega_m-\omega_c)H^*(\omega_m+\omega_c)F(-\omega_c,\omega_m,[\omega_m-\omega_c])F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H^*(-\omega_c)H^*(\omega_m-\omega_c)H^*(\omega_m+\omega_c)F(-\omega_c,\omega_m,[\omega_m+\omega_c])F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H^*(\omega_c)H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m+\omega_c)F(-\omega_c,\omega_m,[\omega_m+\omega_c])F^*(-\omega_c,\omega_m,[\omega_m-\omega_c]) \\
&+H^*(\omega_c)H(\omega_m+\omega_c)H^*(\omega_m+\omega_c)F(\omega_c,\omega_m,[\omega_m+\omega_c])F^*(\omega_c,\omega_m,[\omega_m+\omega_c]) \\
&+H^*(\omega_c)H^*(\omega_m+\omega_c)H(\omega_m-\omega_c)F(\omega_c,\omega_m,[\omega_m+\omega_c])F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H^*(\omega_c)H^*(\omega_m+\omega_c)H^*(\omega_m+\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H(\omega_c)H(-\omega_c)^2H^*(\omega_m+\omega_c)F(\omega_c,\omega_m,[\omega_m+\omega_c])F^*(-\omega_c,\omega_m,[\omega_m+\omega_c]) \\
&+H(\omega_c)H(-\omega_c)^2H^*(\omega_m+\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H(\omega_c)H(-\omega_c)^2H^*(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(\omega_c)H(-\omega_c)^2H(\omega_m+\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H(\omega_c)H(-\omega_c)^2H(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m-\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(-\omega_c,\omega_m,[\omega_m+\omega_c]) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m-\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m-\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(-\omega_c,\omega_m,[\omega_m+\omega_c]) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m-\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m-\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(-\omega_c,\omega_m,[\omega_m+\omega_c]) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m-\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m-\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(-\omega_c,\omega_m,[\omega_m+\omega_c]) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m-\omega_c)F(\omega_m-\omega_c,\omega_m,\omega_c)F^*(\omega_m-\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H^*(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c) \\
&+H(-\omega_c)H(\omega_m+\omega_c)H(\omega_m+\omega_c)F(\omega_m+\omega_c,\omega_m,\omega_c)F^*(\omega_m+\omega_c,\omega_m,\omega_c)
\[ H(-\omega_c)H^*(\omega_c)H(\omega_c-\omega_m)H^*(-\omega_c-\omega_m)F([-\omega_c-\omega_m],\omega_m,-\omega_c)F^*([\omega_c-\omega_m],\omega_m,\omega_c) \]
\[ H(-\omega_c)H^*(-\omega_c)H(-\omega_c-\omega_m)F([-\omega_c-\omega_m],\omega_m,-\omega_c)F^*([\omega_c-\omega_m],\omega_m,-\omega_c) \]

because
\[ H^*([\omega_c-\omega_m],\omega_m,\omega_c)=H(\omega_c,-\omega_m,[-\omega_c-\omega_m]) \]

If
\[ H(\omega_c+\omega_m)\approx H(\omega_c), \text{ and, } F((\omega_c-\omega_m),\omega_m,\omega_c)\approx F(\omega_c,\omega_m,\omega_c) \]

the power becomes,
\[
\begin{align*}
i^2(\omega)R_L &= A_c^4\mu^2 R_L G_2(\omega_m)G_2^*(\omega_m)H^*(\omega_c)H(\omega_c)H(\omega_c)F(\omega_c,\omega_m,\omega_c)F^*(\omega_c,\omega_m,\omega_c) \\
&= \frac{8}{4}G_aA_d^2 \mu^2 R_L G_2(\omega_m)G_2^*(\omega_m)H^*(\omega_c)H(\omega_c)H(\omega_c)F(\omega_c,\omega_m,\omega_c)F^*(\omega_c,\omega_m,\omega_c) \\
&= R_L \frac{8}{4}G_aA_d^2 \omega_c(2+\mu^2)H_{\text{OGN}} \\
&= R_L \frac{8}{4}G_aA_d \omega_c(2+\mu^2)H_{\text{OGN}} \\
&= \frac{5.140}{(5.142)}
\end{align*}
\]

The assumption, made above, is valid for photo detectors at modulation frequencies that are many orders of magnitude lower than the frequencies of light [4],[36]. This is generally the case. However, if the light frequencies are found in the far infra-red and optical frequencies in the hundred GigaHertz range, the actual frequencies of the incoming light will be significant [46].

The OGN is calculated by summing the contributions of each delta function in the frequency domain using (5.123).
\[
\begin{align*}
i^2(\omega)R_L &= \frac{\omega c(2+\mu^2)}{4}H_{\text{OGN}} \\
&= \frac{5.140}{(5.142)}
\end{align*}
\]

for \( \omega_m \ll \omega_c \), which is usually the case.
The electrical shot noise contribution is found using (5.108)

\[ i_{SN}^2(\omega) = 2R_L q_e I_D \left( 1 + \exp \left[ \frac{q_e V_B}{kT} \right] \right) H_{SN} \]  

(5.143)

The Johnson noise contribution, from contact and lead resistance, is

\[ i_{JN}^2(\omega) = 2kT G_{c, L} H_{JN} \]  

(5.144)

Using (5.141), (5.142), (5.143) and (5.144), the AM signal to noise ratio is calculated.

\[
\text{SNR}_{AM} = \frac{4E^2 G_A d \mu^2 G_1(\omega_m) G_2(\omega_m)^* H(\omega_c)^* H^{*}(\omega_c) H(\omega_c) H(\omega_c) F(\omega_c, \omega_m, \omega_c) F^*(\omega_c, \omega_m, \omega_c)}{\pi E^2 G_A d \omega_c (2+\mu^2) H_{GN} + 4q_e I_D \left( 1 + \exp \left[ \frac{q_e V_B}{kT} \right] \right) H_{SN} + kT G_{c, L} H_{JN}}
\]  

(5.145)

5.4.3 The Photo-Detector as a Mixer

From the discussions in the preceding section, it is clear that detection of AM is a mixing process. The desired signal is generated by the mixing of each sideband with the carrier. The two mixer products are then summed to produce the modulating signal at the baseband. Other mixer products, either fall outside the electrical response of the SPD and are absent at the output, or they are distorting signals of significantly smaller power than the desired signal at modulation depths below 100%. Such distorting signals may be ignored at modulation depths below 100% [5],[6], because the power in the desired signals is proportional to twice the square of the modulation depth [5],[6], that is,

\[ P_m \propto 2\mu^2 \]  

(5.146)
and power in the distorting signal is proportional to the fourth power of the modulation index [5],[6], that is,

\[ p \propto \mu^4 \]  

(5.147)

The mixing property of AM recovery indicates that the photodiode may be used as a mixer of other optical signals. By exploiting the mixing properties of the photo detector, methods of improving the SNR of detection using coherent detection, and other modulation types, such as frequency and phase shift keying, that give better signal to noise performance than AM, may be employed.

In section (5.4.1), it was found that the optimum detection efficiency obtainable for an AM system is 1/3, that is, the ratio of power in the sidebands to total transmitted power is 1/3. By selecting a modulation scheme which exhibits a detection efficiency of unity, the regenerator spacings, on optical fibre communications systems, could be increased by a factor of 3.

The mixing process involved in the demodulation of AM were discussed in the preceding section. The carrier and sideband amplitudes of AM are fixed through the modulation index, therefore detection of AM is a rather specific case of mixing. Mixing properties of the photo detector may be more generally understood if the SNR of the mixer products, of two general light signals, is considered.

Let

\[ S_1(t) = E_1 \sqrt{A_d} \cos(\omega_1 t - \theta_1) \]  

(5.148)
represent two arbitrary signals, incident upon SPD.

Using elementary trigonometry, these may be written in terms of in-phase and quadrature components.

\[ S_1(t) = E_1 \sqrt{G_A d} \cos \theta_1 \cos \omega_1 t + E_1 \sqrt{G_A d} \sin \theta_1 \sin \omega_1 t \]  

(5.150)

and

\[ S_2(t) = E_2 \sqrt{G_A d} \cos \theta_2 \cos \omega_2 t + E_2 \sqrt{G_A d} \sin \theta_2 \sin \omega_2 t \]  

(5.151)

The Fourier transform of these signals may be written as [42],

\[ S_1(\omega) = \pi E_1 \sqrt{G_A d} (\cos \theta_1 - j \sin \theta_1) \delta(\omega - \omega_1) + \pi E_1 \sqrt{G_A d} (\cos \theta_1 - j \sin \theta_1) \delta(\omega + \omega_1) \]

\[ = A_1 \delta(\omega - \omega_1) + A_1^* \delta(\omega + \omega_1) \]  

(5.152)

where

\[ A_1 = \pi E_1 \sqrt{G_A d} (\cos \theta_1 - j \sin \theta_1) \]

and

\[ S_2(\omega) = A_2 \delta(\omega - \omega_2) + A_2^* \delta(\omega + \omega_2) \]  

(5.153)

where

\[ A_2 = \pi E_2 \sqrt{G_A d} (\cos \theta_2 - j \sin \theta_2) \]

Using equation (5.89) the current is derived,

\[ i(\omega) = \frac{G_2(\omega) A_1^* A_1^*(\omega_1) H(\omega + \omega_1) F(\omega_1, \omega, \omega + \omega_1, \delta(\omega)}}{2\pi} \]
\[ +A_1^*A_1H^*(\omega_1)H(\omega+\omega_1)F(\omega_1, \omega, [\omega+\omega_1])\delta(\omega+2\omega_1) \]

\[ +A_1^*A_2H^*(\omega_1)H(\omega+\omega_1)F(\omega_1, \omega, [\omega+\omega_1])\delta(\omega+\omega_2+\omega_1) \]

\[ +A_1^*A_2^*H^*(\omega_1)H(\omega+\omega_1)F(\omega_1, \omega, [\omega+\omega_1])\delta(\omega+\omega_2+\omega_1) \]

\[ +A_1A_1^*H^*(\omega_1)H(\omega-\omega_1)F(-\omega_1, \omega, [\omega+\omega_1])\delta(\omega-2\omega_1) \]

\[ +A_1A_1^*H^*(\omega_1)H(\omega-\omega_1)F(-\omega_1, \omega, [\omega+\omega_1])\delta(\omega) \]

\[ +A_1A_2^*H^*(\omega_1)H(\omega-\omega_1)F(-\omega_1, \omega, [\omega+\omega_1])\delta(\omega_1-\omega_2) \]

\[ +A_1A_2^*H^*(\omega_1)H(\omega-\omega_1)F(-\omega_1, \omega, [\omega+\omega_1])\delta(\omega-\omega_1+\omega_2) \]

\[ +A_1A_2^*H^*(\omega_1)H(\omega+\omega_2)F(\omega_2, \omega, [\omega+\omega_1])\delta(\omega_2+\omega_1) \]

\[ +A_1^*A_2^*H^*(\omega_1)H(\omega+\omega_2)F(\omega_2, \omega, [\omega+\omega_1])\delta(\omega+\omega_2+\omega_1) \]

\[ +A_1^*A_2^*H^*(\omega_2)H(\omega+\omega_2)F(\omega_2, \omega, [\omega+\omega_1])\delta(\omega) \]

\[ +A_2^*A_2^*H^*(\omega_2)H(\omega+\omega_2)F(\omega_2, \omega, [\omega+\omega_2])\delta(\omega+2\omega_2) \]

\[ +A_1A_2^*H^*(\omega_1)H(\omega-\omega_2)F(-\omega_2, \omega, [\omega-\omega_2])\delta(\omega-\omega_2-\omega_1) \]

\[ +A_1^*A_2H^*(\omega_2)H(\omega-\omega_2)F(-\omega_2, \omega, [\omega-\omega_2])\delta(\omega-\omega_2+\omega_1) \]

\[ A_2^*A_2H^*(\omega_2)H(\omega-\omega_2)F(-\omega_2, \omega, [\omega-\omega_2])\delta(\omega-2\omega_2) \]

\[ A_2^*A_2H^*(\omega_2)H(\omega-\omega_2)F(\omega_2, \omega, [\omega-\omega_2])\delta(\omega) \]  (5.154)
The arguments of the impulse functions indicate that mixer products appear at DC, \( \pm |\omega_2 - \omega_1| \), \( \omega_2 + \omega_1 \), \( -\omega_2 - \omega_1 \), \( +2\omega_1 \), \( +2\omega_2 \). The signals at \( \pm (\omega_1 + \omega_2) \pm 2\omega_1 \) and \( \pm 2\omega_2 \), are removed by the electrical filter function. Moreover, the terms at D.C. are of little interest. The only remaining signals are those at the difference frequencies \( \pm |\omega_2 - \omega_1| \),

\[
\begin{align*}
\text{i}(\omega) &= A_1^* A_2 G_2^*(\omega_2 - \omega_1) \left[ H^*(\omega_1) H(\omega_2) F(\omega_1, [\omega_2 - \omega_1], \omega_2) + H^*(\omega_2) H(-\omega_1) \right] \\
&\quad \times F(-\omega_1, [\omega_2 - \omega_1], -\omega_1) \delta(\omega - \omega_1 - \omega_2) \\
&\quad + A_1^* A_2 G_2^*(\omega_1 - \omega_2) \left[ H^*(-\omega_1) H(-\omega_2) F(-\omega_1, [\omega_1 - \omega_2], \omega_2) + H^*(\omega_2) H(\omega_1) \right] \\
&\quad \times F(\omega_2, [\omega_1 - \omega_2], \omega_1) \delta(\omega - \omega_1 + \omega_2) \\
&= \frac{\text{i}^2(\omega) R_\text{i}}{8} \left[ G_2(\omega_2 - \omega_1) G_2^*(\omega_2 - \omega_1) + G_2(\omega_1 - \omega_2) G_2^*(\omega_1 - \omega_2) \right] \\
&\quad + H^*(\omega_1) H(\omega_1) H^*(\omega_2) H(\omega_2) F^*(\omega_1, [\omega_2 - \omega_1], \omega_2) F(\omega_1, [\omega_2 - \omega_1], \omega_2) \\
&\quad + H^*(\omega_1) H(-\omega_2) H^*(-\omega_1) H(\omega_2) F(\omega_1, [\omega_2 - \omega_1], \omega_2) F^*(-\omega_2, [\omega_2 - \omega_1], -\omega_1) \\
&\quad + H^*(\omega_2) H(\omega_1) H^*(\omega_2) H(-\omega_1) F(-\omega_2, [\omega_2 - \omega_1], -\omega_1) F^*(\omega_1, [\omega_2 - \omega_1], \omega_2) \\
&\quad + H^*(-\omega_2) H(-\omega_2) H^*(-\omega_1) H(-\omega_1) F(-\omega_2, [\omega_2 - \omega_1], -\omega_1) F^*(-\omega_2, [\omega_2 - \omega_1], -\omega_1) \\
&= \frac{\text{i}^2(\omega) R_\text{i}}{8} \left[ G_2(\omega_2 - \omega_1) G_2^*(\omega_2 - \omega_1) + G_2(\omega_1 - \omega_2) G_2^*(\omega_1 - \omega_2) \right] \\
&\quad + H^*(\omega_1) H(\omega_1) H^*(\omega_2) H(\omega_2) F^*(\omega_1, [\omega_2 - \omega_1], \omega_2) F(\omega_1, [\omega_2 - \omega_1], \omega_2) \\
&\quad + H^*(\omega_1) H(-\omega_2) H^*(-\omega_1) H(\omega_2) F(\omega_1, [\omega_2 - \omega_1], \omega_2) F^*(-\omega_2, [\omega_2 - \omega_1], -\omega_1) \\
&\quad + H^*(\omega_2) H(\omega_1) H^*(\omega_2) H(-\omega_1) F(-\omega_2, [\omega_2 - \omega_1], -\omega_1) F^*(\omega_1, [\omega_2 - \omega_1], \omega_2) \\
&\quad + H^*(-\omega_2) H(-\omega_2) H^*(-\omega_1) H(-\omega_1) F(-\omega_2, [\omega_2 - \omega_1], -\omega_1) F^*(-\omega_2, [\omega_2 - \omega_1], -\omega_1) \\
&= \frac{\text{i}^2(\omega) R_\text{i}}{8} \left[ G_2(\omega_2 - \omega_1) G_2^*(\omega_2 - \omega_1) + G_2(\omega_1 - \omega_2) G_2^*(\omega_1 - \omega_2) \right] \\
&\quad + H^*(\omega_1) H(\omega_1) H^*(\omega_2) H(\omega_2) F^*(\omega_1, [\omega_2 - \omega_1], \omega_2) F(\omega_1, [\omega_2 - \omega_1], \omega_2) \\
&\quad + H^*(\omega_1) H(-\omega_2) H^*(-\omega_1) H(\omega_2) F(\omega_1, [\omega_2 - \omega_1], \omega_2) F^*(-\omega_2, [\omega_2 - \omega_1], -\omega_1) \\
&\quad + H^*(\omega_2) H(\omega_1) H^*(\omega_2) H(-\omega_1) F(-\omega_2, [\omega_2 - \omega_1], -\omega_1) F^*(\omega_1, [\omega_2 - \omega_1], \omega_2) \\
&\quad + H^*(-\omega_2) H(-\omega_2) H^*(-\omega_1) H(-\omega_1) F(-\omega_2, [\omega_2 - \omega_1], -\omega_1) F^*(-\omega_2, [\omega_2 - \omega_1], -\omega_1) \\
&\quad \left[ (5.156) \right] \\
\end{align*}
\]

If \( H(\omega_1) \approx H(\omega_2) \) and \( F(\omega_1, [\omega_2 - \omega_1], \omega_2) \approx F(-\omega_2, [\omega_2 - \omega_1], -\omega_1) \) then,
The light dependent noise-power is found in a similar way to the AM case, (5.142)

\[ i^2_R = \frac{\mathbb{E}_1^2 + \mathbb{E}_2^2}{2} A_d G_{1G} R_1 G(\omega_2 - \omega_1) G^*(\omega_2 - \omega_1) H^*(\omega_1) H(\omega_1) H(\omega_1) H(\omega_1) \]

\[ x F^*(\omega_1, [\omega_2 - \omega_1], \omega_1) F(\omega_1, [\omega_2 - \omega_1], \omega_2) \quad (5.157) \]

The electrical sources of noise are unchanged from the AM case. Using (5.143), (5.144), (5.157) and (5.159), an expression for optical SNR can be derived.

\[ \frac{i^2}{\Omega_{GN}} R_l = \frac{R_l h G d (E_1^2 + E_2^2)}{2} H_{OGN} \quad (5.158) \]

But \( \omega_1 \approx \omega_2 \) thus

\[ \frac{i^2}{\Omega_{GN}} R_l = \frac{R_l h G d \omega_1}{2} (E_1^2 + E_2^2) H_{OGN} \quad (5.159) \]

The electrical sources of noise are unchanged from the AM case. Using (5.143), (5.144), (5.157) and (5.159), an expression for optical SNR can be derived.

\[ \text{SNR}_{\text{AM}} = 2\mathbb{E}_1^2 + 2\mathbb{E}_2^2 A_d G_2 G^*_2 (\omega_2 - \omega_1) G^*_2 (\omega_2 - \omega_1) H^*(\omega_1) H(\omega_1) F(\omega_1, \omega_2 - \omega_1, \omega_1) \]

\[ FR^*(\omega_1, [\omega_2 - \omega_1], \omega_1) \]

\[ A_d h G d \omega_1 (E_1^2 + E_2^2) H_{OGN} + 2q_e |B|^2 \left( \frac{1 + \exp \left[ \frac{q_e B}{K T} \right]}{K T} \right) H_{SN} + 4K T G c H_{CN} \quad (5.160) \]

Equation (5.160) illustrates some important properties of the detection process. Firstly, if the signals are very large, the noise processes at the detector output are dominated by the OGN [1],[15],[16],[17],[18],[25],[27],[29],[30],[32],[36], and the SNR becomes,

\[ \text{SNR}_{\text{AM}} = 2\mathbb{E}_1^2 + 2\mathbb{E}_2^2 A_d G_2 G^*_2 (\omega_2 - \omega_1) G^*_2 (\omega_2 - \omega_1) H^*(\omega_1) H(\omega_1) F(\omega_1, \omega_2 - \omega_1, \omega_1) \]

\[ x F^*(\omega_1, [\omega_2 - \omega_1], \omega_1) \]

\[ A_d h G d \omega_1 (E_1^2 + E_2^2) H_{OGN} \quad (5.161) \]
Alternatively, if the electrical noise sources are sufficiently small to be ignored, the SNR has the form of (5.161). When equation (5.161) applies, the detection process is quantum noise limited. If (5.161) applies when the actual signal power is very low, the signal will eventually disappear into the quantum noise because the numerator power is proportional to the square of the signal power and the denominator is proportional to the noise power [1],[15],[16],[17],[18],[25],[27],[29],[30],[32],[36].

The case in which \( E_1 \) represents the electric field of a signal, and \( E_2 \) is the electric field of a strong local oscillator, which is sufficiently large that detection is quantum noise limited, the SNR becomes, the same as that of equation (5.161), however \( E_1^2 + E_2^2 \). \( E_1^2 \) therefore the SNR is

\[
\text{SNR} = \frac{2E_1^2 G_d^2 A_d G_2^2 (\omega_2 - \omega_1) G^*(\omega_2 - \omega_1) H^*(\omega_1) H(\omega_1) F^*(\omega_1, [\omega_2 - \omega_1], \omega_1) F(\omega_1, (\omega_2 - \omega_1) \omega_1)}{R G_d A_d \omega_1 H_{\text{OGN}}}
\]

(5.162)

In this case the noise power is constant, the detected signal to noise ratio is dependent only on the signal strength of the weak signal. By using a detector of this type, it is possible to detect signals at the quantum signal to noise limit, even when the incoming signal is so weak that it would be overcome by the electrical noises of the photo detector.

Finally, consider an AM signal arriving at an SPD. Added to this is a local oscillator with the same frequency and phase as the signal carrier. Analysis proceeds the same way as the analysis of section 5.4.2, with a carrier amplitude of \( A_c + A_{LO} \). The SNR in the case is
\[ SNR_{AM} = 4E_c^2 (E_c + E_{LO})^2 G_a^2 G_2^2 (\omega_m) G_2^* (\omega_m) H^* (\omega_c) H(\omega_c) F(\omega_c, \omega_m, \omega_c) \]

\[ xF(\omega_c, \omega_m, \omega_c) \]

\[ A_G \omega_c [2(E_c + E_{LO})^2 + E_c^2 \mu^2] H_{GN} + 4q_e \left[ 1 + \exp \left( \frac{-e E_c}{KT} \right) \right] H_{SN} + 8KT G_c H_{JN} \]  

Where \( E_c \) is the electric field of the carrier and the \( E_{LO} \) the electric field of the local oscillator.

If the local oscillator is much larger than signal carrier, equation (5.163) becomes a quantum limited SNR.

\[ SNR_{AM} = 2E_c^2 G_a^2 G_2^2 (\omega_m) G_2^* (\omega_m) H^* (\omega_c) H(\omega_c) F(\omega_c, \omega_m, \omega_c) \]

\[ nG_a \omega_c H_{GN} \]

Which is an improvement in the detection signal to noise ratio for AM.

Deception of AM, by coherent means, improves the AM SNR to the quantum limit [1],[15],[16],[17],[18],[25],[27],[29],[20],[32],[36].
5.5 DETECTION OF MODIFIED-AMPLITUDE-MODULATED, FREQUENCY-SHIFTED AND PHASE-SHIFTED SIGNAL

Selection of an optimally detected modulation scheme requires a scheme in which all of the transmitter power is contained in the information waveform. As little of the transmitter power as possible should be wasted in a carrier [5],[6].

In some systems, however, it may be necessary to transmit a pilot carrier to lock the local oscillator, as is done in radio systems [5],[6],[47], however this must be of very small amplitude compared with the information signals, to improve the SNR over AM Systems [5],[6],[47].

Suitable candidates for modulation techniques, which can be optimally detected, are: Frequency-Shift keyed, Phase-Shift keyed and modified forms of AM, where the carrier is either suppressed altogether or reduced to a very low level, used only to lock the local oscillator [5],[6],[47]. Each modulation scheme may be detected by mixing the incoming modulated signal with a local oscillator, from which an optimal output signal can be produced. Detection efficiencies approaching 100% may be achieved with such detection systems [5],[6],[47].

In the following sections important examples of the detection of optical modulation systems, that are candidates for application in commercial transmission systems, will be discussed. Firstly in section 5.5.1, the detection of a single sideband suppressed carrier (SSB) signal is discussed. Such an analysis is suitable, even though the emphasis in this dissertation is on digital-transmission, because it is possible to generate the other forms of signalling mentioned using SSB signals [5],[6],[47],[48].
Sections 4.5.2 and 4.5.3 are concerned with Frequency and Phase-Shift keying respectively. It is believed that such signalling techniques will be used in the future; these are therefore examined in some detail.

Expressions for the SNRs of the 3 modulation types, discussed, are derived, and comparisons made, with the SNRs, of the AM systems, discussed in the preceding section.

5.5.1 Detection of Single Sideband Suppressed Carrier Signals

Detection of SSB signals is similar, in nature, to the general mixing case discussed in the previous section. In the frequency domain, a single sideband suppressed carrier signal may be represented by two delta functions [5],[6],[42],

\[ S_{SSB}(\omega) = \pi E_s \sqrt{G_a d} \delta(\omega - \omega_c - \omega_m) + \pi E_s \sqrt{G_a d} \delta(\omega + \omega_c + \omega_m) \]  

(5.165)

\( \omega_m \) is negative for a lower sideband signal, and positive for an upper sideband signal.

To detect SSB, a carrier must be mixed with the signal, returning the modulation to the base-band. The local oscillator is represented by [5],[6],[42].

\[ S_{LO}(\omega) = \pi E_{LO} \sqrt{G_a d} (\omega - \omega_c) + \pi E_{LO} \sqrt{G_a d} \delta(\omega + \omega_c) \]  

(5.166)

Using (5.155) with \( \omega_1 = \omega_c + \omega_m \) and \( \omega_2 = \omega_c \), the current is given by,
Similarly, the recovered signal power is,

\[
i^2(\omega)R_c = \left| E_s \right|^2 G_{A_d} R_t \left[ \delta(-\omega_m)G_2(-\omega_m) + G_2(\omega_m)G_2(\omega_m) \right]
\]

Using (5.158), the OGN is found,

\[
i^2_{\text{OGN}}(\omega)R_h G_{A_d} E_s^2(\omega_c + \omega_m) + E_{LO}^2(\omega_c) H_{\text{OGN}}
\]
The corresponding SNR is,

\[
\text{SNR}_{SSB} = 2 \left| \frac{E_s}{E_L} \right| E_L \left\{ G_a G_2 \left( -\omega_m \right) G_2^* \left( -\omega_m \right) H^* \left( \omega_c \right) H \left( \omega_c \right) F^* \left( \omega_c, -\omega_m, \omega_c \right) \right\} + \frac{x F \left( \omega_c, -\omega_m, \omega_c \right)}{h G \alpha d \omega_c \left( E_s^2 + E_L^2 \right)} + q_e \left( 1 + \exp \left[ \frac{q_e V_b}{K T} \right] \right) H_{SN} H_{TG} H_{JN}
\]

(5.171)

If the signal is weak compared with the local oscillator,

\[
\text{SNR}_{SSB} = 2 \left| \frac{E_s}{E_L} \right| E_L \left\{ G_a G_2 \left( -\omega_m \right) G_2^* \left( -\omega_m \right) H^* \left( \omega_c \right) H \left( \omega_c \right) F^* \left( \omega_c, -\omega_m, \omega_c \right) \right\} \frac{h G \alpha d \omega_c \left( E_s^2 + E_L^2 \right)}{h G \alpha d \omega_c \left( E_s^2 + E_L^2 \right)}
\]

(5.172)

5.5.2 Detection of Frequency-Shift-Keyed Signals

Frequency Shift keying (FSK) conveys digital information by transmission of 1 frequency to represent a logical 1 and a second to represent a logical zero. The frequency shifts may occur at any time in unsynchronized FSK, in which case the oscillators are cut off at non zero values of carrier [5],[6]. Harmonics are generated by this process. FSK frequency shift may also occur when the time varying fields, produced by the oscillators, reach zero [5],[6]. In this case harmonic generation is reduced. The latter form of FSK is a more desirable modulation type because bandwidths required at the detector are smaller than for non synchronized FSK.
An FSK signal may be represented by [6],

\[ s_1(t) = \begin{cases} 
A \cos m \omega_m t & \frac{T_b}{2} < t < \frac{T_b}{2} \\
0 & \text{elsewhere} 
\end{cases} \]

and

\[ s_2(t) = \begin{cases} 
A \cos n \omega_m t & -T_b < t < -\frac{T_b}{2} \\
0 & \text{elsewhere} 
\end{cases} \]

where \( T_b \) is the bit period

Synchronized FSK will occur if the bit period \( T \) starts at time,

\[ t_1 = \frac{2k \pi}{m \omega_m} \]

and finishes at \( [6], \)

\[ t_2 = \frac{2(k + 1) \pi}{m \omega_m} \]

Therefore \( T_b = \frac{\pi}{m \omega_m} \)

Where \( k \) and \( l \) are arbitrary integers.

Equations (5.174) and (5.175) apply to both FSK frequencies, thus,

\[ T_b = \frac{P \pi}{m \omega_m} \]

where \( P \) is an integer.

Therefore, synchronized FSK occurs if the ratios,

\[ \frac{l}{m} = \frac{P}{n} \]

(5.177)
If the digital signal consists of alternate 1's and 0's, the bit stream will be a square wave, which may be represented in the time domain by the convolution [42].

\[ S_D \Theta \sum_{k=-\infty}^{\infty} \delta(t-2kT_b) \]  
(5.178)

Where \( S_D = \begin{cases} 1 & \frac{-T_b}{2} < t < \frac{T_b}{2} \\ 0 & \text{elsewhere} \end{cases} \)

The transmitted signals may be represented by,

\[ S_1(t) = A_1[ S_D(t) \Theta \sum_{k=-\infty}^{\infty} \delta(t-2kT_b)] \cos(m\omega_m t) \]  
(5.179)

representing 1's and

\[ S_2(t) = A_2[ S_D(t) \Theta \sum_{l=-\infty}^{\infty} \delta(t-(2l+1)T_b)] \cos(n\omega_m t) \]  
(5.180)

representing the 0's.

The bit interval \( T_b \) may be written in terms of the modulation frequency

\[ T_b = \frac{\pi}{\omega_m} \]  
(5.181)

The Frequency domain representation is found by taking the Fourier transform of each term [6],[42], of (5.179) and (5.180), thus,

\[ S_1(\omega) = \left( \frac{\pi}{2\omega_m} \right) \sin \left( \frac{\pi}{2\omega_m} \right) \left( \frac{2\pi}{\omega_m} \right) \left[ \sum_{k=-\infty}^{\infty} \delta(\omega-k\omega_m) \right] \Theta \left[ A_1 \pi \delta(\omega-m\omega_m)+\delta(\omega+m\omega_m) \right] \]  
(5.182)

and

\[ S_2(\omega) = \left( \frac{\pi}{2\omega_m} \right) \sin \left( \frac{\omega\pi}{2\omega_m} \right) \left( \frac{2\pi}{\omega_m} \right) \left[ \sum_{l=-\infty}^{\infty} \delta(\omega-l\omega_m) \exp[-jl\pi] \right] \Theta \left[ A_2 \pi \delta(\omega-n\omega_m)+\delta(\omega+n\omega_m) \right] \]  
(5.183)
The combined signal \( S(\omega) = S_1(\omega) + S_2(\omega) \) is shown in Fig 5.12 [6].

**Fig 5.12 The Spectrum of an FSK Signal**

The FSK spectrum contains harmonics of the square modulating wave. However the information is completely contained in the AM spectrum about the signal frequencies, \( n\omega_m \) and \( m\omega_m \). For analysis purposes, other harmonics will be ignored. Each frequency of the FSK signal is then, an amplitude modulated wave, the oscillator at frequency \( n\omega_m \) is modulated 180° out of phase with the other frequency \( m\omega_m \) [42].

FSK in radio systems may be detected in a variety of ways. For example, a bandpass filter centred at each frequency, followed by an envelope detector may be used [6]. A fluctuating waveform is produced by each envelope detector, the output of one is in antiphase with the other. The full modulation signal is extracted by subtracting one output from the other [6].

Optically, however, such detection schemes are not possible, as it is impossible to produce sufficiently narrow bandpass filters. Fortunately, coherent or quasi-coherent techniques can be used to detect FSK, in a similar way to coherent detection of FSK radio signals [6].
To optically detect FSK, two SPD's are used. Local oscillators at the two FSK frequencies, \( \omega_m \) and \( \omega_{-m} \), are applied to the detectors, as shown in Fig 5.13 [6].

![Fig 5.13 Coherent Optical Detection of FSK](image)

Using this apparatus, two coherent ASK detectors are produced. The output of the signal at carrier \( \omega_m \) is subtracted from the signal at \( \omega_{-m} \). The combined amplitude of the two detector outputs is presented to a digital detector [6]. Digital pulses are recovered and transmitted into a local digital network.

Timing recovery may be made by mixing the FSK signal with a local oscillator at

\[
\omega_c = \left( \frac{n+m}{2} \right) \omega_m
\]

(5.184)

Producing a timing signal at frequency,

\[
\omega_t = \left( \frac{n-m}{2} \right) \omega_m
\]

(5.185)

The equipment for timing recovery is shown in Fig 5.14.
The SNR, for a coherently detected FSK signal is given by

$$\text{SNR}_{\text{FSK}} = \frac{E_c^2 (E_c^2 E_L^0)^2 A_1^2 A_2^2 (\omega_1) G_2^2 (\omega_m) H^*(\omega_c) H(\omega_c) F^*(\omega_m, \omega) F(\omega, \omega_m, \omega) + \text{2}\pi \alpha G A_1 A_2 \text{OGN} + 32q_e |B| \left( 1 + \exp \left[ \frac{q_e V_B}{KT} \right] \right) H_{\text{SN}} + 32K T G A_1 A_2 \text{OGN} \right]$$

Where the electric fields of both signals are equal, that is

$$A_1 = A_2 = E_c \sqrt{\frac{G A_1}{a d}}$$

the electric field strength of both local oscillators is $E_L^0$ and

$$\omega_c = \frac{\omega_1 + \omega_2}{2}$$

If the local oscillator electric fields are much larger than the signal electric fields, (5.186) becomes quantum limited,

$$\text{SNR}_{\text{FSK}} = \frac{E_c^2 G A_1^2 A_2^2 (\omega_1) G_2^2 (\omega_m) H^*(\omega_c) H(\omega_c) F^*(\omega_m, \omega) F(\omega, \omega_m, \omega) + \text{2}\pi \alpha G A_1 A_2 \text{OGN} \right]}{4\pi \alpha G A_1 A_2 \text{OGN}}$$

Both (5.186) and (5.187) show an SNR improvement of 3dB over an ASK system transmitting the same amount of power [5],[6].
5.5.3 Detection of Phase-Shift-Keyed Signals

The most efficient form of information transmission is Bi-phase Shift Keying (BPSK) \([5],[6]\). It yields the highest SNR of the techniques discussed in section 5.5 \([5],[6]\). A BPSK modulator changes the phase of a carrier by \(\pi\) radians. One phase is transmitted to represent a logical zero, the other is transmitted to represent a logical 1.

The BPSK signal is represented by \((5.188)\) in the time domain \([6]\),

\[
s(t) = AS_D(t)\sum_{k=-\infty}^{\infty} \delta(t-2kT_b) \cos \omega_m t - AS_D(t) \sum_{k=-\infty}^{\infty} (2k+1)T_b \cos \omega_m t
\]

\((5.188)\)

where \(S_D(t) = \begin{cases} 1 & \frac{-T_b}{2} \leq t \leq \frac{T_b}{2} \\ 0 & \text{elsewhere} \end{cases} \)

and the bit interval is \(T = \pi/\omega_m\).

The spectrum of BPSK is similar to the FSK spectrum \([6],[42]\),

\[
S(\omega) = \left[ \left( \frac{\pi}{\omega_m} \right) \sin \omega_m \left( \frac{\pi}{\omega_m} \right) \sum_{k=-\infty}^{\infty} \delta(\omega-k\omega_m) \right] \odot \left[ \left( \frac{\pi}{2\omega_m} \right) \sin \omega_m \left( \frac{\pi}{2\omega_m} \right) \sum_{k=-\infty}^{\infty} \delta(\omega-k\omega_m) \exp[-j1\pi] \right] \odot \left[ \left( \frac{\pi}{2\omega_m} \right) \sin \omega_m \left( \frac{\pi}{2\omega_m} \right) \sum_{l=-\infty}^{\infty} \delta(\omega-l\omega_m) \exp[-j1\pi] \right] \odot \left[ \left( \frac{\pi}{\omega_m} \right) \sin \omega_m \left( \frac{\pi}{\omega_m} \right) \sum_{k=-\infty}^{\infty} \delta(\omega-k\omega_m) \right]
\]

\((5.189)\)

and is shown in Fig. 5.15.
The carrier of a BPSK signal, at $m\omega_m$, is suppressed. The information is contained in the two sidebands, at $(n-1)\omega_m$ and $(n+1)\omega_m$.

BPSK is detected optically, by either inserting a carrier at the detector, producing a signal similar to a coherently detected AM signal, as shown in Fig 5.15 [5],[6],

or by using two detectors, splitting the incoming signal between the two detectors and mixing the local oscillator with the signal, at one detector, and a $180^\circ$ phase shifted local oscillator with the signal at the other, as shown in Fig 5.17.
In both cases, an additional low pass filter is connected to the terminals of the detector. It has a bandwidth of $\omega_m$ [6].

The output signals are amplitude modulated signals. They are threshold detected and digital pulses regenerated.

Both forms of detection shown in Fig 5.15 and Fig 5.16 are analysed in a similar manner to the AM signals of section 5.4.3.

The SNR of PSK, detected by one detector, (Fig 5.15) is given by

$$SNR_{PSK} = \frac{E_C^2 E_{LO}^2 G^2 A^2 G^2_2(\omega_m)G_2(\omega_m)H^*(\omega_m)H(\omega_m)H(\omega_m)F^*(\omega_c, \omega_m, \omega_c)F(\omega_c, \omega_m, \omega_c)}{2hG_A d_c [2E_{LO}^2 + E_c^2]H_{OGN} + 32q_e T G_{H_{SN}} + 32kT G_{H_{JN}}} \left(1 + \exp \left[\frac{q_e V_B}{eKT}\right]\right)\left(1 + \exp \left[\frac{q_e V_B}{eKT}\right]\right)$$

(5.190)

For a large local oscillator,

$$SNR_{PSK} = \frac{E_C^2 G^2 A^2 G^2_2(\omega_m)G_2(\omega_m)H^*(\omega_c)H(\omega_c)H(\omega_c)F^*(\omega_c, \omega_m, \omega_c)F(\omega_c, \omega_m, \omega_c)}{4hG_A d_c [2E_{LO}^2 + E_c^2]H_{OGN}}$$

(5.191)
If the detection process of Fig 5.17 is used, the SNR ratio is

\[
\text{SNR}_{PSK} = \frac{E_c^2 (E_c + 2E_L)^2 G_a^2 G_d^2 G_m^2 \omega_m G_s^2 \omega_m^* H^* (\omega_m) H^* (\omega_c) H(\omega_c) H(\omega_c) F^* (\omega_c, \omega_m, \omega_c, \omega_c)}{\pi G_a \omega_c \omega_m [2 (E_c + 2E_L)^2 + E_c^2] G_OGN + 16 q_e \ln (1 + \exp \left[ \frac{q_e V_B}{kT} \right]) H_{SN} + 16KTG_c H_{JN}}
\]

For a large local oscillator, the SNR becomes

\[
\text{SNR}_{PSK} = \frac{E_c^2 G_a^2 G_d^2 G_m^2 \omega_m G_s^2 \omega_m^* H^* (\omega_m) H^* (\omega_c) H(\omega_c) H(\omega_c) F^* (\omega_c, \omega_m, \omega_c, \omega_c)}{2hG_a G_d G_c G_OGN}
\]

\[(5.192)\]

\[(5.193)\]
5.6 IMPROVING THE SIGNAL TO NOISE RATIO OF DETECTION

From the analyses of section 5.5 it is apparent that, for a given transmit power, FSK and PSK offer improvement in the SNR of detection over ASK methods [5],[6]. Moreover, using 1 photo detector, it is possible to exploit a 3dB gain in signal to noise performance of the coherent PSK detection techniques, of Fig 5.15, over the equivalent FSK and ASK systems [5],[6].

Few additional improvements in SNR are possible using the other types of modulation and detection used in radio systems. Fortunately it is possible to improve the detector SNR by, either post detector correlation techniques, or pre-modulation light-squeezing techniques.

In this section both techniques of improving the detector performance beyond that described in section 5.5.

In section 5.7.1, techniques of correlating the input signals from two detectors is discussed [36]. The incorrelated noise of each detector is reduced, by the technique, until the quantum limit is reached [36].

Using the theory of the squeezed states of light it is possible, theoretically at least, to reduce the quantum noise of the detection process below the quantum limits given in the preceding analyses [33]. The theory of the squeezed state of light is reviewed briefly in section 5.7.2.
5.6.1 Correlative Reduction of Detection Noise

Consider two devices which generate noise, or receive noise signals, denoted by $n_1(t)$ and $n_2(t)$ in the time domain [5],[6],[36],[44]. If the sources are subject to independent conditions, or receive noise independently, then the noise signals $n_1(t)$ and $n_2(t)$ are uncorrelated [5],[6],[36],[44] and so

$$
l \lim_{T_i \to \infty} \frac{1}{T_i} \int_{-\frac{T_i}{2}}^{\frac{T_i}{2}} n_1^*(\tau)n_2(\tau + t)d\tau = 0 \quad (5.194)$$

Where $T_i$ is the period of observation over which the correlation is made and $\tau$ is a dummy variable of integration with respect to time.

The mathematical process described by (5.194) may be implemented using an analogue circuit as shown in Fig 5.18 [5],[6],[36].

**Fig 5.18 Correlation of Two Signals**
Which consists of a multiplier, an integrator and a scaling circuit. If the integration period $T_1$ is very large, the output of the integrator, of Fig 5.18, will be insignificant [36].

If, however, the integration time is reduced the output of the integrator will rise, as those components of noise which move through less than half a cycle, during the integration period, will contribute a measurable D.C. to the integrator sum. In this case the limit of (5.194) is removed and so (5.194) becomes [42].

\[
C_n(t, \tau) = \frac{1}{T_1} \int_0^{T_1} n_1^*(\tau) n_2(t + \tau) d\tau
\]  

(5.195)

Since much of the preceding analysis has been undertaken in the real frequency domain, it is instructive to take the Fourier transform of (5.195) [42].

\[
C_n(\omega) = \int_{-\infty}^{\infty} \frac{1}{T_1} \int_0^{T_1} n_1^*(\tau) n_2(t + \tau) d\tau \exp[j\omega t] dt
\]  

(5.196)

Unfortunately there is no obvious transform of the argument. However, it is known that the observation period $T_1$ is finite and that the integrator output is discharged to 0 after each observation period. Therefore (5.195) may be written as,

\[
C_n(t, \tau) = Q_i(t) \left[ \int_{-\infty}^{\infty} n_1^*(\tau) n_2(t + \tau) d\tau \right]
\]  

(5.197)
Where
\[ Q_i(t) = \begin{cases} 1 & 0 \leq t \leq T_i \\ 0 & \text{elsewhere} \end{cases} \]
in which case (5.196) becomes [42],

\[ C_n(\omega) \int_{-\infty}^{\infty} Q(t) \int_{-\infty}^{\infty} n^*(\tau)n(t+\tau)d\tau \exp[j\omega t] dt \]  

(5.198)

But the integral expression of (5.197) is readily transformed to the product of \( N_1(\omega) \) and \( N_2(\omega) \) where,

\[ N_1(\omega) = F\{n_1(t)\} \quad \text{and} \quad N_2(\omega) = F\{n_2(t)\} \]

and \( F\{\} \) denotes Fourier transformation.

Since the integral of (5.197) is multiplied by \( Q(t) \) in \( T_i \)
the time domain, its Fourier transform is convolved with \( Q(\omega) \) where,

\[ Q(\omega) = F\{Q(t)\} \]

Taking the Fourier transform of \( Q(t) \) gives [42],

\[ Q(\omega) = \exp\left[\frac{j\omega T_i}{2}\right] T_i \text{sinc}\left[\frac{\omega T_i}{2\pi}\right] \]

(5.199)

where \( \text{sinc} x = \frac{\sin \pi x}{\pi x} \)

and so,

\[ C_n(\omega) = \exp\left[\frac{j\omega T_i}{2}\right] \text{sinc}\left(\frac{\omega T_i}{2\pi}\right) N(\omega)^* N(\omega) \]  

(5.200)

Where \( \odot \) denotes convolution.
Evaluating the convolution gives,

\[ C_n(\omega) = \int_{-\infty}^{\infty} \exp \left[ j(u - \omega)T_i \right] \frac{\text{sinc}\left(\frac{(u - \omega)T_i}{2\pi}\right) N(u)N^*(u)du}{2} \quad (5.201) \]

But \(N(u)N^*(u)\) represents the power spectral density of the noise [5],[6]. It is assumed that \(N(u)\) and \(N^*(u)\) are independent of frequency, (the spectral density is constant) in which case the argument \(u\) is dropped and (5.201) becomes,

\[ C_n(\omega) = \frac{N^*}{T_i} \int_{-\infty}^{\infty} \exp \left[ j(u - \omega)T_i \right] \frac{\text{sinc}\left(\frac{(u - \omega)T_i}{2\pi}\right) du}{2} \quad (5.202) \]

Solving the integral gives [42],

\[ C_n(\omega) = \frac{N^*}{T_i} \quad (5.203) \]

When \(T_i\) is very large, the RHS of (5.203) vanishes, when \(T_i\) is very small the RHS of (5.203) tends towards infinity.

Now, consider the correlation of two voltages present over one data-bit interval, in which case the correlation becomes,

\[ C_s(t, \tau) = \frac{1}{T_i} \int_{0}^{T_i} s_b(\tau)s_b(\tau + t)dt \quad (5.204) \]

Where

\[ s_b(t) = \begin{cases} \text{i_b} & 0 \leq t \leq T_b \\ 0 & \text{elsewhere in the integration period.} \end{cases} \]
$i_b$ is the current associated with the signal and $T_b$ is the bit period.

Solving the integral of (5.204) [42] gives a correlation of

$$
C_s(t) = \begin{cases} 
\frac{i_b^2}{T_i} (T_b + t) & -T_b \leq t \leq 0 \\
\frac{i_b^2}{T_i} (T_b - t) & 0 \leq t \leq T_b \\
0 & \text{elsewhere}
\end{cases}
$$ (5.205)

In this case $t$ represents a delay. If $t = 0$ the integral is a maximum, which is the desired delay to optimize the signal at the integrator output, thus,

$$
C_s(t) = \frac{i_b^2 T_b}{T_i}
$$ (5.206)

$C_s(t)$ is also optimized if the integration period equals the bit interval, in which case,

$$
C_s(t) = i_b^2
$$ (5.207)

If the output of the integrator represents a current, then the SNR is readily found,

$$
\text{SNR}_i = \frac{|T_b|^2 i_b^4}{N^2 N^*}
$$ (5.208)

Implicit in the derivation, is the assumption that the signal is sufficiently large that one can ignore the multiplicative effects of the signal and noise, as is done in the analysis of amplitude modulated systems [6].
As the bit interval is increased, the SNR is also increased, as it is decreased the SNR falls. Therefore as the transmission speed of the data is increased the SNR of (5.208) falls.

Using the variables of (5.208), the input SNR is, \( \text{SNR}_I \),

\[
\text{SNR}_I = \frac{i^2_b}{\text{NN}^*}
\]  

(5.209)

Substituting (5.209) into (5.208) gives,

\[
\text{SNR}_I = (\text{SNR}_I)^2 \left| T_b \right|^2
\]  

(5.210)

The correlative method of detection only affords an SNR improvement if

\[
\text{SNR}_I \geq \text{SNR}_I
\]

and so

\[
\left| T_b \right|^2 \text{SNR}_I^2 \geq \text{SNR}_I
\]

(5.211)

In which case \( \text{SNR}_I \geq \left| T_b \right|^2 \).

Using the preceding analysis, the improvements in detection SNR can be found, for optical signals.

Consider an incoming optical signal which would be detected using one of the methods described in section 5.5. The SNR in this case is \( \text{SNR}_I \). If this same signal is split between two arms of a correlative detector as shown in Fig 5.19 [36].
The SNR of each detector arm will fall, depending upon the nature of
the dominant noise source of detection.

If the noise is dominated by the thermal noise sources of the
detectors, the SNR will fall by 1/2, in which case the SNR is improved by a
correlative detector if,

\[ \text{SNR}_i > \frac{\text{SNR}_i}{2} \]  

(5.212)

thus,

\[ \left| T_b \right|^2 \frac{\text{SNR}_i^2}{4} > \text{SNR}_i \]  

(5.213)

in which case the SNR must be greater than \( 4\left| T_b \right|^2 \) to achieve any
significant gain in SNR.

If the detector noise is predominantly OGN due to a large local
oscillator. The SNR will also be decreased by a factor of two and (5.213)
would again apply.

Finally, if the detection noise is dominated by optical granular noise
and the local oscillator strength is small, the detector SNR will not
change thus to gain an improvement in SNR the input SNR must be

\[ \text{SNR}_i > \frac{1}{\left| T_b \right|^2} \]  

(5.214)
Examining equations (5.213) and (5.214), it is evident that at high data rates where $|T_b|^2$ is small, there is little advantage in using correlative methods of SNR improvement, as the input signal SNR's required to achieve a significant improvement are very large indeed.

For example, the SNR required for an improvement at a data rate of 1 Gbps is a very large $1 \times 10^{18}$ or 180 dB.

5.6.2 Reducing the Quantum Noise Limit - The Squeezed State.

The ultimate limit on the detection SNR of an optical detection system is due to the granular noise associated with the discrete nature of photons [1],[2],[3],[4],[7],[8],[16],[17],[18],[27],[28],[29],[30],[32],[36]. Until recently, it was believed that one could not reduce the noise involved in detection, below the quantum noise limit due to the uncertainty in photon number and phase.

However, recent analysis of two photon correlation processes, and careful re-examination of quantum field theory, have indicated the possibility of transmitting information in a field-state which exhibits noise statistics with sub-Poissonian variance. Such a field state is known as a squeezed state of the field [33].

The squeezed state of an optical field is significantly different from the photon anti-bunching effect noticed in coherent light at very low light levels. Photon anti-bunching [12],[33] occurs when the assumptions used in section 3.2.2, to derive the photon number-phase uncertainty [8], cannot be
made. The numbers of photons, per unit time, flowing in a coherent optical beam are significantly altered by making measurements of photon number or phase \([12],[33]\).

Moreover, in a finite period of time, a fixed number of photons will flow out of a laser. At low light levels the correlation between the arrival of two photons \([12],[33]\) will exhibit an uncertainty lower than that given by the photon number-phase uncertainty derived in section 3.2.2 \([8]\).

Photon anti-bunching is a characteristic of coherent light at low levels \([12],[33]\). The phenomenon disappears as the light level is increased. Moreover photon anti-bunching is not experienced in chaotic light, which in fact, exhibits the opposite effect of photon bunching \([12]\); the second order correlation exhibits greater uncertainty, than expected due to photon number-phase uncertainty of section 3.2.2 \([8]\).

On the other hand, state squeezing is unaffected by the light level. It is a characteristic of the quantum mechanical description of light, as the sum of an ensemble of simple harmonic oscillators \([9],[10],[11],[12]\) which does not require the simplifications used in section 3.2.2 \([8]\).

Conceptually, state squeezing involves states which are orthogonal to each other, they are therefore independent of each other. However the uncertainties in each orthogonal state are interrelated \([33]\). By increasing the certainty, in photon number, in one state, the uncertainty in the other states is increased \([33]\). By transmitting information in the state with lower uncertainty, which can be detected by heterodyne means independently of other states, the quantum noise limit may be reduced below the traditional levels given by photon numbers phase uncertainty \([33]\).
It is possible to gain a better view of the squeezed state concept if one considers the common practice of noise analysis in narrow band systems using in-phase and quadrature components of noise [5],[6].

When a signal is coherently detected, the quadrature component of noise makes no contribution to the noise at the detector output [5],[6]. Suppose, that in addition, the mean square noise power of each component is dependent in some way upon the nature of the transmitted signal itself. Moreover the RMS noise voltages in each noise component is related to the other such that their product is a constant.

If, by some means, the RMS noise voltage in the in-phase component can be lowered at the expense of the RHS noise voltage of the quadrature component, the mean squared noise, at the output of the detector, will be smaller than the normal mean squared noise component [5],[6]. However the overall noise envelope will be unaffected [5],[6]. The behaviour of squeezed state information transmission seems entirely analogous [33].

Using the tools developed in section 3.2.1 the characteristics of state squeezing are readily shown mathematically.

The electric field of a single coherent state may be written in terms of its quantum mechanical creation and annihilation operators, [9],[12],[33].

\[ E(t) = \lambda (a \exp[-j\omega t] + a^+ \exp[j\omega t]) \] (5.215)

Where \(a\) and \(a^+\) are the creation and annihilation operators respectively, and \(\lambda\) is a complex quantity containing spatial wavefunctions and quantum mechanical constant [50].
The order in which the creation and annihilation operators are applied to a system effects the state of the system. That is, the creation and annihilation operators are non-commutative \([50]\).

Let \([x, y]\) represent the commutator of two general operators \(x\) and \(y\), then

\[
xy - yx = z
\]

(5.216)

Where \(z\) is a fixed quantity dependent upon the nature of \(x\) and \(y\).

If \(x\) and \(y\) commute, \(z = 0\) otherwise \(z\) is a constant known as the commutator \([50]\),

\[
[x, y] = z
\]

(5.217)

The commutator of two quantum mechanical operators represents the uncertainty with which two measurements or operations, on the system, can be made. By applying operator \(x\) the system is disturbed such that when operator \(y\) is applied the system it disturbs an entirely different system \([50]\).

For example, in Gaussian distributed uncertainty of position and momentum operators, of a multi-particle system, the commutator for the momentum and position operators; \(p\) and \(x\) respectively, is \([50]\),

\[
[x, p] = \frac{\hbar}{2}
\]

(5.218)

The commutator for creation and annihilation operators is \([9],[12],[33],[50]\)}
\[ [a, a^+] = \frac{1}{2} \]  

(5.219)

but \(a\) and \(a^+\) are, in general, complex quantities which may be written in terms of their real and imaginary parts, \(X_1\) and \(X_2\) respectively, thus

\[ a = x_1 + jx_2 \]  

(5.220)

Using (5.220) and the commutation relation, (5.219) one can show that \(X_1\) and \(X_2\) obey the commutation [9],[12],[33],[50]

\[ [x_1, x_2] = \frac{i}{2} \]  

(5.221)

The time varying electric field can also be written in terms of \(X_1\) and \(X_2\) [50],

\[
E(t) = \frac{\hbar}{2} (X_1 \cos(\omega t) + X_2 \sin(\omega t))
\]  

(5.222)

Therefore, \(X_1\) and \(X_2\) are the amplitudes of two quadrature phases of the electric field.

Using commutation relation (5.221), the minimum uncertainty in \(X_1\) and \(X_2\) can be derived [9],[12],[33],[50]

\[
\Delta X_1 \Delta X_2 = \frac{1}{4}
\]  

(5.223)

Where \(X_1\) and \(X_1\) are the uncertainties in \(X_1\) and \(X_2\) respectively.

Implicit in the analysis, of coherent states, is the assumption that \(X_1 = X_2 = 1/2\) [4],[12]. However these are not the only possible uncertainty values associated with (5.223) [4]. If the uncertainty in \(X_1\), \(X_1\) is smaller than 1/2, the uncertainty in \(X_2\), \(X_2\) will be larger than 1/2 [4]. States in which the uncertainties, in (5.223), are of different values, are the squeezed states.
Recall equation (3.85) which describes a coherent state in terms of a
displacement operator, \( D(\alpha) \), on the vacuum state.

\[ |\alpha\rangle = D(\alpha) |0\rangle \]

Where

\[ D(\alpha) = \exp(\alpha a^+ - \alpha^* a) \]

A squeezed state \( |\alpha, \xi\rangle \) may be written as the product of a squeezing
operator and a displacement operator acting upon the vacuum state [33],

\[ |\alpha, \xi\rangle = D(\alpha)S(\xi) |0\rangle \]

Where \( S(\xi) = \exp\left[ \frac{1}{2} \xi^* a^2 - \frac{1}{2} \xi a^+ a \right] \)

and \( \xi = r \exp[j\theta] \)

The variances in the squeezed state are [33],

\[ V(Y_1) = \frac{1}{4} \exp[-2r] \]

and \[ V(Y_2) = \frac{1}{4} \exp[2r] \]

Where

\[ Y_1 + jY_2 = (X_1 + jX_2) \exp\left[ -\frac{j\theta}{2} \right] \]

Which is a rotated complex amplitude [33].

The mean number of photons in the squeezed state \( |\alpha, \xi\rangle \) is

\[ \langle n \rangle = |\alpha|^2 + \sinh^2 r \]
In order to generate a squeezed state, part of a single mode field is mixed with its phase conjugate to produce a new mode b such that [33],

\[ b = \mu a + \nu a^+ \]  

(5.228)

where

\[ \mu^2 + \nu^2 = 1 \]

For mode a in a coherent state mode b is in a squeezed state.

Theoretically, at least, squeezed states may be generated using either the phase conjugate mirror or a parametric amplifier [33].

Squeezed states may be detected using homodyne or heterodyne detection. The presence of a squeezed state is measured by changing the local oscillator phase through 90°. If the noise statistics change from sub-Poissonian to super-Poissonian, state squeezing exists [33]. Other methods of detecting squeezed states also exist [33].

As indicated earlier, information can be transmitted in the state exhibiting subpoissonian statistics. Using either homodyne or heterodyne detection, the signal in the squeezed state may be detected with reduced quantum noise. The quadrature phase with increased noise does not contribute to the detection SNR.

According to Robinson [49], the Squeezed State research group at Bell Laboratories has succeeded in generating a 7% quantum noise reduction by state squeezing. The reduction is very small indeed and must be of an order comparable with the experimental error. However it verifies the possibility of state squeezing as a means of noise reduction in optical transmission systems.
5.7 A COMPARISON OF HOMODYNE AND HETERODYNE DETECTION OF OPTICAL SIGNALS

In the correspondence to the proceedings of the IRE of 1962 [31] B.M. Oliver compared the processes of homodyne and heterodyne detection of optical signals in a quantum limited noise environment [31],[32]. He reached the surprising, and incorrect, conclusion, that it is possible to achieve a 3dB gain in SNR over heterodyne detection using homodyne methods of detection.

This incorrect conclusion was reached because Oliver [31] did not include the principle of quantum mechanical complementarity [8] in his analysis of the quantum mechanical noise. The principle of complementarity shows that it is not possible to determine the photon number and phase, of a system, simultaneously [8]; as discussed in section 3.2.2.

Moreover, Oliver's [31] approach to signal detection is itself strange. In heterodyne detection, Oliver [31] analyses the SNR based upon the power contained in the Intermediate Frequency (IF) Carrier. In other words, he does not introduce modulation, which contains the actual signal power received.

Similarly, homodyne detection of a signal involves converting a modulated signal directly to the baseband [5],[6],[48]. It is a special case of the heterodyne system. In homodyne detection the IF is at DC. Oliver derives the expression for detection SNR based on the DC power at the detector output. However, in any information transmission system, the modulation is of greatest interest. It is usually contained in sidebands distributed about the carrier. If the modulation is a steady signal at one frequency, the process of detection becomes very similar to the heterodyne
situation. The signal power is the power contained in the alternating current at the detector output, not in the direct current produced by the carrier.

In the signal to noise analyses of section 5.5, homodyne systems were treated as special cases of heterodyne detection. It was assumed that the signal at the transmitter was modulated by a single frequency sinusoid. The process of homodyne detection recovers this sinusoid which is the signal to be detected. Its power is therefore the detected signal power [5],[6].

By analysing homodyne detection systems in this way, errors due to quantum mechanical complementarity do not occur. Homodyne and heterodyne detection systems exhibit the same Signal-to-Noise performance.

Conceptually of course, the heterodyne and homodyne techniques of detection should exhibit the same noise behaviour. Moreover when detecting radio signals, the Signal-to-Noise performance of heterodyne and homodyne systems are the same [5],[6],[48].

When analysing modulated signals which are homodyne detected, the noise due to quantum mechanical phase uncertainty need not be considered, except at very low light levels. The phase uncertainty does not effect the power in the recovered signal unless it is at DC [32].

By examining 3 idealized examples, the effects of phase uncertainty can be deduced.
Consider an unmodulated carrier of angular frequency $\omega_c$, and electric field,

$$E_s(t) = E_s \cos (\omega_c t)$$  \hspace{1cm} (5.229)

This signal is detected using a local oscillator of angular frequency $\omega_0$ and electric field,

$$E_o(t) = E_o \cos (\omega_0 t)$$  \hspace{1cm} (5.230)

Then the mean square IF current at the detector output is [31],[32]

$$\overline{i_s^2} = \frac{q_e E_s^2 E_o^2 A_d^2}{c_s \epsilon_0 a_d^2} \cos^2[(\omega_0 - \omega_c)t + \varnothing(t)] \hspace{1cm} (5.231)$$

Where $G_a$ is the atmospheric electromagnetic conductance [41]. $\varnothing(t)$ accounts for the phase difference between the signal and local oscillator. $\varnothing(t)$ will be effected by the quantum mechanical phase uncertainty [33]. Without extra knowledge, it is assumed that the uncertainty in phase is distributed evenly over $2\pi$ radians [32].

The cosine term of (5.231) may be rewritten in a more convenient form,

$$\overline{i_s^2} = q_e E_s^2 E_o^2 A_d^2 \cos^2[(\omega_0 - \omega_c)t]\cos^2[\varnothing(t)] - 2 \cos[\omega_0 - \omega_c]t \sin[(\omega_0 - \omega_c)t]$$

$$\times \cos \varnothing(t) \sin \varnothing(t) + \sin^2(\omega_0 - \omega_c)t \sin^2 \varnothing(t)$$  \hspace{1cm} (5.232)

Taking the average value yields, a signal power of [32],

$$\overline{i_s^2} = q_e E_s^2 E_o^2 A_d^2 \frac{1}{\frac{1}{4} + \frac{1}{4}}$$  \hspace{1cm} (5.233)
Consider now, the mean-square signal current of detection without the introduction of the phase uncertainty parameter, $\theta$,

$$
\overline{i^2_s} = \frac{q^2 E^2 G^2 A_d}{2 \hbar^2 \omega_c \omega_0} \cos^2[(\omega_0 - \omega_c)t] \tag{5.235}
$$

By taking the mean value, the mean square noise current is,

$$
\overline{i^2_s} = \frac{q^2 E^2 G^2 A_d}{\hbar^2 \omega_c \omega_0} \cos[\Phi(t)] \tag{5.236}
$$

Which is exactly the same as the equation (5.234), the phase uncertainty term $\Phi(t)$ has no effect on the mean square signal power.

Now consider heterodyne detection with a local oscillator at the same frequency as the unmodulated carrier, that is

$$
\omega_c = \omega_0
$$

In this case (5.233) becomes,

$$
\overline{i^2_s} = \frac{q^2 E^2 G^2 A_d}{\hbar^2 \omega_c \omega_0} \tag{5.237}
$$

Taking the average [32],

$$
\overline{i^2_s} = \frac{q^2 E^2 G^2}{\hbar^2 \omega_c \omega_0} \tag{5.238}
$$

However, if the phase term is omitted, the mean square noise current is [31]
$$\overline{i_s^2} = \frac{q^2 E_s^2 \rho^2}{\hbar^2 \omega_c \omega_0}$$ \hfill (5.239)

Which is twice the value given by (5.234), (5.236) and (5.238).

In all cases discussed above, it is assumed that $\cos^2 \theta(t)$ averages to 1, which is reasonable for any signal varying in frequency and time about which little more is known.

When analysing a homodyne system SNR using the techniques of section 5.5, the recovered modulation has a power expression similar to (5.234), the phase uncertainty is therefore unimportant in the analysis.
5.8 DETECTOR ALIGNMENT

The process of optical mixing is similar to an in-line hologram. The frequency difference between the two beams being mixed produces beats or fringes which are detected by the intensity sensitive detector. At any point within the optical wavefront, the optical intensity rises and falls at the beat frequency.

Unfortunately, a photo detector does not have a point aperture. It has a finite aperture, over which the intensity is summed, to produce the detector output current.

When the two beams being mixed are perfectly aligned with each other and the surface of the detector, the nodes and anti-modes associated with the two beams will occur at the same time at all points on the detector surface.

When the two wavefronts form a node the detector produces no signal current. When they form an anti-node the detector current rises to a maximum value. In this case the maximum fluctuations in photo detector output current, occur. The mixer product, therefore, has its greatest amplitude.

If the light beams are not aligned with each other and/or they are not perfectly orthogonal to the surface of the detector, the amplitude of the photo-detector output current falls [27],[29],[30]. This effect occurs because anti-nodes appear at some point on the detector surface at the same time as nodes at other points. Therefore the total amount of light falling upon the detector surface changes by a smaller amount [27],[29],[30]. The
total amount of light neither reaches the minimum or maximum values that occur when the incident light beams are perfectly aligned. Moreover at certain values of misalignment angle, the total detector power, at the surface of the detector, does not change with time \[27],[28],[30]. Only the positions of nodes and anti-nodes, on the detector surface, move across the surface with time.

For these reasons it is essential that the light beams, to be mixed, are perfectly aligned and orthogonal to the detector surface, if optimum mixing or detection is required \[27],[29],[30].

If the light beams, incident at the detector surface are misaligned, the photo currents predicted in sections 5.4, 5.5 and 5.6 will be greater than those obtained by measurement.

In order to gain an accurate theoretical prediction of the photo current due to misaligned incident optical wavefronts, the beam misalignment must be included in the equations describing the detection process. However misalignment is not a property of the quantum efficiency of the detector, this is fixed by the properties of the material from which the detector is made. This beam misalignment must be included with the electric fields of the signals incident upon the detector.

Mathematically, the beam misalignment is described using retarded electric fields of the form \[27],[29],[30].

\[
E(t, r) = E \cos(\omega t - k \cdot r)
\]  

(5.240)
Where \(E(t, r)\) is the time varying retarded electric field, \(E_0\) is the peak electric field and \(r\) is the direction of field measurement.

Equation (5.240) may be written in terms of a unit wave direction vector \(\hat{k}\).

\[
E(t, r) = E \cos(\omega t - \frac{\omega}{c} \hat{k} \cdot r)
\]

(5.241)

Where \(k\) is a unit wave vector.

Consider two wavefronts of the form of (5.241) incident upon the surface of an ideal detector at the origin of co-ordinates as shown in Fig 5.20. Each wavefront is travelling in a different direction as indicated by \(k_1\), and \(k_2\), thus, [27],[29],[30]

\[
E_1(t, r_1) + E_2(t, r_2) = E_1 \cos(\omega_1 t - k_1 \cdot r) + E_2 \cos(\omega_2 t - k_2 \cdot r)
\]

(5.242)

Fig 5.20 Beam Misalignment at the Optical Detector Surface
In order to be detected, the light incident upon the surface of the crystal must penetrate into the crystal and photons must be absorbed [4]. Therefore the component of the propagation directions of each wavefront orthogonal to the surface of the crystal are of interest.

Let \( \mathbf{r} \) represent the normal to the surface of the crystal. Using the co-ordinate axes of Fig 5.20 vectors \( \mathbf{k}_1 \) and \( \mathbf{k}_2 \) and \( \mathbf{r} \) may be resolved into their components,

\[
\mathbf{k}_1 = k_{1x} \mathbf{i} + k_{1y} \mathbf{j} + k_{1z} \mathbf{k}
\]

\[
\mathbf{k}_2 = k_{2x} \mathbf{i} + k_{2y} \mathbf{j} + k_{2z} \mathbf{k}
\]

and

\[
\mathbf{r} = x \mathbf{i} + y \mathbf{j} + z \mathbf{k}
\]

Moreover, the surface of the detector is located at the origin of co-ordinates, therefore \( z = 0 \) in \( \mathbf{r} \). Whence at the surface, the combined electric field, penetrating the surface of the detector crystal is,

\[
E_1(t,x,y) + E_2(t,x,y) = E_1 \cos(\omega_1 t - k_{1x} x - k_{1y} y) + E_2 \cos(\omega_2 t - k_{2x} x - k_{2y} y)
\]

Equation (5.246) therefore describes the field distribution over the detector surface.

Using (5.246) the instantaneous intensity of the light at the detector surface is found [41],

\[
|E_1(t,x,y) + E_2(t,x,y)|^2 G_a = G_a \left\{ E_1^2 + E_1^2 \cos[2\omega_1 t - 2k_{1x} x - 2k_{1y} y] + E_2^2 + E_2^2 \cos[2\omega_2 t - 2k_{2x} x - 2k_{2y} y] + E_1 E_2 \cos[(\omega_1 + \omega_2) t - (k_{1x} + k_{2x}) x - (k_{1y} + k_{2y}) y] \right\}
\]
The electrical frequency response of the ideal photo detector will eliminate all but the DC and frequency-difference terms [4],[27],[29],[30]. Therefore interest is confined to these terms. Whence the optical intensity of interest is,

\[ I_0(x,y) = E_1^2 + E_2^2 + E_1 E_2 \cos[(\omega_1 - \omega_2)t - (k_{1x} - k_{2x})x - (k_{1y} - k_{2y})y] \]

(5.249)

Where \( I_0(x,y) \) is the optical intensity at the detector surface.

However, the DC term is of little interest; it contains no information, therefore,

\[ I_0'(x,y) = E_1 E_2 \cos[(\omega_1 - \omega_2)t - (k_{1x} - k_{2x})x - (k_{1y} - k_{2y})y] \]

(5.249)

represents the signal optical intensity.

The instantaneous optical power, of interest, at the photo detector surface, may be found by integrating \( I_0(x,y) \) over the surface [27],[29],[30]

\[ P_0(x,y) = \int_A I_0'(x,y) dA \]

(5.250)

Where \( A \) is the detector surface.

If the photo detector consists of a square crystal of dimensions \( x_0 \times y_0 \) then the optical power, penetrating the detector surface and generating a photo-current, is
The coefficient of the difference signal, at frequency \((\omega_1 - \omega_2)\), is the product of two sinc functions \([42]\) and the electric fields \(E_1\) and \(E_2\).

Using (5.241), (5.251) may be expressed in terms of the frequencies \(\omega_1\) and \(\omega_2\),

\[
P_0 = \frac{4E_1 E_2 \sin \left[ \frac{(k_{1x} - k_{2x})x_0}{2} \right] \sin \left[ \frac{(k_{1y} - k_{2y})y_0}{2} \right] \cos[(\omega_1 - \omega_2) t]}{(k_{1x} - k_{2x})(k_{1y} - k_{2y})}
\]

(5.251)

As the misalignment of beams increases the amplitude of the mixer product falls. The mixer product has zero amplitude when the arguments of either sinc function, in (5.252), go to integer values of \(180^\circ\).

The maximum value of (5.252) occurs when the arguments of both sinc functions are zero simultaneously, in which case

\[
P_0 = x_0 y_0 E_1 E_2 \cos(\omega_1 - \omega_2) t
\]

(5.253)

It is also clear from (5.252) that the amplitude of the mixer product, produced by the misaligned beams, is heavily dependent on the individual frequencies of the light beams. Therefore, misalignment introduces an additional optical bandwidth limit.
REFERENCES - CHAPTER 5

   "Optoelectronics - An Introduction."

   "Optical Communications."
   John-Wiley and Sons, 1976, pp 107 - 140.

   "Coherent Light."

[4] KINGSTON, R.H.
   "Detection of Optical and Infrared Radiation."

[5] HAYKEN, S.
   "Communications Systems."

[6] STREMLER, F.G.
   "Introduction to Communication Systems."

[7] CAVES, C.M.
   "Quantum Mechanical Noise in an Interferometer."
"Limits on Electromagnetic Amplification Due to Complementarity."
First International Conference on Quantum Electronics, High View, N.Y., 1939.

[9] GLAUBER, R.J.
"Coherent and Incoherent States of the Radiation Field."

"Equivalence of Semiclassical and Quantum Mechanical Description of Statistical Light Beams."

"Equivalence Classes of Minimum Uncertainty Packets."

[12] LOUDON, R.
"The Quantum Theory of Light."

[13] KITTEL, C.
"Introduction to Solid State Physics."

"Optical Communications."

[16] TAKENAKA, T., TANAKA, K. and FUKUMITSU, O.
"Signal-to-Noise in Optical Heterodyne Detection for Gaussian Fields."

[17] NICHOLSON, G. and CAMPBEL', J.C.
"Performance Estimates for Heterodyne and Homodyne Optical Fibre Communication Systems."

[18] TARG, R.
"Optical Heterodyne Detection of Microwave Modulated Light."

[19] YOSHIMURA, T., HARA, K. and WAKABAYASHI, N.
"Modified-Triggered Photocounting and its Application to Optical Superhetrodyne Spectrum Analyses."

[20] CHURNSIDE, J.H. and McINTYRE, C.M.

[21] CHURNSIDE, J.H. and McINTYRE, C.M.
"Signal Current Probability Distribution for Optical Heterodyne Receivers in the Turbulent Atmosphere. 2 Experiment."
[22] TEICH, M.C. and YEN, R.Y.  
"Three-Frequency Nonlinear Heterodyne Detection.  
1. C.W. Radar and Analog Communications."  

[23] TEICH, M.C. and YEN, R.Y.  
"Three-Frequency Nonlinear Heterodyne Detection.  
2. Digital Communications and Pulsed Radar."  

[24] HEALEY, P.  
"Fading in Heterodyne OTDR."  

"Two-Channel Heterodyne-Type Transmission-Type Experiment."  

[26] FRIED, D.L.  
"Optical Heterodyne Detection of an Atmospherically Distorted  
Signal Wave Front."  

[27] COHEN, S.C.  
"Heterodyne Detection: Phase Front Alignment, Beam Spot Size, and  
Detector Uniformity."  
[28] FINK, D. and VODOPIA, S.N.

[29] DEGAN, J.J. and KLEIN, B.J.
"Optical Antenna Gain 1:"

[30] DEGNAN, J.J. and KLEIN, B.J.
"Optical Antenna Gain 2:"

[31] Oliver, B.M.
"Signal-to-Noise Ratios in Photoelectric Mixing."

"Comments on Noise in Photoelectric Mixing."
Proc. IRE (Correspondence), Vol. 50, June 1962, pp 1544 - 1545.

[33] WALLS, D.F.
"Squeezed States of Light."

[34] HARVEY, A.F. (As for [3]) pp 80 - 84.

[35] BROWN, R.
"Lasers - A Survey of their Performance and Applications."
[36] HANBURY-BROWN, R.
"The Intensity Interferometer its Application to Astronomy."

[37] GOUGH, P.T.
"Notes on Optical Communications."
Notes of a Seminar Delivered to the Engineering Staff of the New Zealand Post Office, NZPO Seminar, 1982.

[38] JORDAN, E.C. and BALMAIN, K.G.
"Electromagnetic Waves and Radiating Systems."

[39] PANKOVE, J.I.
"Optical Processes in Semiconductors."

[40] WILSON and HAWKES
"Optoelectronics - An Introduction."

[41] JORDAN and BALMAIN (as for [38]) pp 162 - 176.

[42] BRACEWELL, R.N.
"The Fourier Transform and its Applications."
[43] KUCERA, V.
"Discrete Linear Control - The Polynomial Equation Approach."
ACADEMIA publishing house - Czechoslovak Academy of Sciences,
Prague, 1979.

[44] SMITH, R.J.
"Circuits Devices and Systems."

[45] HOLT, C.A.
"Electronic Circuits Digital and Analogue."

[46] MOTOROLA TECHNICAL INFORMATION CENTRE
"Motorola Optoelectronic Device Data."

[47] LAUMAN, L.E. (President)
"1985 D.A.T.A. Book (Optoelectronics)."

[48] DEMAW, D.
"The Radio Amateurs Handbook."
American Radio Relay League, Newington, CT, USA 06111.
pp 6.1 - 7.27.

[49] ROBINSON, A.L.
"Bell Labs Generates Squeezed Light."
[50] MARCUSE, D.

"Light Transmission Optics."

6. ELECTRONIC DEVICES FOR MODULATION AND DETECTION

In section 5.7 of the preceding chapter, techniques of post detection processing, intended to improve the SNR of detection, are discussed. In order to implement these techniques active equipment, such as amplifiers, mixers and integrators, is required.

In this chapter, the specification design and construction of modulation and detection equipment is discussed in detail. Although the equipment is intended to process digital signals, the detector output is generally analogue [1],[2],[3],[4],[5],[6],[7],[8], it can vary over a wide range of detector output currents. The design of detection equipment between the detector and the comparator [6],[7],[8], which decides whether a logical one or zero is present, is the design of analogue equipment. For analogue signals, which will eventually lead to the regeneration of digital signals, the design is more difficult than it would first appear.

Useful optical systems operate at high data rates, therefore wide equipment bandwidths are required. Moreover, the phase response of such equipment must be linear, to avoid reduction in bandwidths due to pulse smearing [9],[10],[11].

Wide bandwidth and linear phase are design requirements not present in analogue systems, where phase is unimportant, and different SSB signals are stacked, with differing amplitudes in the frequency domain, to overcome the bearer bandwidth limitations.

In section 6.1, the specifications for individual devices needed for modulation and detection, to reduce the SNR, are derived. The
specification selection may appear initially inconsistent, however the Acousto-Optic modulator was available some time after the detectors were purchased and the specifications fixed.

After examining a number of product lists, an SPD suitable for optical mixing experiments was selected, Section 6.2 outlines the specifications and manufacturers recommendations for mounting and connecting the detectors to the supply. From these specifications and recommendations, an electrical design was produced.

Using stripline techniques a photo receiver, suitable for use on the optical bench, was designed and built. Full details of the mechanical layout of the photo receiver are also available in section 6.2.

In order to modulate the light, an AO modulator was purchased from Isomet Corporation. The modulator arrived as two separate devices. A lead Molybdate AO cell with sonic transducer and absorber. An input signal to the AO cell is applied via an SMA connector mounted on one end of the cell. The AO cell was mounted on a platform with 3 degrees of spatial movement under screw gauge control so that it could be aligned precisely.

The second device was a sonic driver packaged in a case. To prevent its overheating, the driver was mounted on a heatsink in accordance with the manufacturers data.

The mounting details of both AO devices is discussed in section 6.3, together with adjustment to optimum operation for the 633nm Helium Neon Laser Wavelength.
A number of small experiments were conducted to test the performance of the photo receiver discussed in section 6.2, using a cooling fan to modulate the laser and later the AO modulator. The electrical bandwidth of these receivers was much lower than expected. Moreover the manufacturers pin configuration for the CD10 detectors was incorrect. With few details available about the CD10 detectors, from the manufacturer, their use was discontinued until the faults could be rectified. A second temporary photo receiver was built for experimentation. Its design and construction is discussed in section 6.4 together with its use in frequency mixing applications.

Frequency mixing experiments illustrated the critical aspects of detector alignment and its frequency dependence when modulating the Laser with the AO cell. Methods of increasing the optical mixer bandwidth, by reducing the frequency dependent deflection effects of the AO modulator, are discussed in section 6.5.

Typical output levels for optical mixer products were found in the experiments of section 6.4. These levels are generally of the same order as those encountered at the output terminals of a radio antenna [12]. Extensive amplification of the detector output signals is, therefore, required before any form of SNR enhancement, by post detector processing means, can be made. In section 6.6 the design of amplifiers suitable for this application is covered. That theory of broadband design is introduced together with some of its failings. Building on the knowledge acquired, from the amplifiers actually built, a new technique of broadband amplifier design was developed. The new technique of broadband circuit design is discussed in chapters 7, 8 and 9.
In order to exploit post detector processing methods of reducing the SNR of detection, a frequency multiplier is required. Section 6.7 outlines the design of a broadband differential mixer suitable for the correlative application. Measurement of its stability and frequency response are also given.

Finally, techniques of integration, required for correlation applications, are reviewed in section 6.8. An integrator for long term correlative measurement was designed. Its design and principles of operations are discussed. The design of correlators for short term measurements is also discussed in section 6.8.
6.1 THE GENERAL REQUIREMENTS OF ELECTRONIC DEVICES IN OPTICAL APPLICATIONS

In order to produce useful experimental results, involving optical mixing and coherent detection, a SPD which can generate electric currents, at radio frequencies, is necessary. Many common SPD's can operate up to bandwidths of only a few KiloHertz or MegaHertz. Such devices would be unsuitable for experiments which should reflect typical optical communications applications, so that useful comparisons between experiments and the characteristics of commercial systems can be made. Equipment should be selected, therefore, with electrical bandwidths in the hundreds of MegaHertz range of commercial optical communications systems.

Opto-Electronics Incorporated, produce a series of detectors which will detect radiation modulated at frequencies up to 1.2 GHz. These devices are suited to a wide range of experiments; thus two Opto-Electronics CD10 silicon photo detectors were purchased.

The CD10 photo detectors consist of SPD's mounted in thick film circuits to match the diodes to a standard 50Ω electrical system. A full specification list for the CD10 photo detector is given in appendix 14.

As it is desired to mix optical signals, the optical source used should have the narrowest possible optical bandwidth. Lasers available in the electrical engineering department are Spectra Physics 135 Helium Neon lasers. These are unstabilized, and have a coherence length of only one or two metres. It was therefore decided to purchase a Spectra Physics 117 HeNe laser, which is activity stabilized, it has a coherence length of a few hundred metres. The Spectra Physics 117 laser puts out 4mW of optical power, which was considered adequate for most experimentation.
The output voltage levels expected, for a CD10 photo detector loaded with 50 Ω and excited by the 177 laser, are low therefore some degree of amplification is required between the detector output and any signal processing equipment. Measurements of output signals would reveal the actual detector output levels, from which gain requirements could be derived. In order to determine how many amplifier stages would be necessary, to achieve the desired gain, it would be necessary to build a suitable amplifier and measure its gain when loaded.

The gain required before signal processing can be undertaken, is fixed by the input levels required by the mixer to produce optimum conversion efficiency. Moreover, based upon some initial experiments carried out on a temporary differential mixer, described in section 6.7, a system bandwidth figure was fixed at 500MHz. It is the bandwidth over which the mixer generated significant mixer products.

The 500MHz bandwidth was also suitable because measurement equipment was available to cover this frequency range.

The optimum data speed associated with a bandwidth of 500MHz is 1G bps, therefore data rates, similar to those of commercial systems, could be accommodated.

Equipment was therefore designed to operate over a 500MHz bandwidth. This presented a great design challenge, from which insights into new techniques of equipment design were gained. Possibly advancing the post optical detection design procedures.
In order to modulate the laser, an Acousto Optic modulator was purchased from Isomet Corporation. The cheaper series 230, 30MHz bandwidth device was purchased. Its full specifications are given in appendix 5. Although the information bandwidth is small, using combinations of the various deflection orders, frequencies over a 140MHz range can be generated at the photo detector output. Therefore the 500MHz design bandwidth is reasonable if such mixer products and harmonics are to be measured.
6.2 DETECTOR DESIGN

The SPD selected for use in the experiments, discussed in this chapter, is the Opto-Electronics CD10 Silicon photodiode. This device is based on the Motorola MRD500 silicon photodiode.

The Opto Electronics device incorporated the MRD500 into a thick film matching circuit for use in 50Ω transmission-line-impedance applications. It is capable of recovering optical modulation frequencies in excess of 1.2GHz.

Selection of the CD10 was a compromise between choice of bandwidth and availability of the device. With a bandwidth in excess of 1.2GHz the SPD could be used in experiments at realistic modulation frequencies. Moreover, its bandwidth would permit the experimenter to view a number of harmonics of the modulation frequency.

In addition to its Electrical and optical characteristics, the physical construction of the CD10 was attractive. It is mounted in a TO-3 case with a lens protruding from the top. This could be mounted easily, onto the side of a suitable case, and used on the optical bench.

For these reasons, two CD10 SPD's were purchased for a number of proposed experiments.

The condensed specifications of the CD10 SPD are given in table 6.1, and an equivalent circuit diagram in Fig. 6.1.
Table 6.1 Specifications of the CD10 SPD

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth -3dB</td>
<td>DC to &gt; 2.2GHz</td>
</tr>
<tr>
<td>Risetime</td>
<td>&lt; 250 ps</td>
</tr>
<tr>
<td>FWHM Pulse Response</td>
<td>&lt; 400 ps</td>
</tr>
<tr>
<td>Peak Flux Responsivity</td>
<td>0.15 A/W</td>
</tr>
<tr>
<td>Spectral Response Range</td>
<td>300 - 1100 nm</td>
</tr>
<tr>
<td>Active Area</td>
<td>500 x 500</td>
</tr>
<tr>
<td>Noise Equivalent Power</td>
<td>$1 \times 10^{-10}$ V/$\sqrt{\text{Hz}}$</td>
</tr>
<tr>
<td>Bias Voltage</td>
<td>15 V</td>
</tr>
<tr>
<td>Maximum Average Optical Input Before Damage</td>
<td>100 mW</td>
</tr>
<tr>
<td>Output Impedance</td>
<td>50</td>
</tr>
</tbody>
</table>

Fig 6.1 Equivalent Circuit for CD10 SPD

![Equivalent Circuit Diagram]

Fig 6.2 Pin Configuration for CD10 SPD

![Pin Configuration Diagram]

The TO-3 Manufacturers Pin configuration is shown in Fig 6.2.
The manufacturers mounting instructions are the same as common radio design practice at frequencies above 30MHz. The leads of the TO-3 package should be kept as short as possible. The output should be connected directly into a 50Ω stripline.

In order to ensure optimum detector performance, the manufacturers instructions were followed, and a photo detector mounting suitable for use on the optical bench, produced. This section discusses the design of the photo detector mounting and circuitry developed to produce a device that could be mounted on the optical bench standard equipment, and that would be electrically robust to prevent electrical damage to the detectors, which are expensive and involved long delivery periods.

The design and construction of a suitable stripline, using local materials and equipment, is discussed in section 6.2.1. However, due to the sparsity of information given by the manufacturer, the need for a stripline, in preference to other types of transmission line, is not understood.

After the design and manufacture of suitable stripline, a complete electrical circuit design was produced, for the CD10. The design was intended to protect the CD10 against over voltage or application of reverse voltage, as any error during the use of the device would be costly. Electrical design of the detector is covered in section 6.2.2.

Section 6.2.3 discusses the physical construction of the photo detector circuitry, and the optical mounting aspects.

Unfortunately, a number of problems were encountered with the CD10 photo-detector, which rendered the devices useless for experimentation.
The manufacturers information was extremely limited, and the given pin configuration incorrect (see Fig 6.2), therefore after a number of electrical and optical tests, outlined in section 6.2.4, it was decided that the active devices, in the CD10, were mounted incorrectly. The photo-receivers would need extensive modification, in a sterile thick film laboratory, before they would operate in the manner specified.

As a result of the problems summarised above, an alternative device was used for detection. It was found that the active device used in the CD10, the MRD 500, was available in New Zealand. This device was used for initial experiments. More details on its use are available in section 6.3.

6.2.1 Photo-Receiver Physical Design

It was mentioned in the preceding section that the detector would be mounted on the side of a suitable case. An Eddystone diecast aluminium box is suitable to mount the CD10, as shown in Fig 6.3.
Using a TO-3, insulating and mounting kit, the CD10 SPD was mounted on the side of the box in the position shown in Fig 6.3, with the TO-3 pins protruding into the box. The lower pin is the output pin of the CD10 (refer to Fig 6.1); it is soldered directly to the stripline as shown in Fig 6.4. The stripline is mounted in the box using 6BA plastic bolts and plastic standoffs. The bottom of the box was tapped to accommodate the plastic screws. Plastic bolts and plastic standoffs were used to insulate the $\mu$-stripline from the metal of the Eddystone box.

Using a washer arrangement, a bulk-head BNC coaxial socket was mounted in the end opposite the CD10. Its centre pin was soldered directly to the $\mu$-stripline as shown in Fig 6.4.

The CD10 was connected, using copper straps, to the groundplane of the stripline via the mounting bolts, and 8 mm tags. At the BNC socket end,
the groundplane was connected to the barrel of the BNC socket using two lugs and copper straps as shown in Fig 6.4.

The voltage regulator and power input socket are located on the bottom of the Eddystone box. Without crossing over the top of the stripline, the capacitors and radio frequency chokes (RFC's) were mounted between each component using their strength and solder as shown in Fig 6.5.

Finally, a block of aluminium was bolted onto the bottom of the diecast to accommodate the threads of the optical bench stands, which are tapped with an optical thread.

Fig 6.4 shows an exploded view of the optical receiver. Fig 6.5 shows the internal electrical connections of the power supply to the CD10 SPD via the RFC's.
Fig 6.5 Electrical Connection Inside the Photo-Receiver

Fig 6.6 gives a view of the SPD, stripline, voltage regulator and connection inside the photo-receiver.

Fig 6.6 Inside the Photo-Receiver
Front and rear views of the photo-receiver are shown in Fig 6.7 and Fig 6.8 respectively. The Photo-receiver is mounted on a typical optical-bench stand.

Fig 6.7 Front View of Photo-Receiver
To test the optical-receiver, a Spectra Physics 135 Helium-Neon Laser was used. The power supply was connected to the photo-receiver, and the laser aimed at the detector lens. No measurable DC level change was noticed on a Hewlett Packard 1740A oscilloscope (HPO). Yet, when the power supply was removed, the oscilloscope registered a change, produced by the SPD in photo-voltaic mode.

Using a lens as shown in Fig. 6.9, the laser light was concentrated at the lens of the CD10.
When the power was connected, no output voltage could be measured. Once the power supply was disconnected, a photo-voltaic voltage, of greater magnitude than in the previous case, was measured.

Using the oscilloscope, measurements of voltage inside the detector were made. No measurable voltage could be found between the bias point and output point of the CD10 (see Fig 6.1 and Fig 6.2). On removal of the bias connection of the CD10, the output voltage of the voltage regulator rose to the required 15 V. Yet when reconnected to the CD10 bias pin, the output voltage of the voltage regulator fell to 0.

Suspecting optically, or electrically, induced damage, of the CD10, an AVO meter was used to test the SPD using Fig 6.1 as a guide. No junction damage was found, however the measured diode forward bias direction was the opposite of the direction stated in the manufacturers data.

Assuming that the manufacturers data was incorrect, the CD10 was reversed, and built into a new prototype optical receiver. When the power supply was connected, in this case, a definite increase in detector output
voltage was measured by the HPO. With the fault corrected, further tests on the photo-receiver could be conducted.

Using a SANYO SANACE-25 cooling fan, the laser output was chopped as shown in Fig 6.10 (the Acousto optic modulator had not yet arrived).

![Modulation of Light Using a Hitachi Fan](image)

A square wave of 40 μV amplitude was measured about a 20 μV DC level using the HPO.

Using additional lenses, to focus the incoming light onto the detector lens, tends to complicate alignment of the detector with the beam. Small variations in lens positioning cause large changes in the beam position at the detector due to the magnifying properties of the lens. As the lens number in any experiment in the optical bench is increased, the alignment of the detector becomes more difficult. For this reason, the lens was subsequently removed from one photo-detector to enable the experimenter to focus the beam directly onto the naked crystal. Details, of the lens removal and experimental results obtained with it, are discussed in more detail in section 6.4.
The subsequent arrival of the AO modulator, its assembly and mounting, which are discussed in section 6.3, enabled testing of the photo-receiver at much higher modulation frequencies than were possible with the SANYO fan.

Using the AO modulator, the frequency response of the optical detector was tested over a 30MHz range. The light, from a Spectra Physics 117 activity stabilised laser, is shone through the aperture of the AO modulator and onto a white card, as shown in Fig 6.11.

The AO modulator driver is connected to the power supply. The AO cell, and mount, are rotated until a second spot appears on the card. Further adjustments of the AO cell are made until the two spots are of equal intensity. When this occurs, the input light is incident with the acoustic wavefront of the Bragg angle. (Refer to Chapter 4 and Appendix 5.)

The detector is aligned with the deflected beam; it is adjusted until the maximum output voltage, due to the presence of light, is reached.
Using a Wavetek Model 193 variable frequency oscillator (VFO) a 1 V peak to peak sinusoid was supplied to the modulation input of the AO driver. The deflected beam was thus amplitude modulated by the input sinusoid. The output voltage of the detector rose and fell in sympathy with the modulation.

By varying the frequency of the VFO, the modulation frequency can be varied, and the frequency response of the photo receiver found at a series of spot frequencies. The relative frequency response of the optical receiver was measured using a Hewlett Packard Spectrum analyser. The results are given in table 6.2.
Table 6.2 The Frequency Response of a Photo-Receiver

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Relative Power Output dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 KHz</td>
<td>0</td>
</tr>
<tr>
<td>30 KHz</td>
<td>0</td>
</tr>
<tr>
<td>40 KHz</td>
<td>0</td>
</tr>
<tr>
<td>50 KHz</td>
<td>0</td>
</tr>
<tr>
<td>60 KHz</td>
<td>0</td>
</tr>
<tr>
<td>70 KHz</td>
<td>0</td>
</tr>
<tr>
<td>80 KHz</td>
<td>0</td>
</tr>
<tr>
<td>90 KHz</td>
<td>0</td>
</tr>
<tr>
<td>100 KHz</td>
<td>0</td>
</tr>
<tr>
<td>200 KHz</td>
<td>0</td>
</tr>
<tr>
<td>300 KHz</td>
<td>0</td>
</tr>
<tr>
<td>400 KHz</td>
<td>0</td>
</tr>
<tr>
<td>500 KHz</td>
<td>0</td>
</tr>
<tr>
<td>600 KHz</td>
<td>0</td>
</tr>
<tr>
<td>700 KHz</td>
<td>0</td>
</tr>
<tr>
<td>800 KHz</td>
<td>0</td>
</tr>
<tr>
<td>900 KHz</td>
<td>0</td>
</tr>
<tr>
<td>1 MHz</td>
<td>0</td>
</tr>
<tr>
<td>2 MHz</td>
<td>-1</td>
</tr>
<tr>
<td>3 MHz</td>
<td>-2</td>
</tr>
<tr>
<td>4 MHz</td>
<td>-4</td>
</tr>
<tr>
<td>5 MHz</td>
<td>-4</td>
</tr>
<tr>
<td>6 MHz</td>
<td>-7.5</td>
</tr>
<tr>
<td>7 MHz</td>
<td>-10</td>
</tr>
<tr>
<td>8 MHz</td>
<td>-15</td>
</tr>
<tr>
<td>9 MHz</td>
<td>-15</td>
</tr>
<tr>
<td>10 MHz</td>
<td>-16</td>
</tr>
<tr>
<td>11 MHz</td>
<td>&lt;-20</td>
</tr>
</tbody>
</table>

The noise level of the photo-receiver and following circuitry is at 20dB relative to the maximum signal power. The relative photo-receiver frequency response is shown graphically in Fig 6.12.
On examination of Fig 6.12 it is clear that the -3dB frequency of the photo receiver is well below the 1.2GHz given by the manufacturer, it is in fact at 3.5MHz. Naturally, the initial reaction of any designer is that a design oversight had been made, effectively reducing the detector bandwidth to the 3.5MHz measured. Moreover, the sensitivity of the photo-receiver seemed very low producing an output voltage of only 13mV (unloaded), with no modulation present. For these reasons, it was decided to measure the actual power sensitivity of the photo-receiver, by measuring the optical power of the photodiode the corresponding electrical output power, and calculating the photo-receiver sensitivity. The CD10 sensitivity could be compared with the sensitivity stated by the manufacturer, and the sensitivities of other SPD's, of similar electrical bandwidth, given in manufacturers catalogues [13],[14],[15],[16],[17].

The spectral content of laser output light was discussed in Chapter 3. Generally the output spectra cover many Megahertz of optical bandwidth.
During the detection process, each component mixes with every other frequency component. Therefore the optical current ranges from DC to the limits of the electrical response of the detector. Even the stabilized Spectra Physics laser produces output over approximately a 1 MHz frequency range. Therefore, in order to measure the total electrical power produced by the photo-receiver for a given optical power, a device which measures the output power over the entire frequency range of the electrical filter is needed. It was decided to use an Hewlett Packard HP 3400A thermocouple based true-RMS meter to measure the responsivity of the photo-receiver. Unfortunately, the true RMS volt-meter is sensitive only up to frequencies of 10MHz. Moreover its input is terminated in a balanced 600 Ω. To use this device with the 50 Ω photo-receivers, some matching would be necessary. For simplicity it was decided to match the 600 Ω true RMS meter to the 50 Ω detector at 10MHz, and ignore the mismatch at lower frequencies, where its significance would decrease for short lengths of transmission line.

Matching was achieved using the circuit in Fig 6.13.
Coaxial cable was connected to the optical receiver and the free end to a general radio unbalanced "T" section. A length of coax was connected from the through piece of the "T" section via a "π" coupler consisting of a 450pF variable capacitor, a General Radio GR 874XL coaxial inductor and a 570nF polycarbonate capacitor in a General Radio component mount. The output of the "π" coupler was connected to a Termaline 50Ω dummy load.

Coaxial cable, with banana plugs on one end, was connected to the "vertical" connection of the "T" section, and the banana plugs connected to the true RMS voltmeter. An HPO was also connected between the two inputs to the true RMS voltmeter.

Using the vector voltmeter, capacitor c2 was adjusted to give a 50Ω impedance at the point where the photo-receiver is connected.

The photo-receiver was connected into the circuit ready for measurements of the photo-receiver output.
The Spectra Physics 135 HeNe laser was mounted on the optical bench. Using a 10x microscope objective lens, a 12 μm pin hole and a 147F x 42D convex lens, a spatial filter was formed and placed in front of the laser as shown in Fig 6.14.

![Fig 6.14 Optical Equipment for Power Measurement](image)

Using a MACAM radiometer and detector head with a preceding 1.0 and 0.5 neutral density filter combination, the total power incident upon the detector head was measured. The chosen lens combination used in the spatial filter, produced a broad light beam which over-filled the detector head, in which case the optical intensity is found by dividing the power measurement by the effective area of the detector head [18].

Measurements of incident power were made firstly with no modulation present, then with the laser modulated by the Sanyo fan. In each case, 10 measurements were made at 30 second intervals. The results were then averaged, and the intensity calculated in each case. The results are summarised in table 6.6.
Table 6.3 Intensity Measurements

<table>
<thead>
<tr>
<th>Mode</th>
<th>Modulation Frequency</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmodulated</td>
<td>0</td>
<td>17.885 W/H HA²</td>
</tr>
<tr>
<td>Modulated</td>
<td>208 Hz</td>
<td>9.1079 W/M²</td>
</tr>
</tbody>
</table>

Removing the radiometer detector head, the photo receiver was placed such that its lens was in the centre of the beam. The unmodulated light output could only be measured by the oscilloscope, as the true RMS voltmeter does not respond below 10Hz, therefore accurate measurement of the true RMS voltage is impossible. When the light is modulated however, a true RMS measurement may be obtained. The results are given in table 6.4.

Table 6.4 Optical Receiver Measurement

<table>
<thead>
<tr>
<th>Mode</th>
<th>Modulation Frequency</th>
<th>Voltage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Oscilloscope</td>
</tr>
<tr>
<td>Unmodulated</td>
<td>0</td>
<td>0.45 mV +</td>
</tr>
<tr>
<td>Modulated</td>
<td>208 Hz</td>
<td>0.425 mV</td>
</tr>
</tbody>
</table>

denotes peak voltage

A second photo-receiver was built. It was also tested, the results of which are given in table 6.5.
Table 6.5 Optical Receiver Measurements for Receiver 2

<table>
<thead>
<tr>
<th>Mode</th>
<th>Modulation Frequency</th>
<th>Voltage Oscilloscope</th>
<th>Voltage RMS Voltmeter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmodulated</td>
<td>0</td>
<td>0.5375 V</td>
<td>NA</td>
</tr>
<tr>
<td>Modulated</td>
<td>208 Hz</td>
<td>0.5625 V</td>
<td>0.555 mV</td>
</tr>
</tbody>
</table>

The receiver load is approximately 50 Ω, therefore the receiver RMS current is given by (6.5).

\[ I_{LRMS} = \frac{V_{LRMS}}{RL} \]

Where \( I_{LRMS} \) is the RMS load current, and \( V_{LRMS} \) is the output voltage.

Using the RMS measurement of voltage, the detector output current for both detectors is obtained as given in table 6.6.

Table 6.6 Detector Output Current

<table>
<thead>
<tr>
<th>Detector</th>
<th>RMS Current</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.05 A</td>
</tr>
<tr>
<td>2</td>
<td>11.1 A</td>
</tr>
</tbody>
</table>

It is assumed that light incident anywhere in the aperture of the photo-detector lens is focused onto the detector surface. The aperture
diameter is 3.8 mm, therefore the area of the aperture is $11.34 \times 10^{-6}$ m$^2$. In which case the incident power is 103.3 $\mu$W. The corresponding flux responsivities for both optical-receivers in given in table 6.7.

<table>
<thead>
<tr>
<th>Detector</th>
<th>Flux Responsivity</th>
<th>Measured</th>
<th>Manufacturers Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.08761 A/W</td>
<td>0.1095 A/W</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.1075 A/W</td>
<td>0.1095 A/W</td>
<td></td>
</tr>
</tbody>
</table>

Both measurements are within 20% of the manufacturers specifications, therefore, it seems unlikely that the photo diodes themselves were damaged. Any fault must lie in, either the thick film circuits of the CD10 SPD's, or in the bias-voltage supply circuits and output circuits, of the photo-receivers.

By measuring the complex impedance of a darkened CD10 photo-receiver, it should be possible to find which component combination, in the photo-receiver, is producing the undesired low frequency roll off.

The vector voltmeter is connected to the electrical output of a photo-receiver. The power supply to the photo-receiver is also connected. Using a General Radio 1208B unit oscillator, the output reflection coefficient of the photo-receiver was measured at regular intervals from 1MHz to 50MHz. The results are given in table 6.8.
Table 6.8 Energised Photo-Receiver Output Reflection Coefficient

<table>
<thead>
<tr>
<th>Frequency MHz</th>
<th>Reflection Magnitude</th>
<th>Coefficient Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-1.96°</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-4.7°</td>
</tr>
<tr>
<td>5</td>
<td>1.111</td>
<td>-7.35°</td>
</tr>
<tr>
<td>7</td>
<td>1.041</td>
<td>-11.01°</td>
</tr>
<tr>
<td>9</td>
<td>0.9672</td>
<td>-8.94°</td>
</tr>
<tr>
<td>10</td>
<td>0.9703</td>
<td>-13.75°</td>
</tr>
<tr>
<td>15</td>
<td>0.9653</td>
<td>-19.8°</td>
</tr>
<tr>
<td>20</td>
<td>0.9264</td>
<td>-20.4°</td>
</tr>
<tr>
<td>30</td>
<td>1.084</td>
<td>-9.7°</td>
</tr>
<tr>
<td>50</td>
<td>0.9206</td>
<td>-28.47°</td>
</tr>
</tbody>
</table>

The reflection coefficient measurements of Table 6.7 are plotted in Fig 6.15.
Reading the normalised impedance values off the Smith Chart of Fig 6.15, it is evident that the 50Ω impedance, expected at the output, is not reached at any frequency. With some minor variations the impedance measurements behave like those of an open-circuit or short-circuit transmission line [19],[20],[31]. The CD10 SPD does not present the 50Ω impedance expected at the input end of the stripline. It is obvious that the manufacturer's brief application note is misunderstood. Unfortunately the manufacturer has provided insufficient information to find the conception fault in the circuit. At a guess, it may be necessary to provide an SPD bias with a 50Ω impedance to correctly terminate the combined transmission lines of the SPD and stripline.
6.2.2 Stripline Design for the CD10 SPD Silicon-Photo-Diode

In this section, the manufacture of a RF stripline, for connection to the CD10 SPD to satisfy the manufacturers recommendations, is discussed.

The most suitable di-electric material for use from DC to 1.2 GHz is Teflon, which introduces minimum loss at 1.2 GHz [22]. However Teflon is an expensive material, therefore it was decided to construct a stripline from cheaper materials initially, and measure the characteristics of the cheaper stripline. If the stripline, made in this way, satisfies the impedance requirements, at least at frequencies below 200MHz, then the same manufacture process could be used with Teflon di-electric, if necessary.

The manufacture of full stripline is a complicated process, because a full stripline consists of a conductor of appropriate width, sandwiched between two slabs of di-electric, which in turn are sandwiched between two metallic groundplanes, as shown in Fig 6.16 [20].

![Fig 6.16 The Structure of Full Stripline](image-url)
However, it seemed unnecessary to use full stripline, as shown in Fig 6.2, when a micro stripline ($\mu$-stripline) could be manufactured much more easily [20]. For this reason, a microstripline was made to connect the SPD external equipment.

$\mu$-Stripline consists of one centre conductor, of appropriate width, separated by a dielectric from the groundplane, as shown in Fig 6.17 [20].

![Fig 6.17 The Structure of $\mu$-Stripline](image)

Using double sided circuit board the $\mu$-stripline could readily be made by removing metal from one side, to produce a centre conductor of appropriate width. The experimental $\mu$-stripline was built out of double sided Epoxy dielectric circuit board, available locally.

Impedance of a $\mu$-stripline is dependent upon 3 parameters [20],

1. Di-electric thickness
2. Conductor width
3. Di-electric constant

These parameters are shown if Fig 6.18.
For a given set of parameters, the ratio of the conductor width, $w_c$ to dielectric thickness $h_d$, is related to the desired stripline impedance $Z_{so}$ by either (6.1) or (6.2) [20],

\[
\frac{w_c}{h_d} = \frac{8 \exp[A_s]}{\exp[2A_s] - 2} \quad \text{for } A \geq 1.52
\]  

or,

\[
\frac{w_c}{h_d} = \frac{2}{\pi} \left\{ \frac{B_s - 1 - \ln(2B_s + 1) + \epsilon_r - 1}{2\epsilon_r} \right\} \times \left[ \ln(B_s - 1) + 0.39 - \frac{0.61}{\epsilon_r} \right]
\]

Where

\[
A_s = \frac{z_{so} (\epsilon_r + 1)}{60 \left( \frac{1}{2} \right)^{1/2}} \frac{1}{\epsilon_r + 1} \left( \frac{0.23 - 0.11}{\epsilon_r} \right)
\]

and

\[
B_s = \frac{60 \pi^2}{z_{so} \sqrt{\epsilon_r}}
\]

$z_{so}$ is the desired characteristic impedance.
In order to use these design equations the dielectric constant, \( \varepsilon_r \), of the Epoxy is needed.

By making a capacitance measurement between the two layers, of a known area of the circuit board, an estimate of the Epoxy dielectric constant can be obtained using the relation (6.3) for an ideal capacitor [23],

\[
c = \frac{\varepsilon_0 \varepsilon_r a_b}{h_s}
\]  

(6.3)

Where \( a_b \) is the area of the conducting plates.

By connecting a 56Ω resistor across a piece of double sided circuit board 15mm x 40mm, and connecting one plate to the centre conductor of a BNC type connector and the other to the outer conductor, as shown in Fig 6.19, a measurement of the parallel impedance could be made by using a Hewlett Packard HP 8405A vector voltmeter. From the impedance measurement, the capacitance of the parallel plates can be calculated.

---

**Fig 6.19 Measurement of Di-Electric Constant**
The vector-voltmeter was connected to a directional coupler, as shown in Fig 6.20, such that measurement of forward and reflected voltage, and their relative angles, could be made [24].

Using a General Radio GR-1208B unit RF oscillator, the length of the adjustable transmission line was varied until a short circuit, at the test point, produced a 180° frequency independent phase shift. The test circuit was then connected to the test point, and a series of forward and reverse voltage measurements and relative angle measurements made, at suitable intervals, over a 65MHz to 500MHz range. The ratio of reverse to forward voltage, and relative phase angle produce a frequency dependent reflection coefficient of the unknown impedance. Reflection coefficients measurements at the spot frequencies are given in table 6.9.
Table 6.9 Unknown Impedance Reflection Coefficients

| Frequency (MHz) | $|\sigma|$  | $\angle \sigma$ |
|----------------|------------|---------------|
| 100            | 0.1852     | -84°          |
| 150            | 0.2744     | -106.0°       |
| 200            | 0.3470     | -124.0°       |
| 250            | 0.444      | -139.4°       |
| 300            | 0.5333     | -153.3        |
| 350            | 0.5893     | -166.5        |
| 400            | 0.6652     | -177.5        |
| 450            | 0.7000     | +172.4        |
| 500            | 0.6667     | +161.5        |

From the data in table 6.9, it is clear that some inductance is also present in the connection to the test circuit.

The inductance was found by disconnecting the coaxial cable from the circuit and shorting it together. A second set of measurements of the reflection coefficient, at various spot frequencies, was made. The results are given in table 6.10.
Table 6.10 Measurement of Connection to Test Piece

<table>
<thead>
<tr>
<th>Frequency (MHz)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.975</td>
<td>162.7</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>0.9688</td>
<td>155.0</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.9893</td>
<td>148.4</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.9767</td>
<td>141.4</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>0.9833</td>
<td>135.2</td>
<td></td>
</tr>
<tr>
<td>350</td>
<td>1.0</td>
<td>129.7</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.9943</td>
<td>124.0</td>
<td></td>
</tr>
<tr>
<td>450</td>
<td>1.0</td>
<td>119.0</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.9667</td>
<td>113.3</td>
<td></td>
</tr>
</tbody>
</table>

Therefore the equivalent circuit, of the test piece and the connection to it, is given in Fig 6.21.

**Fig 6.21 Equivalent Circuit for Test Piece and Connection**

By plotting the reflection coefficients given in tables 6.8 and 6.9, on a Smith Chart, the capacitance of the circuit board test piece, may be found in the following way [19].
Reading the normalized impedance values of circuit board and test piece off the Smith Chart [19], a third plot of total impedance minus the test piece inductance is obtained [19], as shown in Fig 6.22 (plot C).

The resulting capacitive-reactive impedance is converted to an admittance by reflection in the $1+j0$ point of the Smith Chart [19], (plot D) is thus produced in Fig 6.22 [19]. By shifting these points along lines of constant admittance to the perimeter of the Smith Chart, the capacitive susceptance is found [19] (plot E). At each spot frequency, the capacitance may be found using equation (6.3).

$$c = \frac{B_C}{2\pi f_t} \quad (6.4)$$

Where $c$ is the capacitance, $B_C$ is capacitive susceptance and $f_t$ is the frequency at which the measurement was made.
At each value of frequency, a slightly different value of capacitance was found, however the variation in capacitance values was only of the order of 20%. The decision as to which value should be used was made, with the ultimate application in mind. The dimensions of the stripline would be most critical at the highest frequency [19],[20], it was therefore decided that the capacitance value selected would be the 500MHz value, which is 11.47\(\mu\)F.

Using 6.2, the di-electric constant is \(\varepsilon_r = 3.24\).

Equation 6.2 does however, ignore the effects of fringing fields [23], yet the results obtained using 6.2 seem to be sufficiently accurate for the stripline, as was experimentally verified.

The remaining parameter needed, to use 6.1 or 6.2, is the di-electric thickness \(h_0\), which was measured with a graduated jewellers lens, therefore \(h_0 \approx 1.5\)mm.

Using \(\varepsilon_r\) and \(h_0\) the value of \(A_s\) is 0.1692, therefore (6.2) is used, giving a \(\mu\)-strip width, \(\omega_c = 3.58\)mm [20].

It was decided to mount the TO-3-CD10 on the front of an Eddystone 87134 diecast aluminium box, with the \(\mu\)-stripline inside, in which case the maximum width of stripline di-electric and ground plane should be less than 112mm.

Using the computer aided design program MCAP given by Gupta et al [25], \(\mu\)-stripline, of various total groundplane widths, was simulated. The characteristic impedance, of the stripline, was unaltered even for widths
of 70mm. Therefore the total μ-stripline width chosen was 70mm, which would fit into an Eddystone box, and would be easily mounted. Further details of the physical design of the detector are given in section 6.3.2.

The plans for detector μ-stripline are given in Fig 6.23.

![Micro-Stripline Design](image)

Rather than sending the circuit board to a circuit etching establishment, for the manufacture of a prototype, it was decided to mill away the excess copper, to produce the stripline, because of its simplicity. Thus, if the μ-Strip width was unsatisfactory, it could be further reduced rapidly and accurately.

Using the mill, a prototype was produced according to the specifications of Fig 6.23.

The μ-stripline was tested using the vector voltmeter. A 56Ω resistor was connected at one end between the centre conductor and groundplane. The free end was connected to the test coax, used for capacitance measurement. At 50MHz intervals, the impedance of the stripline was measured. The results are given in Table 6.11.
Table 6.11 Micro-Stripline Reflection Coefficients

| Frequency (MHz) | $|\sigma|\quad |\sigma|$ |
|-----------------|------------------|
| 100             | 0.041            | - 0°   |
| 150             | 0.074            | - 85.0°|
| 200             | 0.11             | -104.0°|
| 250             | 0.15             | -122.0°|
| 300             | 0.21             | -148.0°|
| 350             | 0.23             | -176.0°|
| 400             | 0.25             | +158.8°|
| 450             | 0.23             | +129.0°|
| 500             | 0.244            | +96.4°  |

Using the Smith Chart in the same way as in Fig 6.22, the impedance of the terminated stripline was calculated [19]. (See Fig 6.24.)
From the reflection coefficient plot of Fig 6.24, it is clear that the stripline impedance measurement varies, to some extent, with frequency. However, the worst deviation in impedance causes a power loss of the resistor of only 6%, which is not significant [19]. Moreover, the amount of power reflected was insufficient to drive the vector voltmeter adequately, therefore the measurements involve greater error than Fig 6.24 would indicate. Fig 6.24 also indicates that the stripline functions adequately even at 500MHz, well above frequencies at which Epoxy insulation is normally used [22].

Using the $\mu$-stripline, an electronic design was produced. By using an LM7965 three terminal voltage regulator, the photo diodes could be protected from overvoltage from the adjustable power supplies [26]. Unfortunately 3 terminal voltage regulators fail in the presence of high frequency electrical signals. Therefore the use of radio frequency chokes and Ferrite beads was essential, in addition to the usual tantalum capacitors needed with an LM7905 [26].

Bearing in mind these observations, a complete detector circuit, including power supply was produced; it is shown in Fig 6.25.
Gp is the Microstripline groundplane

L1 and L3 are radio frequency chokes

L2 and L4 are ferrite beads

C1 is a 0.1 F Disc ceramic capacitor

C2 and C3 are 0.22 μF Tantalum electrolytic capacitors.

All grounded components are connected to the groundplane of the μ-stripline. No connections are made to the chassis of the optical receiver. It is intended that the receiver chassis should be grounded, by a separate wire, to a common system-earth-point to reduce the effects of external interference [12],[27],[28], such as stray radio signals, and electromagnetic emissions from the nearby high voltage laboratory.

The BNC connectors were mounted using plastic washers machined from plastic dowel to insulate them from the diecast aluminium box.
6.3 MOUNTING AND ADJUSTMENT OF THE ACOUSTO-OPTIC MODULATOR

The Acousto optic modulator, supplied by Isomet Corporation comprises two parts, a Lead Molybdate AO cell, and a modulator-driver. The two must be connected via a coaxial cable with a BNC plug on the driver end and a SMA plug on the AO cell end.

The AO cell was mounted on a Daedal adjustable stage, using two bolts and a length of 15mm aluminium as shown in Fig 6.26. A thin layer of cork was glued to the aluminium to protect the AO cell from damage. The top of the stage can be moved using a finely threaded screw, providing 1 horizontal degree of freedom.

The bottom of the stage was bolted to the bottom of a precision optical mount, with screw adjustment, giving a second direction of horizontal movement, both horizontal degrees of movement are controlled by adjustment screws, as shown in Fig 6.27.
Using a 1.9mm brass rod, the assembly of Fig 6.27 is mounted on a base suitable for the optical bench.

The complete AO cell mounting assembly is shown in Fig. 6.28.
The acousto optic driver delivers 1W of electrical power to the AO cell, drawing a current of $I_{AOD} = 400mA$ at 28 volts. The driver input power is therefore 11.2W, 1W of which is delivered to the AO cell. The remainder 10.2W must be dissipated in the driver circuitry. Moreover the AO driver case temperature must not exceed 70°C. (Refer appendix 5).

In order to ensure that the case temperature remains below 70°C, the AO driver must be mounted on a heatsink.

By selecting a 20% derating of temperature, and an ambient temperature of 23°C, the required heat flux is, $\Phi_H = 3.236k/w$. The required length of black anodised aluminium extrusion type 56230, is 25.5 mm using free convection (Refer to Appendix 6) for a point heat source.
The AO driver, however, is not a point heat source, it has a base area of 28.6 mm x 101.6 mm. A heatsink, of the same length as the base, that is 101.6 mm long, of the type discussed, should be adequate to cool the driver. By mounting it on a longer heatsink the driver could be kept still cooler. It was decided, therefore to mount the AO driver on a 150 mm heatsink 150 mm in length so that it would fit inside a 200 mm x 200 mm x 300 mm box available. Heat sealing compound was used to optimize the transfer of heat from the driver to the heatsink.

The AO driver and mounting is shown in Fig 6.29.

**Fig 6.29 Mounting of the AO Driver**

In their manual on the AO modulator (appendix 5), Isomet Corporation recommend alignment of the modulator to the specific wavelength of light
with which it is to be used. Two adjustment points are accessible by the removal of the two snap-in plugs shown in Fig 6.35. One gives access to a power adjustment, the other to a bias adjustment.

By rotating the potentiometers behind the snap-in plugs, the power was adjusted to quarter of its peak level with no bias applied. The AO cell was aligned to optimize the intensity of the first order beam. Once aligned the power was increased to maximize the intensity of the first order beam, as measured by the photo-receiver of section 6.2. The bias control was rotated to reduce the first order intensity by half, rendering the AO modulator ready to use.

Little difference, in the beam intensities and number of deflected orders, could be seen between the zero and first order deflected beams before and after alignment.
6.4 A TEMPORARY PHOTO-DETECTOR

Until the fault with the photo-receivers discussed in section 6.2 could be found. An alternative method of photo detection was used. A TO-3 cover of one photo detector had been poorly fitted and during testing had fallen off, revealing the basic construction of the CD10 detector. It was observed immediately that the photodiode used in the CD10 was an MRD-500 silicon photodiode; manufactured by Motorola.

Fortunately, a number of these devices were available in New Zealand, and were therefore purchased.

The MRD-500 is packaged in a TO-18 case with a bi-convex lens on the top. Many of its specifications are similar to those of the CD10. The full specifications for the MRD-500 are given in appendix 7, the specifications corresponding to those of the CD10 are repeated in table 6.12.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Time</td>
<td>1.0 ns</td>
</tr>
<tr>
<td>Flux Responsitivity</td>
<td>6.6 A/mw/cm²</td>
</tr>
<tr>
<td>Spectral Response Range</td>
<td>300 - 1100 nm</td>
</tr>
<tr>
<td>Lens Diameter</td>
<td>4.735 mm</td>
</tr>
<tr>
<td>Dark Current $i_0$</td>
<td>0.05 nA $V_B=-15V$</td>
</tr>
<tr>
<td>Bias Voltage $V_B$</td>
<td>15V</td>
</tr>
</tbody>
</table>
Other reverse bias voltages can be selected, however these introduce more dark current and therefore more noise.

To operate the MRD-500 in a 50 Ω electrical environment, the manufacturer recommends the use of the circuit shown in Fig 6.30.

![Fig 6.30 Circuit for Operation of MRD-500](image)

At a bias voltage of 15V, the MRD-500 has an infinite incremental impedance, it may therefore be matched to the transmission line with a 50 Ω resistor as shown in Fig 6.30.

Using the MRD-500, a temporary detector was built on a small piece of veroboard, as shown in Fig 6.31. An extra 0.1 μF disceramic capacitor was included across the supply connection, and the power supply leads were twisted together to reduce noise, at the photo-receiver output, due to external electromagnetic sources [12],[27],[28]. The complete circuit diagram is shown in Fig 6.31.
In order to reduce the effects of multiple lens combinations, and reduce the difficulties of detector alignment, the lens of one MRD-500 detector was removed to expose the photo sensitive silicon, making it possible to focus incident light directly onto the silicon using an external lens. Using the MRD-500 detector, a second temporary photo-receiver was built.

6.4.1 Producing Mixer Products

Using the temporary photo-receiver with the lens removed from the MRD-500, it was decided to attempt detection of an optical mixer product. Unfortunately, as mentioned in section 3.1.4; the frequency content of a laser output is broad. If two such laser signals are mixed, a band of frequencies would be produced across the electrical bandwidths of the photo detector. It would be extremely difficult to find a definite mixer product. However, if one laser is used, and two separate beams generated by a beam splitter, the instantaneous frequencies of the two beams will be equal; if their path length difference is less than the laser temporal
coherence length [5]. A change in the frequency of 1 beam is matched by a change in frequency of the other. Therefore no frequency difference, between the two beams should be measured. Moreover, if 1 beam is shifted in frequency, using the AO modulator, it should differ in frequency, from the unmodulated beam, by a constant amount.

The spatial properties of light emerging from the AO modulator were discussed in section 4.2.2. When sinusoidal modulation is applied to the AO modulator driver, a sinusoidally amplitude modulated soundwave is launched into the AO crystal. Light incident upon the crystal is deflected into two orders each consisting of 3 separate beams with angular separation given by 4.25. Using a simple spatial filter, the centre beam, or carrier, of an order, can be removed from the ensemble leaving just the upper and lower sidebands. By shining these two beams onto the optical detector, a mixer product should be generated at the difference frequency of the two beams. An electrical output at that frequency would verify that the MRD-500 SPD was mixing the two optical beams.

In the first mixing experiment, it was intended to remove the central carrier from the first order deflection ensemble using a simple spatial filter. Using two lenses it was then intended to recolumnate the sidebands sufficiently close together that they would both be incident upon the detector crystal.

The activity stabilized laser was turned on and shone through the AO modulator crystal. Power was connected to the AO modulator driver, which, without modulation, produces an 80 MHz acoustic carrier. The AO cell was rotated until the intensity of the beam of first order deflection was of equal intensity to that of the undeflected beam.
By connecting a sinusoidal oscillator to the modulation input of the AO driver the 6 light beams, associated with a amplitude modulated sound wave, were produced.

Using a simple screen, the zero order ensemble was removed, as shown in Fig 6.32. The remaining first order ensemble was magnified with a D25-F37 lens and recolumnated using a D52-F220 lens; also shown in Fig 6.32.

Using an F140-D33 lens, and a x10 microscope objective lens, the upper sideband USB₁, and lower sideband LSB₁, beams were columnated into a fine pencil, which was directed at the surface of the MRD 500 SPD crystal, as shown in Fig 6.32.
Unfortunately, it was extremely difficult to align the lenses to produce the fine pencil beam described above. The X10 microscope objective, used prior to the MRD-500, was particularly difficult to align. Alignment difficulties were further increased by the different locations of the points of deflection of the USB, and LSB, sidebands and the focii of the beams themselves. (When columnated the two beams become parallel but each focusses to a separate point.) For these reasons the experiment failed to produce the desired optical mixer products. It did however illustrate the alignment requirements for optical mixing to occur.
To ensure that the optical beams can be mixed, they must be completely colinear, the points of deflection and focus must be coincident, or apparently so (refer to section 5.8).

Using a Mach-Zehnder Interferometer [29] it is possible to combine the two light beams, to be mixed, such that they emerge colinearly from the interferometer as shown in Fig 6.33.

Fig 6.33 Producing Colinear Light Beams Using the Mach-Zehnder Interferometer

The actual adjustment of the Mach Zehnder interferometer for this purpose is difficult, fortunately its rectangular characteristics may be exploited to make an initial alignment.

Placing two screens about the interferometer recombination point, as shown in Fig 6.34, one 2m away the other 1m away, spots corresponding to the beams, to be aligned, can be seen on both cards.
Spots on the alignment cards, corresponding to unwanted beams, may be removed by placing screens in each arm of the interferometer. For example, the upper and lower sideband spots, of the first order deflection ensemble, are selected by placing a screen in one arm of the interferometer, to remove beams corresponding to the carrier $c_1$ and the lower sideband LSB$_1$. A second screen is placed in the remaining arm to remove carrier $c_1$, and upper sideband USB$_1$. The upper sideband spot, on the alignment cards, then emerges from one arm, and the lower sideband spot from the other.

The zero order spots are removed by a screen as shown in Fig 6.39. The azimuth and elevation of each mirror is then adjusted using the screw guage, so that the spots on both alignment cards merge into a single spot. Each mirror must be adjusted only by small amounts otherwise the task of alignment becomes almost impossible.
The two beams are completely aligned, when the single spot appears to be stationary as each arm of the interferometer is blocked alternately. The alignment cards are then removed and a lens is placed in line with the interferometer output. The photo detector is placed beyond the lens and adjusted to focus the light onto the SPD crystal.

The Hewlett Packard HP 405A RF Spectrum analyser was connected to the photo-receiver output and power supply connected to the bias terminals.

Minor adjustments to the interferometer mirrors produced an impulse at 24 MHz on the spectrum analyser, corresponding to a modulation frequency of 12 MHz. Moreover no impulse at 12 MHz was visible, indicating that no carrier was present. The 24 MHz electrical output signal, from the photo-receiver, was therefore due only to the mixing of the USB and LSB light beams.

By blocking out light in either arm of the Mach-Zehnder interferometer, the 24 MHz impulse vanished, furnishing further proof that the output of the photo-receiver was due only to the mixing of the two light beams.

The frequency of the modulation signal generator was varied between 1 MHz and 30 MHz without adjustment of the interferometer. The results are given in table 6.13.
Table 6.13 Mach Zehnder Frequency Mixing Without Adjustment

<table>
<thead>
<tr>
<th>Modulation Frequency (MHz)</th>
<th>dBr</th>
<th>Modulation Frequency (MHz)</th>
<th>dBr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-16*</td>
<td>11</td>
<td>-3</td>
</tr>
<tr>
<td>2</td>
<td>-16</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>-16</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>-16</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>-16</td>
<td>15</td>
<td>-4</td>
</tr>
<tr>
<td>6</td>
<td>-16</td>
<td>16</td>
<td>-16</td>
</tr>
<tr>
<td>7</td>
<td>-16</td>
<td>17</td>
<td>-16</td>
</tr>
<tr>
<td>8</td>
<td>-16</td>
<td>18</td>
<td>-16</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>19</td>
<td>-16</td>
</tr>
<tr>
<td>10</td>
<td>-6</td>
<td>20</td>
<td>-16</td>
</tr>
</tbody>
</table>

* Noise Floor at -16dBr

The corresponding response plot is given in Fig 6.35.
The frequency response, plotted in Fig 6.35, is due to two effects. As the modulation frequency is altered, the angles at which the mixed beams emerge from the AO cell change and the beams become misaligned at the detector. Moreover as the deflection angles change the beams tend to move out of the receive lens aperture, reducing the actual amount of optical power arriving at the detector.

The Mach-Zehnder interferometer illustrates the most important difficulty encountered in the modulation and detection of light. In order to produce a system, in which light beams are to be mixed, the designer must ensure that the angles at which light beams emerge from the modulator are rendered independent of the frequency of modulation. Using the unmodified Mach-Zehnder interferometer, mixing can be conducted only over a very limited frequency range.
Techniques of expanding the range of frequencies over which the Mach-Zehnder interferometer can be used are discussed in section 6.5.

6.4.2 Frequency Sweep of Temporary Photo-Receiver

Once it had been verified that the MRD-500 can be used as an optical frequency mixer, it was decided to try to produce mixer products using various combinations of the six beams produced by an AM sonic wave. It is possible to frequency sweep the photo-receiver response from 1 MHz to 140 MHz using the various beams produced by the AO modulator.

The AO driver produces an 80 MHz sonic carrier which is amplitude modulated by the input modulating signal from 0 to 15 MHz, and in fact, beyond the manufacturers specifications.

The MRD-500 based photo-receiver was frequency swept using the same apparatus as used in Fig. 6.39.

Initially the Mach-Zehnder interferometer was aligned, as described in section 6.4.1. However after alignment, the mirrors in the interferometer were adjusted to produce a peak impulse at the difference frequency. After each measurement was made, the modulation frequency was altered. Again the mirrors of the Mach-Zehnder interferometer were altered until a peak impulse height was produced.

The beam combinations chosen to frequency sweep the photo-receiver are tabulated in table 6.14.
Table 6.14 Beam Combinations Covering the Frequency Range 0-140 MHz

<table>
<thead>
<tr>
<th>Beam Combination</th>
<th>Frequency Range for Modulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>USB₁ + LSB₁</td>
<td>0 - 60 MHz</td>
</tr>
<tr>
<td>USB₀ + LSB₁</td>
<td>20 MHz - 80 MHz</td>
</tr>
<tr>
<td>USB₀ + C₁</td>
<td>50 MHz - 80 MHz</td>
</tr>
<tr>
<td>LSB₀ + USB₁</td>
<td>80 MHz - 140 MHz</td>
</tr>
</tbody>
</table>

Making measurements at each frequency of the combination produced the results given in table 6.15.

Table 6.15 Frequency Sweep of Temporary Photo-Receiver

<table>
<thead>
<tr>
<th>Beam Combination</th>
<th>Frequency Measured (MHz)</th>
<th>Relative Strength (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSB₁ + USB₁</td>
<td>2</td>
<td>Unable to remove carrier with spatial filter-measurements unreliable.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-34</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-33</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-31</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-31</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>-31</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>-31</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>-30</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>-30</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>-42</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>-22</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>-30</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>-28</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>-25</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>-38</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>-38</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>-36</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>-37</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>-36</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>-36</td>
</tr>
</tbody>
</table>
Table 6.15 Frequency Sweep of Temporary Photo-Receiver (cont.)

<table>
<thead>
<tr>
<th>Beam Combination</th>
<th>Frequency Measured (MHz)</th>
<th>Relative Strength (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USB₀ + LSB₁</td>
<td>20</td>
<td>-50</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>-43</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>-30</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-33</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>-35</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>-20</td>
</tr>
<tr>
<td>USB₀ + C₁</td>
<td>50</td>
<td>-42</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>-45</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>-44</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>-43</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>-41</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>-39</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>-39</td>
</tr>
<tr>
<td></td>
<td>57</td>
<td>-38</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>-38</td>
</tr>
<tr>
<td></td>
<td>59</td>
<td>-38</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>-39</td>
</tr>
<tr>
<td>LSB₀ + USB₁</td>
<td>80</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>-34</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-33</td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>-35</td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>-38</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>-34</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>-34</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>-36</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>-40</td>
</tr>
<tr>
<td></td>
<td>135</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>*</td>
</tr>
</tbody>
</table>

* Alignment was difficult, unwanted signals could not be spatially removed therefore measurements could not be obtained.

The results given in table 6.15 are graphed on log-linear axes in Fig 6.36.
The sharp peak at 80 MHz is a result of the mixing of the two carriers $C_0$ and $C_1$ in addition to USB$_0$ and LSB$_1$, with no modulation present, there is no angular separation between carrier and sidebands therefore the carrier cannot be eliminated. Removing 6dB from the figure obtained at 80 MHz would give a closer estimate of the height [30],[31], due to two beams of the same intensity as USB$_0$ and LSB$_1$; assuming a modulation depth of 100%.

Individual peaks and troughs, in the measurements, are probably due to poor beam alignment. As the experiment progressed, the techniques of alignment became more familiar, however at the very high and very low frequencies of each sweep, alignment was very difficult. Moreover, the response roll-off of the AO modulator was encountered at modulation frequencies above 15 MHz.

Fortunately, despite the peaks and troughs of measurement, a general trend for mixing response is evident as shown in Fig 6.36.

In each measurement group, a true estimate of mixing results is gained by referring each measurement to the noise floor at -50dB.
6.5 INCREASING THE MIXING BANDWIDTH BY REDUCING THE EFFECTS OF DEFLECTION

The frequency sweep of detector response, given in section 6.4.2, illustrated a major problem encountered in frequency mixing experiments using the AO modulator. Each measurement was made after adjusting the Mach-Zehnder interferometer to realign the two beams being mixed at the new frequency. Without adjustment of the interferometer, the mixing of two beams can be performed only over a very small range of frequencies. Therefore it is desirable to reduce the effects of frequency upon deflection angle.

Consider a light beam incident upon an amplitude modulated sonic wave such that the incident wave is split into six beams, emerging at differing angles, as shown in Fig 6.37.
To simplify the analysis, as far as the external environment is concerned, it is assumed that beam deflection occurs within the AO crystal about a point as shown in Fig 6.37.

By placing a lens of appropriate aperture behind the AO crystal, such that its focal point is coincident with the point of deflection, rays emerging from the AO crystal will become parallel to the ray incident with the AO crystal as shown in Fig 6.38.

As the modulation frequency varies, the beams emerging from the AO cell move laterally in space, however their angles, relative to the original axis, are un-altered. Using a second lens, the light beams emerging from the lens may be focussed onto the photo detector, where the position will be independent of the modulation frequency, as shown in Fig 6.39.
By placing the second lens after the recombination point of the interferometer, frequency independent optical mixing can be produced.

Unfortunately, the simple two lens approach is suitable only if the light beam, incident upon the AO cell, is very narrow; the beam emerging from the Spectra Physics HeNe laser tends to diverge and is of approximately 1.5 mm in diameter as it emerges from the laser cavity. The lens placed after the AO modulator tends to focus each beam to a point on the other side. As the beams reach the second lens they are no longer pencils of light but broad diverging cones, as shown in Fig 6.40.

Fig 6.40 Angular Divergence Introduced by the Post Modulator Lens

A pre-modulator, lens positioned between the laser and the AO modulator, such that the incident light is focussed at the modulator point of deflection, will eliminate the post modulator beam divergence. The post-modulator lens will then columnate the emerging, deflected light and also remove the angular frequency variation of light beams, as shown in Fig 6.41.
Considerable difficulty was encountered in separating the beams emerging from the post-modulator lens so that spatial filtering could be used to remove unwanted deflection orders. Moreover, the alignment of lenses was difficult. Precise alignment of the lenses could only be made by removing the modulator and adjusting the distances between the two lenses until the spot projected onto a card remained the same size at any distance from the post-modulator lens.

The AO modulator was then placed between the lenses with the common focii of the lenses at the geometric centre of the AO cell. By covering the post-modulator lens aperture with a card, the AO cell could be adjusted to deflect the light at the Bragg angle.

In order to increase the spatial separation between zero and first order spots emerging from the post-modulator lens, the pre-modulator - post-modulator lens combinations were tried.
A pre-modulator lens of larger focal length would produce beams of smaller solid angle which could be more easily separated with a spatial filter. The spots due to the beams emerging from the post-modulator lens would be smaller.

If a post-modulator lens of greater focal length is used, the spots due to beams emerging from the post-modulator lens would be larger but much further apart.

Each lens combination was adjusted to produce a common focus with the AO crystal in the same way as was described above. The AO cell was adjusted using a card covering the post-modulator lens aperture.

After considerable experimentation, a lens combination was reached, which separated the two beams of the zero and first deflection orders. The pre-modulator lens is 34D-65F and post-modulator lens 52D-220F. Each beam did not measurably diverge over a distance 2.6 m from the post-modulator lens.

This result verified the possibility of using the two lens combination to remove the angular separation of beams emerging from the modulator, producing two spatially separated yet columnated output light beams.

Unfortunately when modulation was applied, the lens combination could not spatially separate and columnate the two sideband beams associated with each deflection order.

Further experimentation is needed to find a lens combination which will spatially separate the six beams emerging from the AO modulator, recolumnate them and eliminate the angle of divergence.
6.6 AMPLIFIER DESIGN

The signal levels present at the output of the photo-receiver, discussed in section 6.4, are of the same magnitude as those received at the output of a radio antenna [12]. Therefore, before any signal processing and detection can be undertaken, the signal must be amplified to a level at which mixing will occur and digital decisions can be made.

In radio, such amplification is conducted to a limited degree at the receive frequency, and further at one or more standardized intermediate frequencies. Some final stages of amplification are used at the baseband to drive a speaker or other transducer [12].

Frequency mixing is used to produce an intermediate frequency, usually after some initial radio frequency amplification [12]. Optically however, mixing occurs usually before any amplification of the signal. The mixing process could be used to produce an intermediate frequency, which could be amplified by standardized circuits, or it could reproduce the baseband signal which could also be amplified.

At high data rates, receive amplifiers must have very wide bandwidths, whether used at intermediate frequencies or at baseband [30],[31],[32]. In the subsequent sections, design techniques for amplifiers used at high frequencies and over wide bandwidths are discussed, with a view towards the design of broadband receive amplifier systems.

In section 6.6.1, the technique of narrow band amplifier design using scattering parameters is discussed. It will form the basis for the discussion of broad-band amplifier design techniques using the Smith Chart,
which are covered in section 6.6.2. The difficulties encountered in both sections have led to the development of a new broad band amplifier design strategy, which unfortunately, has yet to be tested and is therefore introduced in chapter 9.

Using the techniques of design of sections 6.6.1 and 6.6.2, the designer may produce single stage amplifiers which meet the design criteria, however coupling of individually designed stages is difficult; often introducing losses which negate the individual stage gain of the amplifier. Moreover, these techniques offer little control of amplifier phase response [21],[33],[34],[35],[36],[37],[38],[39],[40],[41],[42],[43],[44],[45],[46],[47],[48],[49],[50]. If the amplifier phase response is itself non-linear, its useful bandwidth for digital signals will be less than -3dB bandwidth, as it will tend to smear digital pulses making pulse detection difficult [9],[10],[11].

Finally, the design of a Broadband amplifier, using a Motorola hybrid device, will be discussed. The resulting amplifiers exhibit excellent amplitude response qualities, however their phase response is largely uncontrolled. Using the techniques of the next two chapters, it is possible to compensate for deviations of phase from linearity and produce linear phase amplifiers with flat gain response.

6.6.1 Narrow-Band Amplifier Design Using the Smith Chart

At frequencies where circuit dimensions become significant parts of a wavelength, the measurement of the common two port parameters, such as
impedance, admittance or hybrid parameters becomes difficult. Mismatches, between the circuit under test and the measuring equipment, cause reflections of voltage and current waves. Voltage and current readings become dependent on the physical location of the measurement point in the same circuit [19],[20].

A more useful representation of circuit parameters, when measured at frequencies above 100 MHz, is in terms of the complex reflection or scattering coefficients, which relate the magnitude and phase of a wave scattered from an impedance, to the wave incident upon it [19],[20],[21], [23],[36],[40],[51],[52],[53],[54]. Moreover, the scattering parameter concept may be extended to describe the behaviour of any circuit in terms of a matched and unmatched component of current or voltage [19],[20],[21], [23],[36],[40],[51],[52],[53],[54].

Consider a voltage source, of voltage $V_s$ and internal impedance $Z_S(j\omega)$, which is connected to a load impedance $Z_L(j\omega)$, as shown in Fig 6.42.

![Fig 6.42 Scattering of a One Port Network](image)

The current through $Z_L(S)$ is given by [54],
\[ I_L = \frac{V_S}{Z_S(j\omega) + Z_L(j\omega)} \]  
(6.5)

and the corresponding voltage by [54],

\[ V_L(s) = \frac{V_S Z_L(j\omega)}{Z_S(j\omega) + Z_L(j\omega)} \]  
(6.6)

The voltage and current may be separated into two components; the incident and reflected waves. The incident component which travels from the source to the load is interpreted as the matched component, the reflected component is the mismatched or correction component [54]. The sum of these components must yield the voltage \( V_L \) and current \( I_L \), thus

\[ V_L = V_i + V_r \]  
(6.7)

and

\[ I_L = I_i - I_r \]  
(6.8)

The matched components are given by [54],

\[ I_i = \frac{V_S}{Z_S(j\omega) + Z_S^*(j\omega)} \]  
(6.9)

and

\[ V_i = \frac{V_S Z_S^*(j\omega) = I_i Z_S^*(j\omega)}{Z_S(j\omega) + Z_S^*(j\omega)} \]  
(6.10)

The mismatched components are given by rearranging (6.7) and (6.8) and substituting for \( V_L, I_L, V_i \) and \( I_i \) using (6.5), (6.6), (6.9) and (6.10),

\[ I_r = \frac{Z_L(j\omega) - Z_S^*(j\omega)}{Z_L(j\omega) + Z_S^*(j\omega)} \]  
(6.11)
and

\[ V_r = \frac{Y_S(j\omega) - Y_L(j\omega)}{Y_S(j\omega) + Y_L(j\omega)} V_i \]  \hspace{1cm} (6.12)

Where \( Y_S(j\omega) \) and \( Y_L(j\omega) \) are admittances,

\[ Y_S(j\omega) = \frac{1}{Z_S(j\omega)} \quad \text{and} \quad Y_L(j\omega) = \frac{1}{Z_L(j\omega)} \]

Therefore, the current scattering parameter \( S_1(j\omega) \) is given by [54],

\[ S_1(j\omega) = \frac{Z_L(j\omega)-Z_S^*(j\omega)}{Z_L(j\omega)+Z_S^*(j\omega)} \]  \hspace{1cm} (6.13)

and the voltage scattering parameter [54],

\[ S_V(j\omega) = \frac{Y_S^*(j\omega) - Y_L(j\omega)}{Y_S(j\omega) + Y_L(j\omega)} \]  \hspace{1cm} (6.14)

If an n-port network is to be described by its scattering parameters, the reflected voltages and currents are described by scattering matrices which comprise scattering parameters between ports [54].

\[ [I_r] = [S_1] [I_i] \]  \hspace{1cm} (6.15)

and

\[ [V_r] = [S_V] [V_i] \]  \hspace{1cm} (6.16)

Where \([ ]\) denotes matrix or vector.

In terms of impedance, the current scattering matrix is given by [54],

\[ [S_1] = \left( [Z_S(j\omega)] + [Z_L(j\omega)] \right)^{-1} [Z_L(j\omega)] - [Z_S^*(j\omega)] \]  \hspace{1cm} (6.17)

and the voltage scattering parameter is given in terms of admittances [54],
By normalizing the scattering parameters suitably, a general scattering parameter is obtained [54].

Let

\[ \frac{1}{[a]} = 1 - \frac{1}{\gamma} \]

and

\[ [b] = 1 - \frac{1}{\gamma} \]

such that

\[ \frac{1}{[S(j\omega)]} = \frac{1}{[Z_s(j\omega)]} + \frac{1}{[Z^*(j\omega)]} \]

and

\[ \frac{1}{[S(j\omega)]} = \frac{1}{[Z_s(j\omega)]} + \frac{1}{[Z^*(j\omega)]} \]

such that

\[ [b] = [S(j\omega)] \frac{1}{[a]} \]

Then \([S(j\omega)]\) is related to \([S_1(j\omega)]\) by [54],

\[ [S(j\omega)] = \left( [Z_s(j\omega)] + [Z^*(j\omega)] \right)^{1/2} [S_1(j\omega)] [Z_s(j\omega) + [Z_s(j\omega)]^{-1/2} \]

and to \([S_V(j\omega)]\) by [54],

\[ [S(j\omega)] = \left( [Z_s(j\omega)] + [Z^*(j\omega)] \right)^{1/2} [Z_s(j\omega)]^{-1} [S_V(j\omega)] [Z_s^*(j\omega)] \]

\[ \times \left( [Z_s(j\omega)] + [Z_s^*(j\omega)] \right)^{-1/2} \]

\[ [a] \) and \([b]\) are chosen such that their squared magnitudes correspond to the incident and reflected powers, respectively [54]. The power scattering matrix is therefore [54],
\[ R(j\omega) = [S(j\omega)] [S^*(j\omega)] \]  

(6.24)

Using scattering parameters a transistor may be represented as a two point network represented by a 2x2 element matrix [54],

\[
\begin{bmatrix}
  b_1 \\
  b_2
\end{bmatrix} =
\begin{bmatrix}
  S_{11}(j\omega) & S_{12}(j\omega) \\
  S_{21}(j\omega) & S_{22}(j\omega)
\end{bmatrix}
\begin{bmatrix}
  a_1 \\
  a_2
\end{bmatrix}
\]

(6.25)

where the function of each element is clarified with the signal flow graph of Fig 6.43 [54].

![Fig 6.43 Two Port Scattering Parameter Signal Flow Graph](image)

When a port is terminated in \( Z^*(j\omega) \), the reflection goes to zero, and the load impedance is matched to the source [54].

It is common to refer the scattering parameters to unit source resistance, which is usually a standard system impedance such as 50 \( \Omega \) or 600 \( \Omega \), by dividing \([a]\) and \([b]\) by the source resistance \( Z_S(j\omega) \), in which case \( S(j\omega) \) is multiplied by \( Z_L(j\omega) \).

Now consider a transistor, driven by a generator of voltage \( V_S \) source impedance \( Z_S(j\omega) \), connected to a load impedance \( Z_L(j\omega) \), as shown in Fig 6.44 [19],[20],[21],[35],[36],[40].
The standard impedance for the system is $Z_S(j\omega) = Z_0$ which is given by the transistor manufacturer, it is usually 50\,\Omega. Using the signal flow graph of Fig 6.43, the reflection coefficient of the amplifier, terminated in $Z_L(j\omega)$, is [40],

$$
S_{11}(j\omega) = \frac{S_{11}(j\omega) + S_{21}(j\omega)S_{12}(j\omega) \Gamma_L(j\omega)}{1-S_{22}(j\omega) \Gamma_L(j\omega)}
$$

(6.26)

Where $\Gamma_L(j\omega)$ is the frequency dependent load reflection coefficient for load $Z_L(j\omega)$ [40].

$$
\Gamma_L(j\omega) = \frac{Z_L(j\omega)-Z_0}{Z_L(j\omega)+Z_0}
$$

(6.27)

Where $Z_0$ is the normalizing impedance.

Similarly, the output reflection coefficient for an input terminated in $Z_S(j\omega)$ is given by [40],

$$
S_{22}(j\omega) = \frac{S_{22}(j\omega) + S_{21}(j\omega)S_{12}(j\omega) \Gamma_S(j\omega)}{1-S_{11}(j\omega) \Gamma_S(j\omega)}
$$

(6.28)

Where $\Gamma_S(j\omega)$ is the, frequency dependent, source reflection coefficient for a generator impedance $Z_S(j\omega)$ [40],

---

**Fig 6.44 Linear Transistor Amplifier Circuit**

![Linear Transistor Amplifier Circuit Diagram]
\[ \Gamma_S(j\omega) = \frac{Z_S(j\omega) - Z_0}{Z_S(j\omega) + Z_0} \quad (6.29) \]

Using (6.26), (6.27), (6.28) and (6.29), an expression for transducer gain \( G_T(j\omega) \) is derived for the terminated transistor [40],

\[
G_T(j\omega) = \frac{|S_{21}(j\omega)|^2 \left( 1 - |\Gamma_S(j\omega)|^2 \right) \left( 1 - |\Gamma_L(j\omega)|^2 \right)}{\left( 1 - S_{11}(j\omega)\Gamma_S(j\omega) \right) \left( 1 - S_{22}(j\omega)\Gamma_L(j\omega) \right) - S_{21}(j\omega)S_{12}(j\omega)\Gamma_S(j\omega)\Gamma_L(j\omega)} \quad (6.29)
\]

If \( S_{12}(j\omega) \) is very small, (6.30) simplifies to [40],

\[
G_{TU} = \frac{|S_{21}(j\omega)|^2 \left( 1 - |\Gamma_S(j\omega)|^2 \right) \left( 1 - |\Gamma_L(j\omega)|^2 \right)}{\left( 1 - S_{11}(j\omega)\Gamma_S(j\omega) \right) \left( 1 - S_{22}(j\omega)\Gamma_L(j\omega) \right) \Gamma_S(j\omega)\Gamma_L(j\omega)} \quad (6.31)
\]

where \( G_{TU}(j\omega) \) is the simplified transducer gain.

Equation (6.32) gives the unilateral transducer gain for the transistor. In the common-emitter transistor configuration, (6.31) is sufficiently accurate for most calculation, however in other transistor configurations considerable error is introduced by ignoring \( S_{12}(j\omega) \), and (6.30) must be used for transducer gain calculations.

Equation 6.31 comprises 3 parts, as illustrated by (6.32).

\[
G_{TU} = \left( |S_{21}(j\omega)|^2 \right) \left( \frac{1 - |\Gamma_S(j\omega)|^2}{1 - S_{11}(j\omega)\Gamma_S(j\omega)} \right) \left( \frac{1 - |\Gamma_L(j\omega)|^2}{1 - S_{22}(j\omega)\Gamma_L(j\omega)} \right) \quad (6.32)
\]

From left to right, the parameters are associated with the transistor matched gain, which is fixed, the source matching and the load matching [40].
The transistor parameters $S_{21}(j\omega)$, $S_{11}(j\omega)$ and $S_{22}(j\omega)$ are fixed by the transistor frequency response and cannot be altered. It is however possible to vary $\Gamma_S(j\omega)$ and $\Gamma_L(j\omega)$ thereby varying the transducer gain of the transistor [40]. $\Gamma_S(j\omega)$ and $\Gamma_L(j\omega)$ are varied by varying the values of $Z_S(j\omega)$ and $Z_L(j\omega)$ respectively [40].

If $\Gamma_S(j\omega)$ and $\Gamma_L(j\omega)$ are plotted on separate Smith Charts for specific values of gain $G_{TU}$, the normalized values of $Z_S(j\omega)$ and $Z_L(j\omega)$ required to achieve that gain may be read directly off the chart, as the impedance plane is mapped to the Smith plane by the bilinear transform [19].

\[
\Gamma(j\omega) = \frac{Z(j\omega) - Z_0}{Z(j\omega) + Z_0} = \frac{Z(j\omega) - 1}{Z(j\omega) + 1}
\]

where $\Gamma(j\omega)$ is a general reflection coefficient and $Z(j\omega)$ is a general impedance.

Optimum transducer gain occurs when $\Gamma_S(j\omega) = S_{11}^*(j\omega)$ and $\Gamma_L(j\omega) = S_{22}^*(j\omega)$ in which case [40],

\[
G_{TU}(j\omega) = \left| \frac{1}{1 - |S_{11}(j\omega)|^2} \right|^2 \left| \frac{1}{1 - |S_{22}(j\omega)|^2} \right|^2
\]

$\Gamma_S(j\omega)$ and $\Gamma_L(j\omega)$ are specific points on the Smith Chart. They correspond to a value of maximum gain.

At a fixed frequency $\omega_0$, the transducer gain is [40],
Let $G = S_{11}^*$ and the transducer gain $G_{TU} = G_0$ where,

$$G_0 < G_{TU}$$

in which case

$$G_0 = \left| S_{21} \right|^2 \frac{1 - \left| \Gamma_S \right|^2}{1 - \left| S_{11} \right|^2} \frac{1 - \left| \Gamma_L \right|^2}{1 - \left| S_{22} \right|^2}$$

Rearranging (6.36) gives an equation describing a circle in the Smith Plane of radius $[40]$,

$$R_1 = \sqrt{1 - g_2} \frac{1 - \left| S_{22} \right|^2}{1 - \left| S_{11} \right|^2} (1 - g_2)$$

and centre $[40]$,

$$d_1 = g_2 \frac{\left| S_{22} \right|^2}{1 - \left| S_{22} \right|^2} (1 - g_2)$$

where

$$g_2 = G_0 (1 - \left| S_{22} \right|^2) = \frac{G_0}{G_0}$$

and $G_0$ is the peak gain value at frequency $\omega_0$.

A number of gain values between unity and $G_0$ are selected and circle of constant gain produced $[40]$. 
Similarly, circles of constant gain are produced for $\Gamma_S$. Each circle of constant gain is plotted on a Smith Chart in rectangular coordinates [40]. The actual values of $Z_L(j\omega_0)$ and $Z_S(j\omega_0)$ are read directly off the impedance coordinates of the Smith Chart.

Fig 6.45 shows a plot of a number of constant gain circles for an MRF-901 transistor, in common-base configuration, assuming that $S_{12}(j\omega)$ can be ignored. In fact in the common base configuration of an MRF-901, $S_{12}(j\omega)$ cannot be ignored.

Fig 6.45 Constant Gain Curves for Common Base MRF-901

The full manufacturers specifications for the MRF-901 are given in appendix 8. Using the manufacturers data, the common base scattering
parameters were calculated from the given common-emitter scattering parameters at 500 MHz.

By selection of input and output impedances, which produce less than optimum gain, controlled amounts of input and output mismatch can be used to stabilize the transistor and minimize input noise contributions.

The amplifier may exhibit one of 3 stability conditions [40]:

1. Unstable
2. Conditionally Stable
3. Unconditionally Stable

If an incorrect choice of source or load impedance is made, the transistor amplifier will oscillate in an uncontrolled way. It is therefore necessary to select input and output impedances which will guarantee that the amplifier is stable.

The amplifier is conditionally stable if at a specified frequency, a source or load impedance can be chosen that will produce a stable amplifier.

Unconditional stability of the amplifier occurs if the amplifier is stable regardless of the source and load impedances chosen.

Potential instability of a transistor occurs if the real part of its terminated input scattering parameter $S_{11}(j\omega)$, given by (6.26), or the real part of its terminated output scattering parameter $S_{22}(j\omega)$, given by (6.28), is greater than unity [40].
Using these conditions, boundaries between regions of stability and instability can be found, by solving for $\Gamma_S(j\omega)$ and $\Gamma_L(j\omega)$ when $S_{11}(j\omega)$ and $S_{22}(j\omega) = 1$ in (6.26) and (6.28) [40]. The solutions are circles in the Smith plane located at a distance from the centre of the Smith Chart of [40],

$$r_{S1}(j\omega) = \frac{S_{22}(j\omega) - \left( S_{11}(j\omega)S_{22}(j\omega) - S_{21}(j\omega)S_{12}(j\omega) \right) S_{11}^*(j\omega)}{\left| S_{22}(j\omega) \right|^2 - \left| S_{11}(j\omega)S_{22}(j\omega) - S_{21}(j\omega)S_{12}(j\omega) \right|^2}$$

(6.40)

for source impedance variation and,

$$r_{S2}(j\omega) = \frac{S_{11}(j\omega) - \left( S_{11}(j\omega)S_{22}(j\omega) - S_{21}(j\omega)S_{12}(j\omega) \right) S_{22}^*(j\omega)}{\left| S_{11}(j\omega) \right|^2 - \left| S_{11}(j\omega)S_{22}(j\omega) - S_{21}(j\omega)S_{12}(j\omega) \right|^2}$$

(6.41)

for load impedance variation.
The respective radii of the source and load impedance stability boundaries are [40],

\[
R_{S1}(j\omega) = \left| \frac{S_{12}(j\omega)S_{21}(j\omega)}{S_{22}(j\omega)} \right|^{2} - \frac{S_{11}(j\omega)S_{22}(j\omega) - S_{21}(j\omega)S_{12}(j\omega)}{S_{11}(j\omega)^{2} - S_{11}(j\omega)S_{22}(j\omega)S_{21}(j\omega)S_{12}(j\omega)}^{2}
\]

and

\[
R_{S2}(j\omega) = \left| \frac{S_{12}(j\omega)S_{21}(j\omega)}{S_{11}(j\omega)} \right|^{2} - \frac{S_{11}(j\omega)S_{22}(j\omega) - S_{21}(j\omega)S_{12}(j\omega)}{S_{11}(j\omega)^{2} - S_{11}(j\omega)S_{22}(j\omega)S_{21}(j\omega)S_{12}(j\omega)}^{2}
\]

The centre of the source stability circle lies on a line from the centre of the Smith Chart through \( S_{11}^*(j\omega) \). Similarly the centre of the load stability circle lies on a line from the centre of the Smith Chart through \( S_{22}^*(j\omega) \) [40].

The stability boundaries for an MRF-901 in common base mode are given in Fig 6.46, together with the original circles of constant gain of Fig 6.45.

Fig 6.46 Gain and Stability of the Common Base MRF-901
In the Common Base configuration, the MRF-901 is conditionally stable. Source and load impedance values must be chosen such that the impedances lie outside the circles of stability [40]. Generally, the source and load impedances are chosen to optimize the stage gain, for a particular margin of stability.

Finally, the noise performance of the amplifier is considered by plotting circles of constant noise figure on the Smith Chart. Their centres and radii are derived using the expression for noise variation of a linear two port, with source admittance; written in terms of the source reflection coefficient $\Gamma_S(j\omega)$ [40].

$$F_N(j\omega) = \frac{\hat{F}_N(j\omega) + 4R_N}{1 - |\Gamma_S(j\omega)|^2} \frac{|\Gamma_S(j\omega) - \hat{\Gamma}_S(j\omega)|^2}{1 - |\hat{\Gamma}_S(j\omega)|^2}$$

(6.44)

Where $F_N$ is the noise figure, $\hat{F}_N$ is the optimum noise figure, $R_N$ is the noise resistance factor and $\hat{\Gamma}_S(j\omega)$ is the reflection coefficient at which optimum noise figure $\hat{F}_N$ occurs.

For various values $F_N(j\omega) < \hat{F}_N(j\omega)$, circles of constant noise figure are plotted on the Smith Chart, centred at [40],

$$C_{\hat{F}}(j\omega) = \frac{\hat{\Gamma}'(j\omega)}{1 + N_N(j\omega)}$$

(6.45)

where

$$N_N(j\omega) = \frac{F_N(j\omega) - \hat{F}_N(j\omega)}{4R_N} \left| 1 + \hat{\Gamma}'(j\omega) \right|^2$$

with radius [40],

$$R_{\hat{F}}(j\omega) = \frac{N_N^2(j\omega)}{1 + N_N(j\omega)} \frac{\left| \hat{\Gamma}'(j\omega) \right|^2}{1 + N_N(j\omega)}$$

(6.46)
The circles of constant noise figure are plotted onto the Smith Chart with the circles of constant gain and the stability boundaries, as shown in Fig 6.47 for the common base MRF-901.

Fig 6.47 Gain, Stability and Noise Figure for the MRF-901 in Common Base Configuration.

The input and output impedances are chosen such that the transistor is stable, with say a 20% stability margin with maximum gain and minimum noise figure using the Smith Chart plots of Fig 6.47. By optimizing the noise figure, in the presence of the constant gain circles, there is little chance of producing an amplifier of optimum noise figure with unit gain, which is a danger if noise figure alone is optimized [55].

A point of suitable operation for the common base MRF 901 is shown in Fig 6.47.

Using the techniques described, it is possible to design a broadband amplifier by plotting the circles of constant gain, stability, and constant noise figure at a number of spot frequencies. Input and output impedances are chosen to produce the same gain and noise figure, as the frequency is
varied over the range. That is, frequency varying impedances are chosen such that they produce reflection coefficients that are tangential to the circles of constant gain and noise figure, as the frequency is varied. Generally this is an extremely difficult task.

Using the techniques of controlled input and output mismatch to optimize gain, stability and noise performances of an amplifier, the designer encounters restrictions imposed by the impedance of loads, input transducers and other amplifier stages; these are usually fixed. In order to match them to the desired optimum mismatch impedances the designer must include additional resistive and reactive components. Unfortunately, the addition of such components can reduce the power transferred to the original load. The amplifier gain may be increased but the power arriving at the input, or supplied to the load is reduced, and no overall improvement in the transducer gain is experienced. Therefore, in many cases the mismatch design technique offers an insufficient number of degrees of design freedom.

Moreover, the unilateral approximation of equation (6.30) may introduce considerable error in design, as in the case with the common base MRF-901. If it is desired to use the unilateral design approach outlined, the transistor must be unilateralized in some way [12].

Equation (6.30) can also be used to plot circles of constant gain without the error introduced by the unilateral approximation.

When the unilateral approximation is not used independent circles cannot be plotted. Equation (6.30) cannot be uncoupled into the form of equation (6.34).
By fixing the load scattering coefficient, equation (6.30) may be rearranged to describe circles of constant gain and functions of input impedance.

Recall equation (6.30)

\[
G_T(j\omega) = \frac{\left| S_{21}(j\omega) (1 - |\Gamma_S(j\omega)|^2 \right| (1 - \left| \Gamma_S(j\omega) \right|)^2 - S_{11}(j\omega) S_{12}(j\omega) \Gamma_S(j\omega) \Gamma_{-1}(j\omega)}{\left(1 - S_{11}(j\omega) \Gamma_S(j\omega) (1 - S_{22}(j\omega) \Gamma_S(j\omega)) - S_{21}(j\omega) S_{12}(j\omega) \Gamma_S(j\omega) \Gamma_{-1}(j\omega) \right)^2}
\]

Equation (6.30) may be rearranged giving an equation for a circle of constant gain in the Smith complex frequency plane,

\[
G_T(j\omega) - \frac{\left| \Gamma_{-1}(j\omega) S_{11}(j\omega) S_{22}(j\omega) \Gamma_{-1}(j\omega) S_{12}(j\omega) S_{21}(j\omega) - S_{11}(j\omega) \right|^2}{\left| \Gamma_S(j\omega) + \frac{1 - S_{22}(j\omega) \Gamma_{-1}(j\omega)}{\Gamma_{-1}(j\omega) S_{11}(j\omega) S_{22}(j\omega) - \Gamma_{-1}(j\omega) S_{12}(j\omega) S_{21}(j\omega) - S_{11}(j\omega)} \right|^2}
\]

Which has a centre located at the point,

\[
\Gamma_0(j\omega) = \frac{\Gamma_A(j\omega) \text{ Re} \{\Gamma_B(j\omega)\} + j \Gamma_A(j\omega) \text{ Im} \{\Gamma_B(j\omega)\}}{\Gamma_A(j\omega) + \Gamma_C(j\omega)}
\]

and Radius \( R_0(j\omega) \)

\[
R_0(j\omega) = \sqrt{\left( \frac{\Gamma_A(j\omega) + \Gamma_C(j\omega)}{\Gamma_A(j\omega) + \Gamma_C(j\omega)} \right)^2 - \frac{\Gamma_A(j\omega) \Gamma_C(j\omega) - \Gamma_B(j\omega) |^2}{\Gamma_A(j\omega) + \Gamma_C(j\omega)}}
\]

Where

\[
\Gamma_A(j\omega) = G_T(j\omega) \left| \Gamma_{-1}(j\omega) S_{11}(j\omega) S_{22}(j\omega) - \Gamma_{-1}(j\omega) S_{12}(j\omega) S_{21}(j\omega) - S_{11}(j\omega) \right|^2
\]
\[ \mathbf{\Gamma}_B(j\omega) = \frac{1 - S_{22}(j\omega) \mathbf{\Gamma}_L(j\omega)}{\mathbf{\Gamma}_L(j\omega)S_{11}(j\omega)S_{22}(j\omega) - \mathbf{\Gamma}_L(j\omega)S_{21}(j\omega)S_{21}(j\omega) - S_{11}(j\omega)} \]

\[ \mathbf{\Gamma}_C(j\omega) = |S_{21}(j\omega)|^2 - |S_{21}(j\omega)|^2 |\mathbf{\Gamma}_L(j\omega)|^2 \]

\[ \mathbf{\Gamma}_D(j\omega) = |S_{21}(j\omega)|^2 - |S_{21}(j\omega)|^2 |\mathbf{\Gamma}_L(j\omega)|^2 \]

The stability and circles of constant noise figure discussed, described by equations (6.40), (6.41), (6.42), (6.43) and (6.45), (6.46) respectively, remain the same.

By fixing the load scattering coefficient \( \mathbf{\Gamma}_L(j\omega) \) to value \( \mathbf{\Gamma}_L(j\omega) = 0.9729 + j0.0857 \) and \( \mathbf{\Gamma}_L(j\omega) = 0.9329 + j0.1257 \) circles of constant gain stability and noise figure for a common base MRF-901 transistor amplifier can be plotted as shown in Fig 6.48.

Fig 6.48 Non Unilateral Gain, Stability and Noise Figure of a Common Base MRF-901 Transistor Amplifier
To select a suitable source and load impedance, a large number of load reflection coefficients must be selected and associated constant gain circles plotted. To calculate these by hand calculator is slow and laborious. To speed up the design process, a computer can be used to find circles of constant gain for any values of load reflection coefficient $\Gamma_L(j\omega)$.

Using external feedback, it is possible to alter the input and output scattering parameters, of the transistor and feedback, to suit the application more closely [21],[38],[39],[56]. Using the techniques discussed, the amplifier may then be designed with more suitable loads. Moreover, if the feedback is of the correct phase, the transistor may be unilateralized, which is a process known as neutralization [12]. It is used extensively in power stages of radio transmitters [12].

Once the transistor is neutralized the, unilateral design techniques could be used to produce much more accurate results.

Using the two port scattering parameter signal flow representation for a transistor and a feedback network, an overall signal flow graph is obtained as shown in Fig 6.49 [40].

![Fig 6.49 Feedback Amplifier Primitive Signal Flow Graph](image-url)
Using the techniques of graphical reduction outlined by KUHN [58], the graph may be simplified to a two port representation, as shown in Fig 6.50. The actual steps of the reduction are given in appendix 9.

The new two-port scattering parameters are written in terms of the scattering parameters of the transistor and feedback.

\[
S_{11}(j\omega) = \frac{(S_{21}(j\omega)S_{22}(j\omega) + 2A(j\omega)S_{21}(j\omega))(S_{12}(j\omega) + 2B(j\omega)S_{11}(j\omega) + 2B(j\omega)S_{11}(j\omega)S_{11}(j\omega))}{(1 - 8S_{21}(j\omega)B(j\omega) - 2S_{22}(j\omega)S_{22}(j\omega))}
+ S_{11}(j\omega) + A(j\omega)S_{11}(j\omega) + A(j\omega)S_{11}(j\omega)S_{11}(j\omega)
\]

\[
S_{21}(j\omega) = \frac{(S_{21}(j\omega)S_{22}(j\omega) + 2A(j\omega)S_{21}(j\omega))(S_{22}(j\omega) + S_{22}(j\omega)S_{22}(j\omega) + 2S_{21}(j\omega)B(j\omega))}{(1 - 8S_{21}(j\omega)B(j\omega) - 2S_{22}(j\omega)S_{22}(j\omega))}
+ S_{21}(j\omega) + S_{21}(j\omega)S_{22}(j\omega) + 2S_{21}(j\omega)B(j\omega)
\]
\[ S_{12}(j\omega) = \frac{1 + S_{11}(j\omega) + S_{21}(j\omega)C(j\omega) + B(j\omega)C(j\omega)S_{12}(j\omega) + 2B(j\omega)S_{11}(j\omega) + 2B(j\omega)S_{11}(j\omega)}{1 - 8S_{21}(j\omega)B(j\omega) - 2S_{22}(j\omega)S_{22}(j\omega)} \]

\[ + \frac{S_{12}(j\omega) + S_{12}(j\omega)S_{11}(j\omega) + C(j\omega)S_{11}(j\omega) + C(j\omega)S_{11}(j\omega)S_{11}(j\omega)}{1 + 2S_{11}(j\omega)S_{11}(j\omega)} \]

\[ S_{22}(j\omega) = \frac{(1 + S_{11}(j\omega)S_{21}(j\omega)C(j\omega) + B(j\omega)C(j\omega))(S_{22}(j\omega) + S_{22}(j\omega)S_{22}(j\omega) + 2S_{21}(j\omega)B(j\omega))}{1 - 8S_{21}(j\omega)B(j\omega) - 2S_{22}(j\omega)S_{22}(j\omega)} \]

\[ S_{22}(j\omega) + A(j\omega)S_{21}(j\omega) \]

Where

\[ A(j\omega) = \frac{1 + S_{11}(j\omega)}{1 + 2S_{11}(j\omega)S_{11}(j\omega)} \]

\[ B(j\omega) = \frac{S_{21}(j\omega)}{1 + 2S_{11}(j\omega)S_{11}(j\omega)} \]

\[ C(j\omega) = \frac{S_{11}(j\omega)S_{12}(j\omega)}{1 + 2S_{11}(j\omega)S_{11}(j\omega)} \]

On examination of the steps of flow graph reduction in appendix 9 it is clear that the reduction of even simple flow graphs is extremely complicated and time consuming. The reduction task may be simplified if a number of pre-reduction assumptions are made about the amplifier performance.

If the feedback network is not being used to unilateralize the amplification stage, then the feedback will be dominated by the feedback network, thus \( S_{12}(j\omega) \) may be ignored. Moreover, the amount of power fed forward by the feedback network will be insignificant compared with the amount fed back therefore \( S_{12}(j\omega) \) may also be ignored, producing a simplified flow graph as shown in Fig 6.51.
Which is more easily reduced to the two port form of Fig 6.50 [58], where

\[
S_{11}(j\omega) = A(j\omega) + S_{21}(j\omega)S_{11}(j\omega) - 2S_{21}(j\omega)S_{11}(j\omega)S_{11}^\prime(j\omega)
\]

\[
2S_{12}(j\omega)(1 + S_{11}(j\omega))(1 - 3S_{22}(j\omega)S_{22}(j\omega)(1 - 4A(j\omega))
\]

\[
S_{12}(j\omega) = \frac{S_{12}(j\omega)(1 + S_{11}(j\omega)) + 2A(j\omega)S_{12}(j\omega)(1 + S_{11}(j\omega))}{1 - 3S_{11}(j\omega)S_{11}^\prime(j\omega)}
\]

\[
S_{22}(j\omega) = \frac{S_{21}(j\omega)S_{21}(j\omega)S_{11}(j\omega) - 2S_{21}(j\omega)S_{11}(j\omega)S_{11}(j\omega)(1 + S_{22}(j\omega))}{(1 - 3S_{11}(j\omega)S_{11}(j\omega))(1 - 3S_{22}(j\omega)S_{22}(j\omega)(1 - 4A(j\omega))}
\]

\[
S_{22}(j\omega) = S_{22}(j\omega) + A(j\omega)(1 + S_{22}(j\omega))
\]

(Refer appendix 10)
6.6.2 Broadband Amplifier Design Using the Smith Chart

Using the controlled mismatched technique of section 6.6.1, a method of designing a broadband amplifier was discussed briefly. This method is very complicated, and, at best, would require a large number of iterations, and the synthesis of a circuit complying with the required frequency response. Moreover the controlled mismatch technique is poorly suited to the design of feedback amplifiers, or indeed, amplifiers in which the reverse scattering parameter $S_{12}(j\omega)$ cannot be ignored.

In application note AN-406, from Motorola [39] a much more easily applied design technique, for broadband UHF amplifiers is discussed. The technique uses "Y" parameters of the transistor and a feedback network to select the magnitude of a feedback impedance at the two frequency extremes at which the amplifier is to be used [39]. Intermediate values of frequency are also used to determine the variation of feedback impedance required, to produce a flat gain, over the desired bandwidth [39].

In addition to the feedback network, some interstage coupling is also required to satisfy the gain requirements [39]. Selection of interstage coupling networks in Kwok's technique [39] is largely heuristic and can lead to amplifiers which, when coupled together, produce no more gain than a single stage.

Using Kwok's technique [39], a Broadband amplifier based on an MRF 901 was built and tested. However major problems were encountered in the selection of interstage coupling networks, rendering the amplifier stage unsuitable for use in multi-stage amplifier chains.
Kwok's [39] design procedure can be divided into 4 steps:

1. Determine the required bandwidth, gain, input and output specifications.

2. Determine the feedback network using:
   (a) A low frequency transistor model, to calculate the value of feedback conductance to meet the gain and input impedance requirements.
   (b) A high frequency model to produce the same gain.
   (c) Constant curves of feedback conductance and admittance to select a suitable feedback inductor.

3. Realize a feedback network.

4. Determine the interstage coupling networks.

The 4 preceding steps were followed to produce a broadband amplifier using an MRF-901 transistor.

The design specifications are given per gain stage,

1. Gain: Maximum Per Stage
2. Bandwidth: 500 MHz
3. Input Impedance termination 50 \( \Omega \)
4. Output Impedance termination 50 \( \Omega \)
5. Noise Figure: Minimum per Stage
Using the manufacturers data given in appendix 8, the "Y" - parameters at 200 MHz and 500 MHz were calculated from the scattering parameters using the conversion tables given by Ghausi [54]. These are presented in table 6.16.

Table 6.16  Y - Parameters of the MRF-901

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Y Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 MHz</td>
<td>Y_{11}</td>
<td>15.67 x 10^{-3} + j19.41 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>Y_{12}</td>
<td>-151.5 x 10^{-6} - j1.299 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>Y_{21}</td>
<td>127.1 x 10^{-3} - j109.1 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>Y_{22}</td>
<td>82.01 x 10^{-6} + j3.510 x 10^{-3}</td>
</tr>
<tr>
<td>200 MHz</td>
<td>Y_{11}</td>
<td>4.224 x 10^{-3} + j9.791 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>Y_{12}</td>
<td>-5.5269 x 10^{-6} - j441.8 x 10^{-6}</td>
</tr>
<tr>
<td></td>
<td>Y_{21}</td>
<td>165.9 x 10^{-3} - j46.68 x 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>Y_{22}</td>
<td>372.8 x 10^{-6} + j1.740 x 10^{-3}</td>
</tr>
</tbody>
</table>

Frequency dependence of the Y parameters is assumed in the following analysis.

Unfortunately, the manufacturer gives little information about the characteristics of the MRF-901 at low frequencies, therefore an MRF-901 was connected to a Hewlett Packard transistor curve tracer and the characteristics measured. The low frequency base characteristic is given in Fig 6.52,
and the low frequency collector characteristics are given in Fig 6.53.

Using the base characteristic, Fig 6.57 [56],[57],

\[ Y_{11} = \left. \frac{\Delta I_1}{\Delta V_1} \right|_{V_2 = 0} \] \hspace{1cm} (6.52)

\( Y_{22} \) is found from the collector characteristic and \( Y_{11} \) [56],[57]

\[ Y_{21} = \left. \frac{\Delta I_2}{\Delta V_1} \right|_{V_2 = 0} \] \hspace{1cm} (6.53)

\( Y_{22} \) and \( Y_{12} \) are insignificant [54].
In order to produce an amplifier with lower input and output impedances than the transistor alone, shunt-shunt feedback is used as shown in Fig 6.54 [39].

The common emitter shunt-shunt feedback configuration is given by the matrix sum of the $Y$ parameters for the transistor and the feedback [59],

$$
\begin{bmatrix}
  Y_{11}' & Y_{12}' \\
  Y_{21}' & Y_{22}'
\end{bmatrix} = \begin{bmatrix}
  Y_{11} & Y_{12} \\
  Y_{21} & Y_{22}
\end{bmatrix} + \begin{bmatrix}
  Y_F & -Y_F \\
  -Y_F & Y_F
\end{bmatrix}
$$

(6.54)

Where $Y_{11}'$ denotes total admittance parameter and $Y_F$ denotes feedback admittance.

The specifications for the amplifier fix the stage load-conductance at $20 \text{ m } \Omega^{-1}$, in which case the amplifier gain at low frequencies is [39],
6.98

\[ G = \left( \frac{G_m - G_F}{G_F + G_L} \right)^2 \frac{G_L}{G_1 + G_F \left( 1 + \frac{G_m - G_F}{G_F + G_L} \right)} \]  

(6.55)

Where \( G_m = \text{re}(Y_{21}), \ G_1 = \text{Re}(Y_{11} + Y_{12}) \ \text{Re}(Y_{11}), \)

\( G_F \) is the feedback conductance and \( G_L \) the load conductance.

The associated stage input-admittance is [39],

\[ Y_{in} = G_1 + G_F \left( 1 + \frac{G_m - G_F}{G_F + G_L} \right) \]

(6.56)

Where \( Y_{in} \) is the input admittance.

With \( G_L \) fixed at 20 m\( \Omega \) by the amplifier specifications, the input admittance and stage gain may be graphed as functions of feedback conductance \( G_L \), as shown in Fig 6.55.

**Fig 6.55** Input Admittance and Stage Gain of an MRF-901 Transistor with Shunt Shunt Feedback

![Graph showing input admittance and stage gain as functions of feedback conductance](image-6.55.png)
Using the manufacturers data, the maximum transistor gain that can be produced over the bandwidth is found. For the MRF-901, the maximum gain attainable over 500 MHz is 17.5dB. An error margin of 20% was introduced and the gain derated by 20%, giving a gain of 16.23dB. The corresponding input conductance is \(102 \times 10^{-3} \ \Omega^{-1}\) and feedback conductance \(6.2 \times 10^{-3} \ \Omega^{-1}\). Normalizing this to \(Y_0 = 10 \times 10^{-3} \ \Omega^{-1}\), the conductance is plotted, on a Smith Chart as shown in Fig 6.56, as a circle of constant gain.

Similarly, a circle of constant gain, as a function of the magnitude of the feedback admittance, is plotted on the Smith Chart for the High Frequency gain given by [39],

\[
G_{oo} = \frac{(\text{Re}(Y_{22}) - G_F)^2 (\text{Im}(Y_{21}) - B_F)^2}{4(\text{Re}(Y_{11})+G_F (\text{Re}(Y_{22}) + G_F) - 2(\text{Re}(Y_{12}) - G_F (\text{Re}(Y_{21}) - G_F) - (\text{Im}(Y_{12}) - B_F) \text{Im}(Y_{21}) - B_F))}
\]

(6.57)

Where \(B_F\) is the susceptance of the feedback admittance. Assuming that \(Y_L = Y_{22}\).

The constant gain circle at high frequencies is centred at [39],

\[
G_{Fo} = \frac{-B}{2A} \quad B_{Fo} = \frac{-D}{2A}
\]

(6.58)

Where \(A = G_{oo} - 1\)

\[
B = 4G_{oo} \text{Re}(Y_{11}) + 4G_{oo} \text{Re}(Y_{22}) + 2G_{oo} \text{Re}(Y_{12}) + (2G_{oo} + 2) \text{Re}(Y_{21})
\]

\[
C = 4G_{oo} \text{Re}(Y_{11}) \text{Re}(Y_{22}) - 2G_{oo} \text{Re}(Y_{12}) \text{Re}(Y_{21}) + 2G_{oo} \text{Im}(Y_{12}) \text{Im}(Y_{21}) - \text{Re}^2(Y_{21}) - \text{Im}^2(Y_{21})
\]

\[
D = 2(1 - G_{oo}) \text{Im}(Y_{21}) - 2G_{oo} \text{Im}(Y_{12})
\]
\begin{equation}
\frac{1}{r} = \frac{1}{2A} \sqrt{B^2 + D^2 - 4AC}
\end{equation}

Appropriately Normalized to \( Y_0 \).

Using (6.58) and (6.59) a 16.23 dB circle of constant gain at 500 MHz, and one at 200 MHz, are plotted in Fig 6.56 also.

**Fig 6.56** Circles of Constant Gain for a Given Feedback Admittance

For the MRF-901

The intersection of the low frequency and 500 MHz curves give an approximate value required for the feedback conductance \([39]\). For the MRF-901 the value is,

\[ Y_F = 0.32 - j0.32 \]

The approximate feedback circuit required for the MRF-901 broadband amplifier is given in Fig 6.57.
Fig 6.57 Approximate Feedback Circuit for Broadband MRF 901 Amplification

\[ L_f = 49.75 \text{ nH} \quad \text{and} \quad R_f = 156.3 \Omega. \]

At 200 MHz, the high frequency circle of constant gain intersects the low frequency circle of constant gain at \( Y_f = 3.226 \times 10^{-3} \). At 200 MHz the feedback resistance is 156.3 \( \Omega \) and the inductive reactance 155 \( \Omega \). At 200 MHz the inductance must be 49.75 nH and 124.14 nH at 500 MHz. Therefore the inductance must vary with frequency.

Using the vector voltmeter, a number of individual inductors, wound on a variety of powdered iron formers, were laboriously tested. Using a second air-cored inductor and a parallel resistor, a feedback circuit was derived which almost produced the required inductance and resistance at 200 MHz and 500 MHz. The inductive and resistive part of the complete feedback circuit is shown in Fig 6.58.
In order not to alter the bias conditions of the amplifier stage, the feedback network was connected into circuit using 0.1 \( \mu \)F disc ceramic capacitors.

Finally, the interstage coupling network must be considered. It is assumed that the bias resistors are sufficiently large, that they can be ignored, as yet bias has not been discussed, however the truth of the assumption will become apparent in subsequent discussion.

Again, to avoid disturbing the bias conditions each stage will be coupled, to the source, load and following stage, using a capacitor. At this stage it is assumed that the capacitive reactance is negligible in the frequency range of interest [39].

The low frequency input conductance, of an amplifier stage, is

\[ Y_{in} = 102 \times 10^{-3} \ \Omega^{-1} \] or the input impedance is \( 9.803 \ \Omega \). At high frequencies \( Y_{in} \) is given by [39],

\[
Y_{in} = Y_{11} + Y_F - (Y_{12} - Y_F) \frac{(Y_{21} - Y_F)}{Y_{22} + Y_F + Y_L} \tag{6.60}
\]

At 500 MHz, the amplifier input admittance is
\[ \text{Yin} = 139.9 \times 10^{-3} - j43.09 \times 10^{-3} \text{ which corresponds to an} \]
\[ \text{input impedance of } 6.529\Omega - j2.011\Omega. \]

Yet at low frequencies, the amplifier is loaded with resistance of 50\(\Omega\), and at high frequencies with an impedance of \(\frac{1}{Y_{22}^*} = 6.653 + j284.8\Omega^{-1}. \) At this point in the design, it is apparent that Kwok's technique is inconsistent [39]. The low frequency and high frequency load impedance values should be the same; because they are not, the interstage coupling networks will be complicated.

In his design, Kwok [39] is concerned chiefly with the high frequency response of the interstage coupling. Following his lead, the input impedance of a second stage need only be increased by a reactance of \(j286.8\Omega\) requiring an inductor of 91.3 nH. The real parts of \(\frac{1}{Y_{22}^*}\) and \(\frac{1}{Y_{\text{in}}}\) are almost equal.

To match the input impedance of the first stage, to a 50\(\Omega\) source, would require a series impedance of \(43.47\Omega + j2.01\Omega\); requiring a series resistance of 43.47\(\Omega\) and an inductor, of 640.1 pH, which is negligible.

Finally, the output matching network must present \(6.653\Omega + j234.8\Omega\) to the output of the last stage. This is largely reactive and could be approximately presented with a 90.65 nF inductor in parallel with the load.

Unfortunately, at low frequencies, some of these matching techniques will be quite unsatisfactory. Much more complicated matching is necessary to satisfy the stage, source and load requirements, at all frequencies of interest.
Finally, having considered the small-signal design of the amplifier, the biasing circuits of the amplifier must be designed. These must not significantly alter any impedance seen by the transistors, input and load equipment.

For this reason, the transistors are biased using an extra base resistor, which can be of very large value, as shown in Fig 6.59.

![Fig 6.59 MRF-901 Bias Circuits](image)

With no emitter resistance and $R_{b3} = 0$ the value of $R_{b2}$ will be small and will significantly alter the input impedance of the amplifier. However, the base current is very small, in which case a large value of $R_{b3}$ can be chosen, which will produce a very small voltage drop, yet present a large bias resistance to the input.

Using the bias configuration shown, and assuming a DC forward current gain $h_{FE}$ of $h_{FE}=80$, a collector-emitter voltage of 7.5V is selected for a supply of 15V. Whence,
The design is now complete, and is shown in Fig 6.60.

Feedthrough capacitors are used with the power supply and bias circuits, to produce a low impedance ground at the supply and bias inputs. Should these be omitted, the power supply points, which are assumed to be AC grounds, may exhibit significant impedance at high frequencies. Moreover, the feedthrough capacitors isolate one stage from another, reducing interaction between stages through the power supply line.
The amplifier of Fig 6.60 was built on double sided, epoxy dielectric, circuit board. The layout of a two stage amplifier is shown in Fig 6.61, the transistor side in Fig 6.61 (a) and the power supply side in Fig 6.61 (b).

Using the Hewlett Packard Hp 8405A Spectrum analyser in the configuration shown in Fig 6.62, the frequency response of one amplifier stage was measured.
The second stage was disconnected from the first. A general radio GR-1208B 65 MHz - 500 MHz unit RF oscillator was used as the signal source. Its output tends to vary with frequency, therefore correction was made for the variations in the frequency response measurements, which are presented in table 6.17.

Table 6.17 Frequency Response of MRF-901 Amplifier Stage

<table>
<thead>
<tr>
<th>Frequency (MHz)</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>12 dB</td>
</tr>
<tr>
<td>100</td>
<td>12 dB</td>
</tr>
<tr>
<td>150</td>
<td>12 dB</td>
</tr>
<tr>
<td>200</td>
<td>12 dB</td>
</tr>
<tr>
<td>250</td>
<td>12 dB</td>
</tr>
<tr>
<td>300</td>
<td>17 dB</td>
</tr>
<tr>
<td>350</td>
<td>31 dB</td>
</tr>
<tr>
<td>400</td>
<td>13 dB</td>
</tr>
<tr>
<td>450</td>
<td>8 dB</td>
</tr>
<tr>
<td>500</td>
<td>14 dB</td>
</tr>
</tbody>
</table>
The MRF-901 amplifier single-stage response is plotted in Fig 6.63.

Unfortunately, the amplifier gain at 350 MHz is very large, possibly due to a region of negative input resistance. Resistors, of various values, were connected, in parallel with the non linear inductor L3, (Fig 6.60) until the gain of 350 MHz fell to 10 dB, which was accomplished using a 470Ω resistor. Using the apparatus of Fig 6.62, the amplifier frequency response was again measured. The results are given in table 6.18 and plotted in Fig 6.64.
<table>
<thead>
<tr>
<th>Frequency (MHz)</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>13 dB</td>
</tr>
<tr>
<td>100</td>
<td>11 dB</td>
</tr>
<tr>
<td>150</td>
<td>11 dB</td>
</tr>
<tr>
<td>200</td>
<td>12 dB</td>
</tr>
<tr>
<td>250</td>
<td>9 dB</td>
</tr>
<tr>
<td>300</td>
<td>10 dB</td>
</tr>
<tr>
<td>350</td>
<td>10 dB</td>
</tr>
<tr>
<td>400</td>
<td>15 dB</td>
</tr>
<tr>
<td>450</td>
<td>10 dB</td>
</tr>
<tr>
<td>500</td>
<td>7 dB</td>
</tr>
</tbody>
</table>
By making minor adjustments to the spacing of turns on the air cored coils, an optimum frequency response was obtained. The optimum response is tabulated in table 6.19 and plotted in Fig 6.65.

Table 6.19 Optimum Frequency Response of MRF-901 Amplifier Stage

<table>
<thead>
<tr>
<th>Frequency (MHz)</th>
<th>Gain (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>12</td>
</tr>
<tr>
<td>100</td>
<td>12</td>
</tr>
<tr>
<td>150</td>
<td>11</td>
</tr>
<tr>
<td>200</td>
<td>8</td>
</tr>
<tr>
<td>250</td>
<td>6</td>
</tr>
<tr>
<td>300</td>
<td>8</td>
</tr>
<tr>
<td>350</td>
<td>16</td>
</tr>
<tr>
<td>400</td>
<td>18</td>
</tr>
<tr>
<td>450</td>
<td>10</td>
</tr>
<tr>
<td>500</td>
<td>9</td>
</tr>
</tbody>
</table>

Fig 6.65 Optimum Frequency Response of MRF-901 Amplifier Stage
Modifying the second stage in the same way as the first, it was connected to the output of the first stage and the combined frequency response measured, in the same way as those for the single stage. Unfortunately the gain of the two stages was the same as the gain for each individual stage; the second stage was loading the first. Although a number of minor modifications were tried, the amplifier gain could not be improved. Moreover, the gain of the individual stages was less than the 16.23 dB expected, yet the measurements of the feedback network indicate that it is close to that required for the 16.23 dB gain specified. Therefore, the fault in the amplifier must lie in the interstage, input and output coupling networks.

Methods of broadband stage matching were examined, and a new theory of broadband matching produced. Initially with a view towards improving the performance of the MRF-901 amplifier, designed using Kwok's techniques [39], and later with a view to a new technique of broadband amplifier design, in which the designer has control of magnitude and phase response of the amplifier. The details of Broadband matching are discussed in the following chapter. Based upon those techniques, a new technique of broadband amplifier design is introduced in chapter 9.

6.6.3 Broadband Hybrid Amplifier Design

The difficulties encountered, in the design of broadband amplifiers in the last section, prompted the search for broadband amplifier equipment which could be built rapidly for experimentation, while the problems, with the transistor amplifier, discussed in the previous section, were solved.
The Motorola electronics corporation produces a series of broadband hybrid amplifiers, which are readily used without much additional circuitry. Out of the series, the devices of greatest interest are the MWA-210, MWA-220 and MWA-230 Wideband Hybrid amplifiers. The full specifications of these devices are given in appendix 11, however the salient specifications for the MWA-210, which was actually purchased, are given in table 6.20.

Table 6.20 Important Electrical Specifications for MWA-210

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Typical Value or Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency Response</td>
<td>0.1 MHz - 600 MHz</td>
</tr>
<tr>
<td>Power Gain</td>
<td>10 dB</td>
</tr>
<tr>
<td>Response Flatness</td>
<td>+1 dB</td>
</tr>
<tr>
<td>Output at 1 dB gain Compression</td>
<td>+1.5 dBm</td>
</tr>
</tbody>
</table>

The MWA-210 was chosen because it covers the desired 500 MHz frequency range, and the input voltages are sufficiently low, that even with a large number in cascade the gain compression is not reached. After proving that a broadband amplifier could be built using these devices, MWA-220 and MWA-230 amplifiers would be purchased to operate, at the higher voltage levels obtained, after amplification by the MWA-210 amplifiers.

With adequate power supply decoupling, the MWA series devices are readily cascaded without significant instability, a problem often encountered with other hybrid amplifiers.
The general configuration of the MWA-210 internal circuit is shown in Fig 6.66.

Using the MWA-210, the amplifier is connected into a circuit as shown in Fig 6.67.
For a 15V power supply, the decoupling is found using (6.61).

\[
R_D = \frac{V_{cc} - V_D}{I_D}
\]  \hspace{1cm} (6.61)

Where \(R_D\) is the decoupling resistance, \(V_{cc}\) is the supply voltage, \(V_D\) is the device voltage which is 1.75 V for an MWA-210, and \(I_D\) is the MWA-210 operating current. \(I_D = 10\, \text{mA}\).

Thus \(R_D = 1325 \Omega\), in which case a 1.2 k\(\Omega\) resistor is used. The input and output capacitors, shown in Fig 6.72, are selected to give the lowest possible frequency response without introducing significant inductance. These values were set therefore at \(C_{BLOCK} = 0.1 \mu\text{F}\), the maximum disc ceramic value available, in which case the minimum frequency is 31.83 KHz calculated from,

\[
f_{LFC} (\text{Hz}) = \frac{1}{100 \times C_{BLOCK} (\text{F})}
\]  \hspace{1cm} (6.62)

The power supply bypass capacitor should be as large as possible. To provide some degree of flexibility in connection of the amplifiers, it was decided to cascade two amplifiers in one Eddystone B7134 diecast aluminium box and produce a number of separate gain modules.

Based upon the manufacturers cascade circuit-diagram and the component values chosen for each stage, a design for each two stage amplifier gain module was produced. It is shown in Fig 6.68.
To achieve the same coupling capacitance as would be achieved when one gain module is all connected to the next, $C_2$ is a disc ceramic capacitor of value $0.05 \mu F$. $C_4$ and $C_6$ are $0.1 \mu F$ disc ceramic capacitors. These capacitors would tend to swamp the capacitors of the feed through capacitors $C_5$ and $C_7$. Therefore their actual value is unimportant. They are used as a convenient means of connecting the power supply through the circuit board.

Inductors $L_1$ and $L_2$ are made using a ferrite bead. Since their main purpose is decoupling of the power supplier between stages, their values are also of little importance. The main purpose of $L_1$, $L_2$, $L_3$ and $L_4$ is to introduce significant amounts of RF resistive loss at the higher frequencies. The ferrite does this very well at frequencies above 100 MHz. $L_1$ and $L_3$ are made of 1 strand of wire through the centre of the bead. $L_2$ and $L_4$ are made from the same ferrite bead with the wire looped through twice.
Connections between each stage of the gain module and to the outside via BNC connectors is effected using a 50Ω stripline. The epoxy based double sided circuit board, used earlier, would be used in construction, therefore the stripline width calculated in section 6.2.3 can be used.

Considerable difficulty had been encountered, in earlier designs of equipment, by building it into the bottom of the diecast Eddystone Box. Moreover, connections to the BNC connectors was difficult, requiring additional earth straps from the circuit ground plane to the barrel of the connector. For these reasons, it was decided to build the circuit into the lid of the Eddystone Box. The BNC Bulkhead connectors would be mounted on the groundplane of the stripline, their centre pins would project through the groundplane into the stripline, where they could be soldered directly into the circuit. The Barrels of the BNC connectors would project through the top of the box, together with a socket for connection to the power supply. The circuit board would be mounted on insulated standoffs to prevent the BNC socket flanges and wiring beneath the groundplane, touching the box lid.

The mechanical design of the gain module is shown in Fig 6.69.

---

**Fig 6.69** Mechanical Design of Two Stage MWA-210 Gain Module

COMPONENTS ON UNDERSIDE SHOWN DOTTED

**PLASTIC BOLTS**

**CIRCUIT BOARD**

**EDDYSTONE BOX LID**
Using a mill, a stripline transmission path was cut into one side of the double sided epoxy circuit board. Breaks were made in the stripline to accommodate the capacitors and hybrid amplifiers. Two holes were drilled into the stripline to accommodate the amplifier input and output pins. Pin 3 of the MWA-210 was completely removed, as it was intended to solder the device case directly to the stripline groundplane, with its leads protruding through the stripline.

The stripline was also broadened near the holes for the BNC centre pins, so that these could be soldered directly to the stripline.

Some copper was left on the stripline side of the circuit board to accommodate the feedthrough capacitors.

The circuit board design is shown in Fig 6.70.

![Circuit Board Design for MWA-210 Gain Module](image)

The feedthrough capacitors were mounted by soldering them to both sides of the circuit board. The lugs, protruding from the groundplane side of the board, were bent over. The power supply socket was mounted on the groundplane side of the circuit board, and the power supply connections made via the inductors L1, L2, L3 and L4 (refer to Fig 6.69).
The Bulkhead BNC connectors were mounted on brass standoffs and the pins soldered to the stripline. Resistors $R_1$ and $R_2$ (refer to Fig 6.68) were connected between the feedthrough capacitors and the stripline. Capacitors $C_1$, $C_2$ and $C_3$ were surface mounted across the appropriate gaps in the stripline.

Finally the MWA-210 packages were soldered into position and their pins soldered to the stripline.

The completed circuit board is shown in Fig 6.71.

Fig 6.71 Components of MWA-210 Gain Module in Position

(a) Stripline Side
Using 6BA plastic bolts and nuts, with 8mm plastic standoffs, the circuit board was mounted in the lid of the Eddystone box, as shown in Fig 6.72.
Using a general radio GR 1208-B 65 MHz - 500 MHz unit oscillator, and the Hewlett Packard HP 8405A RF spectrum analyser, the frequency response of the MWA-210 gain module was measured in the same way as the MRF-901 amplifier (refer to section 6.6.2), the results are given in table 6.21 and plotted in Fig 6.73.
Table 6.21 Frequency Response of MWA-210 Gain Module

<table>
<thead>
<tr>
<th>Frequency (MHz)</th>
<th>Gain (dB)</th>
<th>Frequency (MHz)</th>
<th>Gain (dB)</th>
<th>Frequency (MHz)</th>
<th>Gain (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>20</td>
<td>260</td>
<td>21</td>
<td>460</td>
<td>20</td>
</tr>
<tr>
<td>70</td>
<td>21</td>
<td>270</td>
<td>21</td>
<td>470</td>
<td>20</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
<td>280</td>
<td>20</td>
<td>480</td>
<td>19</td>
</tr>
<tr>
<td>90</td>
<td>19</td>
<td>290</td>
<td>20</td>
<td>490</td>
<td>19</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>300</td>
<td>20</td>
<td>500</td>
<td>19</td>
</tr>
<tr>
<td>110</td>
<td>20</td>
<td>310</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>19</td>
<td>320</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>20</td>
<td>330</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>20</td>
<td>340</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>20</td>
<td>350</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>19</td>
<td>360</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>20</td>
<td>370</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>20</td>
<td>380</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>190</td>
<td>19</td>
<td>390</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>400</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>20</td>
<td>410</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>20</td>
<td>420</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>230</td>
<td>20</td>
<td>430</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>19</td>
<td>440</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>20</td>
<td>450</td>
<td>21</td>
<td></td>
<td>Error + 1 dB</td>
</tr>
</tbody>
</table>
Fig 6.73 Frequency Response of MWA-210 Gain Module

Clearly the MWA-210 gain module is satisfactory over the desired frequency range of approximately 500 MHz, yielding a 20 dB gain. The gain module will not operate at DC, but it will operate at frequencies as low as 31.83 KHz.
6.7 MIXER DESIGN

Correlative methods of reducing the detector noise output were discussed in section 5.8. Such techniques require the use of an electronic multiplier or mixer [5], [56], [57], [60]. The process of electrical mixing is a non-linear process and is achieved by exploiting the non-linear large-signal behaviour of semiconductor devices [12], [56], [57], [61], [62]. The process of optical frequency mixing has been discussed in section 5.4 to 5.6. The sensitivity of a SPD to optical intensity rather than the electric field, as is the case with radio antennae, is the mechanism by which optical mixing occurs.

The process of electrical mixing is somewhat similar. The non-linear behaviour of active electronic devices generates cross terms in the mathematical representation of the device behaviour [12], [56], [57], [61], [62]. As the models generally reflect physical behaviour of the device, such cross terms do occur and electrical mixing can take place [12], [56], [57], [61], [62].

A variety of active devices may be used as mixers. It is common to use semiconductor diodes in a ring configuration [12]. The non-linear large-signal behaviour, of the diode, mixes the signals fed into the ring, generating a large number of mixer products. Using a filter the desired mixer product is extracted for further processing. The other products of the ensemble are ignored.

Generally, diode mixers are supplied with input signals using transformer coupling, and cannot therefore be used at very low frequencies. Diode mixers are therefore generally unsuitable as frequency mixers in
correlative applications [12]. The multiple mixer products which they
generate are also undesirable in correlative applications. In order to
emulate a true mathematical multiplication, the mixer itself must eliminate
the multiple mixing and remixing of products and produce one product signal
for two input signals.

The transistor based differential mixer configuration may be used
without coupling transformers, it will produce mixer products right down to
DC [61],[62]. Its high frequency response is limited only by the coupling
of the input and output, and the frequency responses of the transistors
used. Using the differential mixer as a building block, it is also
possible to produce a true multiplier by using the common mode behaviour of
the configuration to effectively phase out all unwanted multiple mixer
products.

In section 6.7.1, the theory of the differential mixer is discussed in
detail. Using this theory it is possible to deduce a method of eliminating
the unwanted mixer products using the common mode rejection of the
differential pair.

Some early informal tests on the differential mixer were conducted by
building a temporary circuit. The temporary differential mixer is
discussed in section 6.7.2.

Using the knowledge of the behaviour of the differential mixer, a
stabilized current differential mixer was designed and built. The
stabilized current differential mixer is discussed in section 6.7.3.
In section 6.7.4, a full multiplier based upon the differential mixer is discussed.

6.7.1 Differential Mixer Theory

The mixing process, of the differential mixer, occurs as a consequence of the exponential relationship between the transistor base emitter voltage, and the emitter current [61],

\[ I_{E_0} = \left( I_S \exp\left[ \frac{qV_{BEO}}{RT} \right] - 1 \right) \]  
(6.63)

Where \( I_{E_0} \) is the emitter current, \( I_S \) is the transistor emitter saturation current and \( V_{BEO} \) is the base emitter voltage.

The differential mixer configuration is constructed essentially from two differential amplifiers, as shown in Fig 6.74 [61].

The collector current of \( Q_3 \) must equal the sum of the emitter currents of \( Q_1 \) and \( Q_2 \) [61],
\[ I_{c3} = I_{e1} + I_{e2} \]  \hspace{1cm} (6.64)

Where \( I_{c3} \) is the collector current of \( Q3 \), \( I_{e1} \) and \( I_{e2} \) are the emitter currents of \( Q1 \) and \( Q2 \) respectively.

The difference between the input voltages, \( V_1 \) and \( V_2 \), is given by the difference between the base emitter voltages of \( Q1 \) and \( Q2 \) [61].

Using (6.63), (6.64) and expressing the base emitter voltages of \( Q1 \) and \( Q2 \) in terms of the input voltage difference [61],

\[ I_{e1} = I_{c3} \exp \left[ \frac{q e (V_1 - V_2)}{K T} \right] \] \hspace{1cm} (6.65)

The transistor collector current is related to the emitter current by the forward transfer efficiency \( \alpha \) [56],[57],[61],

\[ I_c = \alpha I_e \]

and so

\[ I_{c1} = \alpha I_{c3} \exp \left[ \frac{q e (V_1 - V_2)}{K T} \right] \] \hspace{1cm} (6.66)

Using a similar analysis, \( I_{c3} \) may be expressed in terms of \( I_0 \), the current of the constant current source [61],

\[ I_{c1} = \alpha^2 I_0 \exp \left[ \frac{q e (V_1 - V_2)}{K T} \right] \exp \left[ \frac{q e (V_3 - V_4)}{K T} \right] \] \hspace{1cm} (6.67)
Using series expansions, $I_{c1}$ may be written in power series form [61]. Therefore,

$$I_{c1} \approx \frac{1}{4} \alpha^2 I_0 \left[ 1 + q_e (V_1 - V_2) \right] \left[ 1 + q_e (V_3 - V_4) \right]$$  \hspace{1cm} (6.68)

Truncating each series after the first two terms.

Using a similar analysis, the collector current of $Q_2$ may be expressed in terms of $I_0$ [61],

$$I_{c2} \approx \frac{1}{4} \alpha^2 I_0 \left[ 1 + q_e (V_1 - V_2) \right] \left[ 1 + q_e (V_3 - V_4) \right]$$  \hspace{1cm} (6.69)

Equations (6.68) and (6.69) contain a voltage multiplication between $(V_1 - V_2)$ and $(V_3 - V_4)$. Using collector resistors, voltages may be generated from the collector currents of $Q_1$ and $Q_2$ which are in antiphase [61].

6.7.2 **Informal Differential Mixer Tests**

Using UHF transistors, a differential mixer was built as shown in Fig 6.75, and some informal tests on the mixer behaviour, were performed.
Using two general radio 1208B RF unit oscillators and a Hewlett Packard 8554B RF Spectrum analyser, the UHF mixer was tested. Mixer products could be generated over a 500 MHz frequency range. Unfortunately the transistors were not well matched and the currents, of each transistor in the differential pairs could not be balanced. Attempts to balance the current in the differential pair Q1 and Q2, using a rheostat in place of RL2, caused Q2 to overheat before any results were recorded.
6.7.3 A Current Stabilized Differential Mixer

Using MRF-901 transistors, it was decided to build a differential mixer using current mirrors [63] to force the equality of collector currents, that is \( I_{c1} = I_{c2} \) and \( I_{c3} = I_{c4} \).

A typical current mirror is shown in Fig 6.76 [63].

![Typical Current Mirror Diagram](image)

The collector current, \( I_{c1} \), is related to the collector current, \( I_{c2} \) of \( Q2 \) by [63]

\[
I_{c2} \approx \frac{B_1 I_{c1}}{B + 2}
\]  

(6.70)

By adding additional transistors as shown in Fig 6.77, the collector current of \( Q1 \) may be reproduced in many more circuits [61].
6.130

Fig 6.77 Extending the Current Mirror Beyond 2 Circuits

If \( R_1 = R_2 \), (refer to Fig 6.76), multiples of the collector current, \( I_{c1} \), of \( Q_1 \) may be forced through the second circuit. This property was exploited in the mixer design. The forward current collection efficiency \( \alpha \) is very close to unity for a MRF-901, therefore the collector current \( I_{c4} \) of \( Q_4 \) will equal twice the collector current of \( Q_1 \) and \( Q_2 \). Using the current mirror, \( I_{c4} \) can be set equal to twice \( I_{c1} \) or \( I_{c2} \). It was decided to set the current mirror in an external circuit, and use transistors to force the collector currents of \( Q_1 \) and \( Q_2 \) to equal this figure. The third transistor would be used to force \( I_{c4} = 2I_{c1} = 2I_{c2} \).

The complete current mirror circuit for a differential mixer is shown in Fig 6.78.
An independent current source would be used in the common-emitter circuit of Q3 and Q4, of Fig 6.74, in order to produce the highest possible common mode rejection of the Q3 - Q4 transistor pair. The common-emitter constant-current source is shown in Fig 6.79.
The constant current source must deliver a current \( I_0 \approx 4I_{c1} \approx 4I_{c2} \). This requirement fixed the value of \( R_b = \) in Fig 6.79. In order to provide a high degree of current stability, a zener diode was connected between the base of \( Q_1 \) of Fig 6.79 and the supply.

It was decided to use the current mirror transistors as load resistors for the collectors of \( Q_1, Q_2 \) and \( Q_4 \). If any series resistors were included they would merely reduce the voltages over which the current mirror transistors could operate to give the desired currents. Therefore the stabilized-current differential mixer was designed as shown in Fig 6.80.
Variable resistors were included with each current mirror transistor, to give limited current adjustment to each, to compensate for values of $\alpha$ below unity. Each current mirror transistor was decoupled from the MRF-901 transistors using a radio frequency choke consisting of a Ferrite bead and a 1 mH indicator, as the transistors of the current mirror become RF short circuits at high frequencies.

The differential mixer was built in two parts, a low frequency part comprising the current mirror and constant current source, and a high frequency part which was the differential mixer itself.

The low frequency current stabilization circuits were built on veroboard and mounted on one side of a diecast aluminium box using 6BA bolts, nuts and standoffs.

Two 15V 3 pin regulator was used to supply positive and negative voltage to the stabilization circuits. Power was externally connected to the mixer via a plug mounted on the side of the box opposite the low frequency circuit board.

The MRF-901 transistors and associated circuits were mounted on the surface of a piece of double sided circuit board, which was mounted in the bottom of the box on 8 mm standoffs. Connections were made between the high frequency circuits and the current stabilizers via the decoupling elements.

The physical layout of the mixer inside the box is shown in Fig 6.81.
The layout of low frequency components and the breaks in the tracks of the veroboard are shown in Fig 6.82 (a) and Fig 6.82 (b) respectively.
The high frequency components were arranged, on the circuit board, to ensure that the interconnections were as short as possible. Although sufficient space, between components, was left so that copper shields could be placed between transistors, to eliminate instability due to radiative RF feedback. The physical layout of these components can be seen in Fig 6.83.

Input and output RF connections were made using 50Ω coaxial cable from the high frequency circuitry to BNC connectors mounted on the sides of the box as shown in Fig 6.81. The BNC sockets were mounted on the box in insulated washers to isolate the earthed circuitry of the mixer from the box, which would be earthed to a common point by an extra external wire.

The completed stabilized-current mixer is shown in Fig 6.84.
Once assembled, a 20V power supply was connected to the mixer. Via a capacitor, each input was connected to a 50Ω dummy load.

Using an analogue voltmeter, the transistor voltages were checked. By adjusting the variable resistors R1, R2 and R3, the MRF-901 collector-emitter voltages could be altered. (See Fig 6.80.) Unfortunately the adjustment of the variable resistors was critical. Minor changes in resistance caused severe swings in the collector voltages of the circuit. The desired operating point could not be reached, as it was a point of instability. The only collector voltages could only be moved between saturation values and the voltage supply values. The transistor combinations tended to turn either current mirror transistor off, or move it into its non-linear region of operation.
The attempts to adjust the mixer to the desired operating point resulted in the overheating and destruction of transistor Q4, which was replaced a number of times.

The quality of the current mirror and constant current source are dependent upon the DC current gain of the transistors used in them [63]. The only way in which the stabilized circuitry could be adjusted by hand, is by reducing the quality of the stabilization circuits so that their interaction, through the differential transistor pairs, could be reduced [63]. The quality reduction was achieved by bypassing each current source with a 10kΩ resistor, introducing some error into the actual value of current supplied. Resistor values were experimentally reduced until the mixer could be adjusted into its operating condition. QS1, QS2 were bypassed with QS3 was bypassed with a resistor and QS4 with a resistor yielding the modified mixer circuit of Fig 6.85.

**Fig 6.85 The Modified Differential Mixer**
The values of the bypass resistors are very large, they would, therefore, introduce much less error in the mixer collector currents than the resistors of the differential mixer of Fig 6.75.

A further adjustment problem, due to external electromagnetic interference, was also encountered. Using the manufacturers values for the hybrid parameters of the current stabilizing transistors, a computer circuit-simulation, using the circuit simulation program SPICE [64], was undertaken. Using SPICE suitable capacitor values were selected to bypass the current mirrors, at a frequency where the collector RF chokes have a significant impedance. The bypass value chosen is 0.1 μF, which produces the impedance curve shown in Fig 6.86.

Fig 6.86 The Impedance of a Current Mirror Connection

Having adjusted the mixer to its operating condition, the mixer response was tested over a 400 MHz frequency range. Connecting the two General Radio unit-oscillators, and the Hewlett Packard RF spectrum analyser, as for the UHF transistor mixer, the response of the mixer, for a number of input frequencies between 100 MHz and 500 MHz, was measured. The results of the mixer frequency sweep are given in table 6.22 and graphed as a series of level curves in Fig 6.87.
Table 6.22 Measurements of Stabilized Differential Mixer Response

<table>
<thead>
<tr>
<th>Input Port Frequency (MHz)</th>
<th>Difference Frequency at Output Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port 1</td>
<td>Port 2 Frequency</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>500</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>500</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>500</td>
</tr>
<tr>
<td>400</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>500</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>500</td>
</tr>
</tbody>
</table>

* Unable to Measure the Power at DC.
If the differential mixer is to be used over the 500 MHz frequency range, the transfer function between the two input signals and the mixer product desired must be equalized. This involves the design of two preconnected equalizers at the mixer inputs, and an equalizer connected to the output. A full treatment of equalization is given in section 7.2.

6.7.4 A Full Multiplier Using the Differential Mixer

Using two differential mixers, it is possible to produce a true analogue multiplier which could be used over a very wide range of frequencies.

Recall equation (6.69) which described the collector current of transistor Q2 (refer to Fig 6.74).
\[ I_{c2} \approx \frac{1}{4} I_0^2 \left[ \frac{1-q_e(V_1 - V_2)}{2KT} \right] \left[ 1 + \frac{q_e(V_3 - V_4)}{2KT} \right] \]

Which may be expanded,

\[ I_{c2} \approx \frac{\alpha^2 I_0}{4} - \frac{2I_0 q_e(V_1 - V_2)}{8KT} + \frac{2I_0 q_e(V_3 - V_4)}{8KT} \]

\[ -\frac{\alpha^2 I_0 q_e^2(V_1 - V_2)(V_3 - V_4)}{16KT} \quad (6.71) \]

By swapping the input signals between the two inputs of a second mixer, a current \( I_{c2}' \) corresponding the \( I_{c2} \) of (6.71) is produced,

\[ I_{c2}' \approx \frac{\alpha^2 I_0}{4} + \frac{\alpha^2 I_0 q_e(V_1 - V_2)}{8KT} - \frac{\alpha^2 I_0 q_e(V_3 - V_4)}{8KT} \]

\[ -\frac{\alpha^2 I_0 q_e^2(V_1 - V_2)(V_3 - V_4)}{16KT} \quad (6.72) \]

By summing \( I_{c2} \) and \( I_{c2}' \), the input signals are eliminated from the output,

\[ I_{c2} + I_{c2}' \approx \frac{\alpha^2 I_0}{2} - \frac{\alpha^2 I_0 q_e^2(V_1 - V_2)(V_3 - V_4)}{8KT} \quad (6.73) \]

\[ \frac{\alpha^2 I_0}{2} \] represents a DC bias, and \[ -\frac{\alpha^2 I_0 q_e^2(V_1 - V_2)(V_3 - V_4)}{8KT} \] is the desired mixer product.

The basic circuit, to produce true multiplication, is shown in Fig 6.88.
Fig 6.88 A True Multiplier Based Upon the Differential Mixer
6.8 DESIGN OF AN INTEGRATOR FOR LONG TERM CORRELATION APPLICATIONS

If correlative methods of noise reduction are used in the detection of optical information, a method of integration of the multiplier output must be found. The designer has the choice of two technologies either of which he may use to produce an integrator [56],[57],[60],[62]. Integration may be readily performed by analogue means, in which charge is accumulated on a capacitor [56],[57],[60],[62]. The capacitor itself may be used as an integrator of charge, as described by

$$Q(\tau) = \int_{0}^{\tau} I(t) \, dt \quad (6.74)$$

Where $I(t)$ is a time varying current, $Q(\tau)$ is the charge at time $\tau$ and $\tau$ is the time at which the charge in the capacitor is observed.

Unfortunately, the single capacitor is unsuited to the integration of voltage because the accumulation of charge makes it difficult to further apply charge [57]. To overcome this problem an amplifier is used to isolate the input voltage from the capacitor, applying charge independently of the input voltage, until the capacitor is charged to the supply voltage [57].

It is also possible to build an integrator from digital circuitry, by quantizing the input voltage using an analogue to digital (A/D) converter [60]. The resulting digits are added to the sum held in an accumulator. Integration may continue until the accumulator overflows [60].

By considering the specific application to which the correlator is put, the designer may decide whether an analogue or digital integrator is
to be used. In section 6.8.1, the application of the correlator, in communications systems and measurement systems, is discussed, and suitable integrators chosen for the application required. Various types of integrators are also discussed.

It was decided to build an analogue integrator for some long term measurement applications for use over long time periods. Unfortunately the stability of analogue integrators varies with temperature. After a short period of time with no input present they tend to accumulate charge due to changes in operating point and leakage current [65]. In section 6.8.2, the mechanisms that accumulate unwanted charge, in analogue integrators, are examined in detail. The operational amplifier based integrator is of particular interest.

In section 6.8.3, a technique of overcoming the effects of the drift mechanisms, in an operational amplifier-based integrator, is introduced. The new integrator is known as the switched-capacitor-stabilized integrator (SCSI). The switch timing for the SCSI is derived using a timing technique well known in the design of integrated circuits at the Very Large Scale Integration (VLSI) level. The timing mechanism is discussed in detail in section 6.8.4. The way in which the VLSI timing technique is applied to the switched capacitor integrator is the subject of chapter 6.8.5, where a full switched capacitor integrator is produced. Considerable care was taken over the layout of components of the integrator, and over the components used to build it. The details are covered in section 6.8.6.
Once constructed, the response of the switched capacitor integrator was measured over very long time periods. The results of the measurements are given in section 6.8.7.

Finally, integrators suitable for correlation over short time periods, for data receiving applications, are discussed.

6.8.1 Digital Integrators Versus Analogue Integrators

Selection of a particular integrator type is a compromise between dynamic range and bandwidth of the signal, and the cost of instrument. Generally, digital implementation of the process is complicated and expensive requiring very high speed digital circuits for signals of wide bandwidth. Firstly, the input signal must be band limited and quantized [60]. The latter process introduces noise, dependent upon the number and distribution of the quantization thresholds [31]. Quantization noise is reduced by increasing the number of quantum levels and therefore the number of bits [31]. Ultimately the number of bits is fixed by the amount of quantization noise that may be tolerated in an application.

The output of the quantizer is connected to a numerical accumulator, where it is added to the total, present. The number of bits, available in the accumulator, sets the dynamic range over which the integrator may be used before it overflows [60]. If the integrator is used over large time periods, the number of bits in the accumulator will be very large indeed; even for signals of small bandwidth which may be sampled at a low rate [30].
Consider the integrator shown in Fig 6.89.

The analogue input is limited by the band pass filter to 100 KHz, which is sampled at a rate of 200 KHz as required by the sampling theorem [9],[30],[31]. The samples are quantized and 8 bits of information are generated every 5 µs. If the integrator is to sum all input over a period of just 1 hour, the accumulator would need to consist of 30 bits. Moreover, to obtain the required speed, TTL digital circuitry would be required [66]. A 30 bit TTL accumulator is a very large device indeed.

The accumulator size increases drastically for each extra bit of accuracy produced by the quantizer.

An analogue integrator, to perform the same task, consists of an amplifier, a capacitor and an input resistor as shown in Fig 6.90 [56], [57].
The value of the capacitor, resistor and supply voltage fix the period over which the integrator could be used. The actual complexity of the integrator is unaffected by the integration period required for a input voltage equal to the supply voltage [56],[57].

The advantage offered to the designer, by the digital integrator, is the stability of the value stored in the accumulator over long periods of time. Unlike the analogue integrator, the value stored in the accumulator will remain there until the power supply is disconnected. Unfortunately the charge stored on the capacitor, of an analogue integrator, will decay due to leakage currents and offset voltages of the amplifier. After only a short period of time, the charge will be completely removed, or indeed altered if the offset voltage tends to add charge to the capacitor. In either case, the sum is lost.

Digital integration techniques may also be simplified in order to reduce the complexity of the equipment. If the Digital integrator of Fig 6.89 is described as a Pulse Code Modulation integrator, since it has the characteristics of a PCM device [30],[31], then it is reasonable to assume that a simpler delta modulation integrator [30],[31] could be used to integrate the signal.
Consider the block diagram shown in Fig 6.91.

**Fig 6.91 An Integrator Based on the Delta Modulator**

The input voltage enters an analogue summer, where a voltage, corresponding to the numerical value in accumulator 1, is subtracted from it. The difference voltage is fed into a hard limiter, which produces a fixed positive output voltage for any value of positive input voltage and a fixed negative output voltage for any value of negative input voltage. The output of the hard limiter [56],[57],[60] is added to the value contained in accumulator 1 when a clock pulse CLK1 arrives. The output of accumulator 1 is passed to a second accumulator and to a digital-to-analogue (D/A) converter [56],[57]. The D/A analogue output is subtracted from the input voltage [56],[57]. The input to accumulator 2 is added to the value in accumulator 2 on the arrival of clock pulse CLK2. The output of accumulator 2 is the integral of the input.

Suitable clock pulses for accumulators 1 and 2 are shown in Fig 6.92.
The rate at which the input waveform is sampled is determined by the bit interval $T_b$ as shown. The actual clock pulse period is $T_b/4$, which sets the maximum frequency $f_d$ at which the digital circuitry must operate,

$$f_d = \frac{4}{T_b}$$  \hspace{1cm} (6.75)

At first glance the integrator of Fig 6.91 is more complicated than that of Fig 6.89, however it does not need an expensive A/D convertor. The digital to analogue converter in the feedback line can be made of cheap materials [67].

The actual operating speed of the integrator is fixed by the maximum slope of the waveform it is intended to follow. This is determined by the maximum frequency waveform under which the area must be found.

The maximum slope that the integrator can follow is given by the scale factor [56]. $K_{AD}$ relating the numerical value contained in accumulator 1 to the output voltage of the D/A convertor,

$$V_{AD} = K_{AD} N_{AD}$$  \hspace{1cm} (6.76)
where \( V_{AD} \) is the output voltage of the D/A convertor and \( N_{AD} \) is the numerical input to the D/A convertor.

and the integrator bit interval \( T_b \).

The maximum slope achievable by the integrator is given by,

\[
S_I = \frac{k_{AD}}{T_b} \tag{6.77}
\]

where \( S_I \) is the integrator maximum slope.

The maximum slope of the sinusoid of frequency \( f \) occurs at zero with a value \( S_s \) of [9],[56],

\[
S_s = 2\pi Vf \tag{6.78}
\]

where \( V \) is the maximum voltage of the input waveform.

Equating (6.77) and (6.78), the minimum bit interval is found,

\[
T_b = \frac{k_{AD}}{2\pi Vf} \tag{6.79}
\]

and the maximum digital operating frequency is found by substituting (6.79) into (6.75),

\[
f_d = \frac{8\pi Vf}{k_{AD}} \tag{6.80}
\]

The D/A proportionality constant, \( k_{AD} \), is selected by the number of bits required to specify the maximum input voltage \( V \), which gives a value of voltage represented by 1 bit.
Where \( n_b \) is the number of bits required to represent \( V \).

Substituting (6.81) into (6.80) gives,

\[
f_d = 8\pi f_n b
\]

(6.82)

If 8 bits are chosen to represent 15V at a maximum frequency of 500 MHz, the maximum operating frequency of the digital circuitry requires a maximum frequency of 3.217 THz, a figure that is quite impossible to meet.

Yet, using a broadband amplifier, an analogue integrator could be built to integrate such signals.

The simple analysis given for the two digital integrators, and the assumption about the required bandwidth for analogue integrator amplifiers, overlook an important characteristic of integrators in general.

An integrator sums all direct current contributions to a waveform. Over a sufficiently long period of time alternating currents contribute nothing to the integrator total. Therefore, an integrator need not be sensitive to alternating voltages of short period when compared with the proposed period of integration. The period of integration is, itself, determined by the application of the correlator. If the integration is to sum voltage contributions over periods of hours, the bandwidth need only be a few Hertz, so that contributions of unequal numbers of upper and lower
excursions of voltage are insignificant compared with these after periods of minutes or hours.

Let $T_I$ denote the proposed integration period and $T_s$ denote the period of a particular Fourier component of the integrator input waveform. The worst case error is introduced if the integration is terminated so that the Fourier component consists of an odd number of half cycles. If the amplitude of the $i$th Fourier Component is $V_i$, the direct current of the half cycle is given by the average,

$$V_i = \frac{V_i}{\pi}$$

or, in terms of the period of the $i$th Fourier component,

$$V_i = \frac{V_i T_i}{\pi}$$

Where $T_i$ is the period of the $i$th Fourier Component.

The error introduced by the half cycle relative to the integration period, where the worst case voltage would be equal to the voltage of the $i$th Fourier component, is

$$\epsilon_i = \frac{T_i}{T_I}$$

Where $\epsilon_i$ is the quality of the band limited integrator.

A one percent error is introduced in the worst case if
Equation (6.85) may be expressed in terms of bandwidth $f_I$ where

$$\Delta T_i = \frac{1}{\Delta f_i} \tag{6.86}$$

$$\Delta f_i = \frac{100}{T_i} \tag{6.87}$$

Integrating over a period of 1 hour with a % error requires a bandwidth of

$$f_i = 27.78 \times 10^{-3} \text{Hz}$$

which is very small indeed.

The integrator requirements are thus simplified for both the Digital and Analogue implementations. Unfortunately even with the reduction in integrator bandwidth requirements, the digital integrator is still much more complicated than the equivalent analogue integrator. If the designer can reduce the leakage current and offset voltage effects on the charged stored by the analogue integrator, it is a much more attractive method of integration than the digital methods.

Using a technique based upon the assumption that operating conditions of an integrator change over a very long period of time, a method of producing a very stable analogue integrator was found.

Applying the bandwidth requirements to an integrator which is intended to operate over a period of hours, it is found that the amplifier bandwidth required, for very long low measurement errors, is very small; less than 1 Hz in fact. Therefore, an analogue integrator, suitable for the task, can be made using an Operational-Amplifier [56],[57].

In this section, the characteristics of the operational-amplifier, that cause an integrator to drift, are discussed. A typical, operational-amplifier based, integrator is shown in Fig 6.93.

![Fig 6.93 A Typical Analogue Integrator Using an Operational Amplifier](image)

In the analysis of such integrators, it is normally assumed that \( V_d = 0 \), therefore no current flows into the input terminals of the operational-amplifier [56],[57]. In normal applications such an assumption is quite reasonable, as the currents, associated with the charging of the capacitor \( C \), are of many orders of magnitude greater than the Operational Amplifier input current [56],[57].

When the input terminal is grounded, \( V_{in} = 0 \), current flows into both operational amplifier inputs keeping the two transistors of the
differential input forward biased [56],[57]. The input transistors are fabricated on the same substrate by the same process, therefore the input currents are the same, and may be assumed equal for all practical purposes. The input current at the non inverting (+) terminal is supplied from the earth. The input current to the inverting terminal, (-), is supplied by node N₁ drawing current, from the earth through R, and from the capacitor, charging it such that the output voltage tends to rise in a positive direction. Therefore the integrator output tends to drift as the input current, drawn by the inverting input, charges capacitor C. Eventually the output rises to the supply voltage of the operational amplifier yet no input was applied. The time taken to charge the capacitor, due to the operational amplifier input current, is fixed by the parallel combinations of resistor and capacitor, R and C, in series with the input resistance of the inverting terminal [65].

If a finite input voltage is applied to the integrator input, \( V_{in} \), current flows through resistor R to node N, most flows from N₁ into the capacitor C, charging it and producing a rising negative output voltage \( V_{OUT} \). A small amount of current flows into the inverting terminal, increasing its input current, above the zero input voltage terminal current. Due to the differential nature of the operation amplifier input [56],[57], the current flowing into the non inverting terminal is reduced by the same amount. The current flowing into the inverting terminal of the operational-amplifier does not contribute to the charge on the capacitor C, it is, in fact, lost to the integrator.

In general the input currents to the inverting and non inverting terminal currents are
\[ i_- = i_o + \Delta i \]  
(6.88)  
\[ i_+ = i_o + \Delta i \]  
(6.89)

Where \( i_- \) and \( i_+ \) are the respective inverting and non-inverting terminal currents, \( \Delta i \) is the additional input current, due to a non-zero input voltage, and \( i_o \) is the terminal input current at zero bias.

If the capacitor \( C \) is charged, and the input voltage returned to zero, the charge on the capacitor causes more current to flow into the inverting terminal, in which case the current flowing into the non-inverting terminal is reduced, as given by equation (6.89).

Corruption of the charge residing in the capacitor, due to the input terminal current, is known as leakage drift.

Due to small imbalances in the differential input circuit, the output voltage of the operational amplifier may be non-zero for zero input voltage. Generally this offset voltage is removed by the connection of a potentiometer to the offset terminals of the operational-amplifier [56], [57]. The potentiometer is adjusted to give an output voltage of zero when the input is grounded. Unfortunately, the offset tends to change with time and temperature [56],[57]. In the integrator configuration, offsets in output voltage cause charge to build up on the capacitor \( C \) corrupting any charge due to an input signal; eventually saturating the integrator.

Using additional components and the switched capacitor concept, the drift, due to offset voltage and leakage current, may be reduced
sufficiently to use the integrator over periods of many hours. The details are given in the following section.

6.8.3 Leakage Current and Offset Voltage Stabilization

By exploiting the relationship between input currents of the inverting and non inverting terminals of the operational amplifier, summarised by equations (6.88) and (6.89), the integrator drift due to leakage current may be reduced. By connecting a parallel capacitor resistor combination between the non inverting terminal and the ground, as shown in Fig 6.94, charge, due to leakage, current may be accumulated on the capacitor [65].

The charge, lost due to leakage current from $C_1$, is stored on $C_2$ at the same rate; if $R_1=R_2$ and $C_1=C_2$. The voltage, due to the charge on $C_1$, is subtracted from $C_2$. There is therefore, no net change in output voltage of the integrator remains constant [65]. Due to the gain on the transistors, of the differential operational-amplifier input, the change in current flow, into the inverting and non inverting terminals,
+$\Delta i$, is very small, its effect is negligible over a period of many hours [65].

The offset voltage, of the operational amplifier, adds unwanted charge to the capacitor $C_1$. The actual amount of charge added is proportional to the time for which the offset is present. The undesired extra charge may be removed if the capacitor $C_1$ is reversed. Connecting the reversed capacitor, for the same period of time, to the operational amplifier. The charge on the capacitor, due to the offset voltage, is removed; if the offset voltage is unchanged.

Regular reversals of $C_1$, over a long period of time, will prevent the integrator from saturating due to the offset voltage.

Charge of $C_1$, due to the input voltage $V_{in}$, is applied to the capacitor, $C_1$, in a direction dependent only on the polarity of the input by reversing the voltage input connections at the same time as the capacitor, $C_1$, is reversed.

Capacitor $C_2$ must also be reversed, with the reversal of $C_1$, to ensure that the leakage compensation is applied in the correct direction.

Finally, the operational amplifier output must be reversed, with the capacitor $C_1$, to preserve the correct polarity of the output voltage.

The complete integrator, with appropriate switching circuits, is shown in Fig 6.95.
Initially contacts 4AA1 through 4AF1 are closed and the integrator assumes the configuration shown in Fig 6.95. To reverse the connections as described contacts 4AA1 through 4AF1 are opened, and contacts 4BA1 to 4BF1 are closed reversing the input, the connection of $C_1$ and the connection of $C_2$ and $R_2$.

Due to the switching of the capacitor, input and output directions, the integrator is called the Switched-Capacitor Integrator.

The details of the switching circuits required to run the integrator are discussed in section 6.8.4.
6.8.4 Switching Techniques and the Two-Phase Clock

In order to ensure that contacts 4AA1 through 4AF1 and 4BA1 through 4BF1 are not all closed at the same time, shorting the input, output and capacitors C1 and C2, a two-phase non-overlapping clock must be used. A timing diagram of a typical two phase non-overlapping clock is shown in Fig 6.96 [68],[69].

![Fig 6.96 A Two-Phase Non-Overlapping Clock](image)

T₁ is the period of the clock, T₉ the guard interval between the downward voltage transition of one phase and the upward transition of the second phase. T₇ is the duration of the high level of the clock.

The two phase non overlapping clock is a well known technique of exerting clock control of devices in VLSI circuits [68],[69],[70]. The guard interval T₉ is of great importance using the two phase clock scheme, it prevents the circuit from entering a state race-condition by giving one piece of circuitry time to operate before applying a clock pulse to another [68],[69],[70]. In the integrator context T₉ must be sufficient to ensure that contacts 4AA1 through 4AF1 have time to open before contacts 4BA1 through 4BF1 close, and visa-versa.
In VLSI applications the non-overlapping two phase clock is derived from the clock input to the chip [68],[69],[70]. Conceptually it is derived from the input clock, which is usually a square wave, as shown in Fig 6.97 [68],[69],[70].

Using two NAND gates, one can ensure that output $01$ cannot rise to a logical unit value until output $02$ has fallen to zero. Similarly, output $02$ cannot rise to logic 1 until output $02$ has fallen to zero. The guard time, $T_g$, between each phase is fixed by the feedback delay, $T_f$, and propagation time of the NAND gates [68],[69].

In VLSI circuits, the guard time delay required, is introduced by passing the feedback circuits around the longest path taken by the clock signals, so that the guard time exceeds the longest clock propagation delay introduced by the greatest path length [68],[69],[70]. This avoids race conditions leading a finite-state circuit into an unwanted state and ensures that capacitive memory devices are not discharged by a short circuit, which is the same requirement as for the switched capacitor integrator.

Using TTL components a method of generating a two phase clock, with sufficient guard time to ensure correct operation of the switched capacitor integrator, was sought. The solution follows in section 6.8.5.
6.8.5 Adapting the Two Phase Clock to the Switched Capacitor Integrator

The frequency stability of the master switching oscillator is of little importance in the switched capacitor integrator, therefore a crystal controlled master oscillator is unnecessary. Using two 74LS098 monostable flipflops, configured as shown in Fig 6.98, a simple digital oscillator can be built.

![Fig 6.98 Switched Capacitor Integrator Master Oscillator](image)

When the power supply is connected, the corresponding upward voltage transition of Q₂ triggers 098A₁. Output, Q₁, goes high for the period specified by the resistor capacitor combination C₁ and R₁. After the time period Q₁ returns to zero. The transition triggers 98A₂ and output Q₂ goes to zero. After a time-out period fixed by C₂ and R₂, Q₂ returns to 1 which triggers 098A₁ and the process is repeated. The output of Q₂ is therefore a rectangular wave. Due to different propagation delays through the monostable flipflops, the rectangular waveform mark-space ratio differs from unity.
The reversals of the capacitor $C_1$, must be of the same duration as the normal connection, therefore the mark-space ratio of the master oscillator must be unity. By dividing the waveform frequency by 2, using a JK flipflop, alternate marks and spaces are generated by one transition of the clock, which is regularly spaced. Thus a unit mark-space ratio clock is generated as shown in Fig 6.99.

The unit mark-space ratio clock is used as input to a two phase non overlapping clock generator, as shown in Fig 6.100.
The feedback delay is introduced by a capacitor and series resistor, in each feedback path, as shown. Assuming that the input impedance of the feedback inputs to the NAND gates are infinite, the delay introduced, by the series resistor and capacitor, is determined by simple analysis of the circuit shown in Fig 6.101.

\[ Y_{\text{out}} = Y_{\text{in}} \left( 1 - \exp \left( \frac{-t}{CR} \right) \right) \] (6.90)

The state of the input to a TTL gate is said to be unity when the voltage is 2V, in which case the product CR is given by,

\[ CR = 1.958T_b \] (6.91)

The choice of switches, used in the switched capacitor integrator, is determined by the leakage current that may be tolerated through the switches. Initially CMOS solid-state analogue switches were selected, and the switched capacitor integrator built on breadboard. The capacitor used for \( C_1 \) and \( C_2 \), of Fig 6.95, are polyester capacitors, which offer a compromise between large capacitance values and dielectric leakage.
Initially, the capacitors were tested using a high voltage source and electrostatic voltmeter. A drop of less than 1 volt in 400 over an hour period was recorded in medium-humidity ambient conditions.

Unfortunately the analogue switches introduced significant leakage. The charge stored on capacitors $C_1$ and $C_2$ leaked away completely over a period of one or two minutes. The switching speed chosen, produced two forward and reverse capacitor connections per second.

At such low speeds, relays are a viable method of switching the integrator capacitors, input and output. Each switch must operate many times during an integration period, therefore, reed delays are the most suitable relay types for the application. Change-over versions of the reed element are available, unfortunately the relay change-over times vary too widely, therefore these are unsuitable for the application, as there is no way in which to introduce a guard period during changeover.

Using the same two phase clocking scheme, simple NEC TYPE-PRA-4 make-or-break reed relays were used. These are packaged as dual in line (DIL) integrated circuit packages complete with coil. The manufacturers data for these devices is given in appendix 12. The minimum coil voltage, at which the contacts must operate, is 3.8 V$_{dc}$. It is therefore possible to use the relay with TTL components, if they can source or sink the associated current, for a 5V logic level. The required drive current, $I_c$, is found using the coil resistance $R_c = 500\Omega$, whence $I_c = 10$ mA. Standard TTL devices can sink 16 mA, therefore each relay can be driven directly by a TTL gate. Using a 7405 open collector hex inverter, each relay is driven, as shown in Fig 6.102.
Diode, $D_p$, is included to provide a back EMF path, for the coil, when the inverter output goes high [71]. For these relays, a 1N4004 is sufficient for the purpose. Capacitor $C_1$ is present to protect the output circuitry, of the inverter, from voltage spikes as the output voltage rises and falls [71]. A $0.1\mu F$ capacitor is sufficient for this application.

Using this drive technique, the full switched-capacitor integrator was designed. Its full circuit diagram is shown in Fig 6.103.
6.8.6 Switched-Capacitor Integrator Construction

The switched-capacitor integrator was built on double sided Dual-In-Line (DIL) circuit-board available in the department. The ground plane, on the upper side of the board, should reduce the effects of voltage spikes, from the relay coils, on the integrator capacitors.

The component arrangement on the board was chosen to isolate the relays from the analogue circuitry as much as possible. The two major storage capacitors C5 and C6 of Fig 6.95 were mounted on the underside of the board, so that the ground plane would isolate them from the switching components. A diagram of the complete circuit layout is given in Fig 6.104.
Fig 6.104 The Switched Capacitor Integrator Circuit Board

(a) Upper Side

(b) Lower Side
A full interconnection table is available in Appendix 13. Cross-
connections were made by looming wires on the top of the board. Wires
carrying analogue signals and those carrying switching signals were loomed
separately.

The card itself, was mounted in a verobin card cage with plug-in
black-plane connectors. Using veroboard, an extender card was made so that
in-circuit adjustments and measurements of the integrator, could be made.

The complete switched-capacitor integrator circuit is shown in Fig
6.104, and the integrator plugged into the extender card, which is in turn
plugged into the backplane of the verobin card cage, is shown in Fig
6.105.

Fig 6.105 The Vero-bin Card Cage, Exender Board and Integrator
6.8.7 Testing the Switched Capacitor Integrator

Once built, the switched-capacitor integrator was tested. A 5V, and ±15V power supply was connected. A second power supply was used as the input voltage source. A HPO was connected to the operational amplifier output. On connection of the 5V power supply, a ticking could be heard, at a rate of approximately 2 per second, as alternate sets of reed relays were opened and closed.

The two input circuits were short circuited together initially. No measurable change in output voltage occurred over a 5 hour period. Using the second power supply, the integrator was charged to an output of -15V over a period of 5 hours. The voltage at the output had risen by approximately 2V to -13V. The integrator was then charged to an output voltage of +15V, which did not measurably change over a 5 hour period. Voltages of ±7.5 V output were also tried, these did not change measurably over a 5 hour period.

Using a potentiometer, the offset voltage, of the operational-amplifier, was changed via the external-offset pins. Adjustment of the potentiometer over its full range had no observable effect on the integrator output. Therefore, the switched capacitor concept is a good method of removing the effects of operational amplifier offset-voltage errors in integrator applications.
6.8.8. Integrators Suitable for Data Signal Applications

The analysis of section 6.8.1 indicates that the integrator bandwidth requirements are dependent upon the period of integration. In correlative digital applications, the integration period is one bit interval $T_b$. The correlator is then sampled every $T_b$ seconds to determine the logical value of the signal received [56],[57]. The amplifier bandwidth required, for an analogue integrator of such signals, is at least,

$$\Delta f = \frac{1}{2T_b}$$

(6.92)

With no redundancy, an optical system with an electrical bandwidth of 500 MHz can transmit data at a maximum speed of 1Gbps. The minimum bandwidth, of an integrator amplifier for correlative detection of such data, is 500 MHz.

For reasons already discussed in section 6.8.1, a digital integrator is quite infeasible for such an application.

Using a broadband amplifier, an integrator of the desired bandwidth can be built, as shown in Fig 6.106.
Methods of reducing the effects of offset and leakage, over the integration period, must be found.

At very high frequencies, capacitor switching may be an impractical technique of offset stabilization.
REFERENCES - CHAPTER 6

"Optoelectronics - An Introduction."

"Optical Communications."
John-Wiley and Sons, 1976, pp 107 - 140.

"Coherent Light."

[4] KINGSTON, R.H.
"Detection of Optical and Infrared Radiation."

[5] HANBURY BROWN, R.
"The Intensity Interferometer its Application to Astronomy."

[6] PERSONICK, S.D.
"Receiver Design for Optical Fiber Systems."

[7] PERSONICK, S.D.
"Receiver Design for Digital Fibre Optic. Communication Systems (I & II)"
"Optical Detector Package."
The Bell System Technical Journal, Vol. 57, No. 6, July - August
pp 1809 - 1822.

[9] BRACEWELL, R.N.
"The Fourier Transform and Its Applications."

[10] SZENTIRMAI, G.
"The Problem of Phase Equalization."

"Linear Phase Versus Step Response."

[12] DEMAW, D.
"The Radio Amateurs Handbook."
American Radio Relay League, Newington CT, USA 0611, pp 6.18,
6.31, 7.3.

[13] ORTEL CORPORATION PRODUCT INFORMATION
PD050 - OM Photodiode
PD050 - PM
Ortel Corporation 2015, W. Chestnut St., Alhambra, C.A. 91803,
USA.
[14] OPTO-ELECTRONICS LTD
"Ultra High Speed Photo-Detectors and Pulsed Diode Laser Sources"
(Product Catalogue)
Opto-Electronics Ltd, Unit 9, 2538 Speers Road, Oakville, Ontario,
Canada, L6L 5K9.

[15] INFRARED INDUSTRIES, INC (EASTERN DIVISION)
New Products.
Infrared Industries, Inc (Eastern Division), 62 Fourth Ave.,
Waltham MA 02154, USA.

[16] CLARK, R.S.
"Photonics Spectra, The International Journal of Optics,
Electro-Optics, Lasers and Fibre Optics.
Teddi C. Laurin. Publisher. Berkshire Common, Pittsfield MA

[17] MOTOROLA TECHNICAL INFORMATION CENTRE
"Motorola Optoelectronic Device Data."

[18] MACAM LTD
Macam Ltd, Scotland.

[19] SMITH, P.H.
"Electronic Applications of the Smith Chart In Waveguide, Circuit
and Component Analysis."
[20] SUPTA, K.C., GARG, R. and CHADHA, R.
"Computer Aided Design of Microwave Circuits."

[21] HA, T.T.
"Solid-State Microwave Amplifier Design."

[22] McPHERSON, N.L. (Compiler)
"Reference Data For Radio Engineers."

[23] JORDAN, E.C. and BALMAIN
"Electromagnetic Waves and Radiating Systems."
pp 29 - 75.

[24] HEWLETT PACKARD
"Measurement of Complex Impedance 1-1000 MHz."
Hewlett Packard Application Note 77-3, Hewlett Packard
Corporation, Palo Alto, USA, 1 Apr. 1967.


[26] NATIONAL SEMICONDUCTOR CORPORATION
"Voltage Regulators."
National Semiconductor Corporation, Santa Clara, CA 95051, 1980,
pp 10-132 - 10-137.
[27] JESSOP, S.R.

"VHF-UHF Manual."

[28] THE PLESSEY COMPANY LIMITED

"Professional Radio Applications."

[29] HARVEY, A.F.

"Coherent Light."

[30] HAYKEN, S.

"Communications Systems."

[31] STREMLER, F.G.

"Introduction to Communications Systems."

[32] BANCETTI, B. and TARTARA, G.

"Equalization and Quality Predictions in Digital Radio Systems."
G.T.E. Telecommunicazioni, Milan, Italy, pp 3 - 40.

[33] HEJHALL, R.

"RF Small Signal Design Using 2-Port Parameters."
Motorola Application Note, AN-215, Motorola Inc.
[34] Young, G.

Broadband Network Design for UHF Amplifiers."
Motorola Application Note, AN-704, Motorola Inc.

[35] NOLLESEN, D.L.

"UHF Transmission-Line Power-Amplifiers Designed With Smith Chart Techniques."
Motorola Application Note, AN-217, Motorola Inc.

[36] WEINERT, F.

"Scattering Parameters Speed Design of High-Frequency Transistor Circuits.

[37] CHERRY, E.M.

"An Engineering Approach to the Design of Transistor Feedback Amplifiers."

[38] CHERRY, E.M. and HOOPER, D.E.

"The Design of Wide-Band Transistor Feedback Amplifiers."

[39] KWOK, SIANG-PING

"UHF Broadband Amplifier Design."
Motorola Application Note, AN-406, Motorola Inc.

[40] MICROWAVE ASSOCIATES

"Transistor Designers Guide."
[41] BECCIOlINI, B.  
"Impedance Matching Networks Applied to R.F. Power Transistors." 
Motorola Application Note, AN-721, Motorola Inc.

[42] NORRIS, P.M.  
"using Linvill Techniques for R.F. Amplifiers." 
Motorola Application Note, AN-166, Motorola Inc.

[43] GOELL, J.E.  
"Input Amplifier for Optical PCM Receivers." 

[44] CHERRY, E.M.  
"Impedance Mismatch in Wide-Band Transistor Amplifier Design." 

[45] FRODSHAM, D.S. and BAKER, D.J.  
"Computer-Aided Design of Optical Infrared Detector Preamplifiers." (SPIE) 

[46] HULLETT, J.L. and MUOI, T.V.  
"A Feedback Receive Amplifier for Optical Transmission Systems." 

[47] OWEN, B.  
"PIN - GaAs FET Optical Receiver With a Wide Dynamic Range." 
[48] OHLAWA, N., YAMADA, J.I. and HAGIMOTO, K.

"Broadband GaAs FET Optical Front-End Circuit Up To 5.6 GHz."


"Integrated In0.53 Ga0.47As p-1-n FET Photoreceiver."


[50] WALSTON, J.A. and MILLER, J.R. Eds

"Transistor Circuit Design."


[51] YOULA, D.C.

"On Scattering Matrices Normalised to Complex Port Numbers."


[52] BODWAY, G.E.

"Two Port Power Flow Analysis Using Generalized Scattering Parameters."


[53] PEIKARI, B.

"Fundamentals of Network Analysis and Synthesis."

[54] GHAUSI, M.S.
"Principles and Design of Linear Active Circuits."

[55] NATIONAL SEMICONDUCTOR CORPORATION
"Application Notes."
National Semiconductor Corporation, Santa Clara, CA 95051, 1980,
pp

[56] HOLT, C.A.
"Electronic Circuits Digital and Analogue."

[57] MILLMAN, J.
"Microelectronics Digital and Analogue Circuits and Systems."

[58] KUHN, N.
"Simplified Signal Flow Graph Analysis."

[59] SKILLING, H.
"Electronic Network."

[60] EGAU, P.C.
"Correlation Systems in Radio Astronomy and Related Fields."
[61] FRATER, R.H.
"Accurate Wideband Multiplier-Square-Law Detector."
The review of scientific Instruments, Vol. 35, No. 7, July 1964,
pp 810 - 812.

[62] ALLEN, L.R. and FRATER, R.H.
"Wideband Multiplier Correlator."

[63] MILLMAN, J.A. (as for [57]) pp537-540

[64] NAGAL, L.W.
"A Computer Program to Simulate Semiconductor Circuits."
ERL Memo No. ERL-M520, Electronics Research Lab., University of
California, Berkley, May 1985.

[65] NATIONAL SEMICONDUCTOR CORPORATION
"Application Notes."
National Semiconductor Corporation, Santa Clara, CA 95051, 1978,
pp 29-1 - 29-9.

[66] MORRIS, R.L. and MILLER (Editors)
"Designing With TTL Integrated Circuits."

[67] MILLMAN J.A. (as for [57]) pp609-615

[68] MEAD, C. and CONWAY, L.
"Introduction to V.L.S.I. Systems."
[69] WINKEL, D. and PROSSER, F.

"The Art of Digital Design - An Introduction to Top Down Design."
pp 170 - 175.

[70] WALKER, M.K.R. and CHICKWANDA, H.S.

"Report on a Controller for a Controlled A.C. Motor Drive System."
Unpublished.

[71] FINK, D.G.

"Electronic Engineers Handbook."
7.1

7. BROADBAND CIRCUIT DESIGN AND THE TECHNIQUES OF EQUALIZATION

A number of design and matching problems were highlighted in the preceding chapter. Full amplifier gain and bandwidth, of the MRF-901 based feedback amplifier, could not be realized. Moreover two such amplifiers could not be connected to produce a gain approaching the sum of the gains of each stage. Much of the problem was attributed to the interstage coupling. KWOK's technique of broadband design [1] did not even consider the interstage coupling or the input and output coupling.

Difficulty was also encountered in obtaining a suitable bandpass response for the MRF-901 differential mixer. Its conversion efficiency is heavily dependent upon frequency. Moreover, due to its non-linear nature [2], the input and output coupling circuits are even more difficult to consider.

In the amplifier case, the interstage coupling problem could be overcome by using an equalizer to match the input and output impedances to the standard system impedance of 50Ω, presenting the required impedances to the amplifier input and output, and 50Ω to the equipment connected to the amplifiers [3],[4],[5]. If the equalizer is optimized, it would be possible to determine the power lost in interstage coupling. The amplifier design could be altered to reduce the interstage loss in a systematic way.

Unfortunately, equalization generally requires the characterization of the input and output impedances mathematically [3],[4],[5]. The feedback inductor $L_2$ is however non-linear [1]; the common amplifier linear models [4],[6],[7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19],[20],
cannot therefore be used to characterize the amplifier input and output impedances. For this reason, a method of mathematically characterizing the amplifier input and output impedances from measurement data, or indeed calculations using measurement data of the non linear inductor, must be found.

The difficulties of equalizing the conversion efficiency of the MRF-901 mixer are somewhat similar to those of the MRF-901 amplifier. Frequency conversion is itself a non-linear process therefore linearized transistor models cannot be used to characterize input or transimpedances of the mixer. Again, a technique of mathematically characterizing measurement data is required.

A large number of techniques exist for characterizing measurement data [23],[24],[25],[26],[27]. Unfortunately few produce mathematical functions that are physically meaningful; they cannot be used for manipulations based on known physical principles.

In section 7.1, the partial fraction summation method of mathematically describing measurement data are discussed. The physical interpretation of the summation method is examined with a view to characterizing the data so that it may be interpreted as a true impedance that may be synthesised, directly from the description, as a combination of physically realizable electronic components [28],[29],[30],[31],[32], [33], [34],[35],[36],[37],[38],[39],[40],[41].

Once the behaviour of the circuit is characterized, its frequency response must be equalized. Youla's technique of load equalization [3] seems to be the technique most applicable to a wide variety of load
frequency responses. Unfortunately, it is an extremely complicated method and can be used only with simple loads of one or two frequency dependent elements. Based upon Youla's technique, a new method of equalization has been produced which is much more readily applied to complicated loads.

In addition, the amplifiers and mixer will be used with digital signals recovered by the photo detectors. In order to use the full 3dB bandwidth of the amplifiers and mixer without pulse distortion, the phase of the equalized impedances must be linear [42],[43],[44]. Therefore the equalizer must produce a flat amplitude response over the desired bandwidth, with linear phase.

A full discussion of equalization, producing a flat amplitude response, and a linear phase response, is given in section 7.2 together with a new theory of equalization which may be used with very complicated load frequency responses.

The equalization techniques of section 7.2 produce mathematical expressions of impedance. In order to construct the equalizer, a method of synthesizing the impedance must be used. Using Ladder techniques, impedances may be synthesized from mathematical expressions [28],[29],[30],[31],[32],[33],[34],[35],[36],[37],[38],[39],[40],[41]. In order to produce a circuit which is useful as an equalizer, the synthesis process must produce a reactive ladder network terminated in a resistor [3],[4],[5]. This imposes specific constraints on the load synthesis process. Methods of synthesizing ladder networks, which consist of reactive ladder networks terminated in a resistor, are discussed in section 7.3.
It is evident, from the preceding discussion, that equalization of an arbitrary load involves three phases:

1. Fitting a mathematical model to measurements of the load impedance, or indeed its admittance.
2. Finding an equalizer that will produce a constant amplitude, and linear phase response, over the desired bandwidth.
3. Synthesizing the resulting equalizer impedance in the form of a ladder network terminated with a resistor.

It was intended that each of these stages be computerized, therefore the discussion of the 3 equalization phases emphasises the way in which they are implemented on a digital computer.
7.1 CURVE FITTING

The process of fitting mathematical descriptions to measurement data involves selection of coefficients of a chosen mathematical form, for example, a polynomial form [23],[24], to the measurement data; this process is also known as curve fitting.

A large amount of data is available on techniques of curve fitting and many techniques exist [23],[24],[25],[26],[27]. The technique chosen to produce a curve fit is largely dependent upon the application to which the mathematical description is to be put, and the source from which the data are derived.

In this case the source of data are measurements of impedance as a function of frequency. In general an impedance can have any complex value at a given frequency point. As the frequency is varied, the impedance will reach local maxima and minima throughout the frequency range as various resonance combinations, between circuit elements, occur [9],[45].

Such smoothly oscillating functions may be described by a variety of mathematical forms, however the common description is the polynomial which produces smooth undulations in the ordinate, as the abscissa is varied.

Unfortunately, polynomial curve fits inadequately describe measurement data obtained from real devices for a number of reasons. The most important of which is the inability of a polynomial of finite order, to describe poles of high Q, as the polynomial is composed entirely of zeros.

For this reason polynomial curve fitting was rejected as a method of mathematically describing measurement data of impedance.
The poles and zeros, which occur in measurement data, may be more accurately represented using mathematical descriptions which contain poles and zeros, rather than polynomials which contain zeros only. As the ratio of two polynomials, a rational function contains both poles and zeros, and is of the same form as the impedance of a network of passive or active components when described mathematically [30],[34],[39],[45]. The numerator polynomial represents the zeros of the rational function, and the denominator polynomial represents the poles.

Using a rational function, it is possible to produce a much simpler representation, of impedance measurement data, than the polynomial representation. In fact the polynomial representation could be generated from a rational function by using a Laurent series expansion about a pole or zero [46]. Therefore the polynomial representation may be viewed as an approximation to a rational function, about one, or a series of, points in the complex plane; the rational function representation is therefore more precise and general [46].

Unfortunately, the task of finding a rational function representation of measurement data is more difficult than producing a polynomial representation. A number of methods exist for producing rational functions [23],[24],[25],[26],[27], and as in the polynomial case, they fall into one of two categories. A rational function may be generated by, either fitting each rational function ordinate to each measurement ordinate generating a series of simultaneous equations for the coefficients or zeros of the numerator and denominator polynomials [23], or by regression methods to minimize the deviation of the ordinate estimates, from the measured ordinates, by altering the coefficients or zeros of the numerator and denominator polynomials [25],[26].
The method of divided differences is an example of the former technique [23]. The divided difference technique was abandoned because it produces rational functions that are extremely complicated, due to errors in the measurement data.

In order to eliminate the effect of these errors it was decided to apply regression methods to produce rational function curve fits.

A rational function curve fitting subroutine is available in the IBM 360 Scientific Subroutine library [25]. It uses orthogonal polynomials to simplify the curve fitting task [25],[47],[48]. Unfortunately when adapted to fit the real and imaginary parts of the data, it failed to produce a curve fit of adequate accuracy. Therefore the routine was abandoned.

The application of well known techniques, to produce a rational function, were quite unsuccessful. It was therefore decided to recast the regression problem as a minimization problem for which a variety of familiar and powerful methods of solution exist [49],[50],[51],[52],[53],[54],[55],[56],

Initially, the numerator and denominator polynomials were represented as products of their zeros. Using a variety of minimization routines, curve fits to the given measurement data were produced by minimizing the squared deviation, of the function estimates, from the measured ordinates.
Using the zero product polynomial representation a number of accurate curve fits were produced. Unfortunately when it was known that the load, from which measurement data was obtained, comprises passive elements, the rational function produced did not necessarily represent a passive network. It could not be synthesized using real components. If the rational function produced by the curve fit is to be realizable, it must be a positive real (PR) rational function, which imposes constraints on the locations of the poles and zeros [57].

A rational function, \( Z(s) \), is positive real if it obeys the following 3 constraints [57].

1. The coefficients of \( Z(s) \), when written in power series polynomial ratio form, must all be real.
2. All poles of \( Z(s) \) must occur either in the left hand half of the complex frequency plane (LHS) or on the imaginary axis, in which case the poles must be simple with positive residues.
3. When evaluated along the real frequency axis, the real part of the function must be equal to or greater than zero, that is

\[
\text{Re} \, Z(j\omega) \geq 0
\]

Conditions 1 and 3 are readily satisfied using the product method.

Condition 3 may be interpreted on the locations of the poles and zeros of the real part \( \text{Re}(Z(s)) \). The zeros of \( \text{Re}(Z(s)) \), occurring at a finite distance away from the imaginary axis, occur in quadrantal symmetry [57]. Each pole corresponds with 3 others; each located in a quadrant of the complex plane. Let \( Z_0 \) represent the location of a zero, then the zero function is \( (s-Z_0) \). \( Z_0 \) is, of course, complex. Three other zeros occur at \( Z_0^* \), \(-Z_0\) and \(-Z_0^*\). When the zero is located on the imaginary axis,
that is when \( \text{Re}(Z_0) = 0 \), quadrantal symmetry occurs irrespective of the multiplicity of that zero. Rule 2 however imposes the constraint that any zero, located on the imaginary axis, must occur with even multiplicity \([57]\) so that half the multiplicity of zeros may be identified with the zeros of the right hand side of the complex plane (RHS), and the other half with zeros of the LHS.

It was believed, erroneously, that the curve fitting procedure was unlikely to produce zeros on the imaginary axis, in which case condition 2 would never be an active constraint. using the product method, it was found that constraint 3 was readily violated, as the curve fitting procedures did generate purely imaginary axis zeros of \( \text{Re}(Z(s)) \); constraint 3 would indeed be active.

After considerable experimentation with new variations of the product method, it was found that constraint 3 could not be enforced using the product method, without fixing the order of the rational function numerator and denominator polynomials; no general restrictions on the poles and zeros, of the rational function polynomials, could be found.

However using the simple observation that the sum of a series of PR rational functions is itself positive real \([57]\), constraint 3 could be enforced if the rational function is decomposed into a sum of series of rational functions of common form. Using this observation, a summation method of generating rational functions, by regression processes, was developed. All PR constraints could be readily imposed to produce PR rational functions of arbitrary order.
7.10

The generation of rational functions, using the summation method, is covered in section 7.1.1.

7.1.1. Rational Function Curve Fitting Using Techniques of Non-Linear Minimization Subject to Non-Linear Constraints - The Summation Method.

After considerable experimentation, it was found that the PR constraints of section 7.1 could not be effectively enforced using the product expansion of the rational function. In applications where the curve fit produced need not be PR, the produce method of curve fitting is adequate.

For applications in which a positive real rational function is fitted to measurement data a new technique was sought.

A common way in which rational functions are often represented, is as a sum of partial fractions of the rational function [58]. If the pole of a rational function is real, its corresponding partial fraction is the ratio of a constant-to-a-linear, term [58].
Where \( b \) and \( a \) are real numbers.

If the pole is complex, it is associated with its complex conjugate and may be written as the ratio of a linear polynomial-to-a-quadratic polynomial \([66]\), thus,
\[
QPF(S) = \frac{es + f}{s^2 + 2cds + d^2}
\]  
(7.2)

Where \( c, d, e \) and \( f \) are real numbers.

If a rational function has a pole at infinity, the numerator polynomial is of order 1 greater than the denominator. The associated partial fraction is therefore \([66]\),
\[
I_{PF}(S) = hs
\]  
(7.3)

If the numerator polynomial order is equal to or greater than the denominator polynomial order, a constant term may be extracted from the ratio. To write the rational fraction as a sum of partial fractions, one must also include a constant term in the sum \([59]\),
\[
C_{PF}(S) = g
\]  
(7.74)

Using just the 4 partial fraction forms, (7.1), (7.2), (7.3) and (7.4), any rational function may be represented.

The quadratic term, \( QPF(S) \), (7.2), can also be used to represent two real poles and their associated residues, in which case the linear term \( LPF(S) \) is only needed if the denominator polynomial order is odd. In
general therefore, a rational function, \( z(s) \), may be represented by the summation [58],[59],

\[
z(s) = \sum_{i=1}^{q} \frac{e_i s + f_i}{s^2 + 2 c_i d_i s + d_i^2} + \frac{\alpha b}{s + a} + \beta g + \gamma hs
\]

Where the \( d_i \) are the undamped natural frequencies
the \( c_i \) are the pole damping factors
a is the location of a linear pole
b is the linear pole residue
g is a constant term
h introduces a pole at infinity
\( \alpha, \beta \) and \( \gamma \) are constant terms with values of either zero or 1
q is the number of quadratic poles in the rational function

The values of \( \alpha, \beta \) and \( \gamma \) are fixed by the numerator and denominator polynomial orders.

\[
q = \text{INT } M \quad \alpha = M - 2 \text{INT } (M/2) \quad \beta = \text{INT } n-m \quad \gamma = n-m
\]

Where \( M \) is the order of the denominator polynomial, and \( \text{INT} \) denotes the truncation-to-integer function.

If a rational function is PR, then when it is written as a summation of partial fractions, of the form given by (7.5), each term is PR [57]. Alternatively, the sum of a series of PR partial fractions is itself PR [57].
The terms in the summation of partial fractions (7.5) are limited to one of 4 forms [58],[59]. If PR constraints are imposed on the variables of each form, it is possible to guarantee that the resulting rational function is PR. The nature of the PR constraints will be entirely independent of the order of the rational function. The PR constraints for each PR form are now derived.

Consider again the quadratic term $Q_{PF_i}(s)$,

$$Q_{PF_i}(s) = \frac{e_i s + f_i}{s^2 + 2c_i d_i s + d_i^2} \quad (7.6)$$

PR condition 1 is now satisfied if the coefficients $c_i, d_i, e_i$ and $f_i$ are all real. Moreover if $c_i$ and $d_i$ are non-negative, the poles of (7.2) lies either in the LHP or on the imaginary axis. If the poles of (7.2) do lie in the imaginary axis, $c_i = 0$. The corresponding residue expression is $e_i s [58]; f_i = 0$, thus $Q_{RF_i}(s)$ becomes [66],

$$Q_{RF_i}(s) = \frac{e_i s}{s^2 + d_i^2} \quad (7.7)$$

The residue at each complex conjugate pole, of (7.7), is readily found thus [66],

$$Q_{RF_i}(s) = \begin{cases} e_i s + e_i s \\ c_i = 0 \end{cases} \frac{2}{(s+jd_i)} \frac{2}{(s-jd_i)} \quad (7.8)$$

PR rule 2 requires the residues at a purely imaginary pole to be non-negative which is guaranteed if $e_i > 0$. 
Therefore, by ensuring that $c_i$, $d_i$, $e_i$ and $f_i$ are all non-negative, PR conditions 1 and 2 are met.

If $QRF_i(s)$ is to be PR the third PR condition must be met. The real part of $QRF_i$

$$
\text{re}\left(QRF_i(j\omega)\right) = \frac{(2c_i d_i e_i - f_i)\omega^2 + d_i^2 f_i}{\omega^4 + 2d_is(c_i^2 - 1)\omega^2 + d_i^4}
$$

(7.9)

Where $\omega$ is the real frequency variable.

The denominator of (7.9) is merely the magnitude of the quadratic denominator squared, it can never be negative. Therefore $\text{Re}(QF_i(j\omega))$ goes negative at a finite frequency only if the numerator of (7.9) is negative. Thus

$$(2c_i d_i e_i - f_i)\omega^2 + d_i^2 f_i \geq 0
$$

(7.10)

The zeros of (7.84) are located at,

$$
\omega = \pm \sqrt{\frac{-d_i^2 f_i}{2c_i d_i e_i - f_i}}
$$

The sign of (7.10) cannot change if its zeros are non-real in which case,

$$
\frac{d_i^2 f_i}{2c_i d_i e_i - f_i} \geq 0
$$

(7.11)

and so the restriction of zero locations of (7.10) gives a restriction on the values of $c_i$, $d_i$, $e_i$ and $f_i$,

$$
2c_i d_i e_i - f_i \geq 0
$$

(7.12)
for non negative values of \( c_i, d_i, e_i \) and \( f_i \).

If (7.12) is obeyed PR conditions 1 and 2 ensure that (7.10) is obeyed.

PR conditions 1, 2 and 3 are obeyed by \( \text{LRF}(s) \) if \( a \) and \( b \) are both non-negative, by \( \text{IRF}(s) \) if \( h \) is non-negative and by \( \text{CRF}(s) \) if \( g \) is non-negative.

A complete summary of the PR conditions, derived, is given in Table 7.1.

### Table 7.1 Positive Real Conditions

<table>
<thead>
<tr>
<th>PARTIAL FRACTION TERM</th>
<th>PARTIAL FRACTION VARIABLES</th>
<th>POSITIVE REAL CONDITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadratic ( Q(s) )</td>
<td>( c_i, d_i, e_i, f_i )</td>
<td>( c_i &gt; 0 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( d_i &gt; 0 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( e_i &gt; 0 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( f_i &gt; 0 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>((2c_i d_i e_i - f_i) &gt; 0)</td>
</tr>
<tr>
<td>LINEAR</td>
<td>( a, b )</td>
<td>( a &gt; 0 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( b &gt; 0 )</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>( g )</td>
<td>( g &gt; 0 )</td>
</tr>
<tr>
<td>INFINITE POLE</td>
<td>( h )</td>
<td>( h &gt; 0 )</td>
</tr>
</tbody>
</table>

To obtain a curve fit, using the summation of partial fractions, one must minimize the objective function,
Which has the same form as the objective function for the product form. Substituting each partial fraction from (7.5) into (7.13) gives

\[ O_s(s, V) = \sum_{i=1}^{n_d} \left| F(s_i) - R_s(s_i, V) \right|^2 \]  

(7.13)

Unfortunately, (7.14) is prone to overflow during minimization if c_i, d_i or a, give rise to zero denominator values.

Overflow problems may be avoided by defining a new objective function \( O'_s(s, V) \) which is \( O_s(s, V) \) multiplied by the RHS denominator of (7.14).

\[ O'_s(s, V) = \sum_{i=1}^{N} W_i \left| \frac{F(s_i) - 1}{1} \right| (s_i^2 + 2c_i d_i s_i + d_i^2) \]

(7.15)

where \( W_i \) are weights altering the significance of each difference to the overall objective function value. \( N \) is the number of given data points, and \( M \) is the number of quadratic pole polynomials.

\( O'_s(s, V) \) is biased objective function and will give slightly different results to those of \( O_s(s, V) \).
The rational curve fit is now a minimization problem; a curve fit is produced by minimizing the objective function $O'_s(s,v)$.

The best fit is obtained at $O'_{s_m}(s,v) = \min O'_s(s,v)$ subject to the constraints of Table 1.

The objective function form $O'_s(s,v)$ is well defined, thus derivation of objective function gradients is straightforward.

Objective function, $O_s(s,v)$ (7.15) may be written in terms of a weighted difference parameter.

$$O_s(s,v) = \sum_{i} W_i^2 \left| d_i \right|^2 = \sum_{i} (W_i d_i) \cdot (W_i d_i)^*$$

(7.16)

where the difference term $d_i$ is,

$$d_i = F(s_i) \cdot (s_i+a) \prod_{j=1}^{M} (s_i+2c_jd_js_i+d_j^2) - \prod_{h=1}^{M} (s_i^2+2c_pd_p+d_p^2)$$

(7.17)

Let $v$ denote any variable of difference $d_i$, that is $v \in v$

$v \in (a,b,c_i,d_i,e_i,f_i,g,h)$, then

$$\frac{\delta O_s(s,v)}{\delta v} = \sum_{i=1}^{N} (w_i d_i)^* w_i \frac{\delta d_i}{\delta v} + \sum_{i=1}^{N} w_i d_i (w_i \frac{\delta d_i}{\delta v})$$

(7.18)

which is the same for all possible values of $v$. The $\delta / \delta v$ terms must now be found for each variable.
\[ \frac{\delta d_i}{\delta d_r} = F(s_i)(s_i+a)(2c_r s_i + 2d_r) \sum_{j=1}^{M} \begin{pmatrix} s_i^2 + 2c_j d_j s_i + d_j \\ j\neq r \end{pmatrix} - \sum_{k=1}^{M} (e_{k}s_i + f_k)(s_i+a)(2c_r s_i + 2d_r) \times \]

\[ \frac{\delta d_i}{\delta c_r} = F(s_i)(s_i+a)(2d_r s_i) \sum_{j=1}^{M} \begin{pmatrix} s_i^2 + 2c_j d_j s_i + d_j \\ j\neq r \end{pmatrix} - \sum_{k=1}^{M} (e_{k}s_i + f_k)(s_i+a)(2d_r s_i) \times \]

\[ \frac{\delta d_i}{\delta s_i} = -s_i(s_i+a) \sum_{h=1}^{M} \begin{pmatrix} s_i^2 + 2c_h d_h s_i + d_h \\ h\neq r \end{pmatrix} \]
\[ \delta d_i = -(s_i + a) \sum_{p=1}^{M} (s_i^2 + 2c_p d_p s_i + d_p^2) \]  
\[ \delta d_i = -(s_i + a) s_i \sum_{p=1}^{M} (s_i^2 + 2c_p d_p s_i + d_p^2) \]

The constraint gradients are trivial and require little elaboration.

For the jth quadratic term the constraints, denoted by \( c_j \), are,

\[ c_{5j-4} = d_j \geq 0 \quad \delta c_{5j-4} / \delta d_j = 1 \]  
\[ c_{5j-3} = c_j \geq 0 \quad \delta c_{5j-3} / \delta c_j = 1 \]  
\[ c_{5j-2} = f_j \geq 0 \quad \delta c_{5j-2} / \delta f_j = 1 \]  
\[ c_{5j-1} = e_j \geq 0 \quad \delta c_{5j-1} / \delta e_j = 1 \]

\[ c_{5j} = (2c_j d_j e_j - f_j) > 0 \quad \delta c_{5j} / \delta d_j = 2c_j e_j \]  
\[ \delta c_{5j} / \delta c_j = 2d_j e_j \]  
\[ \delta c_{5j} / \delta f_j = -1 \]  
\[ \delta c_{5j} / \delta e_j = 2c_j d_j \]  

For the linear term,

\[ c_a = a \geq 0 \quad \delta c_a / \delta a = 1 \]  
\[ c_b = b \geq 0 \quad \delta c_b / \delta b = 1 \]

The constant term,
In the objective function definition (7.15), a weighting parameter \( W_i \) was introduced. The weighting parameter is a constant that multiplies each difference between each given point and its estimate given by the rational function. The chosen weights alter the contribution, between a given point and its estimate, to the total objective function value.

If the weight is large for a particular difference, the difference will contribute, more than others, to the objective function value. During minimization this difference will be reduced further than the others. If the weight is small then the difference will contribute minimally to the objective function value and the difference will be reduced by only a small amount.

If it is desired that the rational function fit the points given with equally distributed absolute error, the weights are chosen to make the contribution of each difference, to the objective function value, equal.

The choice of weights is therefore dependent upon the requirement of the curve fit.

If it is desired to fit a rational function to measurement data, the nature of the measurement data must be considered. Generally measurement of impedance with frequency is done over equal frequency intervals, thus few points occur around large sharp peaks of poles and zeros. More data
points occur around the shallower turning points and at average levels as shown in Fig 7.1.

![Fig 7.1 Typical Measurement Data](image)

It is clear that there are a large number of points about an average value, and very few points in the peaks of the poles.

During objective function minimization, the fitted curve will tend to match the average value more closely than the peak values.

For this reason a two stage weighting scheme was selected. Finally pre-weight values, dependent upon the input data points only, are calculated. Secondly, the final weighting values are calculated by raising each weight to an index supplied by the user.

The pre-weight values lie between 1 and 2. They are selected by calculating the difference in magnitudes of adjacent points of impedance or admittance. Each difference is the normalized by the largest difference.

One point contributes to two differences; one on its left and the other to its right. The larger difference of the two is used to weight the point.
Once the normalised difference has been selected for each point, unity is added to it such that the pre-weights lie between 1 and 2.

The pre-weights are then raised to the user supplied index giving the actual objective function weights.

The end data points contribute to only 1 difference and are weighted using this difference only.

A set of typical input data values is shown in Fig 7.2.

Fig 7.2 Typical Input Data
Consider the 4 data points shown. The maximum magnitude difference occurs at \(|Z_3| - |Z_2|\). The prescale weight differences are

\[
d_1 = \frac{|Z_2| - |Z_1|}{|Z_3| - |Z_2|} \quad d_2 = \frac{|Z_3| - |Z_2|}{|Z_3| - |Z_2|} \quad d_3 = \frac{|Z_4| - |Z_3|}{|Z_3| - |Z_2|}
\] (7.40)

. The two end points are preweighted with value 2 to ensure a close fit at these points, so

\[
W_1 = W_4 = 2W_u
\] (7.41)

Where \(W_u\) is the user input weighting value.

\[
W_2 = (1 + \max(d_1, d_2))W_u \quad W_3 = (1 + \max(d_2, d_3))W_u
\] (7.42)

Thus, when sharp changes between the magnitudes of the input data points occur, the points on either side of the differences are heavily weighted so that the curve fits well at poles and zeros close to the real frequency axis or at sharp changes in the frequency response of impedance, since these points are generally of greatest interest.

This approach is used because sharp changes in response are represented only by a small number of points. Without weighting the process of minimization will tend to ignore these points, producing an inferior curve fit.

When actually using the data fitting techniques, discussed in the following, the user strategy is usually to start with no weighting, that is \(W_u = 0\), to avoid possible initial computer overflow. Weights are then heavily increased to obtain a good fit and enhance convergence to the correct solution.
Subjective graphical assessment is then used to examine the resulting fit and the weighting-power changed to produce the desired visual goodness of fit.

The minimization process is dramatically effected by the starting points chosen for the curve fit. At best, a good choice of starting points dramatically reduces the number of iterations required to find the minimum of the objective function. At worst, an unwise choice of starting points will lead to minimization of the objective function to a local minimum from which gradient methods will not move. If a linear term occurs in the objective function, it is possible that the minimization will tend to put this in an incorrect position depending on its starting value and whether or not the objective function contains local minima.

If the objective function does contain local minima, then the final solution can be effected by the choice of linear term start location; in particular.

In some cases, a good curve fit will be obtained only if the starting point is suitably chosen and the user may need to try a number of different starting points.

Selection of starting points is best done manually by examining the general nature of the input data plots. Each turning point of the real, imaginary or magnitude plots represents a complex pole, if it is a maximum, or a complex zero if it is a minimum. These observations firstly give some clue as to the orders of numerator and denominator polynomials, of the required rational function.
The locations of the turning points are used to fix the start locations of poles and zeros of the objective function.

In summation form it is difficult to find a clear relationship between residues, and the locations of zeros of the rational function, thus the $f_i$ terms are set to zero, terms $b$, $g$ and $h$ are set to unity.

The locations of the rational function poles is more clear cut, they are the same whether the rational function is in sum or product form. Since the amount of damping at each pole is largely unknown initially and it is difficult to estimate, it is assumed that no damping occurs at any pole. The user then sets the values of $c_i$ and $d_i$ to values close to the pole location, as if the poles lay on the real frequency axis.

Minimization of the objective function is then undertaken. If the result, at the minimum, is not the result desired, then a different set of starting values is chosen. If a linear term is present, then it is swapped with one of the earlier quadratic choices and minimization undertaken again.

If the solution is still unsatisfactory another choice of linear pole start point is made.

Fig 7.3 illustrates the choices of starting point made and the subsequent migration of poles towards the actual solution; away from the real frequency axis.
If solutions are unsatisfactory for all start choices the rational function order should be raised and minimization again attempted, if necessary from several different starting point combinations.

The details of objective function, starting points, weighting and constraints have been discussed, however the actual minimization techniques employed, to minimize the objective function, have not.

Experience has shown that some techniques of optimization are much better suited to the curve fitting problem than others.

Therefore the optimization methods used will be discussed together with their incorporation into a general rational-function curve fitting program.

Generally, gradient techniques are the most rapidly converging optimization techniques [27],[50],[51],[52],[53],[54],[55],[56], thus three
such optimization techniques were tried in order to find the most rapidly converging curve fitting method.

The first technique used the BFGS quasi-Newton optimization technique [53]. The commercially written optimization technique was obtained from the Harwell scientific subroutine library [53].

During the early stages of optimization the BFGS method converges quite rapidly, typically one or two orders of magnitude per iteration, however as the magnitude of the objective function falls the rate of convergence tends to stagnate; the gradient becomes small but are not insignificant. Thousands of iterations are required to reach the desired minimum, even with good starting points.

Moreover, the PR constraint cannot be imposed directly therefore a simple penalty function procedure, based on BFGS, was attempted, however the constraints could not be satisfied without abnormally high penalties [50].

A more suitable technique for minimizing the objective functions is the method of Marguardt [54]. The Marguardt method involves the differences between individual points, so the cost function becomes a cost function matrix, and the summations over \( i \) in the objective and gradient functions, (7.15), (7.17), (7.18), (7.19), (7.20), (7.21), (7.22), (7.23), (7.24), (7.25), (7.26), (7.27) are removed [54].

Using the unconstrained Marguardt [54] method, convergence is very rapid indeed. In the first 10 iterations it is possible to obtain reductions, in the objective function value, of 10 orders of magnitude.
Convergence seldom stagnates however, the matrices can become ill conditioned, especially when using simple penalties for the constraints [50]. Simple penalty functions were again unsuitable, as the final results did not obey the constraints with sufficiently small error.

It was therefore decided to try a Lagrange Multiplier Penalty function Method [55], again from the Harwell Library. Convergence of this method is generally slower than the Marguardt Method, however constraints are always obeyed once the minimum is reached. As in the BFGS [53] case this approach is prone to stagnation.

By using the Marguardt method [54] to produce starting points for the Penalty function Multiplier [55] method one can exploit the rapid convergence of the Marguardt method to obtain an unconstrained solution followed by the accurate application of PR constraints via the VF02AD. The Marguardt method produces a good unconstrained curve fit moving the poles and zeros, of the rational function, into approximately the correct places. Application of the multiplier penalty function [55] moves the poles and zeros slightly until it has satisfied the positive real constraints and typically requires only 20 iterations. The examples of the next chapter will demonstrate the effectiveness of this approach.

Having selected the most appropriate method of curve fitting using the software available from the Harwell library [53],[54],[55],[56], a general curve fitting program structure may be defined. It must facilitate.

1. Initial pole & zero estimation and input
2. Input of weighting information
3. Input of rational function numerator and denominator order
4. Selection of appropriate optimization routine
5. Input of curve fit error termination values
6. Iteration number restrictions
7. Possible internal routine scaling modes
8. Examination of convergence behaviour
9. Listing of input points and corresponding estimates
10. Plots of estimation against input points
11. Repetition procedures

The interactive flow of routine DRFA is somewhat crude, the program cycles through a series of steps which may be bypassed if the old information, from an earlier result, is to be retained. However it is effective and requires a minimum of user interaction. It is proposed to alter the Program from the sequential system of execution to a menu driven system; where the operator may select directly the mode of operation required thereby increasing the flexibility of the routine.

The interactive operator procedure for using DRFA follows a fixed pattern.

1. The operator enters the numerator and denominator polynomial function orders.
2. The weighting power is introduced.
3. Initial estimates of the pole location are entered.
4. The calculation error is entered.
5. The initial minimization routine is selected. (Initially the Marguardt method [54] is selected.)
6. The internal scaling mode is selected.
7. A permitted maximum number of iterations entered. (Usual choice is 100 which should be sufficient for most problems.)

8. If convergence is good the operator continues, if not steps 1 to 7 are repeated with different orders (1) or weights (2).

9. The estimates of the curve fit and the original data are plotted and printed. If the curve fit looks reasonably close the operator continues, otherwise he repeats steps 1 to 9.

10. The operator retains the orders of numerator and denominator

11. Generally he also retains the same weights

12. The calculation error is also retained

13. The initial pole estimates are bypassed and old values for all coefficients retained.

14. The Lagrange multiplier penalty function routine [55] is selected.

15. A maximum iteration number is selected typically 100 which is seldom reached.

16. If the final square error is smaller than the square root of the Marguardt error [54] then the operator terminates the program which prints and plots the actual input and estimated point values scaled and unscaled.

17. The rational function is written out as the ratio of two power series.

18. The program is terminated.

This approach, while crude, is effective for producing rational function curve fits to a variety of input data.
7.2 THE THEORY OF EQUALIZATION

In order to overcome the bandwidth limitations of the amplifier of section 6.6 and the mixer of section 6.7 a method of broadband equalization was sought. Methods often used to produce broadband active devices are piecemeal, in the worst cases, a designer adjusts components in a basic circuit configuration, simulated on a computer, iteratively until the circuit yields the desired bandwidth characteristics \([1],[4],[6],[7],[8],[9],[10],[11],[12],[13],[14],[15],[21],[22],[60],[61],[62],[63],[64],[65],[66],[67],[68],[69],[70],[71],[72]\). The phase response of the circuit is often ignored completely \([1],[4],[6],[7],[8],[9],[10],[11],[12],[13],[14],[15],[21],[22],[60],[61],[62],[63],[64],[65],[66],[67],[68],[69],[70],[71],[72]\).

In circuits where uncoded voice signals are carried the phase characteristics are unimportant, however in applications where digital signals must be amplified and recovered, the phase of these circuits cannot be ignored. In fact it dictates the bandwidth over which the circuit can be used \([42],[43],[44]\). If the phase response is significantly non-linear, pulses will be smeared \([42],[43],[44]\). The recovery of digital information in such cases becomes more difficult as the pulses become less distinct. For this reason, the phase response of analogue circuits used prior to digit recovery, in digital systems, must be linear. It cannot be ignored.

To equalize the amplitude and phase of a device in a manageable way, sound theoretically based equalization techniques are needed.

In his original paper on equalization, D.C. Youla \([3]\) presented a theoretical technique of equalization which can be applied to any PR load.
The load completely does not alter the overall technique of application, it merely alters the complexity of calculation. The Youla \cite{3} technique can equalize a load to any prescribed response that may be written in PR rational function form.

The Youla \cite{3} technique itself involves preconnecting a two port network, to the circuit to be equalized, such that the power transfer from a resistive voltage source to the resistive part of the load, complies with a chosen rational function response. The preconnected two-port network is known as an equalizer.

In section 7.2.1, the background theory of Youla equalization theory \cite{3} is introduced. An understanding of the background theory is essential to understand the LDE equalization technique discussed in sections 7.2.4 to 7.2.7.

Using the background theory, the Youla technique of equalization \cite{3} is explained in detail. The explanation is intended to be easier to understand than Youla's original explanation \cite{3}.

Once Youla's equalization technique is understood, some background theory of LDE's is needed, thus section 7.2.3 briefly covers the theory of LDE's with 2, or more, unknown polynomials.

Using the background theory introduced in section 7.2.1 and the LDE theory of section 7.2.3 the equalization problem is recast as a LDE problem. By solving the resulting LDE, subject to a set of constraints derived in section 7.2.1, an equalizer is found by expressing the equalizer
back impedance in terms of the LDE solutions, as discussed in section 7.2.5.

Imaginary axis zeros of transmission (ZOT) must however be considered separately, when they are present. It is not possible to guarantee that an equalizer can be built unless further solution constraints are introduced. Youla divides these ZOT into 3 separate classes [3] and derives constraints for each class. In section 7.2.6, each ZOT class and its effects on the LDE equalization process is discussed in detail. If necessary additional solution constraints are derived.

Due to the nature of the LDE of equalization, trivial solutions can be found which will not produce a realizable equalizer. Using, a bounding polynomial concept, a new LDE is derived in section 7.2.7, which cannot be satisfied by the trivial solution.

In section 7.2.8 a new techniques of synthetic polynomial division is discussed. Using the matrix pseudo-inverse, it can improve the accuracy of polynomial division which is recast as an LDE problem. Using the matrix pseudo-inverse the LDE's of equalization can also be readily solved using a digital computer.

Using the techniques of LDE solution and synthetic division discussed, an equalization program was written. It is discussed in section 7.2.9. To avoid the need for zero searches of polynomials the Routh array was used extensively; it is a very powerful tool and its use is also covered in section 7.2.9.

The equalizer itself is realized by synthesizing the back impedance expression derived in section 7.2.5. Unfortunately many circuits produced
by synthesis processes are unsuitable for application as an equalizer. In section 7.2.10 the general requirements for the synthesis of an equalizer circuit are discussed.

7.2.1 Background Theory of Youla Equalization

Consider a frequency dependent load \( Z_L(s) \) which may be separated into even and odd rational functions in complex frequency \( s \). That is,

\[
Z_L(s) = r_L(s) + x_L(s) \tag{7.43}
\]

where \( Z_L(s) \) is a PR rational function and \( r_L(s), x_L(s) \) are respective even and parts.

The power dissipation in \( r_L(s) \) is of most interest so a load transmission function is now defined [3].

\[
T(s) = \frac{r_L(s)}{Z_L(s)} \tag{7.44}
\]

If the equalizer network shown in Fig 7.4 has a transfer function of \( E(s) \), the transfer function from resistive source to resistive load becomes,

\[
G(s) = E(s)T(s) \tag{7.45}
\]

where \( G(s) \) is the transducer gain of the equalized load.
Referring now to the two port network shown in Fig 7.56; it is clear that the normalized scattering parameters $[3],[4],[5],[6],[9],[13],[14],[15]$ (Fig 7.56), for resistive source and complex load impedances, can be determined from the input and output conditions as seen by the network.
When the equalizer is lossless, the scattering parameter magnitudes are related by [3],

\[ S_{21}(s)S_{21}(-s) = 1 - S_{22}(s)S_{22}(-s) \tag{7.48} \]

Moreover, the transducer gain is related to the normalized scattering parameters by [3],

\[ G(s)G(-s) = S_{21}(s)S_{21}(-s) \tag{7.49} \]

Thus, given a load that must be equalised so that the resulting equaliser and load transmission has the form of \( G(s) \), the output scattering parameter \( S_{22}(s) \) may be found by rewriting (7.48),
Using the definition of the normalised equaliser output scattering parameter \([3] \), \(S_{22}(s)\), which must be normalised to the load impedance, \(Z_{L}(s)\), the expression for the output scattering parameter of an equaliser is obtained.

\[
S_{22}(s) = \frac{Z_{2}(s) - Z_{L}(-s)}{Z_{2}(s) + Z_{L}(s)} \tag{7.51}
\]

where \(Z_{2}(s)\) is the output impedance of the equaliser with the resistive input voltage source connected.

Let \(Z_{2}(s)\) and \(Z_{L}(s)\) be represented by a ratio of polynomials,

\[
Z_{2}(s) = \frac{N_{2}(s)}{D_{2}(s)} \quad Z_{L}(s) = \frac{C_{n}N(s)}{D(s)} \tag{7.52}
\]

where \(C_{n}\) is chosen such that the coefficient of the highest power of \(N(s)\) is unity.

After some minor rearrangement, \(S_{22}(s)\) (7.51) becomes,

\[
S_{22}(s) = \frac{(N_{2}(s)D(-s) - N(-s)D_{2}(s))}{(N_{2}(s)D(s) + N(s)D_{2}(s))} \cdot \frac{D(s)}{D(-s)} \tag{7.53}
\]

Since \(Z_{2}(s)\) and \(Z_{L}(s)\) must be positive real, for a passive equaliser and load, the only RHP poles of \(S_{22}(S)\) are the zeros of \(D(-s)\).

Now introducing a regular all pass function \(b(s)\) of the form,

\[
b(s) = \prod_{j=1}^{m} \frac{s - d_{j}}{s + d_{j}} \tag{7.54}
\]
where the $d_j$ are the poles of $Z_l(-s)$, which are the zeros of $D(-s)$ to the right of the imaginary axis.

The function,

$$S(s) = S_22(s)b(s) \quad (7.55)$$

is analytic in the RHP, and is thus a bounded real scattering coefficient [3],[5]. Note that $b(s)b(-s)=1$, and hence in view of (7.50)

$$S(s)S(-s) = S_{22}(s)S_{22}(-s) = 1 - G(s)G(-s) \quad (7.56)$$

In general,

$$S(s) = a(s)S_0(s) \quad (7.57)$$

where $S_0(s)S_0(-s) = 1 - G(s)G(-s)$.

Thus to produce an equaliser with transducer gain $G(s)$, the scattering parameter $S(s)$ is studied. Rewriting $S(s)$ in terms of the impedances, $Z_2(s)$ and $Z_l(s)$, $S(s)$ becomes,

$$S(s) = b(s) \frac{Z_2(s) - Z_l(-s)}{Z_2(s) + Z_l(s)} \quad (7.58)$$

Synthetically dividing the denominator of (7.58) into the numerator and subtracting $b(s)$ from both sides equally gives,

$$b(s) - S(s) = 2r_l(s)b(s) \frac{Z_2(s) - Z_l(s)}{Z_2(s) + Z_l(s)} \quad (7.59)$$

which is the basis of the Youla observations [3]; the RHP zeros of the RHS of (7.59) are the same as those of the load transmission function $T(s)$,

$$T(s) = \frac{r_l(s)}{Z_l(s)}$$
The zeros of $T(s)$ are known as zeros of transmission (ZOT).

Using the Youla observation, an equalizer is produced by selecting the scattering parameter $S(s)$ to produce a PR equalizer output impedance $Z_2(s)$, and a transducer gain of $G(s)$.

Generally $G(s)$ has the form [3],[4],[5],

$$G(s) = \frac{kX(s)}{Y(s)} \quad (7.60)$$

where $X(s)$ and $Y(s)$ are polynomials set by the response fo the selected filter, and $k$ is the filter transducer gain.

and $S_0(s) = K(s)/J(s)$ is derived using (7.49) and spectrally factoring $S_0(s)S_0(-s)$ [64].

Let $a(s)$ is a regular arbitrary all pass function of the form,

$$a(s) = \prod_{i=1}^{n} \frac{s - c_i}{s + c_i} \quad (7.61)$$

where $c_i$ are the zeros of $a(s)$ and $\text{re}(c_i) > 0$, but are otherwise of arbitrary value.

Since

$$a(s)a(-s) = 1 \quad (7.62)$$

the most general scattering parameter satisfying (7.50) is,

$$S(s) = a(s)S_{22}(s) \quad (7.63)$$

thus $S(s)$ is related to $S_{22}(s)$ by,
\[ S(s)S(-s) = S_{22}(s)S_{22}(-s) \]  \hspace{1cm} (7.64)

S(s) therefore, represents the variables of (7.59).

In order to produce an equalizer which is realizable, S(s) is selected to solve (7.59) and produce a PR equalizer output impedance \( Z_2(s) \) subject to the restrictions imposed by (7.60) and (7.61). These requirements produce restrictions on the solutions that one can obtain for a given load [3],[4],[5].

In the following section the restrictions derived by Youla [3] are discussed together with the methods of solution proposed by him. Inadequacies in his method of solution are indicated, and a new approach based upon the theory of the LDE is derived.

7.2.2 Youla Equalization

In the preceding section the theory underlying the Youla method of equalization [3] was discussed. Key observations made by Youla [3] were emphasised.

In this section, the methods of finding S(s) to satisfy the requirements of the preceding section are discussed. Details of the Youla restrictions [3] are examined, as an understanding of these is important for derivation of restrictions required for the LDE technique of equalization, discussed in section 7.2.6.

Using (7.59.), one may express \( Z_2(s) \) in terms of the load impedance \( Z_L(s) \) and the equalizer transimpedance.
\[ Z_2(s) = \frac{2r_i(s)b(s)}{b(s) - S(s)} - Z_L(s) \quad (7.65) \]

Where the transimpedance is \( Z(s) \)

\[ Z(s) = \frac{2r_i(s)b(s)}{b(s) - S(s)} \quad (7.66) \]

If \( Z_2(s) \) is to be realized using reactive components terminated in a resistor it must be PR.

It was assumed earlier that, \( Z_L(s) \) is a PR impedance, therefore (7.65) can only be NPR due to the choice of \( S(s) \).

\( Z_2(s) \) is PR of the RHS of (7.65) obeys the 3 PR conditions given in section 7.1.

The coefficients of power series expansions of the component polynomials \( r_i(s), b(s), S(s) \) and \( Z_L(s) \) are all real, therefore condition 1 is satisfied.

If \( Z(s) \) is expressed in terms of its numerator and denominator polynomials \( N_0(s) \) and \( D_0(s) \) respectively, and \( Z(s) \) is expressed as a ratio of its numerator and denominator polynomials, \( Z_2(s) \) becomes,

\[ Z_2(s) = \frac{N_0(s)}{D_0(s)} - \frac{N(s)}{D(s)} \quad (7.67) \]

The poles of \( Z_2(s) \) are the zeros of the product \( D_0(s)D(s) \), \( Z_L(s) \) is, however, positive real therefore any RHP poles of \( Z_2(s) \) are due to those of \( Z(s) \). Therefore violations of PR condition 2 occur if \( Z(s) \) has RHP poles. Condition 2 may also be violated if \( Z(s) \) has an imaginary axis pole coincident with one at \( Z_L(s) \). If the pole has unit multiplicity in
Z(s) then Z_2(s) is PR if the residue of Z_2(s), at the pole, is zero. This case is discussed more fully in the following paragraphs.

By forming the real part of (7.65) condition 3 may be tested,

\[ \text{re}(Z_2(s)) = r_2(s) = r_L(s)1-S(s)S(-s) = \frac{r_L(s)(1-S(s)S(-s))}{(b(s)-S(s))(b(-s)-S(-s))} \]  

(7.68)

using the identity \( b(s)b(-s) = 1 \).

At \( s=j\omega \) the real part is

\[ r_2(j\omega) = r_L(j\omega)(1-S(j\omega)S^*(j\omega)) = \frac{r_L(j\omega)(1-S(j\omega)S^*(j\omega))}{(b(j\omega)-S(j\omega))(b(j\omega)-S(j\omega))^*} \]  

(7.69)

Due to the PR condition of \( Z_L(s) \), \( r_L(j\omega) \geq 0 \). Moreover the equalizer is passive therefore the general output scattering parameter has a magnitude less than unity, that is,

\[ s(j\omega)s^*(j\omega) \leq 1 \]

The denominator is the magnitude of the difference \( b(j\omega)-s(j\omega) \). It is always positive.

From the above analysis therefore, it is necessary, with the exception of the case where \( Z(s) \) and \( Z_L(s) \) have common imaginary axis poles, only to prove that \( Z(s) \) is PR to ensure that \( Z_2(s) \) is PR [3], which is entirely equivalent to prove that \( Y(s) \) is PR where [3],

\[ Y(s) = \frac{1}{Z(s)} \]

In this analysis it is more convenient to work with \( Y(s) \) rather than \( Z(s) \).
Y(s) obeys PR condition 1.

It is also known that $Z_2(j\omega)$ obeys condition 3 therefore,

$\text{Re}(Z_2(j\omega)) \implies \text{Re}(Z(j\omega)) \geq \text{Re}(Z_L(j\omega))$

and as $Z_L(s)$ is PR, $\text{Re}(Z_L(j\omega)) \geq 0$ therefore,

$\text{Re}(Z(j\omega)) \geq 0$

and so $Z(s)$ obeys PR condition 3, in which case $Y(s)$ also obeys PR condition 3.

It is necessary therefore, to ensure that $Y(s)$ obeys PR condition 2 to yield a PR $Z_2(s)$.

The zeros of transmission were given by the zeros of

$$T(s) = \frac{r_L(s)}{Z_L(s)}$$

which may belong to one of 4 classes [3];

Class I ZOT are those RHP ZOT confined to the RHP only, they must not occur on the imaginary axis [3].

Class II ZOT are those ZOT which are zeros of both $T(s)$ and $Z_L(s)$, they occur on the imaginary axis [3].
Class III ZOT are ZOT which occur on the imaginary axis and are neither poles nor zeros of \( Z(\sigma) \).

Class IV ZOT are imaginary axis zeros of \( T(\sigma) \) which are also poles of \( Z_L(\sigma) \).

Restrictions on \( S(\sigma) \) can be found for each ZOT class to ensure that \( Z_2(\sigma) \) is PR and (7.61) is obeyed.

The zeros of transmission are also poles of \( Y(\sigma) \), where

\[
Y(\sigma) = \frac{b(\sigma) - S(\sigma)}{2r_L(\sigma)b(\sigma)} \tag{7.70}
\]

The RHP and even multiplicity zeros, of \( r_L(\sigma) \), cause \( Y(\sigma) \) to violate PR condition 2. \( S(\sigma) \) must therefore be chosen to cancel enough of these poles to ensure that \( Y(\sigma) \) obeys PR condition 2 [3].

ZOT of class I are poles of \( Y(\sigma) \), multiplies as ZOT and Poles of \( Y(\sigma) \) are the same. The difference \( b(\sigma) - S(\sigma) \) must therefore cancel all RHP ZOT, removing the RHP poles of \( Y(\sigma) \). Let \( m_{ZOT} \) be the number of RHP poles of class I. At these ZOT, the RHS of (7.133) is zero, therefore \( m_{ZOT} \) terms of the Laurent series expansion of \( b(\sigma) \) must equal \( m_{ZOT} \) terms of the Laurent series expansion of \( S(\sigma) \) [3]. That is [3],

\[
B_i = S_i \quad i = (0, 1 \ldots m_{ZOT} -1)
\]

Where \( B_i \) and \( S_i \) are the \( i \)th terms of the Laurent series of \( b(\sigma) \) and \( S(\sigma) \), about the ZOT respectively.
ZOT of class II introduce imaginary axis poles of even multiplicity, into \( Y(s) \), therefore it violates PR condition 2. If the multiplicity of these ZOT is \( m_{ZOT} - 1 \), then \( S(s) \) must be chosen to cancel at least \( m_{ZOT} - 1 \) of those ZOT. If only \( m_{ZOT} - 1 \) are cancelled, the residue of \( Y(s) \) at the remaining pole must be positive. Naturally, if \( m_{ZOT} \) of the class II ZOT are cancelled, \( Y(s) \) will cease to violate PR condition 3, due to the class II ZOT, therefore the residue of \( Y(s) \) at the ZOT will be zero. The multiplicity of the ZOT, in \( T(s) \), is actually 1 less than that in \( r_L(s) \) due to cancellation with \( N(s) \), therefore, at least \( m_{ZOT} - 1 \) terms of the Laurent series of \( b(s) \) must equal \( m_{ZOT} - 1 \) terms of \( S(s) \).

Thus,\[ B_i = S_i \quad (i = 0, 1, 2 \ldots m_{ZOT} - 1) \]
The residue of \( Y(s) \) at the ZOT is given by [3], \[ \frac{B_i}{m_{ZOT}} - \frac{S_i}{m_{ZOT} + 1} \geq 0 \] (7.71)

where \( F(s) = 2r_L(s)b(s) \).

The residue must be either larger than zero or equal to zero in which case the ZOT is completely cancelled from \( Y(s) \).

Class III imaginary axis zeros are similar to those of class II, they differ only in the ZOT order obtained from \( T(s) \). Their multiplicity, as a ZOT, is the same as that of their pole multiplicity in \( Y(s) \). The order of the ZOT is \( m_{ZOT} \), at least \( m_{ZOT} - 1 \) terms of the Laurent series expansion of \( S(s) \) must equal \( m_{ZOT} - 1 \) terms of the Laurent series expansion of \( b(s) \) [4],
\[ B_i = S_i \quad (i = 0, 1, 2 \ldots m_{\text{ZOT}} - 2) \]

The residue at the remaining ZOT must be either positive or zero. In the latter case, the ZOT is completely cancelled from \( Y(s) \) and is not therefore a pole of \( Y(s) \) \([3]\).

The residue of a class III ZOT is \([3]\),

\[
\frac{B_{m_{\text{ZOT}} - 1} - S_{m_{\text{ZOT}} - 1}}{F_{m_{\text{ZOT}}}} > 0 \quad (7.72)
\]

ZOT of Class IV are those ZOT discussed earlier, in which case, the positive realness of \( Y(s) \) is not sufficient to guarantee that \( Z_2(s) \) is PR.

If a ZOT of class IV has multiplicity \( m_{\text{ZOT}} \) as a ZOT, then its multiplicity as a zero of \( F(s) \) is \( m - 1 \). It's multiplicity as a pole of \( Y(s) \), or a zero of \( Z(s) \), is \( m_{\text{ZOT}} \). At the pole of \( Z_\perp(s) \) which is the ZOT, the residue of \( Z(s) \) is a non-zero positive value. If \( S(s) \) is chosen to cancel fewer than \( m_{\text{ZOT}} \), the residue of \( Z(s) \) at the pole of \( Z_\perp(s) \) will be zero and \( Z_2(s) \) will violate PR condition 2. Therefore \( S(s) \) must be chosen to introduce a pole at the ZOT into \( Z(s) \), therefore \( m_{\text{ZOT}} \) terms of the Laurent series expansion of \( S(s) \) must equal \( m_{\text{ZOT}} \) terms of the Laurent series expansion of \( b(s) \) \([3]\) thus,

\[ B_i = S_i \quad (i = 0, 1, 2 \ldots m_{\text{ZOT}} - 1) \]

at the remaining class IV pole of \( Z(s) \) the residue is given by \([4]\),

\[
\frac{F_{m_{\text{ZOT}} - 1}}{B_{m_{\text{ZOT}}} - S_{m_{\text{ZOT}}}} \quad (7.73)
\]
and the residue of \( Z_L(s) \) at the pole is \( a_{-1} \), therefore the overall residue of \( Z_2(s) \) at the pole is \([3]\),

\[
\frac{F_{m_{ZOT}}}{B_{m_{ZOT}}} - \frac{1}{S_{m_{ZOT}}} - a_{-1} \geq 0 \tag{7.74}
\]

thus \( S(s) \) must also be selected to ensure that (7.74) is obeyed hence \([3]\)

\[
\frac{F_{m_{ZOT}}}{B_{m_{ZOT}}} - \frac{1}{S_{m_{ZOT}}} \geq a_{-1} \tag{7.75}
\]

The Youla equalization procedure is based upon the 4 classes of ZOT, and is broken down into the following 3 steps \([3]\).

1. The load transmission function, \( T(s) \), is formed. Its ZOT are found and separated appropriately into the 4 classes, given above.

2. Using Laurent series expansions, \( S(s) \) is selected to satisfy the requirements of the particular ZOT class for all ZOT in the RHP or on the imaginary axis.

3. The backend impedance \( Z_2(s) \) is found using (7.65). An equalizer is produced by synthesizing \( Z_2(s) \).

Unless \( S(s) \) is of closed form \([4]\), it can be difficult to find a suitable \( S(s) \) to solve (7.59). However tables of \( S(s) \) for various values of transducer DC gain "k" can be used to successfully produce an equalizer.

Recall the general form of the desired transducer gain \( G(s) \) (7.60)
\[ G(s) = k \frac{X(s)}{Y(s)} \]

\( S(s) \) is related to \( G(s) \) by (7.130),

\[ S(s)S(-s) = 1 - G(s)G(-s) \]

Substituting (7.60) into (7.56) yields an equation relating the DC transducer gain, \( k \), to the scattering parameter \( S(s) \),

\[ S(s)S(-s) = Y(s) - k^2 X(s) \frac{Y(s)}{Y(s)} \]

(7.76)

\( S(s) \) is found by spectrally factorizing (7.76) [73]. In the equalization process, a suitable value of DC gain, \( k \), must be found to produce a PR back impedance \( Z_2(s) \). Moreover, if the all pass function \( a(s) \) is included, \( k \) must be selected such that the poles of \( a(s) \) lie in the LHP, and \( S(s) = a(s)S_0(s) \) satisfies (7.56). If \( k \) can be expressed explicitly in \( S(s) \), it may be included directly in the elements of the Laurent series expansions. For example, \( G(s) \) may represent a Butterworth filter response where [5],

\[ G(s)G(-s) = \frac{k^2}{1 + (-1)^n s^{2n}} \]

(7.77)

Where \( n \) is the filter-polynomial order.

The corresponding value of \( S(s) \) is

\[ S(s) = \frac{1}{(1-k^2)^2} \frac{1 + \Delta(s)(1-k^2)^{2n}}{1 + \Delta_n(s)} \]

(7.78)

Where \( \Delta_n(s) \Delta_n(-s) = 1 + (-1)^n s^{2n} \).
For a given filter order, the numerator coefficients of \( S(s) \) may be stated as function of unknown gain \( k \).\[3\],\[4\],\[5\].

Using the Laurent series expansions, one may solve for \( k \) directly, together with the coefficients of the all pass function \( a(s) \), as shown in the first example of chapter 8 \[3\],\[4\],\[5\].

If a linear phase response is desired, the Bessel polynomial filter function can be used \[74\],\[75\],\[76\],\[77\],\[78\],\[79\],\[80\],\[81\],\[82\],\[83\]. Unfortunately it is not possible to express the coefficients of \( S(s) \), corresponding to the Bessel filter function in terms of the unknown DC gain \( k \).\[74\],\[75\],\[76\],\[77\],\[78\],\[79\],\[80\],\[81\],\[82\],\[83\]. In fact the only way in which \( S(s) \), corresponding to a particular value of \( k \), may be found is to spectrally factorize \( S(s)S(-s) \) iteratively \[73\], each time a new value of \( k \) is selected.

By generating tables of \( S(s) \), for various orders of Bessel polynomial and gain \( k \), a Bessel equalizer may be found using the Laurent series technique. Using the tables of \( S(s) \) for the Bessel filter, Laurent series expansions of \( S(s) \) are formed \[3\]. (7.59) is then solved for \( a(s) \) \[3\]. If the poles of \( a(s) \) occur in the RHP, or on the imaginary axis, a new value, \( k \), is selected and the process repeated until the poles of \( a(s) \) are confined to the LHP. The corresponding value of DC gain, \( k \), may be lower than its optimum value, however if it is within 3dB of the optimum value, the difference can be ignored.

In order to generate the Bessel equalizer scattering tables, a computer program was written to spectrally factorize (7.76) for selected
values of \( k \). The program itself was based on the following iterative procedure.

Let \( L(s) \) represent the squared magnitude of a polynomial,

\[
\alpha(s) = (-1)^n \alpha_{2n}s^{2n} + (-1)^{n-1} \alpha_{2n-2}s^{2n-2} + (-1)^{n-2} \alpha_{2n-4}s^{2n-4} + \ldots + \alpha_4 s^4 + 2s^2 + \alpha_0
\]

(7.79)

It is desired to find the spectral factors \( \beta(s) \beta(-s) \) for \( \alpha(s) \)

\[
\beta(s)\beta(-s) = \alpha(s)
\]

(7.80)

Expanding \( \beta(s) \) and \( \beta(-s) \) in power series form,

\[
(\beta_n s^n + \beta_{n-1}s^{n-1} + \beta_{n-2}s^{n-2} + \ldots + \beta_1 s + \beta_0)
\]

\[
\times ((-1)^n \beta_n s^n + (-1)^{n-1} \beta_{n-1}s^{n-1} + (-1)^{n-2} \beta_{n-2}s^{n-2} + \ldots + \beta_1 s + \beta_0)
\]

Multiplying the two bracketed terms,

\[
\beta(s)\beta(-s) = (-1)^n \beta_n^2 s^{2n} + (-1)^n \beta_n \beta_{n-2} + (-1)^{n-1} \beta_{n-2}^2 s^{2n-2} + \ldots + (2 \beta_0 \beta_2 - \beta_1^2) s^2 + \beta_0^2
\]

(7.81)

The \( i \)th term of (7.81) may be written as,

\[
\alpha_{2i} = 2 \sum_{j=1}^{\min(n-i,i)} (-1)^i \beta_{i+j} \beta_{i-j} + (-1)^i \beta_i^2
\]

(7.82)

Each coefficient of \( s \) in (7.81) must equal the coefficient of the corresponding power of "s" in (7.79).

\[
(-1)^i \alpha_{2i} = 2 \sum_{j=1}^{\min(n-i,i)} (-1)^i \beta_{i+j} \beta_{i-j} + (-1)^i \beta_i
\]

(7.83)
Equation (7.83) may be rearranged to give each coefficient of $\beta(s)$ in terms of past values of these coefficients and the coefficients of $\alpha'(s)$,

$$
\beta_i = \left[ \alpha_{2i} - 2 \sum_{j=1}^{\min(n-i,i)} (-1)^j \beta_{i+j} \beta_{i-j} \right]^{\frac{1}{2}}
$$  \hspace{1cm} (7.84)

Equation (7.84) defines an iteration scheme. For each value of $i$ from 0 to $n$, a new value of each coefficient $\beta_i(k+1)$ is generated from the previous coefficient values $\beta_i(k)$. As each new value is generated, it replaces the previous value and contributes to the generation of the next $i$th coefficient.

The iteration procedure is continued until the differences between the new coefficients of $\beta(s)$, generated from the previous values of $\beta(s)$, differ by less than a tolerance figure. In which case the spectral factor is found. The iteration scheme is started by supplying initial estimates of the coefficient of $\beta(s)$. These are usually the square root of the corresponding 2ith coefficients of $\alpha(s)$.

The iteration termination criterion is the difference,

$$
\sum_{i=1}^{n} \left| \beta_i^{(k+1)} - \beta_i^{(k)} \right|^2 \leq \delta_{SF}
$$  \hspace{1cm} (7.85)

Where $k$ is the iteration number and $\delta_{SF}$ is the iteration tolerance parameter.

Using the spectral factorization algorithm method, the set of Bessel equalizer tables, given in Table 7.2, was produced.
### Table 7.2: Bessel Equalizer Tables

#### 2 Term Bessel Filter Response

<table>
<thead>
<tr>
<th>Transducer Gain</th>
<th>Term 1</th>
<th>Term 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10000 + 01</td>
<td>0.10000 + 01 S²</td>
<td></td>
</tr>
<tr>
<td>0.00000 + 00</td>
<td>0.10000 + 01 S²</td>
<td></td>
</tr>
<tr>
<td>2 Term Bessel Filter Response</td>
<td>0.10000 + 01 S²</td>
<td></td>
</tr>
<tr>
<td>0.00000 + 00</td>
<td>0.10000 + 01 S²</td>
<td></td>
</tr>
</tbody>
</table>

#### 3 Term Bessel Filter Response

<table>
<thead>
<tr>
<th>Transducer Gain</th>
<th>Term 1</th>
<th>Term 2</th>
<th>Term 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15000 + 02</td>
<td>0.15000 + 02 S² + 0.10000 + 01 S³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00000 + 00</td>
<td>0.40000 + 01 S² + 0.10000 + 01 S³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00000 + 00</td>
<td>0.40000 + 01 S² + 0.10000 + 01 S³</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 4 Term Bessel Filter Response

<table>
<thead>
<tr>
<th>Transducer Gain</th>
<th>Term 1</th>
<th>Term 2</th>
<th>Term 3</th>
<th>Term 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10500 + 02</td>
<td>0.10500 + 02 S² + 0.10000 + 01 S³ + 0.10000 + 01 S⁴</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00000 + 00</td>
<td>0.40000 + 01 S² + 0.10000 + 01 S³ + 0.10000 + 01 S⁴</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00000 + 00</td>
<td>0.40000 + 01 S² + 0.10000 + 01 S³ + 0.10000 + 01 S⁴</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 5 Term Bessel Filter Response

<table>
<thead>
<tr>
<th>Transducer Gain</th>
<th>Term 1</th>
<th>Term 2</th>
<th>Term 3</th>
<th>Term 4</th>
<th>Term 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.94500 + 02</td>
<td>0.94500 + 02 S² + 0.10500 + 03 S³ + 0.15000 + 02 S⁴ + 0.10000 + 01 S⁵</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00000 + 00</td>
<td>0.40000 + 01 S² + 0.10000 + 01 S³ + 0.10000 + 01 S⁴ + 0.10000 + 01 S⁵</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 6 Term Bessel Filter Response

<table>
<thead>
<tr>
<th>Transducer Gain</th>
<th>Term 1</th>
<th>Term 2</th>
<th>Term 3</th>
<th>Term 4</th>
<th>Term 5</th>
<th>Term 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95000 + 03</td>
<td>0.95000 + 03 S² + 0.10500 + 03 S³ + 0.15000 + 02 S⁴ + 0.10000 + 01 S⁵ + 0.10000 + 01 S⁶</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00000 + 00</td>
<td>0.40000 + 01 S² + 0.10000 + 01 S³ + 0.10000 + 01 S⁴ + 0.10000 + 01 S⁵ + 0.10000 + 01 S⁶</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Unfortunately only simple loads can be equalized using the Laurent series approach, as the number of simultaneous equations that must be solved increases with load complexity. Moreover numerical accuracy is required to produce sensible results. Therefore a computer must be used to equalize a complicated load.

When using the Laurent series approach, each set of simultaneous equations generated is load specific, and changes form and equation number with the load. A new computer program is needed for each load, unless the program generates the Laurent series, which is a very complicated task and tends to produce cumulative errors. The cumulative errors cause significant loss of accuracy in the results. Thus certain common factors between the numerator and denominator polynomials, of $Z_2(s)$, cannot be found and so the resulting back impedance $Z_2(s)$ ceases to be PR and cannot be realized as a reactive network terminated in a resistor.

<table>
<thead>
<tr>
<th>TRANSDUCER GAIN</th>
<th>$K =$ 0.2500D+00</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1000D+00</td>
<td>0.1816D+02 $S$</td>
</tr>
<tr>
<td>0.0920D+00</td>
<td>0.10000D+00 $S$</td>
</tr>
<tr>
<td>0.0905D+00</td>
<td>0.01000D+01 $S$</td>
</tr>
<tr>
<td>0.0900D+00</td>
<td>0.01000D+01 $S$</td>
</tr>
<tr>
<td>0.0895D+00</td>
<td>0.01000D+01 $S$</td>
</tr>
<tr>
<td>0.0890D+00</td>
<td>0.01000D+01 $S$</td>
</tr>
</tbody>
</table>

### Table 7.2 Bessel Equalizer Tables (cont.)

<table>
<thead>
<tr>
<th>TRANSDUCER GAIN</th>
<th>$K =$ 0.1250D+00</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8840D+00</td>
<td>0.0972D+00 $S$</td>
</tr>
<tr>
<td>0.8820D+00</td>
<td>0.0972D+00 $S$</td>
</tr>
<tr>
<td>0.8800D+00</td>
<td>0.0972D+00 $S$</td>
</tr>
<tr>
<td>0.8780D+00</td>
<td>0.0972D+00 $S$</td>
</tr>
<tr>
<td>0.8760D+00</td>
<td>0.0972D+00 $S$</td>
</tr>
</tbody>
</table>

### 7 TERM BESSEL FILTER RESPONSE

<table>
<thead>
<tr>
<th>TRANSDUCER GAIN</th>
<th>$K =$ 0.1000D+00</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0900D+00</td>
<td>0.01000D+00 $S$</td>
</tr>
<tr>
<td>0.0900D+00</td>
<td>0.01000D+00 $S$</td>
</tr>
<tr>
<td>0.0900D+00</td>
<td>0.01000D+00 $S$</td>
</tr>
<tr>
<td>0.0900D+00</td>
<td>0.01000D+00 $S$</td>
</tr>
<tr>
<td>0.0900D+00</td>
<td>0.01000D+00 $S$</td>
</tr>
</tbody>
</table>

Unfortunately only simple loads can be equalized using the Laurent series approach, as the number of simultaneous equations that must be solved increases with load complexity. Moreover numerical accuracy is required to produce sensible results. Therefore a computer must be used to equalize a complicated load.

When using the Laurent series approach, each set of simultaneous equations generated is load specific, and changes form and equation number with the load. A new computer program is needed for each load, unless the program generates the Laurent series, which is a very complicated task and tends to produce cumulative errors. The cumulative errors cause significant loss of accuracy in the results. Thus certain common factors between the numerator and denominator polynomials, of $Z_2(s)$, cannot be found and so the resulting back impedance $Z_2(s)$ ceases to be PR and cannot be realized as a reactive network terminated in a resistor.
Moreover, if the common factors can be found, division of the numerator and denominator polynomials, by these factors, can render $Z_2(s)$ non-PR due to inaccuracies within the synthetic division process.

For these reasons an alternative method of equalization, more suitable for computer implementation, was sought. Based upon some of Youla's observations [3], in conjunction with the theory of LDEs, a new method of load equalization has been developed.

The LDE is well known in control theory [73], however it has not been used in filter and related theory. LDEs are well suited to numerical solution as errors in solution, are not cumulative [73].

7.2.3 The Linear Diophantine Equation

Before embarking on an explanation of the Diophantine equation method of equalization, some theory of LDE's must be introduced. In this section the general theory of LDE's is discussed.

The following equation is one of many LDEs [73],[84],[85],[86],[87].

\[ L(s)X(s) + M(s)Y(s) = W(s) \]  \hspace{1cm} (7.86)

where $L(s)$, $M(s)$ and $W(s)$ are known polynomials, and $X(s)$ and $Y(s)$ are unknown polynomials.

It is desired to find $X(s)$ and $Y(s)$ to solve (7.86).
The known polynomials may be expressed as products of their associated zero factors. For a solution to exist to (7.86), any common factors of \( L(s) \) and \( M(s) \) must also be common factors \( W(s) \) \([73],[84]\).

Let \( W(s) \) be rewritten as the product of two polynomials, \( F(s) \) and \( H(s) \) containing the common factors of \( L(s) \) and \( M(s) \), and the remaining zeros respectively. That is,

\[
W(s) = F(s)H(s) \quad (7.87)
\]

A further two polynomials \( Q(s) \) and \( R(s) \) can be found to satisfy \([64],[91]\),

\[
L(s)Q(s) + M(s)R(s) \quad \text{and} \quad F(s) \quad (7.88)
\]

Multiplying by the common factor \([73],[84]\),

\[
L(s)Q(s)H(s) + M(s)R(s)H(s) = F(s)H(s) = W(s) \quad (7.89)
\]

Thus one solution to (7.86) is \([73],[84]\),

\[
X(s) = Q(s)H(s) \quad \text{and} \quad Y(s) = R(s)H(s) \quad (7.90) \quad (7.91)
\]

However this is not the only solution, since polynomials \( U(s) \) and \( V(s) \) exist such that \([73],[84]\),

\[
L(s)U(s) + M(s)V(s) = 0 \quad (7.92)
\]
Moreover the general form of (7.92) is unchanged if (7.92) is multiplied by any polynomial A(s) [73],[84],

$$A(s)L(s)U(s) + A(s)M(s)V(s) = 0$$

(7.93)

Adding (7.93) to both sides of (7.89) gives a more general solution to (7.86) [73],[84].

$$L(s)[Q(s)H(s) + A(s)U(s)] + M(s)[R(s)(s) + A(s)V(s)] = W(s)$$

(7.94)

where the general solution to X(s) and Y(s) is [73],[84],

$$X(s) = Q(s)H(s) + A(s)U(s)$$

(7.95)

and

$$Y(s) = R(s)H(s) + A(s)V(s)$$

(7.96)

In order to find a unique solution to the LDE (7.86) further information, about the desired values of X(s) and Y(s), is required.

LDE's are not confined to equations in just two unknown polynomials, they may involve 3 or more unknowns. To solve such LDE's, successive substitutions are made to produce groups of LDE's involving just two unknowns. Each LDE, generated by substitution is solved in the same way as (7.97). The results are then re-assembled as general solutions of the original LDE.

The following equation is an LDE with 3 unknown polynomials.

$$L_1(s)X_1(s) + L_2(s)X_2(s) + L_3(s)X_3(s) = W(s)$$

(7.97)
Let

\[ M(s)Y(s) = L_2(s)X_2(s) + L_3(s)X_3(s) \]  
(7.98)

Where \( M(s) \) is a polynomial representing the common factors of \( L_1(s) \) and \( L_2(s) \). \( Y(s) \) is an unknown polynomial.

Substituting equation (7.98) into (7.97) produces the LDE [84],

\[ L_1(s)X_1(s) + M(s)Y(s) = W(s) \]  
(7.99)

Which has the solution [84],

\[ X_1(s) = Q_1(s)H(s) + A_1(s)V_1(s) \]  
(7.100)

and

\[ Y(s) = R(s)H(s) + A_1(s)V(s) \]  
(7.101)

Where \( H(s) \) is the quotient polynomial after division by the common factors of \( L_1(s) \) and \( M(s) \); it is represented by \( F(s) \). \( A_1(s) \) is an arbitrary polynomial, \( Q_1(s) \) and \( R(s) \) are polynomials solving the equation [84],

\[ L_1(s)Q_1(s) + M(s)R(s) = F(s) \]

\( U_1(s) \) and \( V_1(s) \) are polynomials satisfying the equations [84],

\[ L_1(s)U_1(s) + M(s)V(s) = 0 \]

Using the same technique, (7.98) can be solved [84],

\[ X_2(s) = Q_2(s)Y(s) + A_2(s)V_2(s) \]  
(7.102)

and

\[ X_3(s) = Q_3(s)Y(s) + A_3(s)V_3(s) \]  
(7.103)
Where $A_2(s)$ is arbitrary, $Q_2(s)$ and $Q_3(s)$ solve the equation

\[ L_2(s)Q_2(s) + L_3(s)Q_3(s) = M(s) \]

and $U_2(s)$ and $U_3(s)$ satisfy the equation \[ L_2(s)U_2(s) + L_3(s)U_3(s) = 0 \]

By substituting equation (7.101) into (7.102) and (7.103), $X_2(s)$ and $X_3(s)$ can be expressed in terms of known polynomials and arbitrary polynomials \[ X_2(s) = Q_2(s)R(s)H(s) + Q_2(s)V(s)A_1(s) + A_2(s)U_2(s) \] \[ X_3(s) = Q_3(s)R(s)H(s) + Q_3(s)V(s)A_1(s) + A_2(s)U_3(s) \]

$LDE$ (7.97) is thus solved. The values of $X_1(s)$, $X_2(s)$ and $X_3(s)$ are given by equations (7.100), (7.104) and (7.105) respectively.

If $Y(s)$ is chosen as a sum of two different terms of (7.98), the actual known polynomials of (7.100), (7.104) and (7.105) will be different, however the arbitrary polynomials $A_1(s)$ and $A_2(s)$ may be chosen to give the same solutions.

In order to solve LDE's involving larger numbers of unknown polynomials, one merely repeats the process discussed until the LDE can be solved in groups of LDE's in two unknowns \[ \text{[84].} \]
7.2.4 Load Equalization Using Linear Diophantine Equations

Now returning to the problem of equalization; the intention is to express the complex load \( Z(s) \) such that (7.59), and the Youla observation [3] can be recast as a LDE problem.

Using numerically well behaved techniques, the unknown polynomials can be found. Further constraints are, however, necessary to find a unique solution.

By rewriting the load impedance as the ratio of two polynomials,

\[
Z_L(s) = \frac{C_n N(s)}{D(s)}
\]

(7.106)

where \( N(s) \) and \( D(s) \) are polynomials in complex frequency, \( s \), and \( C_n \) is a constant such that the coefficient, of the highest order term in \( N(s) \), is unity.

The real part of the load becomes,

\[
r_L(s) = \frac{Z_L(s) + Z_L(-s)}{2}
\]

(7.107)

\[
r_L(s) = \frac{C_n (N(s)D(-s) + N(-s)D(s))}{2D(s)D(-s)}
\]

(7.108)

\( Z_L(s) \) is a physically realisable load, therefore the zeros of \( r_L(s) \) are symmetrical about both the real and imaginary axes of the complex frequency plane; any zero in one quadrant of the plane has three other corresponding zeros in the remaining quadrants [57].
Grouping the RHP and LHP zeros together in two separate polynomials, \( C(s) \) and \( C(-s) \) respectively, and the imaginary axis zeros in yet another, \( F(s) \),

\[
\tau_L(s) = \frac{C_n C_r C(s) C(-s) F(s)}{2D(s) D(-s)} \tag{7.109}
\]

Where \( C_r \) is a constant chosen such the coefficient of the highest order term, of \( C(s) \), is unity.

Now recall that the load transmission was defined as (7.44 ),

\[
T(s) = \frac{\tau_L(s)}{Z_L(s)}
\]

In order to simplify the analysis, let \( F(s) = 1 \). Solutions when \( F(s) \neq 1 \) will be considered later, as they give rise to special restrictions corresponding to transmission zeros of classes II to IV; in Youla's theory [3].

When \( F(s) = 1 \), all RHP zeros of \( T(s) \), which are also RHP zeros of \( \tau_L(s) \), are conveniently represented by the new polynomial \( C(-s) \).

In the same way that \( Z_L(s) \) can be expressed as a ratio of two polynomials, it is convenient to express \( S_0(s) \) as,

\[
S_0(s) = \frac{K(s)}{J(s)} \tag{7.110}
\]

where \( K(s) \) and \( J(s) \) are both polynomials in complex frequency \( s \), they are unknown, but in view of (7.60 ) the coefficient values are restricted by the transmission function \( G(s) \).

The LHS of (7.59 ) becomes
\[ b(s) - S(s) = \frac{B(-s)}{B(s)} - \frac{P(-s)K(s)}{P(s)J(s)} \]  

(7.111)

where \( b(s) = B(-s)/B(s) \) and \( S(s) = a(s)S_0(s) \) where \( a(s) = P(-s)/P(s) \)

Recall that the zeros of \( b(s) \) were defined as the RHP poles of \( Z_t(-s) \), which are the zeros of \( D(-s) \), thus (7.59 ) may be rewritten,

\[ b(s) - S(s) = \frac{D(-s)}{D(s)} - \frac{P(-s)K(s)}{P(s)J(s)} \]  

(7.112)

\[ = \frac{D(-s)P(s)J(s)}{D(s)P(s)J(s)} - \frac{D(s)P(-s)K(s)}{D(s)P(s)J(s)} \]  

(7.113)

Equation (7.59 ) requires that (7.113) contain the RHP zeros of \( r_L(s) \); and so [3],

\[ D(-s)P(s)J(s) + D(s)P(-s)K(s) = C(-s)H(s) \]  

(7.114)

where \( H(s) \) is an unknown polynomial containing the remaining zeros of the LHS of (7.113).

The Equalization problem has thus been recast as the solution to an LDE (7.114). The ambiguity is resolved by considering the positive realness of \( Z_2(s) \).

The unknowns are \( P(s) \) and \( H(s) \); \( J(s) \) is known and \( K(s) \) is related to \( J(s) \) through the unknown gain, \( k \), of \( G(s) \), by (7.76 ).

The all pass function \( a(s) \) provides further restrictions on the solution of (7.114). From the definition of \( a(s) \), which is rewritten in (7.114) as the ratio of \( P(-s) \) to \( P(s) \), it is clear that the coefficient of the highest term in \( P(s) \) is unity. Additionally, the zeros of \( P(s) \) must lie to the left of the imaginary axis by some small amount \( \epsilon \).
Bringing the restrictions derived together into a group of solution constraints, a unique solution to (7.114) and the constraints can be found.

The PR restriction parameter \( \epsilon \) is chosen according to the calculation error in the problem. In test samples discussed, it is usually 10 times the error of calculation. The role of \( \epsilon \) will be discussed in the section 7.2.7.

7.2.5 The Equalizer Back-Impedance

Having solved for the transducer gain, \( k \), and the all-pass function \( a(s) \), in the Youla equalisation technique [3], the designer finds the equalizer output impedance. He then attempts to find any common factors between numerator and denominator, deflate the numerator and denominator polynomials, and finally realize \( Z_2(s) \). Unless sufficient precision is carried through the earlier calculations, the common factors cannot be found; the resulting equalizer will not be accurate, and the power transfer to the real part of the load will not be constant, as required.

The LDE theory can also be used to find the common factors between the back impedance numerator and denominator. After deflation a much more accurate estimate of the back impedance is available.

Recall equation (7.59)

\[
b(s) - S(s) = \frac{2r_L(s)b(s)}{Z_2(s) - Z_L(s)}
\]

with some rearrangement, an expression for \( Z_2(s) \) can be found.
Substituting polynomial expressions for each rational function and the solution to \( b(s) - S(s) \) yields,

\[
Z_2(s) = \frac{C_n C_r C(s) J(s) P(s)}{D(s) H(s)} - \frac{C_n N(s)}{D(s)}
\]  

(7.116)

\[
= \frac{C_n C_r C(s) J(s) P(s) - C_n N(s) H(s)}{D(s) H(s)}
\]  

(7.117)

The common factors of (7.117) are not obvious. However since (7.117) is the expression for the equalizer back impedance, it is unlikely that \( D(s) \) should appear in the denominator. Thus it is likely that \( D(s) \) is a common factor between the numerator and denominator of (7.117).

By reconsidering the LDE (7.114), the common factors of (7.117) can be found

\[
D(-s) P(s) J(s) - D(s) P(-s) K(s) = C(-s) H(s)
\]

Multiplying (7.188) by \( C_r C(s) \) gives

\[
C_r C(s) D(-s) P(s) J(s) - C_r C(s) D(s) P(-s) K(s) = C_r C(s) C(-s) H(s)
\]  

(7.118)

The RHS of (7.118) is merely the numerator of \( r_L(s) \) multiplied by the unknown expression \( H(s) \). Substituting the numerator of \( r_L(s) \) (7.109) into the RHS of (7.118), rearranging the result to factor out \( D(s) \) on the RHS, and \( D(-s) \) on the LHS,

\[
(C_r C(s) P(s) J(s) - N(s) H(s)) D(-s) - (C_r C(s) P(-s) K(s) + N(-s) H(s)) D(s)
\]  

(7.119)

From the LDE theory of section 7.2.4, for a solution to exist, \( D(s) \) must also be a factor of the LHS of (7.119). Since it cannot be a factor of
D(-s), D(s) must be a factor of \( C_r C(s)P(s)J(s) - N(s)H(s) \). Similarly, D(-s) must be a factor of \( C_r C(s)P(-s)K(s) - N(-s)H(s) \).

Using (7.119), a second expression for equalizer back impedance may be found by substituting for the numerator of

\[
Z_2(s) = \frac{C_n C_r(s)P(-s)K(s) + C_n N(-s)H(s)}{D(-s)H(s)} \quad (7.120)
\]

Expression (7.120) was used in the equalisation computer program. Using (7.119) it can be shown that D(-s) is the common factor in (7.120).

Solution to the equalisation problem merely involves solving the LDE (7.114), finding the back impedance using (7.120) and dividing the numerator of (7.120) by D(-s). The designer then realises (7.120) as a reactive load terminated in a resistor.

The LDE is solved for an optimum DC gain k, which relates \( K(s) \) to \( J(s) \). The constraints on the zero locations of \( P(s) \) constrain \( K \) through (7.60). The following algorithm is used to solve (7.110).

Step 1. Find the real part of the load \( r_L(s) \) from \( Z_L(s) \)
Step 2. Set \( k = 1 \), count = 1
Step 3. Using current value of \( k \), find \( K(s) \) by spectral factorization of (7.76)
Step 4. Solve (7.114) for \( P(s) \) and thus \( P(-s) \)
Step 5. If \( P(s) \) has no zeros in the RHP or on the Imaginary axis go to step 13.
Step 6. If \( k = 0 \) go to step 9
Step 7. Reduce the absolute value of \( k \) by say 0.1
Step 8. Go to step 3
Step 9. If count = 2 go to step 12
Step 10. Set D(s) = (-1)*D(s)
Step 11. Go to step 3
Step 12. Stop (the gain $k = 0$, over the desired bandwidth, no power transfer to the real part of the load, $r_L(s)$, is possible.)
Step 13. Set $k_{ub} = k + 0.1$ and $k_{1b} = k$
Step 14. Set $k = (k_{ub} + K_{1b})/2$
Step 15. Using current value of $k$, find $K(s)$ by spectral factorization of (7.76)
Step 16. Solve (7.114) for $P(s)$ and thus $P(-s)$
Step 17. If $P(s)$ has no zeros in the RHP or on the Imaginary axis go to step 21
Step 18. Set $k_{1b} = k$
Step 19. If $(k_{ub} - k_{1b}) = 0$ go to step 24
Step 20. Go to step 14
Step 21. $k_{ub} = k$
Step 22. If $(k_{ub} - k_{1b}) = 0$ go to step 24
Step 23. Go to step 14
Step 24. Find the numerator of (7.120)
Step 25. Divide the numerator of (7.120) by $D(-s)$
Step 26. Stop ($Z_2(s)$ has been found)

7.2.6 Imaginary Axis Zeros of Transmission

In order to reduce the complexity of the diophantine equation method of solution, it was assumed that no imaginary axis zeros occurred in $r_L(s)$
(7.109) and also in \( T(s) \); the load transmission. Clearly, such an assumption restricts the load types, for which one can apply LDE based equalisation, to the loads complying with Youla's class I restrictions \([3]\).

The LDE approach is equally applicable to loads with imaginary axis zeros of transmission, however extra solution restrictions are required.

The real part of the load, \( r_\ell(s) \) was given in (7.109),

\[
r_\ell(s) = \frac{C_n C_\ell C(s) C(-s) F(s)}{2d(s)D(-s)}
\]

The load transmission function \( T(s) \) (7.44) may be rewritten,

\[
T(s) = \frac{r_\ell(s)}{Z_\ell(s)} = \frac{C_n C_\ell C(s) C(-s) F(s)}{2N(-s)D(-s)} \tag{7.121}
\]

Youla \([3]\) divided the imaginary axis zeros of \( T(s) \) into three classes, dependent upon the way in which they relate to \( N(s) \) and \( D(s) \). It is appropriate to do the same for theoretical purposes, although in calculation the LDE approach renders the divisions unnecessary.

Using Youla's nomenclature \([3]\), ZOT of class II are zeros of \( r_\ell(s) \) which are also zeros of \( N(s) \), those of class III are zeros of \( r_\ell(s) \), which are neither zeros of \( N(s) \) or \( D(s) \), and those of class IV are zeros of \( r_\ell(s) \) which are also zeros of \( D(s) \).

Rational functions can have three types of imaginary axis zero. Finite imaginary axis zeros are zeros of the numerator polynomial, of the rational function, and are obvious.
Zeros of a rational function at zero are also imaginary axis zeros of the rational function numerator polynomial.

The third, and less obvious imaginary axis zeros are those located at infinity. Their representation is more difficult than those of the types previously discussed. The polynomial representation of all three types of imaginary axis zero must be considered.

Finite imaginary axis zeros occur in complex conjugate pairs and are represented by quadratic factors of the form,

\[ F(s) = (s^2 + e^2) \quad (7.122) \]

Imaginary axis zeros at zero are represented by,

\[ F(s) = (s + e) \quad (7.123) \]

where \( e = 0 \)

The number of zeros, in a polynomial, is equal to the number of lowest order coefficients that are consecutively zero; for example the polynomial,

\[ F(s) = a_7 s^7 + a_6 s^6 + a_5 s^5 + a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0 \quad (7.124) \]

where \( a_3 = a_2 = a_1 = a_0 = 0 \)

has 4 zeros at zero. Computer representation of zeros at zero, is thus quite straightforward.

Zeros at infinity involve a similar concept.

If \( F(s) = \).
\[ a_n s^n + a_{n-1} s^{n-1} + a_{n-2} s^{n-2} + a_{n-3} s^{n-3} + a_{n-4} s^{n-4} + \ldots + a_2 s^2 + a_1 s + a_0 \]  

(7.125)

has \( m \) zeros at infinity, then

\[ \frac{1}{s^n}(a_n + a_{n-1} s + a_{n-2} s^2 + a_{n-3} s^3 = a_{n-4} s^4 + \ldots + a_2 s^2 + a_1 s + a_0 s^n \]  

(7.126)

has \( m \) zeros at zero, which implies that,

\[ a_n = a_{n-1} = a_{n-2} = \ldots = a_{n-m-2} = a_{n-m-1} = 0 \]

i.e. the \( m \) highest-order coefficients of \( f(s) \) are consecutively zero.

It is now possible to represent all zero-types in polynomial form, irrespective of their locations on the complex frequency plane.

Transmission zeros of Youla's three imaginary transmission zero classes [3] can now be considered in the context of LDE equalization.

Consider a load \( Z_L(s) \), which has some zeros located on the imaginary axis.

\[ Z_L(s) = \frac{C N(s)}{D(s)} = \frac{N'(s)L(s)}{D(s)} \]  

(7.127)

where \( N'(s) \) is a polynomial representing the LHP zeros, and \( L(s) \) is a polynomial representing those zeros located on the imaginary axis.

Since \( Z_L(s) \) is PR, the zeros of \( L(s) \) have unit multiplicity. The corresponding real part of the load, \( r_L(s) \), has the form,

\[ r_L(s) = \frac{C r_c C(s) C(-s) F(s)}{2D(s)D(-s)} \]  

(7.128)
In order not to cloud the issue, let $Z_L(s)$ be of the type which gives rise only to imaginary axis zeros in $r_L(s)$. Thus $C(s)=C(-s)=1$. Moreover let $F(s)$ represent zeros of transmission of Youla's class II; confining $F(s)$ to the zeros of $L(s)$ only. The multiplicities of zeros in $F(s)$ must also be even as required by the positive real nature of $Z_L(s)$. Thus let,

$$F(s) = L(s)M(s)$$  \hspace{1cm} (7.129)$$

where $M(s)$ contains those zeros of $L(s)$ which occur more than twice in $F(s)$.

The real part of the load is thus,

$$r_L(s) = \frac{C_nC_F L(s)L(s)M(s)M(s)}{2D(s)D(-s)}$$ \hspace{1cm} (7.130)$$

and the back impedance expression (7.65) is,

$$Z_2(s) = \frac{2r_L(s)b(s) - Z_L(s)}{b(s) - S(s)}$$

Define $Z(s)$,

$$Z(s) = \frac{2r_L(s)b(s)}{b(s) - S(s)}$$ \hspace{1cm} (7.131)$$

Irrespective of its functional form,

$$\text{re} \left\{ Z_2(j\omega) \right\} \geq 0 \quad [3] \hspace{1cm} (7.132)$$

implying that,

$$\text{re} \left\{ Z(j\omega) \right\} \geq \text{re} \left\{ Z_L(j\omega) \right\} \geq 0 \quad [3] \hspace{1cm} (7.133)$$

Using (7.59), (7.60), (7.76) and (7.129), $Z(s)$ may be rewritten as the ratio of two polynomials,
Define the denominator of (7.134) as the product of three polynomials.

\[ Q(s)R(s)D(s) = (D(-s)J(s)P(s) - D(s)K(s)P(-s))D(s) \]  

where \( R(s) \) is a polynomial representing the denominator zeros which cancel with some zeros of the numerator of (7.134), and \( Q(s) \) is a polynomial representing the zeros remaining in the denominator of (7.134) after cancellation by \( R(s) \).

\[ Z(s) = \frac{C_nC_rL^2(s)M^2(s)J(s)P(s)}{R(s)Q(s)D(s)} \]  

Since \( Z_L(s) \) in (7.115) has neither RHP nor imaginary axis poles, and \( \text{re} \{ Z_L(s) \} \geq 0 \), the only violations of the PR requirement can come from \( Z(s) \), which may have either RHP or imaginary axis poles with non-positive residues. It is thus necessary to consider only the positive realness of \( Z(s) \).

The uncancelled numerator of \( Z(s) \) is known, therefore, it is more convenient to consider the admittance \( Y(s) = 1/Z(s) \),

\[ Y(s) = \frac{R(s)Q(s)D(s)}{C_nC_rL^2(s)M^2(s)J(s)P(s)} \]  

\( \text{re} \{ Y(s) \} \geq 0 \) is implied by \( \text{re} \{ Z(j\omega) \} \geq 0 \).
Therefore for $Y(s)$ to be PR; the denominator polynomials must have zeros confined either to the LHP or to the imaginary axis; in which case the poles they represent must be of unit multiplicity with non-negative residues.

$Y(s)$ cannot generally, however be PR since $L^2(s)M^2(s)$ represent imaginary axis poles, which from (7.129), have greater than unit multiplicity. If,

$$R(s) = L^2(s)M^2(s) \quad (7.138)$$

such that terms in the numerator and denominator cancel all imaginary axis zeros, $Y(s)$ is guaranteed PR. If however, $R(s)$ is chosen to cancel $L(s)M^2(s)$,

$$R(s) = L(s)M^2(s) \quad (7.139)$$

then $Y(s)$ will have imaginary axis poles of unit multiplicity. In this case, to ensure positive realness of $Y(s)$, the residues at each pole of $L(s)$ must be tested. Since $D(s)$ is a positive real polynomial, it is necessary only to ensure that $H(s)$ is PR, computationally $H(s)$ can be tested using a Routh array [88],[89],[90],[91].

Now consider a load impedance,

$$Z_L(s) = \frac{C_nN(s)}{D(s)} \quad (7.140)$$

which contains no imaginary axis zeros itself, yet the real part of the load $r_L(s)$, has zeros on the imaginary axis, of arbitrary even multiplicity, and may be written,
As in the previous discussion let \( C(s) = C(-s) = 1 \). \( r_L(s) \) now has zeros of Youla's class III [3].

Since \( Z_L(s) \) contains neither imaginary axis zeros nor imaginary axis poles, \( \text{re} \{ Z_L(j\omega) \} \geq 0 \) (7.121), the only violation of the PR condition for \( Z_2(s) \) (7.115) can occur in \( Z(s) \) (7.131).

Analysis of this case is similar to that of class II. Again to ensure the positive realness of \( Y(s) \), \( R(s) \) must be chosen either to cancel all imaginary axis poles of \( Y(s) \), that is,

\[
R(s) = U^2(s)V^2(s)
\]

(7.142)

or cancel all but unit multiplicity imaginary axis poles,

\[
R(s) = U(s)V^2(s)
\]

(7.143)

in which case, \( H(s) \) must be tested for positive realness.

The LDE, corresponding to the two ZOT classes discussed, is,

\[
D(-s)P(s)J(s) - D(s)P(-s)K(s) = R(s)Q(s)
\]

(7.144)

where \( Q(s) = H(s) \). (refer equation (7.114).

Finally consider a load impedance \( Z_L(s) \) which has poles on the imaginary axis,
The real part of the load $r_L(s)$ is,

$$r_L(s) = \frac{C(s)C(-s)U(s)}{D'(s)D'(-s)W(s)}$$  \hspace{1cm} (7.146)

Let $C(s) = C(-s) = 1$, and $U(s)$ contains zeros of $W(s)$ of arbitrary even multiplicity; producing zeros of transmission of Youla's class IV [3].

Recall $Z_2(s)$ the equaliser back impedance,

$$Z_2(s) = \frac{2r_L(s)b(s)}{b(s) - S(s)} - Z_L(s)$$  \hspace{1cm} (7.147)

$Z_L(s)$ is positive real and has poles located on the imaginary axis, which are simple with positive residues (7.145).

If the denominator of $Z(s)$ has no poles coincident with those of $Z_L(s)$, then it is sufficient to verify the positive realness of $Z(s)$ alone. If however, $Z(s)$ has one or more poles coincident with those of $Z_L(s)$, then the residue of $Z(s)$ at each pole must be greater or equal to that of $Z_L(s)$ at the same pole. The residue of the difference will then be positive and $Z_2(s)$ will be PR.

Rewriting $Z_2(s)$ in polynomial form will reveal the PR requirements more explicitly,

$$Z_2(s) = \frac{N(s)}{D'(s)W(s) - N(s)K(s)}$$  \hspace{1cm} (7.148)
\[ \frac{C_n C_r W^2(s) A^2(s) D'(s) K(s)}{(D'(-s) K(s) - D'(s) J(s)) D'(s) W(s)} - \frac{N(s)}{D'(s) W(s)} \]  

(7.149)

Let the denominator (7.149) be represented by the product of 4 polynomials,

\[ Q(s) R(s) D'(s) W(s) = (D'(s) K(s) - D'(s) J(s)) D'(s) W(s) \]  

(7.150)

where \( Q(s) \) and \( R(s) \) have been defined.

\[ Z_2(s) = \frac{C_n C_r W^2(s) A^2(s) D'(s) K(s)}{Q(s) R(s) D'(s) W(s)} - \frac{C_n N(s)}{D'(s) W(s)} \]  

(7.151)

\[ Z(s) = \frac{C_n C_r W^2(s) A^2(s) D'(s) K(s)}{Q(s) R(s) D'(s) W(s)} \]  

(7.152)

\[ Z(s) = \frac{C_n C_r W^2(s) A^2(s) D'(s) K(s)}{Q(s) R(s) D'(s) W(s)} \]  

(7.153)

To ensure that \( Y(s) \), and therefore \( Z(s) \), is positive real \( R(s) \) must be chosen to cancel either all imaginary axis poles of \( Y(s) \),

\[ R(s) = W^2(s) A^2(s) \]  

(7.154)

or all but those of unit multiplicity,

\[ R(s) = W(s) A^2(s) \]  

(7.155)

in which case \( Y(s) \) has poles of unit multiplicity on the imaginary axis. If the latter choice is made, \( Z(s) \) will have a common factor \( W(s) \) in the numerator and denominator, and will therefore have a zero residue at each zero of \( W(s) \). However, the residue of \( Z_L(s) \) at each of zero \( W(s) \) will be
positive, so that the residue of $Z_2(s)$ at each zero will be negative. Thus $Z_2(s)$ will not be PR.

For this reason $R(s)$ must be chosen to satisfy (7.154). In this case the residue of each imaginary axis pole of $Z(s)$ must be greater than, or equal to, the corresponding residue of $Z_1(s)$.

To ensure that $Z_2(s)$ is PR in the computer solution, it is necessary to test that $Q(s)$ is a positive real polynomial, and that,

$$N_2(s) = C_n C_r K(s) - C_n N(s) Q(s)$$  \hspace{1cm} (7.156)

is also a PR polynomial.

The LDE corresponding to $R(s) = W_2(s) A_2(s)$ is,

$$D'(-s) K(s) - D'(s) J(s) = Q(s) W_2(s) A_2(s)$$  \hspace{1cm} (7.157)

As in the class II and III cases, a further option for solution, guaranteeing positive realness of $Z_2(s)$ is available. The numerator of (7.152) may be chosen to cancel $W(s)$ in the denominator, thereby removing the imaginary axis poles from $Z_2(s)$. This option produces yet another LDE,

$$C_r K(s) - N(s) Q(s) = W(s) I(s)$$  \hspace{1cm} (7.158)

where $I(s)$ is an unknown polynomial representing the remaining zeros of the LHS.

By solving (7.157) and (7.158), a PR solution to $Z_2(s)$ is guaranteed.
Solutions for a mixture of all types of transmission zero may be combined either into two LDEs,

\[ D(-s)P(s)J(s) - D(s)P(-s)K(s) = C(-s)L^2(s)M^2(s)V^2(s)W^2(s)A^2(s)H(s) \]  
(7.159)

and,

\[ C_r C(s)P(-s)K(s) + N(-s)H(s) = W(s)I(s) \]  
(7.160)

which guarantee a PR back impedance solution, or

\[ D(-s)P(s)J(s) - D(s)P(-s)K(s) = C(-s)L(s)M^2(s)U(s)V^2(s)W^2(s)A^2(s)H(s) \]  
(7.161)

in which case \( H(s) \) and \( N_2(s) \) must be tested to ensure that they are PR polynomials.

7.2.7 The Trivial Solution and the Bounding Polynomial

When imaginary axis transmission zeros are present some difficulty may be encountered in solving (7.159), or (7.161). Rewriting (7.159) in simpler form, the LDE becomes,

\[ D(-s)P(s)J(s) - D(s)P(-s)K(s) = C(-s)R(s)H(s) \]  
(7.162)

where \( R(s) \) is a polynomial representing all imaginary axis zeros of transmission.
Recall the definition of \( a(s) \),

\[
a(s) = \prod_{i=1}^{n} \frac{s - c_i}{s + c_i} = \frac{P(-s)}{P(s)} \quad (7.163)
\]

Thus \( P(s) \prod_{i=1}^{n} s + c_i \) and \( \text{re} \ c_i > 0 \) for all values of \( i \).

If \( C(s) = C(-s) = 1 \), one possible solution to (7.162) is the trivial solution,

\[
P(s) = P(-s) = R(s), \quad H(s)C(-s) = D(-s)J(s) - D(s)K(s) \quad (7.164)
\]

moreover if \( C(-s) = 1 \) then the solution to (7.159) may incorporate the trivial solution. In either cases the trivial solution violates the constraints imposed by (7.163).

To avoid violation of constraints, in the computer solution, a method must be found to ensure that possible solutions are forced away from the trivial solution by an amount sufficient to obey (7.163).

The inequality constraint may be rewritten as an equality constraint,

\[
\text{re}(c) + \epsilon \geq 0 \quad (7.165)
\]

where \( \epsilon \) is a positive number.

Choice of \( \epsilon \) must be sufficiently large that the computer program is aware that it is non-zero. In the computer solution of the examples of the next chapter, \( \epsilon \) is ten times the permitted computation error. A similar approach to the constraints of (7.163) would also be necessary in hand solution using Youla's original method [3].
The new constraint must be incorporated into (7.162). \( P(s) \) may be redefined as the sum of a PR polynomial \( \alpha(s) \) and a non-negative real polynomial \( \beta(s) \).

\[
P(s) = \alpha(s) + \beta(s) \tag{7.166}
\]

\( \alpha(s) \) is a minimum positive real polynomial that is fixed by the computer precision, such that its zeros satisfy (7.163). Since \( \beta(s) \) cannot be negative for \( \text{re}(s) \geq 0 \), the minimum permitted solution polynomial is

\[ P(s) = \alpha(s), \]

thus definition (7.166) obeys the constraints of (7.163). \( \alpha(s) \) is thus referred to as the bounding polynomial.

Substituting (7.166) into (7.162) produces a new LDE,

\[
D(-s)J(s)(-s) - D(s)K(s)(-s) - C(-s)R(s)H(s)
= D(s)K(s)(-s) - D(-s)J(s)(s) \tag{7.167}
\]

The new LDE is solved by finding a non-negative real \( \beta(s) \), which may contain a trivial solution. Adding the bounding polynomial \( \alpha(s) \) to \( \beta(s) \) ensures that \( P(s) \) solves (7.162) and satisfies the constraints of (7.163).

The choice of bounding polynomial, which reflects the calculation error, must now be considered. The region of the complex frequency plane to which the zeros of \( P(s) \) must be confined is shown in Fig 7.6.
If \( \beta(s) = 0 \), then the zeros of \( P(s) \) must lie along the constraint as shown in Fig 7.7 [88],[89],[90],[91].

Thus \( \beta(s) \) must be the product

\[
\beta(s) = \prod_{i=1}^{n} (s + \gamma_i + j\delta_i) \tag{7.168}
\]

where \( \gamma_i > 0 \), and the \( \delta_i \) are arbitrary and real.
When written in series form, the coefficients of $\beta(s)$ have smallest value when the $\delta_i = 0$. The non-negative requirement of $\beta(s)$ sets all the coefficients of $\beta(s) \geq 0$.

The smallest coefficient in the series expansion of $\beta(s)$ (7.168) is the constant term,

$$\beta_0 = \prod_{i=1}^{n} \gamma_i$$  \hspace{1cm} (7.169)$$

which must be large enough to give a constant term greater than zero, otherwise constraint (7.163) will be violated. Thus the $\gamma_i$ are chosen,

$$\beta_i = \epsilon^{1/n}$$  \hspace{1cm} (7.170)$$

for all $i$.

The bounding polynomial thus defined has been used in the computer equalisation program used to solve Youla's examples [3]. The results are given in the next chapter.

7.2.8  New Methods of Synthetic Division and Solutions to the Linear Diophantine Equations of Equalization.

The process of synthetic division can lead to a reduction in the accuracy of the equalizer output impedance $Z_2(s)$, making it impossible to realize using the synthesis techniques discussed in section 7.3. Therefore an alternative way of performing polynomial division was sought.
Using the theory of the LDE a new technique of polynomial division was found. By expressing the LDE in vector matrix form, polynomial division can be performed producing results of much greater accuracy than the traditional techniques of synthetic division. Moreover the vector matrix technique of synthetic division indicates ways in which general LDE's may be solved using a digital computer.

Traditionally synthetic division of one polynomial by another is accomplished by performing long division operations between successively decreasing powers of the independent variable [92]. When little is known about the quotient and remainder polynomials the traditional approach is generally quite suitable. However in the special case when the divisor polynomial is a known factor of the Dividend polynomial, the long division approach fails to use all available information about the solution, such information could be used to enhance its accuracy. When dividing a polynomial by a known factor, it is hoped that the remainder is very small. With certain polynomials this is not the case, and errors build up in the quotient.

In earlier series approaches to equalization, the build up of errors in quotient polynomials often made it impossible to find common factors between polynomials, leading to quite inaccurate solutions.

The process of polynomial zero finding is particularly susceptible to the accumulation of error. Polynomial zero finding generally proceeds in two phases. In the first phase a zero is located to within a fixed error parameter [93]. In the second phase, the polynomial is then deflated by the associated zero factor [93].
The error in the zero location is then propagated throughout the coefficients of the quotient polynomial. While it may be quite small, giving rise to small quotient errors, small errors in coefficients can have a major effect on the locations of the remaining zeros and these may be quite some distance from the original zero location.

Phases one and two are then repeated successively until all zeros are found. Each time phase two is applied further errors are introduced into the remaining zero locations.

If one attempts to regenerate the original polynomial from the zeros the series expansion can be quite different from the original.

The vector matrix approach, to be outlined here, forces a redistribution of error amongst the quotient coefficients, by using the knowledge that the remainder must be zero. When deflating a polynomial by its zeros, the vector matrix approach also offers the possibility of scaling to further improved error distribution. In vector matrix form the quotient polynomial is readily found by inverting the non-square matrix \([94],[95],[96],[97],[98],[99]\). This approach is readily coded, for digital computer solution \([97],[98],[99]\).

The vector matrix approach may also be used for cases in which the remainder is unknown, by suitable scaling \([97]\), improvement of error over traditional synthetic division is possible.

The diophantine equations and associated vector matrix solution to polynomial division problems, illustrate a way in which vector matrix solutions may be applied to more general linear diophantine equations; such
as those encountered in diophantine equalization. Generally, sufficient constraints on the solutions to equalization diophantine equations exist for a unique solution to be sought. Thus the LDE general solutions discussed in section 7.2.3, while helpful in theoretical development, are of little use numerically. The vector matrix approach is much more suitable for finding a unique solution to linear diophantine equations, than the general solution.

The vector matrix expressions generated by diophantine equations may represent over determined systems. Using the matrix pseudo-inverse all available information may be used to find the unique solution in a least squares sense [97], [98], [99].

If it is known that one polynomial is a factor of another, then the requirement that the remainder polynomial is zero is implicit in the solution; a constraint that is not enforced in traditional synthetic division [92]. Coupled with suitable scaling, this factor will tend to distribute error evenly into the quotient polynomial coefficients.

The general polynomial division problem may be considered as the sum of a remainder polynomial and a product of known and unknown polynomials,

\[ D(s)Q(s) + R(s) = P(s) \] (7.171)

Where \( s \) is the polynomial independent variable. \( P(s) \) is the dividend polynomial, \( D(s) \) is the divisor polynomial, and \( Q(s) \) and \( R(s) \) are the respective quotient and remainder polynomials.

The traditional approach is to solve for \( Q(s) \) and \( R(s) \) by long division, stopping when the constant term in \( Q(s) \) is found [92].
Equation (7.171) is however a LDE with general solutions [84],[85],[86],[87],[88],

\[ Q(s) = A(s) + C(s)t(s) \quad \text{and} \quad R(s) = B(s)E(s)t(s) \]  

(7.172)

Where \( A(s) \), \( B(s) \), \( C(s) \), and \( E(s) \) are polynomials satisfying,

\[ D(s)A(s) + B(s) = P(s) \]  

(7.173)

and

\[ D(s)C(s)t(s) + E(s)t(s) = 0 \]  

(7.174)

Where \( t(s) \) is arbitrary.

Fortunately, further information about the solution, to (7.171), is available and a unique solution to (7.149) can be found.

Consider the orders of polynomials in (7.171). Let \( P(s) \) be of order \( n \), and \( D(s) \) of order \( m \) such that \( n > m \). The quotient polynomial \( Q(s) \) will then have an order of \( n-m \), and the remainder polynomial \( R(s) \) will have an order of at most \( m-1 \). These restrictions, of the solution order, are sufficient to define a unique solution for \( Q(s) \) and \( R(s) \) of (7.171).

To illustrate the uniqueness of solution consider the case in which \( D(s) \) is a known factor of \( P(s) \), which implies that the remainder polynomial \( R(s) = 0 \).

\[ D(s)Q(s) = P(s) \]

The coefficients of \( P(s) \) may be related to the coefficients of \( D(s) \), which are known, and the coefficients of \( Q(s) \), which are unknown,
\[ p_n s^n + p_{n-1} s^{n-1} + p_{n-2} s^{n-2} + \ldots + p_2 s^2 + p_1 s + p_0 \]  
(7.175)

Relationships between the coefficients of the LHS and RHS of (7.175), may be represented in vector matrix form,

\[ DQ = P \]
(7.176)

\[
\begin{bmatrix}
    d_0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
    d_1 & d_0 & 0 & 0 & 0 & 0 & 0 \\
    d_2 & d_1 & d_0 & 0 & 0 & 0 & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    d_{m-2} & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    d_{m-1} & d_{m-2} & \vdots & \vdots & \vdots & \vdots & \vdots \\
    d_m & d_{m-1} & d_{m-2} & \vdots & \vdots & \vdots & \vdots \\
    0 & d_m & d_{m-1} & \vdots & \vdots & \vdots & \vdots \\
    0 & 0 & d_m & d_{m-2} & \vdots & \vdots & \vdots \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    0 & 0 & 0 & \ldots & 0 & 0 & d_m \\
\end{bmatrix}
\begin{bmatrix}
    q_0 \\
    q_1 \\
    q_2 \\
    \vdots \\
    q_{n-m-2} \\
    q_{n-m-1} \\
    q_{n-m} \\
    \vdots \\
    0 \\
\end{bmatrix} =
\begin{bmatrix}
    p_0 \\
    p_1 \\
    p_2 \\
    \vdots \\
    p_{n-3} \\
    p_{n-1} \\
    p_n \\
\end{bmatrix}
\]

Where \( D \) is a matrix of dimension \([n+1],(n-m+1)\], \( Q \) is an \([(n-m+1),1]\) vector and \( P \) is an \([(n+1),1]\) vector.

The orders of matrix \( D \) and vectors \( Q \) and \( P \) are related by

\[ [\binom{n+1}{n-m+1},\binom{n-m+1}{1}] = (n+1,1) \]
(7.177)

The order of the divisor polynomial must be either equal to or less
than the order of the dividend, thus \( n \) the number rows of matrix \( D \) is either equal to or greater than the number of columns of \( D \).

Moreover, it can be shown that for all but the zero matrix, \( D \) has full row and column rank. Thus (7.176) represents a set of linear equations containing \( m \) more equations than unknowns. If the \( d_i \) and \( p_i \) are known exactly, then any \( ((n-m)+1) \) of these equations will generate the unique quotient vector, \( Q \).

In the special case where \( R(s) = 0 \), a unique solution to (7.162) always exists.

Consider now, the more general case in which \( R(s) = 0 \). In vector matrix form (7.162) becomes,

\[
P = DQ + I'R
\]  

(7.178)

Where \( R \) is an \([m,1]\) vector containing the unknown remainder polynomial coefficients,

\[
R = \begin{bmatrix}
r_0 \\
r_1 \\
r_2 \\
\vdots \\
r_{m-3} \\
r_{m-2} \\
r_{m-1} \\
r_m
\end{bmatrix}
\]

and \( I' \) is a modified \([m,m]\) identity matrix, augmented vertically with \( n-m+1 \) rows of zeros thus,
So $I'$ is an $[(n+1),m]$ matrix, $R$ is an $(m \times 1)$ D vectors, $Q$ and $P$ have the same dimensions as in (7.176).

The vector matrix dimensions of (7.179) are related

$$[(n+1), (n-m+1)], [(n-m+1), 1] = [(n+1), m], [m, 1] = [(n+1), 1] \quad (7.180)$$

The vector matrix sum of the RHS of (7.178) is combined into one augmented vector, augment matrix,

$$P = [D \quad I'] \begin{bmatrix} Q \\ R \end{bmatrix} = \begin{bmatrix} D \quad I \\ 0 \end{bmatrix} \begin{bmatrix} Q \\ R \end{bmatrix} = FG \quad (7.181)$$

$F$ is an augmented matrix with $(n+1)$ rows by $(n+1)$ columns; $G$ is an $[(n+1), 1]$ vector. (7.178) represents a square system of linear equations and the number of equations is equal to the number of unknowns. It can be shown that $F$ always has full row and column rank and is therefore never singular.
To find the solution $G$ to (7.178), matrix $F$ must be inverted,

$$G = F^{-1}p$$  \hspace{1cm} (7.182)

Thus synthetic division of one polynomial by another can be recast as a vector matrix problem which is readily solved. The polynomials equation equivalent to (7.182) is a diophantine equation (7.178). The obvious relationship between the vector matrix expression (7.182) and the diophantine equation (7.178) indicates a vector matrix approach to the solution of more general diophantine equations.

With the exception of the synthetic division problem; the general vector matrix diophantine equation expression involves a non-square matrix. Thus Gauss elimination is, in general, not suitable for solving the diophantine vector matrix problem [96],[97]; unless the linear equation system is overdetermined and one can afford to disregard information, be eliminating rows to obtain a square matrix.

Unfortunately, removing rows to produce a square matrix, which represents a removal of information from the equation system, can lead to almost singular matrices in some cases.

A square matrix expression represents one equation for each unknown variable, which should be sufficient for solution if sufficient computation accuracy is available. However computation accuracy is finite for digital computers. Thus if more information is available, a better estimate of the solution can be made.
If the number of rows of the matrix exceeds the number of columns each variable is confined by more than one equation. Generally using finite arithmetic, each equation defining a variable will give a slightly different value, thus a means of arbitration between solution is necessary.

Consider the vector matrix expression,

\[ Ab = C \quad (7.183) \]

Where \( A \) is an \([n,m]\) matrix, \( n > m \) \( C \) is a known \([n,1]\) vector and \( b \) is a \([m,1]\) vector of unknowns.

If an estimate \( \hat{b} \) of \( b \) is obtained, an error function may be found

\[ \epsilon = \hat{A}\hat{b} - C \quad (7.184) \]

The solution to (7.184) is then defined as the minimum of the squared error \( \| \epsilon \|^2 \)

\[ \min \| \hat{A}\hat{b} - C \| \quad \text{from which } \hat{b} \rightarrow b \quad (7.185) \]

Differentiating (7.185) and equating the result to zero gives the location of the minimum which occurs at \([97],[98],[99]\),

\[ \hat{A}\hat{b} = C \quad (7.186) \]

To solve for \( \hat{b} \), the inverse of \( A \) must be found. Since \( A \) is non-square, the elementary ideas of matrix inverses must be extended, and the
concept of the generalized or pseudo matrix-inverse introduced [94],[95], [96],[97],[98],[99].

The properties of the pseudo-inverse of a matrix are defined by the Moore-Penrose conditions [97]

1. \( AXA = A \)
2. \( XAX = X \)
3. \( (AX)^T = AX \)
4. \( (XA)^T =XA \)

Where \( A \) is a general \([n,m]\) matrix of respective row and column rank \( n \) and \( m \).

\( X \) is the pseudo-inverse of \( A \).

The pseudo-inverse of a matrix will be denoted \( A^+ \). If the row rank of a matrix is equal to the column rank the pseudo-inverse of a matrix is equal to the actual inverse, that is \( A^+ = A^{-1} \) [97].

The most reliable method of finding the pseudo-inverse of a matrix is by the method of singular value decomposition. This method is a generalization of matrix inversion diagonalization of a square matrix by finding its eigen values and eigen vectors, and inverting the diagonal matrix [97],[98].

If two matrices, \( U \) and \( V \), exist such that,

\[
UU^T = I_{[n,n]} \quad \quad VV^T = I_{[m,m]} \tag{7.187}
\]

Where \( U \) is a square matrix of dimensions \([n,n]\),
\( V \) is a square matrix of dimensions \([m,m]\) and \( I_{m,n} \) and \( I_{m,m} \) are respectively the \([n,n]\) and \([m,m]\) identity matrices.

The location of vector \( b \), the solution to (7.185), will be unaffected if (7.185) is rewritten.

\[
\sigma' = \min \left| U^TAVV^Tb - U^TC \right|^2
\]

(7.188)

Where \( \sigma' \) is the new value of \((U^TAVV^Tb - U^TC)\) at the minimum.

Since (7.183) is solved by (7.184),

\[
Ab = C
\]

The solution to (7.188) is given by \( \hat{b} \) satisfying,

\[
U^TAVV^Tb = U^TC
\]

(7.189)

Denote the product \( U^TAV \) by the diagonal matrix \( D \) which has the form.
Where \( n > m \) and
D has dimension \([n,m]\).
The \( d_i \) are the singular values of \( A \).

Substituting \( D \) into (7.189) gives

\[
DV^Tb = U^TC
\]  
(7.190)

Using the pseudo-inverse \([97],[98]\),

\[b = (DV^T)^+U^TC\]  
(7.191)

Since \( V^T \) is square with inverse \( V \) \([97],[98]\), (7.187),

\[b = VD^+U^TC\]  
(7.192)

The pseudo-inverse of the diagonal matrix \( D \) is merely the transpose of
matrix \( D \) with each singular value replaced by its reciprocal thus \([97],[98]\),
\[
\begin{bmatrix}
\frac{1}{d_0} & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\
0 & \frac{1}{d_1} & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & \frac{1}{d_2} & 0 & 0 & 0 & \ldots & 0 & 0 \\
D^T = & 0 & 0 & 0 & \frac{1}{d_{m-2}} & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{d_{m-1}} & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{d_m} & \ldots & 0 & 0 & 0
\end{bmatrix}
\]

\(D^T\) is actually found using Householder's technique of tridiagonalization [94]. Givens rotations [97] are then applied to produce a diagonal matrix, whose elements are the singular values of \(A\).

Matrices \(U^T\) and \(V\) are the products of the matrices representing the Householder and Givens respective row and column operations [97].

An algorithm written by Golub to find singular values by the methods described is given in his paper [98]. The Golub singular value decomposition routine MINFIT was used extensively for the solution of diophantine equalization problems.

The solution to the diophantine equation generated by the synthetic division process has been considered. The new approach indicates a way in which diophantine equations may be solved.
Consider the diophantine equation derived for diophantine equalization equation (7.159) in its simplified form (7.162)

\[ D(-s)P(s)J(s) - D(s)P(-s)K(s) = C(-s)R(s)H(s) \]

\( D(-s) \) is a polynomial of order \( n \), \( J(s) \) and \( K(s) \) are polynomials of an equal order \( m \), \( P(s) \) is a polynomial of order \( p \) and is unknown. The polynomial product \( C(-s)R(s) \) is of order \( q \), and \( H(s) \) is or order \( r \) and also unknown.

The orders of the LHS are related to those of RHS by

\[ n+m+p=q+r \]  

(7.193)

In most solutions to (7.162) encountered, the order of \( P(s) \) is less than or equal to that of the product \( C(-s)R(s) \), that is,

\[ p \leq q \]

Solutions in which this is not the case do occur, but these are few; they lead to undetermined equations systems in which additional solution constraints must be imposed. Underdetermined equation systems will be discussed in the next section.

In vector matrix form (7.162) becomes,

\[ AB-CD=EF \]  

(7.194)
Where \( A \) is an \([(n+m+p+1),(p+1)]\) coefficient matrix of the known product \( D(-s) \)

\( B \) is a \([(p+1),1]\) vector of the unknown coefficients of \((s)\)

\( C \) is an \([(n+m+p+1)(p+1)]\) coefficient matrix of the known product \( D(s)N_s(s) \)

\( D \) is an \([(p+1),1]\) vector of the unknown coefficients of \( \eta(-s) \)

\( E \) is an \([(n+m+p+1),(n+m+p-q+1)]\) coefficient matrix of the known product \( N_1(s)\psi(s) \)

and \( F \) is an \([(n+m+p-q+1),1]\) vector of unknown coefficients of \( \theta(s) \)

Before producing an augmented vector matrix expression of (7.162) it is important to briefly examine the vector matrix expression \( CD \) which represents the polynomial product,

\[
D(s)N_s(s)(-s)=CD
\]  

(7.195)

Consider the polynomials \( P(s) \) and \( P(-s) \). The coefficients of even powers of \( s \) in \( P(s) \) are equal to those in \( P(-s) \), and the coefficient of odd powers in \( s \) of \( P(s) \) are equal to those of \( P(-s) \) in magnitude, but are of opposite sign. Therefore \( P(-s) \) is derived from \( P(s) \) by multiplying all odd coefficients of \( P(s) \) by \((-1)\). Thus \( D \) is related to \( B \).
Thus the general form of the vector matrix product is,

\[
B = \begin{bmatrix}
    b_0 \\
    b_1 \\
    b_2 \\
    b_3 \\
    \vdots \\
    b_{p-2} \\
    b_{p-1} \\
    b_p
\end{bmatrix}, \quad \quad \quad D = \begin{bmatrix}
    b_0 \\
    -b_1 \\
    b_2 \\
    -b_3 \\
    \vdots \\
    (-1)^{p-2} b_{p-2} \\
    (-1)^{p-1} b_{p-1} \\
    (-1)^{p} b_p
\end{bmatrix}
\]

(7.196)

Thus the general form of the vector matrix product is,

\[
\begin{bmatrix}
    c_0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
    c_1 & c_0 & 0 & 0 & 0 & 0 \\
    c_2 & c_1 & c_0 & \ldots & 0 & 0 \\
    \vdots & \vdots & c_2 & \ldots & \ldots & \ldots \\
    0 & \ldots & \ldots & \ldots & \ldots & \ldots \\
    0 & 0 & \ldots & c_n & c_{n-1} & c_{n-2} \\
    0 & 0 & 0 & 0 & c_m & c_{n-1} \\
    0 & 0 & 0 & 0 & 0 & c_n
\end{bmatrix} \begin{bmatrix}
    b_0 \\
    -b_1 \\
    b_2 \\
    -b_3 \\
    \vdots \\
    (-1)^{p-2} b_{p-2} \\
    (-1)^{p-1} b_{p-1} \\
    (-1)^{p} b_p
\end{bmatrix}
\]

Which is entirely equivalent to,
(7.197) illustrates a useful property of the vector matrix method of polynomial product representation.

Vector matrix equation (7.194) may be rewritten as

\[ AB + CB = EF \]  

(7.199)

\[ [A+C]B = EF \]  

(7.200)

Since \( F \) is also an unknown vector (7.200) may be expressed as an homogeneous vector-matrix equation [94],
\[ [A+C]B-EF = 0 \quad (7.201) \]

Let \([A+C]=H\) and \(W=-E\), then the augmented vector matrix expression for (7.201) may be found.

\[ [H;W] \begin{bmatrix} B \\ F \end{bmatrix} = 0 \quad (7.202) \]

The augmented coefficient matrix \([H|W]\) has dimensions \([n+m+p+1], (n+m+2p-q+2)\]. As mentioned \(p \leq q\) usually, so the augmented matrix has one more column than the number of rows. Augmented vector \([B]\) has dimensions \([(n+m+2p-q+2),1]\).

The augmented linear equation system is undetermined, and an infinity of solution to (7.202) exist. Additional information about the solution to (7.202) is required.

The definition of \(\eta(s)\) is given by equation (7.203),

\[ b(s) = \frac{\eta(-s)}{\eta(s)} = \sum_{i=1}^{n} \frac{s-\alpha_{i}}{s+\alpha_{i}} \quad \text{re} (\alpha) > 0 \quad (7.203) \]

From the definition, it is clear that the highest power of \(s\) is unity. Thus \(b_p = 1\). For this reason the pth column of \(H\) may be extracted forming a reduced matrix \(H_r\), a square augmented matrix, and a vector \(H_p\) representing the extracted column,

\[ \begin{bmatrix} H_r ; W \end{bmatrix} \begin{bmatrix} B_r \\ F \end{bmatrix} + H_p = 0 \quad (7.204) \]

Where \(H_r\) is matrix \(H\) with the pth column removed

\(H_p\) is the vector of the pth column of \(H\)

\(B_r\) is the vector \(B\) with the pth element removed.
With minor rearrangement (7.204) becomes

\[
\begin{bmatrix}
    H_r \\
    W
\end{bmatrix}
\begin{bmatrix}
    B_r \\
    F
\end{bmatrix} = -H_p
\] (7.205)

In most cases encountered, \([H_r|W]\) is an \([(n-m+p-l), (p+1)]\) of full row and column and row rank. To find \([B_r|F]\) the pseudo inverse of \([H_r|W]\) is found. Thus,

\[
\begin{bmatrix}
    B_r \\
    H
\end{bmatrix} = \left[H_r|W\right]^+ [-H_p]
\] (7.206)

Cases in which \([H_r|W]^+\) is not of full row and column rank will be discussed in the following section.

In order to avoid the trivial solution discussed in section 7.2.7, the concept of a bounding polynomial was introduced. The same bounding polynomial concept will be considered. In section 7.2.7 \(\eta(s)\) was defined as the sum of a PR bounding polynomial and an unknown NNR polynomial,

\[
p(s) = X(s) + Y(s)
\] (7.207)

Where \(X(s)\) is the bounding polynomial and is of order \(p\).

\(Y(s)\) becomes the unknown polynomial and is of order \((p-1)\).

Equation (7.167) is then rewritten,

\[
D(-s) J(s) X(s) + D(-s) J(s) Y(s) - D(s) K(s) X(-s) - D(s) K(s) Y(-s)
= C(-s) R(s) H(s)
\] (7.208)
Using (7.167) and (7.176), (7.208) may be written in augmented vector-matrix form,

\[
[H_r;W] \begin{bmatrix} Y \\ F \end{bmatrix} + [H_r;W] \begin{bmatrix} X \\ 0 \end{bmatrix} = 0 \tag{7.209}
\]

Where \( X \) is the known bounding polynomial coefficient vector, and \( Y \) is the bounded unknown polynomial. \( H_r, H, W \) and \( F \) have already been defined.

\( X \) is known, thus (7.209) becomes.

\[
[H_r;W] \begin{bmatrix} Y \\ F \end{bmatrix} = -[H_r;W] \begin{bmatrix} X \\ 0 \end{bmatrix} \tag{7.210}
\]

The RHS of (7.210) represents a vector of dimensions \([(n-m+p),1]\), therefore (7.210) is solved in the same way as (7.194)

\[
\begin{bmatrix} Y \\ F \end{bmatrix} = -[H_r;W]^+ [H_r;W] \begin{bmatrix} X \\ 0 \end{bmatrix} \tag{7.211}
\]

In equation (7.200) it was assumed that \( p \), the order of \( \Psi(s) \), is normally less than or equal to \( q \), the order of the polynomial product \( N_1(s)\Psi(s) \). In many practical cases this assumption is true, however there are cases where the assumption is not true. When \( p \leq q \) vector matrix equation (7.210) is either square or over-determined; there are always more equations than unknowns.

If \( p \geq q \), the vector matrix equation (7.210) represents an undetermined equation system, in which there are more unknowns than equations. In order to find a unique solution additional constraints, on possible solutions to (7.210), must be introduced.
In an undetermined system of linear equations, the choices of individual coefficients of $X(s)$ are not arbitrary. Once some are chosen their choice will fix remaining coefficients. For example, if the number of columns of $A$ is one greater than the number of rows, once one coefficient is chosen the values of all others will be fixed. If the number of columns exceeds the number of rows of $A$ by $i$, that is

$$m-n=i \quad (7.212)$$

Then $i$ coefficients may be chosen arbitrarily, and the choice will fix the remaining $n$ coefficients.

Therefore, in order to solve an undetermined equation system, some restriction, on the choice of arbitrary variables, must be found such that the $i$ arbitrary values and $n$ fixed values satisfy any given constraint on the solution.

In the diophantine equation context, it is desired to find the coefficients of $X(s)$ such that $X(s)$ is a NNR polynomial. If $A$ is undetermined, then $i$ coefficients of $X(s)$ may be chosen arbitrarily, the remaining $n$ coefficients will be fixed. The combination of $i$ arbitrary and $n$ fixed coefficients may not produce a NNR polynomial. The NNR nature of $X(s)$ is thus dependent, not only on the numerical choices of the $i$ coefficients but on which $i$ coefficients, of $X(s)$, are arbitrarily chosen.

The possible effects of choosing $i$ coefficients of $X(s)$ arbitrarily are shown graphically in Fig 7.8.
Three possible solution regions are shown in Fig 7.8. Solution set #1 is obtained when the arbitrary choice of coefficients would give rise to a negative real polynomial and the fixed coefficient additionally give rise to a negative real polynomial. Solution set #2 is obtained when the arbitrary coefficients are chosen, from those that may give rise to a NNR polynomial, but those fixed coefficients do not. Solution set #3 represents a possible viable solution, where the arbitrary coefficients are chosen from those producing a NNR polynomial and from the choice, the remaining coefficients give rise to a NNR polynomial.

The Boundary in Fig 7.8 represents the choice of zero coefficients in \( X(s) \), and the corresponding coefficients of \( \eta(s) \) would be those of the bounding polynomial.

It was decided that the least restrictive choice of arbitrary polynomial coefficients would be on the boundary, thus all arbitrary chosen coefficients of \( X(s) \) would be set to 0. The corresponding columns of \( A \) are eliminated until the coefficient matrix is square. Using the new reduced
matrix, the remaining coefficients of $X(s)$ are found. $X(s)$ is then tested to see whether or not it is NNR. If not, a new combination of coefficients is set to zero and the process repeated until a NNR is found. The process is shown diagramatically in Fig 7.9.

Fig 7.9 Searching for Non-Negative Real $X(s)$

If all combinations of arbitrary coefficients are tried and none produce a NNR polynomial. The gain term $k$ is reduced and the process repeated.

If it is found that matrix $V$ is singular, but the number of columns does not exceed the number of rows, the problem is similar to that discussed, and the sequential search process is applicable also.

7.2.9 DEQDIO - A Program For Load Equalization

Based upon the LDE approach to the equalization problem, discussed in sections 7.2.4 to 7.2.7, and the methods of solving the diophantine
equations discussed in section 7.2.8, a computer program was written to automatically equalize an impedance function. Titled DEQDIO (Double precision EQualization by the DIophantine equation technique) it was written to take an arbitrary load, described by a rational function and equalize it over a prescribed bandwidth using a Butterworth response [5], a Chebyschev response [5], an elliptic filter response [5] a Bessel Filter response [87] or an arbitrary input response given by the user.

Once DEQDIO is given the load rational function, the equalization bandwidth, and either the order of the desired Butterworth, Chebyschev, Elliptic or Bessel response, or the user input response, it should automatically solve the equalization problem, optimizing the DC transfer gain [3],[4],[5], k. Once the back impedance Z_2(s) is found, DEQDIO realizes the back impedance using the techniques outlined in section 7.3.

Given the rational function of impedance Z_L(s), DEQDIO scales it to a unit bandwidth [34],[45] and 1Ω DC impedance [34],[45] using the user input equalization bandwidth. During scaling the largest values of coefficients are tested to ensure that overflows do not occur.

Once scaled, a constant term is removed from Z_L(s) to produce a unit coefficient of the highest powers of s in the numerator and denominator polynomials of Z_L(s).

If Z_L(s) contains any imaginary axis poles these must be removed. It is desirable not to resort to zero searches, of D(s), to remove these poles. A complete power series polynomial expression including all imaginary axis poles can be removed from D(s) by forming a Routh array of D(s) [88],[89],[90],[91], as if testing it for stability.
A Routh array produces a Row of Zeros, if zeros of the original polynomial occur with their negative forms, that is of \( D(s) \) contains a group of zeros \([88],[89],[90],[91]\),

\[(s+\alpha)(s+\alpha^*)(s-\alpha)(s-\alpha^*)\]

But \( D(s) \) cannot contain poles in the RHP as \( Z_0(s) \) is PR, therefore the only zeros of \( D(s) \), which can occur with their associated negative forms, are those which occur on the imaginary axis where their complex conjugate is the associated negative value of the zero.

Consider, for example, the polynomial \( D(s) = (s+j)(s-j)(s+2) \). The corresponding series expansion is

\[ D(s) = s^3 + 2s^2 + s + 2, \]

The corresponding Routh array is \([88],[89],[90],[91]\),

| \( s^3 \) | \( 1 \) | \( 1 \) | \( s^3 + s \) |
| \( s^2 \) | \( 2 \) | \( 2 \) | \( 2s^2 + 2 \) |
| \( s^1 \) | \( 0 \) | \( 0 \) | \( 0 \) |
| \( \frac{d}{ds} \) | \( s^1 \) | \( 4 \) | \( 0 \) |
| \( s^0 \) | \( 8 \) | \( 0 \) | \( 0 \) |

The array has a zero row at \( s^1 \). To continue the array the polynomial corresponding to \( s^2 \) is differentiated \([88],[89]\). But the row corresponding to \( s^2 \) can be scaled by a constant to give \( s^2 + 1 \) which is a power series form of the polynomial containing \((s+j)(s-j)\) \([88],[89]\).

In this way, the polynomial containing all imaginary axis zeros of \( D'(s) \) is found. A new value of \( D(s) \) is found by dividing \( D'(s) \) by the imaginary axis polynomial thus
\[ Z_L(s) = \frac{N_L(s)}{D(s)T(s)} \tag{7.213} \]

Where \( T(s) \) denotes the imaginary axis polynomial.

The ZOT are found by forming the real part of the load.

\[ r_L(s) = \frac{N(s)D'(s) + N(-s)D'(s)}{D(s)D'(-s)T(s)} \tag{7.214} \]

The numerator of \( r_L(s) \) contains the ZOT to their full multiplicities, as they would appear in \( Z(s) \). There is no advantage in actually forming \( T(s) \), as further division into ZOT classes would be required to ascertain their full multiplicities in \( r_L(s) \) [3],[4],[5].

The numerator of \( r_L(s) \) is spectrally factorized to produce \( C'(s) \) and \( C'(-s) \). \( C(-s) \) contains the RHP zeros of \( r_L(s) \). If any imaginary zeros of \( r_L(s) \) are present, they must be of even multiplicity due to the PR nature of \( Z_L(s) \). The process of spectral factorization divides them evenly between \( C'(s) \) and \( C'(-s) \). But all imaginary axis ZOT must be included with \( C'(-s) \) when (7.233) is solved. Using the Routh technique discussed earlier [88],[89], the imaginary axis ZOT are removed from \( C'(s) \) producing a quotient \( C(s) \) and a polynomial \( P(s) \) containing the ZOT to half their multiplicity. If it is desired to relax the ZOT requirements and include the imaginary axis zeros once in the equalizer, the polynomial, corresponding to the ZOT occurring only once, must be found. Once again the Routh Array can be used. By differentiating the row of the Routh Array preceding the zero row, one may proceed with the array as shown earlier [88],[89]. If a second row of zeros is obtained, the polynomial corresponding to the row preceding the second zero row contains the zeros with a multiplicity of 2 or greater. For example the polynomial specified...
by the zero product \((s+j)^2(s-j)^2(s+2)\) is used to form the following Routh array.

\[
\begin{array}{cccc}
1 & S^5 & 1 & 2 & 1 & S^5 + 2S^3 + S \\
2 & S^4 & 2 & 4 & 2 & 2S^4 + 4S^2 + 2 \\
3 & S^3 & 1 & 2 & 1 & S^3 + 2S^2 + 1 \\
4 & S^2 & 0 & 0 & & \\
5 & S^3 & 4 & 4 & 4S^3 + 4S \\
6 & S^3 & 1 & 1 & S^3 + S \\
7 & S^2 & 1 & 1 & S^2 + 1 \\
8 & S & 0 & 0 & & \\
9 & S & 2 & & 2S \\
S^0 & 1 & & & &
\end{array}
\]

Row 3 corresponds to the polynomial \(S^4 + 2S^2 + 1\) which has zero factors \((S+1)^2(S-1)^2\). Row 7 corresponds to the polynomial \((S^2 + 1)\) which has zero factors \((S+j)(S+j)\).

The polynomial corresponding to the Routh array row preceding the second row of zeros contains imaginary zeros of multiplicity one less than the row preceding the first row of zeros. By dividing the polynomial corresponding to the row preceding the first row of zeros by that corresponding to the second row of zeros, a polynomial \(P_1(s)\) is generated containing all imaginary axis zeros but of unit multiplicity. The remaining zeros corresponding to those of \(P_1(s)\) of multiplicity of 2 or greater are contained in the polynomial preceding the second row of zeros in the Routh array \(P_2(s)\).
If it is decided to relax the constraints on the difference $b(s) - S(s)$ and permit imaginary axis zeros to occur in $Z_2(s)$ the class IV ZOT must be extracted from $\rho_1(s)$ since these must occur in (7.233) to their full multiplicities [3],[4],[5].

Forming yet another Routh array, any common factors between $\tau(s)$ and $\rho_1(s)$ may be found. In this case the Routh array is formed between $\rho_1(s)$ and $\tau(s)$ [73],[88],[89]. The polynomial of higher order forms the first row, the polynomial of lower order the second. A Routh array is then formed [73],[88],[89]. Each row is divided by the leading term to produce a positive leading term of unit magnitude. If a row of zeros is found, the row preceding it is the polynomial containing the common factors [73],[88],[89]. Once the common factors are found they are removed from $\rho_1(s)$ by division, and multiplied into $\rho_2(s)$.

The following Routh array illustrates the process of finding common factors between the series expansions of $(S+j)(S-j)(S+2)(S+3)$ and $(S+j)(S-j)$. 


The row preceding the zero row corresponds to the equation \((s^2+1)\) which contains the zeros \((s+j)\) \((s-j)\). These zeros are the common factors of \(s^4 + 5s^3 + 7s^2 + 5s + 6\). In this case, \(\tau(s)\) is a factor of \(\rho_1(s)\). However this is not always the case. Let \(\mu(s)\) denote the common factors of \(\tau(s)\) and \(\rho_1(s)\), then the zeros of transmission that must occur in (7.222) are given by the product

\[
C(s) \rho(s) \mu(s)
\]

In which case zeros corresponding to \(\rho_1(s)\) will appear in \(Z_2(s)\). If it is desired to produce a \(Z_2(s)\) in which \(\rho_1(s)\) does not appear, all RHP and imaginary axis zeros must appear in (7.159), in which case the product,

\[
C(s) \rho(s) \mu(s) \rho_1'(s)
\]

where \(\rho_1'(s)\) is given by the quotient \(\rho_1(s)/\mu(s)\), must be present in the RHS of (7.159).

| \(s^4\) | 1 | 5 | 7 | 5 | 6 |
| \(s^4\) | 1 | 0 | 1 | 0 |
| \(s^3\) | 5 | 6 | 5 | 6 |
| \(s^3\) | 1 | 1.2 | 1 | 1.2 |
| \(s^3\) | -0.2 | 0 | -0.2 |
| \(s^2\) | 1 | 0 | 1 |
| \(s^2\) | 1.2 | 0 | 1.2 |
| \(s^2\) | 1 | 0 | 1 |
| \(s\) | 0 | 0 | 0 |
| \(s\) | 2 | 0 | 2s |

etc.
With the RHS of (7.159) complete, it may be solved by selecting a unit value of DC gain \( k \) and producing a corresponding for \( S_0(s) \). Values of \( \beta(s) \) and \( H(s) \) are found. \( \beta(s) \) is then tested using a Routh Array to ensure that it is non-negative real (NNR), that is, its zeros can lie in the LHP or on the imaginary axis. They may be of any multiplicity, but they cannot lie in the RHP. If these requirements are satisfied, the bounding polynomial ensures that the zeros of \( b(s) \), given by (7.159),

\[
b(s) = \alpha(s) + \beta(s) \tag{7.215}
\]

lie within the LHP only, and that none lie on the imaginary axis.

If the value of \( k \) chosen produces a non-NNR \( \beta(s) \) then \( k \) is reduced by 0.1 and the process repeated until either \( k = 0 \) in which case an equalizer cannot be found, or \( \beta(s) \) is NNR in which case the optimum value of \( k \) is bounded by the last value of \( k_{lb} \) which produced a NNR \( \beta(s) \) and the preceding value of \( k \), \( k_{ub} \) which produced a non-NNR polynomial \( \beta(s) \). Thus

\[
k_{lb} \leq k_{opt} < k_{ub}
\]

Using the method of bisection [23], new values of \( k \) are generated, that is

\[
k = \frac{k_{lb} + k_{ub}}{2}
\]

If \( \beta(s) \) is NNR \( k_{lb} \) is updated, \( k_{lb} = k \), if \( \beta(s) \) is Non-NNR, \( k_{ub} \) is updated. This process is contained until the difference between \( k_{lb} \) and \( k_{ub} \) is very small in which case \( k_{lb} \) is the desired optimum value of \( k \), \( k_{opt} \).
In DEQDIO, a check is also conducted on \( H(s) \) to ensure that its zeros lie in the LHP or if on the imaginary axis, they are of unit multiplicity again using a Routh array. However this test is performed only if \( \beta(s) \) is NNR.

In addition, \( Z_2(s) \) is tested when \( \beta(s) \) is NNR and \( H(s) \) complies with PR conditions 2, using a Routh test on the numerator ensuring that the residues at any imaginary axis zero of \( H(s) \) \( \tau(s) \) have positive residues as poles of \( Z_2(s) \), and a Sturm test to ensure that \( \text{Re}(Z_2(j\omega)) \geq 0 \) [57], [88]. The Sturm test is conducted using a modified Routh Array [88].

If the above restrictions are all satisfied and \( k_{opt} \) is found, the back impedance corresponding to an optimum gain has been found. \( Z_2(s) \) can be synthesised using the techniques discussed in the following section 7.2.

During the calculations to obtain a \( Z_2(s) \), error can accumulate in the iteration cycle. Unfortunately errors can accumulate which prevent one from obtaining rows in Routh or Sturm arrays which are identically zero, or indeed below a tolerance about zero. It can also become difficult to decide whether a polynomial coefficient is itself zero. For these reasons estimates of maximum error are made at each addition, subtraction, multiplication or division of one polynomial by another, and carried as a parameter with each polynomial in the same way as the polynomial order.

When it is read into the DEQDIO Program, each polynomial is tested to find the largest coefficient. The largest coefficient is then multiplied by the zero tolerance parameter to produce an absolute polynomial error value. Any coefficient below their value is regarded as zero.
If two polynomials are added or subtracted, the error of the resulting polynomial is the sum of each polynomial error. When two polynomials are multiplied or divided the error value for the new polynomial is calculated by finding the largest coefficients in the factor polynomials, dividing each respective polynomial error by the largest coefficient, and adding the two together to produce a relative result error. The absolute polynomial error, of the result, is found by multiplying the relative error by the largest coefficient of the resultant polynomial.

Finally, if a polynomial is multiplied by a constant with which no error is associated, the polynomial error is also multiplied by that constant.

In the way described above, an estimate of the error in each polynomial coefficient of the result is obtained. The error is particularly important in realizing the back impedance as discussed in section 7.3.

Using the techniques discussed above, the program DEQDIO was written. It is a large program written in Fortran 77. Consisting of some 3000 lines of main program and up to 5000 lines of subroutines.

7.2.10 Realizing the Equalizer Back-Impedance

Youla's technique of load equalization [3] is intended to match a resistive voltage source to an arbitrary load, over a specified frequency
range, such that the power transfer characteristic complies with a desired response over that frequency range.

By realizing the back impedance, the values of components required to build the equalizer are found. The voltage source is naturally not included, as a perfect voltage source is a generalized short circuit [3], [4], [5]. All that is seen of the voltage source is its internal resistance [3].

If a voltage source is to be connected to the equalizer circuit, it should be connected such that it provides the resistance for the equalizer [3], [4], [5]. The equalizer input terminals are those across which the terminating resistor is found. The voltage source is connected in series with the terminating resistor [3], [4], [5]. The terminating resistor itself forms the voltage source internal resistance. If the equalizer terminating impedance is not equal to the driving point impedance, a transformer is used to couple the resistive voltage source to the equalizer with a turns ratio selected to produce the required terminating resistance [3], [4], [5].

If additional resistors appear in the equalizer realization, the optimum power transfer is reduced. Moreover, the resistors will in general be associated with a resistive voltage source. The question then arises as to which voltage source is that corresponding to the desired input point for the equalizer.

To avoid this dilemma therefore, the back impedance $Z_2(s)$ must be realized as a reactive load terminated in a resistor. In section 7.3, the technique of producing a back impedances of the required form are discussed.
7.3 LOAD SYNTHESIS

The rational functions of the preceding two sections represent impedances or admittances which are dependent upon frequency. In both sections, the solutions to curve fitting and equalization problems were confined to PR rational functions only. If the 3-PR conditions are satisfied, the rational function of impedance or admittance can be realized with a passive network of real inductors, capacitors and resistors [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40].

In this section, the process of load realization as an interconnection of passive electronic components is discussed. Guilleman [35], [100], [101], discuss methods of synthesizing loads by means of least squares regression of parallel RLC components connected between the points of grid. Such techniques are entirely computer based and require vast amounts of computer memory and time. Moreover, it is unlikely that such techniques will produce networks with the minimum numbers of components.

In this chapter, therefore, the more traditional analytical techniques of load synthesis [29], [30], [31], [32], [35], [37], [38], [39] are discussed. These were combined into a load synthesis program to realize equalizer back impedances specifically. Naturally however, it may be used to synthesise any PR rational function of impedance or admittance.

The requirements of a suitable network, to realize an equalizer back-impedance function, were discussed in section 6.2.7. The most important single requirement, in the synthesis of an equalizer by this method, is that the resulting circuit must be an interconnection of reactive components only, terminated in a resistor, which is the internal
resistance of a voltage source driving the equalizer and load circuits [3], [4], [5].

The load synthesis techniques discussed in this section produce ladder networks [29], [30], [31], [32], [35], [37], [38], [39], therefore the actual synthesis process applied to the rational function of impedance or admittance must be chosen so that it introduces a resistor as the terminating component only [3], [4], [5].

In section 7.3.1 the techniques of Cauer Load synthesis are discussed. If a low pass or high pass rational function is synthesised, in which no changes in characteristic impedance occur, Cauer synthesis will generate a ladder network terminated in a resistor with a minimum number of reactive components [4], [5], [34], [37], [38], [39]. Unfortunately the Cauer synthesis process can terminate prematurely leaving rational function which remains to be synthesised. Cauer Synthesis cannot continue the realization process in such situations [4], [5], [34], [35], [37], [38], [39].

Fortunately with the use of Brune Synthesis [34], [35], [38], a network can be synthesised to reduce the rational function and permit Cauer synthesis to be continued until it either, again, prematurely terminates or the load is synthesized. Brune synthesis cannot however be used in all premature termination cases, as it may require the insertion of a series resistance [34], [35], [38]. This would violate the equalizer back-impedance synthesis requirements. The techniques of Brune Synthesis are discussed in section 7.3.2.

When Brune synthesis cannot be applied to realise the rational function the more powerful techniques of Darlington synthesis can be
applied [32],[33],[35],[38],[40]. The traditional techniques of applying Darlington synthesis are, however, unsuitable for computer application as they require the knowledge of a number of poles and residues of the rational function [33],[35],[38],[40]. Fortunately using the theory of LDE's a useful closed solution Darlington synthesis technique was developed, which was much more amenable to computer based load synthesis [32]. Aspects of Darlington synthesis and its application to computer load synthesis are covered in section 7.3.3.

Using a combination of Cauer and Darlington synthesis techniques, a load realization computer program was developed. Known as D_REAL, it is reviewed in section 7.3.4. The accumulation of errors during synthesis and ways of overcoming this, are also discussed.

7.3.1 Cauer Load Synthesis

Cauer Load Synthesis involves the removal of imaginary axis poles of impedance or admittance from the given rational function, to specify the value of a particular reactive component or component combination and reduce the order of the rational function polynomials [4],[5],[34],[35],[37],[38],[39].

Consider a general function of the form, (7.216),

\[ Z_1(s) = \frac{N_1(s)}{D_1(s)} \]  

(7.216)
Where $N_1(s)$ and $D_1(s)$ are numerator and denominator polynomials in $s$, respectively.

As discussed in section 7.3, if $N_1(s)$ is of order 1 greater than $D_1(s)$ then $Z(s)$ has a pole at infinity. ($N_1(s)$ and $D_1(s)$ cannot differ in order by more than 1, as $Z(s)$ is PR [57]), therefore at infinity $Z(s)$ is itself infinite. Such a response can be introduced by an inductor connected in series with a unknown load [4],[5],[34],[35],[37],[38],[39]. By dividing $D_1(s)$ into $N_1(s)$ the inductor value and remainder impedance are found [4],[5].

$$Z_1(s) = \frac{N_1(s)}{D_1(s)} = sL + \frac{N_2(s)}{D_1(s)} \quad (7.217)$$

Let

$$N_1(s) = n_m s^m + n_{m-1} s^{m-1} + n_{m-2} s^{m-2} + ... + n_3 s^3 + n_2 s^2 + n_1 s + n_0 \quad (7.218)$$

and

$$D_1(s) = d_{m-1} s^{n-1} + d_{m-2} s^{n-2} + d_{m-3} s^{n-3} + ... + d_3 s^3 + d_2 s^2 + d_1 s + d_0 \quad (7.219)$$

Where $m$ is the order of the numerator polynomial, then

$$L = \frac{n_m}{d_{m-1}}$$

and

$$N_2(s) = \frac{n_{m-1} d_m - n_{m-2} d_{m-2}}{d_{m-1}} s^{m-1} + \frac{n_{m-2} d_m - n_{m-3} d_{m-3}}{d_{m-1}} s^{m-2} + ... + \frac{n_2 d_{m-2} d_m - n_1 d_{m-1} d_m}{d_{m-1}} s + n_0 \quad (7.220)$$

Therefore $N_2(s)$ is of order 1 lower than $N_1(s)$, in general [4],[5],[34],[35],[37],[38],[39].
Similarly if $D_1(s)$ is of order $n$, and 1 greater than the order of $N_1(s)$, then $Z_1(s)$ has a zero at infinity. If the reciprocal of $Z_1(s)$ is taken, the corresponding admittance $Y_1(s)$ is found. It has a pole of admittance at infinity which corresponds to a capacitor in parallel with a remainder of admittance [4],[5],[34],[35],[37],[38],[39], thus,

$$y_1(s) = \frac{D_1(s)}{N_1(s)} = sC + \frac{D_2(s)}{N_1(s)}$$

(7.221)

Using representations similar to (7.218) and (7.219) of appropriate order,

$$C = \frac{d_m}{n_{m-1}}$$

and

$$D_2(s) = \frac{d_{m-1} - d_m n_{m-2}}{n_{m-1}} s^{m-1} + \frac{d_{m-2} - d_m n_{m-3}}{n_{m-1}} s^{m-2} + \ldots + \frac{d_2 - d_m n_1}{n_{m-1}} s^2 + \frac{d_1 + d_m n_0}{n_{m-1}} s + d_0$$

(7.222)

If the leading term in either (7.220) or (7.222) is zero, an impedance or admittance with a pole at infinity generates a remainder function with a zero at infinity. By inverting the remainder the extraction of a component can be repeated [4],[5],[34],[35],[37],[38],[39].

When the leading terms in either (7.220) or (7.222) are non zero the process of extracting poles at infinity ceases.

The process of extracting poles of impedance or admittance generates a low pass reactive ladder network as shown in Fig 7.10 [4],[5],[34],[35],[37],[38],[39].
A general impedance $Z_1(s)$ an admittance $Y_1(s)$ may have a pole at DC, $s=j\omega$, in which case the pole lies on the imaginary axis. An impedance pole at DC corresponds to a capacitor in series with an unknown circuit [32],[35],[37],[38],

$$Z_1(s)=\frac{N_1(s)}{D_1(s)} = \frac{1}{sC} + \frac{N_2(s)}{D_2(s)} \quad (7.223)$$

and a pole of admittance at DC corresponds to a parallel inductor [32],[35],[37],[38].

$$Y_1(s)=\frac{D_1(s)}{N_1(s)} = \frac{1}{sL} + \frac{D_2(s)}{N_2(s)} \quad (7.224)$$

But the residues at DC in (7.223) and (7.224) are $1/C$ and $1/L$ respectively.

Let $\alpha$ represent the residue of impedance at DC, then using (7.223),
\[
\frac{N_1(s)}{D_1(s)} = \frac{\alpha D_2(s) + N_2(s)}{sD_2(s)} \tag{7.225}
\]

Whence

\[
D_1(s) = sD_2(s) \tag{7.226}
\]

and

\[
N_1(s) = \alpha D_2(s) + N_2(s) \tag{7.227}
\]

\(D_2(s)\) is found by dividing \(D_1(s)\) by "s", \(\text{(7.226)}\). Using the resulting polynomial, the LDE \(\text{(7.227)}\) is solved for \(\alpha\) and \(N_2(s)\), giving a series capacitor of value \(C = \frac{1}{\alpha}\) and a remaining impedance \([32],[35],[37],[38]\),

\[
z_2(s) = \frac{N_2(s)}{D_2(s)} \tag{7.228}
\]

If \(Y_1(s)\) has a pole of admittance at zero, \(Y_2(s)\) is found in the same way giving

\[
L = \frac{1}{\alpha}
\]

and \(y_2(s) = \frac{D_2(s)}{N_2(s)} \tag{7.229}\)

If the remaining impedance or admittance has a pole at DC, the process is repeated.

A general impedance \(Z_1(s)\) or admittance \(Y_1(s)\) may also have imaginary axis poles at finite values. Such poles may also be extracted by Cauer Synthesis \([32],[35],[37],[38]\). A pole of impedance at a finite value of frequency corresponds to a lossless parallel resonant circuit in series with the remainder of the network \([32],[35],[37],[38]\). A pole of
admittance at a finite value of frequency corresponds to a series resonant

circuit in parallel with the remainder of the network [32],[35],[37],[38].

Let \((S^2 + \alpha)\) represent a pole of impedance at frequency \(s=\pm j\sqrt{\alpha}\).

Then

\[
\frac{N_1(s)}{D_1(s)} = \frac{\beta_2}{s + \alpha} + \frac{N_2(s)}{D_2(s)}
\]

\[
= \frac{\beta D_2(s) + (s^2 + \alpha)N_2(s)}{D_2(s)(s^2 + \alpha)}
\]

Where \(\beta\) is the residue of impedance at \(S = \pm j\sqrt{\alpha}\).

Equating the numerator and denominator polynomials,

\[
D_1(s) = D_2(s)(s^2 + \alpha)
\]

and

\[
N_1(s) = \beta s D_2(s) + (s^2 + \alpha)N_2(s)
\]

\(D_2(s)\) is found by dividing the imaginary axis pole out of \(D_1(s)\),
as indicated by (7.232). \(N_2(s)\) and \(\beta\) are found by solving LDE (7.233).

The capacitor of the resonant circuit is given by \(C = 1/\beta\) and the
inductor in parallel, with it, by \(L = \beta/\alpha\) [32],[35],[37],[38].

A finite pole of admittance is removed in the same way. The inductor
of the corresponding series resonant circuit is given by \(L = 1/\beta\), and the
capacitor by \(C = \beta/\alpha\) [32],[35],[37],[38].
Once all imaginary axis poles of impedance and admittance have been extracted, a rational function comprising polynomials of equal order containing no imaginary axis poles or zeros remains. Cauer synthesis can proceed no further [32],[35],[37],[38].

If the numerator and denominator of the remaining rational function are constants, synthesis is complete, and the ratio of the two constants gives the resistance or conductance of the terminating resistor [32],[35],[37],[38]. If the remaining rational function polynomials are of order greater than, or equal to, 1, synthesis is incomplete and further synthesis requires application of Brune or Darlington Synthesis, covered in sections 7.3.2 and 7.3.3 respectively.

7.3.2 Brune Synthesis

After application of Cauer Synthesis, the remaining poles of impedance or admittance must lie in the LHP. By removing appropriate components from the impedance function in series with remaining load, a pole of admittance may be introduced into the remainder [34],[35],[38]. This can be realized using Cauer synthesis. The values of components removed must be chosen to ensure that the remainder is PR [34],[35],[38].

Consider the real part \( r_1(s) \) of a general impedance \( Z_1(s) \) which is devoid of imaginary axis poles and zeros [34],[35],[38].

\[
 r_1(s) = \frac{z(s) + z(-s)}{2} \tag{7.234}
\]
As for equalization, \( r_1(s) \) may be written in terms of numerator and denominator polynomials \( N_1(s) \) and \( D_1(s) \) respectively, of the impedance \( Z_1(s) \),

\[
    r_1(s) = \frac{N_1(s)D_1(-s) + N_1(-s)D_1(s)}{D_1(s)D_1(-s)} \quad (7.235)
\]

Due to the PR nature of \( Z_1(s) \), the zeros of \( r_1(s) \) appear in quadrantal symmetry \([34],[35],[38]\). Moreover PR condition 3 requires that any purely imaginary axis zeros of \( r_1(s) \) must be of even multiplicity.

If \( r_1(j\omega) \) is plotted as a function of \( \omega \), the resulting graph will lie above the \( \omega \) axis. If any purely imaginary axis zeros are present in \( r(s) \), \( r_1(j\omega) \) will touch the \( \omega \) axis, it will not cross it however \([34],[57]\).

A typical plot of \( r_1(j\omega) \) is shown in Fig 7.11

**Fig 7.11** A Typical Impedance Plot of the Real Part

![Impedance Plot](image)
One may subtract a constant from the real part of the impedance such that the global minimum of \( r(j\omega) \) touches the \( j\omega \)-axis, in which case 4 purely imaginary axis zeros occur in the remainder \([34],[57]\).

\[
r_1(s) = R + r_2(s) (s^2 + \alpha)^2
\]

(7.236)

Where \( r_2(s) \) is a remainder function which is PR.

If a larger value of resistance is subtracted, the remainder will cease to be PR.

Subtracting \( R \) from \( r_1(s) \) is equivalent to subtracting it from \( z_1(s) \) giving a remaining impedance \( z_2(s) \) \([34],[35],[38]\).

\[
z_1(s) = R + z_2(s)
\]

(7.237)

At frequency \( s = \pm j\sqrt{\alpha} \), the real part of the load \( r_1(s) \) is zero, therefore subtracting \( R \) from \( z_1(s) \) produces an impedance \( z_2(s) \) which is purely imaginary at \( s = \pm j\sqrt{\alpha} \). If an inductor which produces the same impedance as \( z_1(s) \) at \( s = \pm j\sqrt{\alpha} \) is subtracted from \( z_2(s) \), then the remaining function \( z_3(s) \) will have a zero at \( s = \pm j\omega \) \([34],[35],[38]\).

\[
z_2(s) = sL_1 + z_3(s)
\]

\[
= sL_1 + \frac{N_3(s)(s^2 + \alpha)}{D_3(s)}
\]

(7.238)

where

\[
z_3(s) = N_3(s)(s^2 + \alpha)
\]

The value of inductor, required to yield the zero of impedance, is given by \([34],[35],[38]\),
Continuing, then, with Cauer synthesis, will produce a series resonant circuit in parallel with a remaining circuit [34],[35],[38],

\[ y_3(s) = \frac{D_3(s)}{N_3(s)(s^2+\alpha)} \]  

(7.240)

\[ = \frac{D_4(s)}{N_3(s)} = \frac{\beta s}{s^2+\alpha} \]

The inductor of this series resonant circuit has value \( L_2 = 1/\beta \) [34],[35],[38], the associated capacitor is of value \( C = \beta/\alpha \) [34],[35],[38].

The remaining impedance is therefore [34],[35],[38],

\[ z_4(s) = \frac{N_3(s)}{D_4(s)} \]  

(7.241)

If \( N_1(s) \) is of order \( n_1 \) then \( D_1(s) \) is also of order \( n_1 \). By performing the steps above a remainder rational function is produced, its numerator \( N_3(s) \) is of order \( n_1-1 \), its denominator \( D_4(s) \) of order \( n_1-2 \) [34],[35],[38]. Again using Cauer synthesis, a third inductor is extracted and \( N_3(s) \) becomes a polynomial \( N_4(s) \) of order \( n_1-2 \) [34],[35],[38].

Using the above procedure, the order of the original rational function polynomials can be reduced by 2 orders, yielding a circuit of the form shown in Fig 7.12 [34],[35],[38].
The steps described above are typical steps in Brune Synthesis [34], [35], [38]. This process is repeated on remainder rational functions until a constant term is reached. In this way a ladder of passive components terminated in a resistor is generated [34], [35], [38].

If the real part of the impedance $Z(s)$ contains one or more sets of imaginary axis zeros of even multiplicity, the removal of the initial resistor is omitted [34], [35], [38].

Unfortunately the synthesis requirements outlined in section 7.2.7 are satisfied only when Brune Synthesis can be conducted without removing a series resistor, therefore Brune Synthesis is not sufficiently general to produce reactive equalizers terminated by one resistor in all cases.

In an attempt to remove the need for a zero search of $r_1(s)$, an iterative method of inductor extraction, to find the inductor value required to produce a remainder, with an imaginary axis zero, when extracted, was devised.
Consider again, equation (7.238),

\[ z_2(s) = sL_1 + \frac{N_3(s)(s^2 + \alpha)}{D_3} \]

If \( z_2(s) \) is written in terms of its numerator and denominator polynomials,

\[
\frac{N_2(s)}{D_2(s)} = \frac{sL_1 + N_3(s)(s^2 + \alpha)}{D_3(s)}
\]  \hspace{1cm} (7.242)

But \( D_2(s) = D_3(s) \) therefore,

\[
N_2(s) = sL_1D_3(s) + N_3(s)(s^2 + \alpha)
\]  \hspace{1cm} (7.243)

Which is an LDE. If \( N_2(s) \) is a polynomial of order \( n \) then \( sL_1D_3(s) \) is a polynomial of order \( n+1 \) and \( N_3(s)(s^2 + \alpha) \) represents a polynomial of order \( n+1 \) also.

Let \( N_2(s) \), \( D_3(s) \) and \( N_3(s) \) be represented in terms of their power series expansions,

\[
N_3(s) = a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \ldots + a_2s^2 + a_1s + a_0
\]  \hspace{1cm} (7.244)

\[
D_3(s) = b_ns^n + b_{n-1}s^{n-1} + \ldots + b_2s^2 + b_1s + b_0
\]  \hspace{1cm} (7.245)

and \( N_2(s) = c_ns^n + c_{n-1}s^{n-1} + \ldots + c_2s^2 + c_1s + c_0 \)  \hspace{1cm} (7.246)

Substituting (7.244), (7.245) and (7.246) into (7.243) and equating coefficients, an ensemble of \( n+2 \) equations can be derived. Orders of the polynomials in (7.231) indicate that there are \( n+2 \) unknowns, in which case
a unique solution to \((7.231)\) should exist. Unfortunately \((7.231)\) is non-linear, involving the product of 2 unknown quantities.

Starting with the equation relating the coefficients of \(s^{n+1}\), the coefficients of \(N_3(s)\) were eliminated by substitution. As there is no term or order \(s^{n+1}\) in \(N_2(s)\), it is set to zero, that is

\[
C_{n+1} = 0 \quad (7.247)
\]

Using \((7.243)\) and the term for \(a_1\), derived from repeated substitution, an expression for the inductor, to be removed, is derived,

\[
L = c_1 - \frac{\sum_{i=0}^{n-1} \alpha^{2i+2} c_3 2i (-1)^i}{\sum_{j=0}^{n-1} \alpha^{2j+2} b_2 2j (-1)^j} \quad (7.248)
\]

The only unknowns are \(c\) and \(L\), \(c\) is squared to ensure that it remains positive.

Each term in \(N_3(s)\) is related to \(c\) and \(L\) by,

\[
a_i = \frac{\sum_{j=0}^{\text{int}(n-i+1)-1} \alpha^{2j} c_{i+2+2j} (-1)^j}{\sum_{k=0}^{\text{int}(n-i+1)} \alpha^{2k} b_{i+1+k} (-1)^k}
\]

One equation remains from the ensemble,

\[
\alpha^2 a_0 = c_0 \quad (7.250)
\]

By selecting an initial estimate of \(c\), a corresponding value of \(L\) is found using \((7.248)\). Using \(c\) and \(L\) the coefficients of \(N_3(s)\) are generated \((7.249)\). Using equation \((7.250)\) a new value for \(c\) can be generated and the process repeated. Therefore using equations \((7.248)\),
(7.249) and (7.250) a general iteration, to solve for $\alpha$, $L$ and the coefficients of $N_3(s)$, is derived.

Using the iteration discussed above, an inductor can be readily extracted from the impedance to be synthesised by Brunes method. Unfortunately convergence is poor for some rational functions, moreover as the solution is approached the rate of convergence decreases. Of the cases tested, none failed to converge to a correct solution, although convergence rates were very poor.

Despite its slow rate of convergence the iterative technique of inductor extraction, discussed above, was computationally attractive. It uses all information available, to extract the inductor, find the required complex imaginary axis zero, and the coefficients of $N_3(s)$, during the iteration. Zero searches of the real part of $Z_2(s)$ are unnecessary. For these reasons, methods of enhancing the convergence of the iterative inductor extraction procedure were sought.

Taking the difference between $a_0$ generated by (7.249) and $a_0$ generated by (7.238), an error function can be defined,

$$
\epsilon_{a_0} = \left| c_0 \frac{\alpha^{2j}}{\sum_{j=0}^{\text{int}(n+1)-1}} c_{2+2j} (-1)^j - L \frac{\sum_{k=0}^{\text{int}(n+1)-1}}{2} \alpha^{2k}b_{i+k}(-1)^k \right| \tag{7.251}
$$

Where $\epsilon_{a_0}$ is the error in the estimate of $a_0$ (7.250).

Assuming that (7.250) is generally quadratic in form, a quadratic curve which touches the $a_0$ axis was used to generate the next estimate of $\alpha$. 
Consider a quadratic curve which touches the $\alpha$ axis, it may be defined by

$$y_q = k_{qs}\left(\alpha - a_q\right)^2 \quad (7.252)$$

Where $y_q$ is the value of the curve at $\alpha$, $k_{qs}$ is a scaling parameter, it defines the steepness of the curve, and $a_q$ is the first at which the curve touches the $\alpha$ axis.

If a particular value of $k_{qs}$ is selected, a value of $a_q$ may be found corresponding to the error $\varepsilon_{a_0}$, whence,

$$a_q = \alpha \pm \frac{\varepsilon_{a_0}}{\sqrt{k_{qs}}} \quad (7.253)$$

Using the quadratic, second order zero, to estimate the position of the next estimate of $\alpha$, the convergence, of the iterative inductor extraction technique, is enhanced. The new value of $\alpha$ is defined in terms of the old value and a scaling parameter $k_{qs}$.

$$\alpha^{(k+1)} = \alpha^{(k)} \pm \frac{\varepsilon_{a_0}^{(k)}}{\sqrt{k_{qs}}} \quad (7.254)$$

Unfortunately the next estimate of $\alpha$ is dependent on the choice of sign in (7.254) and the value of $k_{qs}$. The choice of sign and $k_{qs}$ value is made by considering the nature of the process. Fig 7.13 shows a curve which should be typical for $\varepsilon_{a_0}$.
Curve 1 of Fig 7 is a quadratic curve fitted by the quadratic equation (7.252). Curve 2 corresponds to use of the positive sign in (7.254). Using the new estimate of $\alpha$ obtained from curve 1, the new value of $e_{a_0}$ is smaller than the old value. If curve 2 is selected the new value of $e_{a_0}$ is greater than the old value. Curve 1 therefore provides the new estimate of $\alpha$.

If however, $k_{qs}$ is too small, curves 3 and 4 would be produced. Neither curve produces a value of $\alpha$ that reduces $e_{a_0}$.

Using the observations above, suitable choices of a new estimate of $\alpha$ can be made. Starting with a value of $\alpha$, $e_{a_0}$ is obtained. Using (7.254) a new estimate of $\alpha$ is made and a second value of $e_{a_0}$ found.

If the new value of $e_{a_0}$ is smaller than the old value, the old value of $\alpha$ is discarded in favour of the new value. If the old value of $e_{a_0}$ exceeds the old value, the new value of $\alpha$ is discarded and a
second estimate of $\alpha$ made using (7.254) by changing the sign. Again, the new value of $\varepsilon_{a_0}$ is smaller than the old value, it replaces the old value otherwise it is discarded. If no reduction in $\varepsilon_{a_0}$ is obtained both signs in (7.254), $k_{qs}$ is increased by a factor of 10 and the above process repeated. In this way the new value of $\alpha$ found by (7.254), replaces the old value only if it reduces $\varepsilon_{a_0}$. Convergence is therefore guaranteed, and much improved over the general iteration scheme using (7.248), (7.249) and (7.250).

The quadratic approximation process is somewhat similar to the process of bisection. However in general it requires fewer iterations, to converge to the desired solution, than the method of bisection. Naturally, one could use Newton methods [23] to find the desired value of $\alpha$. Unfortunately little is known about the form of equation (7.251), therefore Newton methods stagnate or diverge at turning points [23]. It was intended that the load synthesis technique be reliable therefore failure of Newton methods would be unacceptable.

Unfortunately, even with the convergence enhancement methods discussed above, extraction of the inductor in some cases was very slow indeed. This may be due to the continually changing form of (7.252). Therefore still more rapid techniques of Brune synthesis were sought. However with the development of a chapter based method of Darlington synthesis [32], Brune Synthesis was abandoned.
7.3.3 Darlington Synthesis

In the preceding section it was found that Brune synthesis was in general, unsuitable for realizing an equalizer generated by the techniques of section 7.2. Only in a small number of cases can Brune Synthesis be used to realize a lossless stage. In general a series resistor must be extracted from the impedance function before reactive components are removed [34],[35],[38]. Thus, in general, Brune synthesis produces networks which violate the constraints of solution imposed in section 7.2.7. Therefore Brune Synthesis was abandoned; and a more suitable method, for realizing a reactive load terminated in a single resistor, sought.

A variety of techniques to synthesis loads exist, however most of these require at least 2 terminating resistors [28],[29],[30],[31],[32],[34],[35],[36],[37],[38],[39]. Some introduce branching ladder networks, unfortunately one new branch must be introduced each time one cannot apply Cauer synthesis. Moreover each branch must be terminated in a resistor.

The only method encountered in the literature which can be applied, to an irreducible function, to yield a reactive ladder, terminated in a resistor, is the method of Darlington Synthesis [33],[35],[38],[40].

Unfortunately the traditional techniques of Darlington Synthesis make it difficult to include in a load synthesis program. Using observations, made by Ordung and Krauss [32], enable the Darlington synthesis technique to be recast as an LDE problem which is easily included in the synthesis program using the techniques of section 7.2.2.
Full details of the Darlington techniques are given in Guilleman, [35], Darlington [33], Tuttle [38] and Christina [40].

Consider a two port network, representing part of the impedance to be synthesised, as shown in Fig 7.14.

Fig 7.14 Two Terminal Network of the Circuit to be Synthesised

The output of the network is terminated in the remainder impedance \( Z_2(s) \), such that the input impedance of the two terminal network is equal to the impedance to be synthesised. Moreover, at present, it is assumed that the orders of numerator and polynomials is 2 or greater.

The lossless 2 terminal network itself has impedance parameters \( z_{11}(s), z_{12}(s), z_{21}(s) \) and \( z_{22}(s) \) [45]. The input impedance of the network \( z_1(s) \) may be written in terms of the network impedance parameters and terminating impedance,

\[
    z_1(s) = z_{11}(s) + \frac{z_{21}^2(s)}{z_{22}(s) + z_2(s)} \quad (7.255)
\]

because the two part network is passive and so \( z_{21}(s) = z_{12}(s) \).
The transmission to the terminating resistor, of the impedance function $z_1(s)$, is given by (7.256) thus

$$T_1(s) = \frac{z_1(s) + z_1(-s)}{2z_1(s)} \tag{7.256}$$

Writing (7.256) in terms of its numerator and denominator polynomials $N_1(s)$ and $D_1(s)$,

$$T_1(s) = \frac{N_1(s)D_1(-s) + N_1(-s)D_1(s)}{2N_1(s)D_1(-s)} \tag{7.257}$$

It is assumed that neither $N_1(s)$ and $D_1(s)$ contain purely imaginary axis zeros as they could be removed by Cauer synthesis. Therefore no cancellation of zeros can occur between the numerator and denominator of (7.257).

If the lossless two part network is terminated in a resistor $R_L$, its transmission function would be,

$$T_{N_L}(s) = \frac{z_2(s)z_{21}(s)}{z_2(s)z_{11}(s) + z_{11}(s)z_{22}(s) - z_{12}(s)z_{21}(s)} \tag{7.258}$$

The zeros of transmission of (7.258) occur in quadrantal symmetry, in which case complex zeros occur in sets of 4 and real zeros occur in pairs. Darlington synthesis may be used to synthesise a group of 4 complex zeros of transmission or a pair of real axis zeros of transmission.

If the circuit, inside the two part network, is to emulate a quadruplet of complex ZOT, and remove them from the remaining impedance function $z_2(s)$, it must have a load transmission $T_{N_L}(s)$ which contains the transmission zeros to be synthesised. The multiplicity of the
zero synthesised, by (7.255), must be reduced by at least 1 in the transmission \( T_2(s) \) of the remaining impedance \( Z_2(s) \) [32].

The transmission of \( Z_1(s) \) may be represented in terms of the transmission of the 2 port network and the remainder impedance, \( Z_2(s) \) [3],[5],

\[
T_1(s) = T_{N_L}(s)t_2(s)
\]

\[
= \left\{ \frac{z_2(s)z_{21}(s)}{z_2(s)z_{11}(s)+z_{11}(s)z_{22}(s)-z_{12}(s)z_{21}(s)} \right\} \left\{ \frac{z_2(s) + z_2(-s)}{2z_2(s)} \right\} \quad (7.259)
\]

\( z_2(s) \) cannot contain the quadruplet of ZOT if the multiplicity of the ZOT is to be reduced by 1 in \( t_2(s) \), therefore \( z_{21}(s) \) must contain the quadruplet. If this is the case, the open circuit transmission \( T_{N_0}(s) \), of the two port network, also contains the quadruplet,

\[
T_{N_0}(s) = \frac{z_{21}(s)}{z_{11}(s)} \quad (7.260)
\]

Because it is desired that the two port network contain no additional zeros in its loaded transmission \( T_{N_L}(s) \), the numerator polynomial of \( z_{21}(s) \) comprises the transmission zero quadruplet,

\[
N_{21}(s) = K_{21}(s+\alpha)(s+\alpha^*)(s-\alpha)(s-\alpha^*) \quad (7.261)
\]

Where \( \alpha \) is the ZOT of the quadruplet in the lower LHS of the complex frequency plane.

If no additional zeros of transmission are to appear in the loaded transmission \( T_{N_L}(s) \), due to the two port network, the denominator of \( z_{11}(s) \), must equal the denominator of \( z_{21}(s) \), whence
\[ D_{11}(s) = D_{21}(s) \]  \hspace{1cm} (7.262)

Where \( D_{11}(s) \) and \( D_{21}(s) \) are the denominator polynomials of \( z_{11}(s) \) and \( z_{21}(s) \) respectively.

The two port network is therefore reciprocal, in which case the reverse transmission is given by

\[
T_{N_1}(s) = \frac{z_s(s)z_{12}(s)}{z_s(s)z_{11}(s)+z_{11}(s)z_{22}(s)-z_{12}(s)z_{21}(s)} \frac{z_s(s)+z_5(-s)}{2z_s(s)} \]  \hspace{1cm} (7.263)

Where \( z_s(s) \) is the impedance of the network connected to the input port of the lossless two port network.

But \( z_{21}(s) = z_{12}(s) \) due to reciprocity, therefore the open circuit reverse transmission also contains quadruplet zeros of transmission,

\[
T_{N_0}(s) = \frac{z_{12}(s)}{z_{22}(s)} \]  \hspace{1cm} (7.264)

and so

\[ D_{22}(s) = D_{21}(s) = D_{11}(s) \]  \hspace{1cm} (7.265)

Where \( D_{22}(s) \) is the denominator polynomial of \( z_{22}(s) \).

These observations give the designer sufficient information to find \( z_{11}(s) \), \( z_{12}(s) \) and \( z_{22}(s) \), and the remainder polynomial \( z_2(s) \).

In order to synthesise \( z_{11}(s) \), \( z_{12}(s) \) and \( z_{22}(s) \), a circuit of suitable form must be found so that the values of individual
components can be selected. The form must also ensure that any negative component values can be incorporated into suitable realizable transformer configurations.

From the discussion of Brune Synthesis in section 7.3.2 it is clear that the "T" network impedances yields suitable transformers to produce negative valued reactive components [34],[35],[38],[40]. Consider the 3 impedances $z_a(s)$, $z_b(s)$ and $z_c(s)$ connected to form the "T" network of Fig 7.15.

The impedance parameters of the "T" network may be written in terms of $z_a$, $z_b$ and $z_c$ [45],

$$z_{11}(s) = z_a(s) + z_b(s)$$  \hspace{1cm} (7.266)$$

$$z_{12}(s) = z_{21}(s) = -z_b(s)$$

$$z_{22}(s) = z_2(s) + z_b(s)$$

Therefore,

$$z_a(s) = z_{11}(s) + z_{21}(s)$$  \hspace{1cm} (7.267)$$
\[ z_b(s) = z_{21}(s) = z_{21}(s) \]

and

\[ z_c(s) = z_{22}(s) + z_{21}(s) \]

Using Cauer Synthesis, the series connected components of \( z_a(s) \), \( z_b(s) \) and \( z_c(s) \) are found [32], [34], [35], [37], [38]. Their ZOT quadruplet produces a circuit of the form shown in Fig 7.16.

**Fig 7.16 Synthesis of a ZOT Quadruplet**

Negative component values are removed by transforming the circuit to the form of Fig 7.17 [38].
Fig. 7.17 Removal of Negative Components

\[ \frac{1}{L_1} = \frac{(C_1 + C_4)}{C_4} \]  

\[ L_2 = \frac{C_2}{C_4(C_1 + C_4)} \]

\[ L_3 = L_2 + L_3 \]

\[ L_4 = L_3 + L_5 \]

Where [38],

\[ C_1' = C_1 \]

\[ C_2' = \frac{C_2 C_3}{C_2 + C_3} \]

\[ C_3' = C_3 \]

\[ C_4' = C_4 \]
If \( C_2' \) is negative then it must be referred to the opposite side of the transformer, in which case the nomenclature is reversed and Equations (7.268) still apply.

In some cases \( Z_1(s) \) may be the ratio of two linear functions in \( s \). By multiplying the numerator and denominator by \( D_1(s) \) a quadruplet of ZOT may be formed and the procedure described above applied to the augmented rational function \( z_1'(s) \) [35], where

\[
z_1'(s) = \frac{N_1(s) D_1(s)}{D_1^2(s)}
\]

(7.269)

It is also possible to use the same procedure to develop a specific Darlington synthesis technique to realize a rational function comprising the ratio of two linear functions in \( s \) [32].

Darlington Synthesis is not confined to the synthesis of just 4 ZOT's, it may be used to synthesise any number of quadruplets of complex ZOT's or pairs of real axis ZOT's. One may find \( z_{11}(s) \), \( z_{21}(s) \) and \( z_{22}(s) \) for a network of any complexity terminated the resistor, and one need not repeatedly apply the 4 ZOT procedure until the terminating resistor is reached [32],[35].

When Darlington synthesis is carried out for a network terminated in the final resistor there is no need to find any individual ZOT's, it is

\[
\begin{align*}
L_{12} &= L_4 \\
L_{34} &= L_3
\end{align*}
\]
possible to work entirely with the series expansions of the numerator and denominator of $z_1(s)$. The general procedure is outlined by Guilleman [35].

Unfortunately, if Darlington synthesis is used to find the complete network between $z_1(s)$ and the terminating resistor, it can be very difficult to find the appropriate configuration and thus the values of components required, and to realize components with negative values.

If Darlington synthesis is used to realize quadruplet ZOT and doublet ZOT groups, the traditional method of synthesis involves finding the residues of each pole of $z_{11}(s)$, $z_{21}(s)$ and $z_{22}(s)$, to find their numerator polynomials, and expressing each as a sum of partial fractions. These are grouped appropriately to define $z_{22}(s)$ and $z_2(s)$ according to the locations of the poles. Using the residue conditions given by Guilleman [35] $z_{11}(s)$ and $z_{22}(s)$ are found.

This procedure is not regular and would be extremely difficult to program for use on a digital computer. Moreover errors would tend to accumulate as the synthesis proceeds. For these reasons a more regular method of performing Darlington synthesis was sought.

Darlington synthesis may also be performed using LDE techniques, which are much more easy to incorporate into a computer program, than the residue techniques of synthesis, discussed previously.

Unfortunately, it is necessary to find $I_z$ ZOT from the numerator polynomial of the load transmission $T_1(s)$ each time Darlington synthesis is applied to synthesise networks with doublet ZOT's or quadruplet ZOT's.
Each time a doublet or quadruplet ZOT group is to be synthesised it is necessary only to find one member of the group from which \( N_{21}(s) \) can be developed, therefore one can use Lins method of successive approximation \([102]\) to find a quadratic term in complex frequency. It may represent either the product of two complex zeros or a pair of real axis zeros, in which case 1 zero is extracted to produce a quadruplet.

The LDE technique of Darlington synthesis is based on observations made by Ordung and Krauss \([32]\), however it exploits the forms and orders permitted, for the numerator and denominator polynomials of \( z_{11}(s) \), \( z_{12}(s) \) and \( z_{21}(s) \), by Darlington synthesis \([32],[35],[38]\). These restrictions facilitate the formulation of a unique, fully determined, set of simultaneous equations which are readily solved by the LDE techniques of section 7.2.2.

The LDE's for Darlington synthesis are formed in the following manner:

The input impedance equation (7.255) may be rearranged into the form \([32],[35],[38]\),

\[
(z_{11}(s) - z_1(s)) \left( z_{22}(s) + z_2(s) \right) = z_{12}^2(s) \quad (7.270)
\]

To synthesise a quadruplet of ZOT's, \( z_{12}(s) \) the 4 ZOT's,

\[
z_{12}(s) = \frac{K_{21}(s+\alpha)(s+\alpha^*)(s-\alpha)(s-\alpha^*)}{D_{12}(s)} = \frac{K_{21}(C(s)C(-s))}{D_{12}(s)} \quad (7.271)
\]

where \( c(s) = (s+\alpha)(s+\alpha^*) \)
But \( z_1(s) \) and \( z_2(s) \) are all PR impedances therefore the RHP ZOT of \( z_{12}(s) \) can only occur due to the difference \( (z_{11}(s) - z_1(s)) \) \[32\],[35],[38].

Moreover one LHP ZOT must also be included with \( z_{11}(s) - z_1(s) \) in order to reduce the order of the numerator and denominator polynomials of \( z_2(s) \) [32]. By writing \( z_{11}(s) \) in terms of their numerator and denominator polynomials,

\[
\frac{N_{11}(s)}{D_{11}(s)} = \frac{C^2(-s)A(s)}{B(s)}
\]  

(7.272)

Where \( A(s) \) and \( B(s) \) are arbitrary polynomials and

\[
z_{11}(s) = \frac{N_{11}(s)}{D_{11}(s)} \quad \text{and} \quad z_1(s) = \frac{N_1(s)}{D_1(s)}
\]  

(7.273)

By rearranging (7.272), a diophantine equation is found,

\[
N_{11}(s)D_1(s) - N_1(s)D_{11}(s) = C^2(-s)
\]  

(7.274)

\( x, N_1(s) \) and \( D_1(s) \) are known, \( N_{11}(s) \), \( D_{11}(s) \) and \( A(s) \) are unknown.

(7.274) has an infinity of solutions. Fortunately, the forms of the numerator and denominator polynomials of \( z_{11}(s) \) are known from the proof given by Ordung and Krauss [32]. \( z_{11}(s) \) must, in fact, contain 4 imaginary axis poles, one at infinity, one at zero and two at an unknown finite value. It must also be the ratio if an even polynomial to an odd polynomial or visa versa. Therefore [32],

\[
z_{21}(s) = \frac{N_{11}(s)}{D_{11}(s)} = \frac{as^4 + bs^2 + c}{s(s^2 + d)}
\]  

(7.275)
So \( N_{11}(s) = as^4 + bs^2 + c \)
and \( D_{11}(s) = s(s^2+d). \)

The above restrictions on \( N_{11}(s) \) and \( D_{11}(s) \) define a unique value for \( a, b, c \) and \( d \) and \( A(s) \), thus \( z_{11}(s) \) is found from (7.274).

The introductory analysis of Darlington synthesis showed that,
(7.265),
\[
D_{11}(s) = D_{21}(s) = D_{22}(s)
\]
Therefore
\[
z_{21}(s) = \frac{K_{21}C(s)C(-s)}{D_{21}(s)} \quad (7.276)
\]
But the open circuit forward transmission is given by (7.271) which may be rewritten in terms of \( a, b, c \) and \( \alpha \),
\[
T_{N_0}(s) = \frac{K_{21}(s+\alpha)(s+\alpha^*)(s-\alpha)(s-\alpha^*)}{as^4 + bs^2 + c} \quad (7.277)
\]
By setting \( s=0 \), the value for \( K_{21} \) may be found [32]
\[
T_{N_0}(0) = \frac{K_{21}\alpha^2(\alpha^*)^2}{c}
\]
For a passive reciprocal network, \( T_{N_0}(0)=-1 \) thus,
\[
K_{21} = \frac{c}{\alpha^2(\alpha^*)^2} \quad (7.278)
\]
and so \( z_{21}(s) \) and \( z_{12}(s) \) are found.

\( z_{22}(s) \) and \( z_{2}(s) \) are found by rearranging (7.270) to give
\[ z_{22}(s) + z_{22}(s) = \frac{z_{21}^2(s)}{z_{11}(s) - z_1(s)} \]  

(7.279)

Which may be expressed in terms of numerator and denominator polynomials,

\[ \frac{N_{22}(s) + N_2(s)}{D_{22}(s)} = \frac{K_{21}^2 C_2(s) C^2(-s)}{D_{21}(s)} \]

(7.280)

With some rearrangement,

\[ \frac{N_{11}(s) D_2(s) + N_2(s) D_{22}(s)}{D_{22}(s) D_2(s)} = \frac{K_{21}^2 C_2(s) C^2(-s) D_1(s) D_{11}(s)}{(N_{11}(s) D_1(s) - D_{11}(s) N_1(s)) D_{21}(s)} \]

(7.281)

Cancellation occurs due to the equality of denominator polynomials given by (7.265). By substituting (7.274) into the RHS of (7.281),

\[ \frac{N_{22}(s) D_2(s) + N_2(s) D_{22}(s)}{D_2(s)} = \frac{K_{21}^2 C_2(s) C^2(-s) D_1(s)}{C^2(-s) A(s)} \]

(7.282)

and so

\[ D_2(s) = A(s) \]

A(s) was found by solving (7.274).

Equating the numerators of (7.281) produces a second LDE,

\[ N_{22}(s) D_2(s) + N_2(s) D_{22}(s) = K_{21}^2 C_2(s) D_1(s) \]

(7.283)

By solving (7.283), \( N_{22}(s) \) and \( N_2(s) \) can be found.
$N_{22}(s)$ must again be an even polynomial therefore a unique solution to (7.283) is defined.

The network, extracted by LDE Darlington synthesis, is found by evaluating,

$$z_a(s) = z_{11}(s) - z_{12}(s)$$

$$z_b = z_{12}(s)$$

and $$z_c = z_{22}(s) - z_{11}(s)$$

and synthesising each branch of the "T" network of Fig 7.16 using Cauer Synthesis [32].

LDE's (7.274) and (7.280) can be solved using the vector matrix approach discussed in section 7.2.8.

Again consider equation (7.274),

$$N_{11}(s)D_1(s) - N_1(s)D_{11}(s) = C(s)C^2(-s)A(s)$$

Forming a matrix of the form of $D$ in (7.276, equation (7.274) may be written as;

$$\begin{bmatrix} D_1(s) \\ N_{11}(s) \end{bmatrix} \begin{bmatrix} N_{11}(s) \\ N_1(s) \end{bmatrix} - \begin{bmatrix} N_1(s) \\ D_{11}(s) \end{bmatrix} = \begin{bmatrix} C(s)C^2(-s) \end{bmatrix} \begin{bmatrix} A(s) \end{bmatrix}$$

(7.284)

where $[\ ]$ denotes matrix or vector.

$N_{11}(s)$ is a polynomial of order 4 and $D_{11}(s)$ is a polynomial of order 3. $D_{11}(s)$ can also be regarded as a polynomial of order 4 with a zero at infinity, thus, at present, $D_{11}(s)$ will be treated as a
7.148

polynomial of order 4. If \( D_1(s) \) and \( N_1(s) \) are polynomials of order \( n_1 \) then \([D_1(s)]\) and \([N_1(s)]\) are matrices of dimensions \([(n_1+5),5]\).

\[ C(s) \] is a polynomial of second order. If \( A(s) \) is a polynomial of order \( N_2 \) then \([C(s)C^2(-s)]\) is a matrix of order \([(n_2+5),(n_2-1)]\).

\([N_{11}(s)]\) and \([D_{11}(s)]\) are vectors of order \([5,1]\) and \( A(s) \) is a vector of order \([(N_2+1),1]\).

Equating the orders of both sides of (7.284), the order of \( A(s) \) may be expressed in terms of the orders of \( N_1(s) \) and \( D_1(s) \),

\[ n_2 = n_1 - 2 \] (7.285)

By rearranging (7.284) and using augmented matrices a set of homogenous equations is obtained,

\[
\begin{bmatrix}
D_1(s) & -N_1(s) & -C(s)C^2(-s)
\end{bmatrix}
\begin{bmatrix}
N_{11}(s) \\
D_{11}(s) \\
A(s)
\end{bmatrix}
\]

The augmented matrix is of order \([(n_1+5)(n_1+9)]\), the augmented vector is of order \([(n_1+9),1]\).

Let \( \{a_0, a_1, a_2, \ldots, a_{n_1}\} \) represent the coefficients of polynomial \( D_1(s) \), \( \{b_0, b_1, b_2, \ldots, b_{n_1}\} \) represent the coefficients of \(-N_1(s)\), \( \{C_0, C_1, C_2, C_3, C_4\} \) the polynomial coefficients of \(-C(s)C^2(-s)\), \( \{d_0, d_1, d_3, d_4\} \) the coefficients of \( N_{11}(s) \), \( \{e_0, e_1, e_2, e_3, e_4\} \) the coefficient of \( D_{11}(s) \) and \( \{f_0, f_1, f_2, f_3, \ldots, f_{n_2}\} \). The vector matrix equation, (7.284), may be written in terms of the coefficients,
\[
\begin{bmatrix}
\begin{array}{cccccccc}
    a_0 & 0 & 0 & 0 & 0 & b_0 & 0 & 0 & 0 & 0 \\
    a_1 & a_0 & 0 & 0 & 0 & b_1 & b_0 & 0 & 0 & 0 \\
    a_2 & a_1 & a_0 & 0 & 0 & b_2 & b_1 & b_0 & 0 & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    a_{n_1-2} & a_2 & a_1 & a_0 & \ldots & b_{n_1-2} & b_2 & b_1 & b_0 & 0 \\
    a_{n_1-1} & a_{n_1-2} & a_2 & a_1 & b_{n_1-1} & b_{n_1-2} & \ldots & \ldots & b_2 & b_1 \\
    a_{n_1} & a_{n_1-1} & a_{n_1-2} & a_2 & b_{n_1} & b_{n_1-1} & b_{n_1-2} & \ldots & \ldots & \ldots \\
    0 & a_{n_1} & a_{n_1-1} & a_{n_1-2} & 0 & b_{n_1} & b_{n_1-1} & b_{n_1-2} & \ldots & \ldots \\
    0 & 0 & a_{n_1} & a_{n_1-1} & a_{n_1-2} & 0 & 0 & b_{n_1} & b_{n_1-1} & b_{n_1-2} \\
    0 & 0 & 0 & a_{n_1} & a_{n_1-1} & 0 & 0 & 0 & b_{n_1} & b_{n_1-1} \\
    0 & 0 & 0 & 0 & a_{n_1} & 0 & 0 & 0 & 0 & b_{n_1} \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{array}
\end{bmatrix}
\begin{bmatrix}
    c_0 \\
    c_1 \\
    \vdots \\
    c_{n_1-2}
\end{bmatrix}
= 0
\]
However the restrictions of equation (7.275) force $d_1$, $d_3$, $e_0$, $e_2$ and $e_4$ to zero. These values may be eliminated from the vector of (3.286). The rows of the augmented matrix corresponding to those values eliminated from the vector are eliminated. Equation (7.275) also requires $e_3 = 1$, in which case the column corresponding to $e_3$ extracted from the matrix and subtracted from both sides of equation (7.287). Having made these eliminations (7.287) becomes,

\[
\begin{bmatrix}
a_0 & 0 & 0 & 0 & c_0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & b_0 & c_1 & c_0 & 0 & \cdots & 0 \\
a_2 & a_0 & 0 & b_1 & c_2 & c_1 & c_0 & \cdots & c_0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
a_{n_1-2} & a_0 & \vdots & b_{n_1-2} & c_4 & c_3 & \cdots & c_2 & c_1 & 0 \\
a_{n_1-1} & \vdots & \vdots & b_{n_1-1} & c_5 & c_4 & \cdots & c_3 & c_2 & c_1 \\
0 & a_{n_1-1} & \vdots & b_{n_1} & c_6 & c_5 & \cdots & c_4 & c_3 & c_2 \\
0 & 0 & a_{n_1-2} & 0 & 0 & c_6 & \cdots & c_5 & c_4 & c_3 \\
0 & 0 & a_{n_1} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & c_6 \\
\end{bmatrix}
\begin{bmatrix}
d_0 \\
d_2 \\
d_4 \\
e_1 \\
f_0 \\
f_1 \\
f_2 \\
f_{n_1-2} \\
\end{bmatrix}
= -\begin{bmatrix}
0 \\
0 \\
0 \\
b_0 \\
b_1 \\
b_2 \\
\vdots \\
b_{n_1-2} \\
b_{n_1-1} \\
b_{n_1} \\
0 \\
\end{bmatrix}
\]

(7.288)

After elimination of columns of the augmented matrix its order is $[(n_1+5),(n_1+3)]$. The order of the vector is $[(n_1+3, 1)]$. The vector matrix equation therefore represents a non singular system of equations which can be solved by finding the pseudo inverse of the augmented matrix of (7.347). Thus $Z_{11}(s)$ and $A(s) = D_2(s)$ are found.
Equation (7.277) may now be solved for $K_{21}$. Having found $D_2(s), D_{11}(s)$ and $K_{21}$ equation (7.283) can be solved for $N_{22}(s)$ and $N_{2}(s)$. $N_{22}(s)$ has the same form as $N_{11}(s)$ therefore the same logic may be used to develop the vector matrix technique of solution (7.283) is restated,

$$N_{22}(s) D_2(s) + N_{2}(s) D_{22}(s) = K_{21}^2 C(s) D_1(s)$$

which may be written as vector matrix form,

$$\begin{bmatrix} D_2(s) \end{bmatrix} \begin{bmatrix} N_{22}(s) \end{bmatrix} + \begin{bmatrix} D_{22}(s) \end{bmatrix} \begin{bmatrix} N_{2}(s) \end{bmatrix} = \begin{bmatrix} K_{21}^2 C(s) \end{bmatrix} \begin{bmatrix} D_1(s) \end{bmatrix}$$

(7.289)

$D_2(s)$ is of order $(N_1-2)$, $N_{22}(s)$ is of order 4, $D_{22}(s)$ is also of order 4. $D_1(s)$ is of order $N_1$ and $C(s)$ is of order 2. $K_{21}$ is a constant.

Let $N_3$ represent the order of $N_2(s)$ in which case, $[D_2(s)]$ is a matrix of order $[(N_1+3), 5]$, $N_{22}(s)$ is a vector of order $[5 \times 1]$, $[D_{22}(s)]$ is a matrix of order $[(N_3+5), (N_3+1)]$ and so $[N_2(s)]$ is a vector or order $(N_3+1), 1]$. Finally the vector $[K_{21}^2 C(s) D_1(s)]$ is of order $[(N_1+3), 1]$.

Equating the orders of (7.289) gives an order of $n_3=n_1-2$, therefore $[N_2(s)]$ has dimensions $[(N_1+3), (N_1-1)]$ and $[N_2(s)]$ is a vector of order $[(N_1-1), 1]$.

Let $\{g_0, g_1, g_2, g_3, g_4\}$ represent the coefficient if $N_{22}(s), \{L_0, L_1, L_2, L_3 \ldots h_{n-2}\}$ represent the coefficients of $N_2(s)$ and $\{P_0, P_1, P_2, P_3, \ldots P_{n+3}\}$ represent the coefficients of the polynomial $K_{21}^2 C(s) D_1(s)$.
Equation (7.289) may be solved in augmented form,

\[
\begin{bmatrix}
D_2(s) & D_{22}(s)
\end{bmatrix}
\begin{bmatrix}
N_{22}(s) \\
\frac{N_{22}(s)}{N_2(s)}
\end{bmatrix}
= \begin{bmatrix}
k_{21}^2 C(s) D_1(s)
\end{bmatrix}
\tag{7.290}
\]

Which is written in terms of the polynomial coefficients.
\[
\begin{bmatrix}
  b_0 & 0 & 0 & 0 & 0 & f_0 & 0 & 0 & \ldots & 0 & 0 \\
  b_1 & b_0 & 0 & 0 & 0 & f_1 & f_0 & 0 & \ldots & 0 & 0 \\
  b_2 & b_1 & b_0 & 0 & 0 & f_2 & f_1 & f_0 & \ldots & 0 & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  b_{n_1-2} & b_{n_1-3} & b_{n_1-4} & \ldots & b_2 & b_1 & b_0 & f_4 & f_3 & f_2 & f_1 & f_0 & 0 & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  b_{n_1-1} & b_{n_1-2} & b_{n_1-3} & \ldots & b_2 & b_1 & b_0 & f_4 & f_3 & f_2 & f_1 & f_0 & 0 & h_1 \\
  b_{n_1} & b_{n_1-1} & b_{n_1-2} & \ldots & b_2 & b_1 & b_0 & f_4 & f_3 & f_2 & f_1 & f_0 & 0 & h_2 \\
  0 & 0 & b_{n_1} & b_{n_1-1} & b_{n_1-2} & \ldots & 0 & 0 & 0 & \ldots & f_4 & f_3 & f_2 & h_{n_1-4} & p_{n_1+1} \\
  0 & 0 & 0 & b_{n_1} & b_{n_1-1} & 0 & 0 & 0 & \ldots & 0 & f_4 & f_3 & h_{n_1-3} & p_{n_1+2} \\
  0 & 0 & 0 & 0 & b_{n_1} & 0 & 0 & 0 & \ldots & 0 & 0 & f_4 & h_{n_1-2} & p_{n_1+3} \\
\end{bmatrix}
= 
\begin{bmatrix}
  g_0 \\
  g_1 \\
  g_2 \\
  \vdots \\
  g_{n_1-2} \\
  g_{n_1-1} \\
  g_{n_1} \\
  \vdots \\
  p_0 \\
  p_1 \\
  p_2 \\
  \vdots \\
  p_{n_1+1} \\
  p_{n_1+2} \\
  p_{n_1+3}
\end{bmatrix}
\]
The augmented matrix is of dimensions \([n_1+3, n_1+4]\) and the unknown augmented vector is of dimensions \([(n_3+6), 1]\). The same restrictions as those of the coefficients of \(N_{11}(s)\) apply to the coefficients of \(N_{22}(s)\) thus, \(g_1 = 0\) and \(g_3 = 0\). These elements of the augmented vector are eliminated together with the corresponding columns of the augmented matrix giving the reduced-augmented vector matrix equation.

\[
\begin{bmatrix}
    b_0 & 0 & 0 & f_0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
    b_1 & 0 & 0 & f_1 & f_0 & 0 & \ldots & 0 & 0 & 0 \\
    b_2 & b_0 & 0 & f_2 & f_1 & f_0 & \ldots & 0 & 0 & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \\
    b_{n_1-2} & b_2 & b_0 & f_4 & f_3 & f_2 & \ldots & f_0 & 0 & 0 \\
    b_{n_1-1} & b_1 & \cdots & f_4 & f_3 & f_2 & \ldots & f_1 & f_0 & 0 \\
    b_{n_1} & b_{n_1-2} & b_0 & \cdots & f_4 & f_3 & f_2 & \ldots & f_1 & f_0 \\
    0 & b_{n_1-1} & 0 & 0 & \cdots & f_3 & f_2 & f_1 & \ddots & \ddots \\
    0 & b_{n_1} & b_{n_1-2} & 0 & 0 & \cdots & f_4 & f_3 & f_2 & \ddots \\
    0 & 0 & b_{n_1-1} & 0 & 0 & \cdots & f_4 & f_3 & f_2 & \ddots \\
    0 & 0 & b_{n_1} & 0 & 0 & \cdots & 0 & f_4 & f_3 & f_2 \\
\end{bmatrix}
\begin{bmatrix}
g_0 \\
g_1 \\
g_2 \\
g_3 \\
g_4 \\
h_1 \\
h_2 \\
h_{n_1-4} \\
h_{n_1-3} \\
h_{n_1-2} \\
\end{bmatrix}
= 
\begin{bmatrix}
p_0 \\
p_1 \\
p_3 \\
p_3 \\
p_4 \\
p_{n_1+1} \\
p_{n_1+2} \\
p_{n_1+3} \\
p_{n_1+4} \\
p_{n_1+5} \\
\end{bmatrix}
\]  

(7.292)

The augmented matrix now has dimensions \([(n_1+3), (n_1+2)]\) and the augmented vector dimensions, \([(n_1+2), 1]\).

The vector-matrix equation represents an over determined system of equations. Using the pseudo-inverse, \(N_{11}(s)\) and \(N_{2}(s)\) can be found.

Using equations (7.267) the values of components for the circuit shown in Fig 7.16 are found. One cycle of Cascade Darlington synthesis is
therefore complete. The remaining impedance $Z_2(s)$ has also been found therefore further synthesis can be undertaken.

The vector matrix technique of Darlington Synthesis discussed above is readily programmed into a computer and is used to perform the Darlington synthesis of loads in the Computer program DLREAL discussed in section 7.3.4.

7.3.4 DLREAL - A Load Synthesis Program

Based upon the load synthesis techniques, discussed in sections 7.3.1 to 7.3.3, a program to synthesis the equalizer back impedance functions of DEQDIO, was produced.

Given an impedance function, the program initially searches for poles of impedance at infinity by testing the relative order of numerator and denominator polynomial. If a pole at infinity is found it is extracted. The corresponding component value is stored with a label. The old impedance function is overwritten by the remainder.

The impedance function is tested for the presence of a pole at zero, if it is found, the pole is extracted and the value of capacitance stored with an appropriate label, again the impedance function is overwritten by the remainder. Finally, using a Routh test, the impedance is tested for the presence of a finite imaginary axis pole. If the Routh array has a zero row, the preceding row gives the coefficients of a polynomial
describing the imaginary axis poles. If it is of zero order, the impedance contains no finite imaginary axis poles. If it is of order 2, then the polynomial describes one imaginary axis pole and its conjugate which is readily extracted, and the remainder is written over the original impedance function. If the Routh array yields a polynomial of order greater than 2, the program searches for a second order polynomial, describing one imaginary axis zero using Lins method [102]. This pole is extracted, the values of parallel inductor and capacitor found and stored with appropriate labels. The remainder again replaces the original impedance function. This process is repeated until all imaginary axis poles are removed.

The impedance function is then inverted, and the preceding process repeated, until all imaginary poles of admittance are extracted, and component values stored, with appropriate labels.

The section of DLREAL described above is the Cauer Synthesis section. It is necessary to record whether the rational function, from which the imaginary poles are extracted, represents an impedance or admittance.

If the program cannot extract any imaginary axis poles, from either the impedance or the admittance, then, either the reactive part of the load has been synthesised and the terminating resistor remains, or Darlington synthesis must be used to proceed with the synthesis.

The numerator polynomial is divided by the denominator polynomial and a constant extracted. If the remainder is zero the constant is stored as a value or either resistance or conductance and the synthesis terminated.
If the remainder is non-zero then Darlington synthesis is used to complete the load realization.

If the rational function represents an admittance, it is inverted to give an impedance function prior to Darlington synthesis.

The LDE's of section 7.3.3 are then solved and component values produced and stored, again with labels to identify the configuration. The remainder overwrites the original impedance function. The program then returns to the start of the Cauer Synthesis section and the synthesis procedure is restarted.

Cauer synthesis is always attempted first, followed by Darlington synthesis until the remainder polynomial is zero and a resistor value is stored. Synthesis is then terminated.

If the rational function, to be synthesised by Darlington synthesis, is the ratio of two linear functions, the numerator and denominator are augmented by the denominator prior to solving the LDE's of section 7.3.3, (7.274) and (7.283).

The component values are stored in an array. Their labels are stored in a character array. The component values are rescaled to their original values using the equations [45],

\[
R' = R_0 \frac{R_0}{R_0 \omega_0} \quad C' = C \quad L' = \frac{R_0}{\omega_0} \quad (7.293)
\]

Where \( R_0 \) is the impedance normalization value and \( \omega_0 \) is the frequency normalization value.
The component values are then printed together with their labels from which one may derive the circuit configuration.

During each component extraction, estimates of the errors in the remainder are gained using the same techniques as discussed in section 7.3.7, for the equalization program DEQDIO. By carrying an error estimate with each polynomial operation it is possible to judge whether a polynomial coefficient is zero or finite.

If good estimates of the errors, present in the polynomial coefficients, are not used, cumulative errors produce false component values and unnecessarily complicate circuit configurations.
REFERENCES - CHAPTER 7

[1] KWOK, SING-PING

"UHF Broadband Amplifier Design."
Motorola Application Note, AN-406, Motorola Inc.

[2] FRATER, R.H.

"Accurate Wideband Multiplier-Square-Law Detector."

[3] YOULA, D.C.

"A New Theory of Broadband Matching."


"Solid-State Microwave Amplifier Design."

[5] CHEN, WAI-KAI

"Theory and Design of Broadband Matching Networks."

[6] MICROWAVE ASSOCIATES

"Transistor Designers Guide."


"Transistor Circuit Design."
[8] SUPTA, K.C., GARG, R. and CHANDA, R.
"Computer Aided Design of Microwave Circuits."

[9] GHAUSI, M.S.
"Principles and Design of Linear Active Circuits."

[10] CHERRY, E.M.
"Impedance Mismatch in Wide-Band Transistor Amplifier Design."

"RF Small Signal Design Using 2-Port Parameters."
Motorola Application Note, AN-215, Motorola Inc.

[12] YOUNG, G.
Boradband Network Design for UHF Amplifiers."
Motorola Application Note, AN-704, Motorola Inc.

"UHF Transmission-Line Power-Amplifier Designed With Smith Chart Techniques."
Motorola Application Note, AN-217, Motorola Inc.

[14] WEINERT, F.
"Scattering Parameters Speed Design of High-Frequency Transistor Circuits.
[15] YOULA, D.C.

"On Scattering Matrices Normalized to Complex Port Numbers."

[16] FUKUI, H.

"Available Power Gain Noise Figure, and Noise Measure of Two-Ports
and Their Graphical Representation."
pp 137 - 142.

[17] TATOM, C.

"Whats and Whys about Y-Parameters."
Motorola Application Note, AN-158, Motorola Inc.

[18] BODWAY, G.E.

"Two Port Power Flow Analysis Using Generalized Scattering
Parameters."

[19] BECCIOLEINI, B.

"Impedance Matching Networks Applied to R.F. Power Transistors."
Motorola Application Note, AN-721, Motorola Inc.

[20] NORRIS, P.M.

"Using Linwill Techniques for R.F. Amplifiers."
Motorola Application Note, AN-166, Motorola Inc.

[21] HOLT, C.A.

"Electronic Circuits Digital and Analogue."
John Wiley and Sons Ltd., 1978, pp 383 - 776.
[22] MILLMAN, J.
"Microelectronics Digital and Analogue Circuits and Systems."

[23] SCHEID, F.
"Numerical Analysis."

[24] INTERNATIONAL BUSINESS MACHINES CORPORATION.
"System/360 Scientific Subroutine Package."
(360 - cm - 03X) Version III Programmers Manual.

[25] INTERNATIONAL BUSINESS MACHINES CORPORATION (as for [24])
DFRAT, pp 265 - 270.

[26] INTERNATIONAL BUSINESS MACHINES CORPORATION (as for [24])
pp 241 - 288.

[27] SCHMIDT, R.
"Advances in Nonlinear Parameter Optimization."

[28] FAILKOW, A.D.
"A Note on L.C. -R Ladder Synthesis of Filters."
[29] CUONG, D.M. and HASLER, M.

"On the Canonic Cascade Synthesis for Lossless One-Ports."
IEEE Trans. on Circuits and Systems, Vol. CAS-27, No. 2,

[30] FIALKOW, A.D.

"Inductance, Capacitance Networks Terminated in Resistance."
IEEE Trans. on Circuits and Systems, Vol. CAS-26, No. 8,
Aug. 1979, pp 603 - 641.

[31] SZENTIRMAI, G.

"FILSYN - A General Purpose Filter Synthesis Program."

[32] ORDUNG, P.F. and KRAUSS, H.L.

"Synthesis of an L.C. Zero Section for Quadruplet Transmission
Zeros."

[33] DARLINGTON, S.

"Synthesis of Reactance 4-poles Which Produce Prescribed Insertion
Loss Characteristics - Including Special Applications to Filter
Design."
Journal of Mathematics and Physics, Vol. 18, Sept. 1939,
pp 257 - 353.

[34] PEIKARI, B.

"Fundamentals of Network Analysis and Synthesis."
[35] GUILLEMAN, E.A.
"Synthesis of Passive Networks - Theory and Methods Appropriate to the Realization and Approximation Problems."

[36] ANDERSON, B.D.O. and KAILATH, T.
"Passive Network Synthesis Via Dual Spectral Factorization."

[37] CAUER, W.
"Synthesis of Linear Communication Networks."

[38] TUTTLE (JR), D.F.
"Network Synthesis - Vol. 1"

[39] KALMAN, R.E. and DECLARIS, N.
"Aspects of Network and System Theory."

[40] CHRISTIAN, E.
"L.C. Filters Design, Testing and Manufacturing."

[41] HA, T.T.
"Computer-Aided Cascade Synthesis of Passive Networks."
[42] MORRISON, R.
"Linear Phase Versus Step Response."

[43] SZENTIRMAI, G.
"The Problem of Phase Equalization."

[44] YOSHIDA, N.
"Transfer Function of Maximally Flat Group Delay Low-Pass Filters With Equal-Ripple Attenuation in the Stop-Band and Flat Attenuation in the Passband."

[45] SKILLING, H.
"Electric Network."

[46] ABRAMOWITZ, M. and STEGUN, I.A. (Ed's)
"Handbook of Mathematical Functions."

[47] BRAESS, K.D.
"Uber Damptung bei Minimalis cirunysverfahren."
7.166

[48] MARQUARDT, D.W.

"An Algorithm for Least Squares Estimation of Non-Linear Parameters."

[49] KUESTER, J.L. and MIZE, J.H.

"Optimization Techniques with FORTRAN."

[50] BEIGHTLER, C.S., PHILLIPS, D.T. and WILDE, D.J.

"Foundation of Optimization."

[51] INTERNATIONAL BUSINESS MACHINES CORPORATION (as for [24])

pp 221 - 224.

[52] INTERNATIONAL BUSINESS MACHINES CORPORATION (as for [24])

pp 225 - 230.

[53] HOOPER, M.J.

"Harwell Scientific Subroutine Library - A Catalogue of Subroutines."

[54] HOOPER, M.J. (as for [53])

VA07AD, April 1971.
[55] HOOPER, M.J. (as for [53])
VF02AD, Feb. 1978.

[56] HOOPER, M.J. (as for [53])

[57] TEMES, G.C. and LA PATRA, J.W.
"Introduction to Circuit Synthesis and Design."

[58] ANTON, H.
"Calculus with Analytic Geometry."

[59] D'AZZO, J. and HOUPIS, C.H.
"Linear Control System Analysis and Design."

[60] PERSONICK, S.D.
"Receiver Design for Optical Fiber Systems."

[61] PERSONICK, S.D.
"Receiver Design for Digital Fibre Optic Communication Systems
(I & II)."
The Bell System Technical Journal, Vol. 52, No. 6, July - August

[62] HADLEY, M.
"Amplifier Gains 10dB Over Nine Octaves."
Motorola Device Data, EB 37, Motorola Inc., pp 8-62 - 8-63.
[63] OWEN, B.
"PIN - GaAs FET Optical Receiver With a Wide Dynamic Range."

[64] FRODSHAM, D.G. and BAKER, D.J.
"Computer-Aided Design of Optimal Infrared Detector
Preamplifiers." (SPIE)

[65] HULLETT, J.L. and MUOI T.V.
"A Feedback Receive Amplifier for Optical Transmission Systems."

[66] GOELL, J.E.
"Input Amplifier for Optical PCM Receivers."
The Bell System Technical Journal, Vol. 53, No. 9, Nov. 1974,
pp 1771 - 1793.

[67] OHKAWA, N., YAMADA, J.I. and HAGIMUTO, K.
"Broadband GaAs FET Optical Front-End Circuit Up to 5.6 GHz."

"Optical Detector Package."
The Bell System Technical Journal, Vol. 57, No. 6, July-August,
pp 1809 - 1822.

"The Design of Wide-Band Transistor Feedback Amplifiers."
[70] LEHENY, R.E., NAHMORY, R.E., POLLACK, W.A. BALLMAN, A.A., BEEBE, E.D., DeWINTER, J.C. and MARTIN, R.J.
"Integrated In_{0.53}Ga_{0.47}As p-l-n F.E.T. Photoreceiver."

[71] RUNGE, P.K.
"An Experimental 50 Mb/s Fiber Optic PCM Repeater."

[72] CHERRY, E.M.
"An Engineering Approach to the Design of Transistor Feedback Amplifiers."

[73] KUCERA, V.
"Discrete Linear Control - The Polynomial Equation Approach."

[74] CUONG, D.M. and NEIRYNCK, J.
"Constant Group Delay Approximation by Series of Bessel Polynomials."

[75] LITOVSKI, V.B.
"Synthesis of Monotonic Passband Sharp Cutoff Filters With Constant Group Delay Response."
[76] JOHNSON, J.R., JOHNSON, D.E., BOUDRA, P.W. (JR) and STOKES, V.P.
    "Filters Using Bessel-Type Polynomials."
    IEEE Trans. on Circuits and Systems. Vol. CAS-23, No. 2,

[77] JOHNSON, D.E., JOHNSON, J.R. and ESKANDAR, A.
    "A Modification of the Bessel Filter."
    pp 645 - 648.

[78] STORCH, L.
    "Synthesis of Constant-Time-Delay Ladder Networks Using Bessel
    Polynomials."

[79] HALPERN, P.H.
    "Solution of Flat Time Delay of First Frequencies."
    IEEE Trans. on Circuit Theory, Vol. CT-18, No. 2, March 1971,
    241 - 74.

[80] McGee, W.F.
    "Band-Pass Bessel Polynomial Filters and Delay Approximation."

[81] MARSHAK, A.H., JOHNSON, D.E. and JOHNSON, J.R.
    "A Bessel Retained Filter."
[82] MARTINEZ, J.R.

"Transfer Functions for Generalized Bessel Polynomials."
IEEE Trans. on Circuits and Systems, Vol. CAS-24, No. 6,
June 1977, pp 325 - 328.

[83] KOMIAK, J.J.

"A New Technique for the Design of Linear Phase Bandpass Filters
with Amplitude Selectivity."
1981, pp 253 - 256.

[84] AGNEW, J.

"Explorations in Number Theory."

[85] GUPTA, H.

"Selected Topics in Number Theory."

[86] GRINSTEAD, C.M.

"On a Method of Solving a Class Diophantine Equation."
Mathematics of Computation, Vol. 32, No. 143, July 1978, pp 936 -
940.

[87] MORDELL, L.J.

'Diophantine Equations."
[88] BARNETT, S.

"Polynomials and Linear Control Systems."

[89] D'AZZO, J.J. and Houpis, C.H.

"Linear Control System Analysis and Design."

[90] ROUTH, E.J.

"Stability of a Given State of Motion."

[91] STURM, C.

"Other Demonstrations of the Same Theorem."

[92] INTERNATIONAL BUSINESS MACHINES CORPORATION (as for [24])

PDIV, p 172.

[93] INTERNATIONAL BUSINESS MACHINES CORPORATION (as for [24])

POLRT, PP 181 - 182.

[94] HOUSEHOLDER, A.S.

"The Theory of Matrices in Numerical Analysis."
[95] SONTAG, E.D.
"On Generalised Inverses of Polynomial and Other Matrices."
pp 514 - 517.

[96] AYRES (JR), F.
"Schaums Outline Series - The Theory and Problems of Matrices -
S.I. (metric) Edition."

[97] GOLUB, G.H. and VAN LOAN, C.F.
"Matrix Computations."
Singular Value Decomposition.

[98] GOLUB, G.H. and REINSCH, C.
"Singular Value Decomposition and Least Squares Solutions."

[99] GOLUB, G. AND KAHN, W.
"Calculating the Singular Values and Pseudo Inverse of a Matrix."

[100] TAPIA, M.A. and MYERS, B.R.
"Generation of Concave Node-Weighted Trees."
pp 229 - 230.
[101] BOESCH, F.T. and YOULA, D.C.

"Synthesis of N + 1 Node Resistor N-Ports."
pp 515 - 520.

[102] DEL TORO V., and PARKER, S.R.

"Principles of Control Systems Engineering."
8. USING THE THREE PHASES OF EQUALIZATION

The computer programs discussed in the preceding chapter were tested with a number of examples. Using measurement data obtained from three sonar transducers. Curve fits for each transducer were generated using DRFA. The results of each curve fit and the rates of convergence are given in section 8.1. Each solution is PR which can be readily verified.

The equalization program of section 7.2 was then tested using the same examples as those used by Youla [1] in his original paper.

Using DEQDIO, an equalizer was produced for the rational function curve fit, found for the STRAZA SB4L measurement data [2].

The equalizer back-impedance function, thus produced, is also given in section 8.2. Finally, in section 8.3, the equalizer back impedance function, derived for the STRAZA sonar transducer in section 8.2, is synthesised using the load realization program DLREAL.
8.1 Curve Fit Examples

The rational function curve fitting technique, is applicable to measurement data from a variety of unknown loads. Three different examples have been chosen for curve fitting illustration. In all cases the data has been obtained from input terminal measurements of sonar transducer impedances or admittances. Although each set of measurement data is taken from the same type of transducer, the response of each is quite different. Example 1 was taken from the manufacturers frequency dependent impedance data for the STRAZA SB45C sonar transducer [2].

Spot measurements of the manufacturers real and imaginary impedance curves were made in 2.5 KHz steps giving 20 data points. A curve fit, to the 20 points, was generated using 100 iterations of the Marquardt method and 79 iterations of the VF02AD routine, which generated a curve fit with minimum objective function value $1.027 \times 10^{-3}$ and rational function,

$$Z(s) = \frac{(3.958 \times 10^{-16})s^3 + (5.724 \times 10^{-12})s^2 + (3.324 \times 10^{-6})s + 1.693 \times 10^{-2}}{(1.288 \times 10^{-24})s^4 + (1.99 \times 10^{-16})s^3 + (1.177 \times 10^{-14})s^2 + (7.255 \times 10^{-11})s + 2.539 \times 10^{-9}}$$

(8.1)

The 20 data points and corresponding rational function estimates are shown in Fig 8.1.
Fig 8.1 Curve Fit to Commercial Sonar Transducer Impedance

a) **Real**

![Real Impedance Graph](image)

b) **Imaginary**

![Imaginary Impedance Graph](image)
From poor starting values the Marquardt [3] minimization converged very rapidly but stagnated after 100 iterations, as shown in Fig 2(a). Initially the VF02AD [4] method diverged as the constraints were satisfied, but slowly converged to the minimum after 52 iterations, as shown in Fig 2(b), at which point the gradients were less than $1 \times 10^{-30}$.

Fig 8.2 Convergence of Curve Fit to Commercial Sonar Transducer Data

a) Marquardt

b) VF02AD
Example 2 is a curve fit to the impedance of a transducer made at the University of Canterbury. The transducer is intended for use under water, but 31 measurements of its terminal admittance were made in air. Applying 100 iterations of the Marquardt routine [3] and 19 iterations of the VF02AD routine [4], a minimum objective function value of $1.054 \times 10^{-2}$ and the rational function

$$Y(s) = \frac{(5.304 \times 10^{-22})s^4 + (2.737 \times 10^{-17})s^3 + (1.308 \times 10^{-11})s^2 + (3.99 \times 10^{-7})s + 1.615 \times 10^{-2}}{(1.623 \times 10^{-13})s^3 + (2.113 \times 10^{-9})s^2 + (3.077 \times 10^{-3})s}$$

were obtained.

The experimental measurements and corresponding rational function estimate are shown in Fig 8.3.
Fig 8.4 gives the rates of convergence of Marquardt [3] and VF02AD [4] minimization methods.

The transducer of example 2 was submerged and measurements of admittance repeated, yielding data used in example 3. A curve fit was generated with minimum objective function value $5.328 \times 10^{-3}$ and rational function.

$$Y(s) = \frac{(7.79 \times 10^{-22})s^4 + (8.102 \times 10^{-17})s^3 + (2.874 \times 10^{-11})s^2 + (7.315 \times 10^{-8})s + 1.77 \times 10^{-2}}{(2.531 \times 10^{-13})s^3 + (2.632 \times 10^{-8})s^2 + 6.367 \times 10^{-3}}$$
Fig 8.4 Convergence of Curve Fit to Marine Transducer in Air

a) Marguardt

b) VFO2AD
Fig 8.5 shows the given input data and corresponding rational function estimates.

Fig 8.5 Admittance of Submerged Marine Sonar Transducer

a) Real

b) Imaginary
Minimization routine rates of convergence are shown in Fig 8.6.

Fig 8.6 Convergence Rates of Curve Fit to Submerged Marine Transducer

Admittance

a) Marguardt

Unit weights were used in the three preceding examples.
Inadequacies of conventional methods of polynomial ratio rational function data fitting, and the additional difficulty of imposing positive real constraints on coefficients, has led to the development of a least squares rational function curve fitting technique using computer based minimization routines. The square deviation objective function, to be minimized, is defined from the partial fraction expansion of a rational function, which contains 4 basic partial fraction forms.

A limited number of explicit PR constraints, on the possible parameter values, can be obtained from the 4 basic partial fraction forms. To avoid overflow during minimization, a biased objective function, derived from the original, is actually used in the minimization process. From a number of results it is evident that the biased objective function is suitable for obtaining good curve fits.

Using the biased objective function the rate of minimization, and the accuracy of solution, may be enhanced by selection of suitable difference weights and suitable starting points for the pole polynomials.

Rational function curve fits obtained are illustrated with 3 quite different examples. In all cases convergence is rapid using the unconstrained Marquardt method of minimization, although the result is not necessarily positive real. The Marquardt [3] result is, however, a good starting point for application of the constrained Lagrange Multiplier penalty function method of minimization [4], which reaches a positive real minimum least squares rational function after only a few iterations.

Of the 3 examples considered, the closest curve fit was obtained in example 1 using the manufacturers impedance data and unit weights. In this
case the measurement error was better than 0.2% which is significantly smaller than the 10% of the other 2 examples.

Examples 2 and 3 illustrate the effects of resonance damping on the rate of convergence and quality of results. In example 3 the number of poles and zeros, and their approximate locations is clear from the data; good starting points can be selected and an accurate curve fit obtained. When the transducer is submerged, as in example 3, it is difficult to see where, by eye, the poles and zeros might be located. It is clear however, that a good curve fit can be obtained in this case. Moreover subsequent synthesis of the circuit components and values as those obtained for a canonic equivalent circuit.

Therefore the partial fraction objective function minimized by well known optimization techniques, and the imposition of appropriate constraints, is suitable for deriving polynomial-ratio-rational-describing-functions from experimental data.
The techniques of equalization can be applied to a variety of PR load impedances. In order to test its validity four equalization examples were considered.

The first three examples are taken from Youla’s original paper [1]. Initially a parallel resistor capacitor combination was equalized using the program DEQDIO.

Consider a capacitor and resistor connected in parallel as shown in Fig. 8.7.

**Fig 8.7 The Bode Parallel R-C Load**

![Diagram of parallel R-C load with C = 1F and R = 3.1416]
It is desired to produce a non-degenerate equalizer, such that the equalizer and load will have a Butterworth transducer gain characteristic with \(1/\text{rad/s}\) cut frequency and an optimum DC gain.

Using Laurent series expansion techniques, Youla derives an expression for the maximum attainable DC gain for equalizer and load transfer

\[
k^2 \leq 1 - \frac{2 \sin(\pi/2n)}{w_c} \left( \frac{1}{RC} - \sum_{m=1}^{m} a_r \right)^{2n} \tag{8.2}
\]

where the filter order is \(n\), \(w_c\) is the 3dB frequency of the filter and the \(a_r\) are the zeros of an arbitrary all pass function.

Using (8.2), the maximum attainable DC gain for a second order Butterworth equalizer is \(k = 0.9085\).

The computer solution has a maximum gain of \(k = 0.9085\).

Example 2 is a Darlington type "C" load shown in Fig 8.7 \(R_1 = R_2 = 10^6\) ohms, and \(C = 1\) pF. The desired response cut frequency of a second order Butterworth equaliser is \(2 \times 10^6\) rad/s.

Darlington's "C" type load is shown in Fig 8.8.
From Youla's calculation, the maximum attainable gain is $k = 0.9959$, and the corresponding inverse scattering parameter $S(s)$ is,

$$
S(s) = \frac{s^2 + 2.653 \times 10^6 s + 3.520 \times 10^{12}}{s^2 + 8.886 \times 10^6 s + 3.946 \times 10^{12}}
$$

(8.3)

The computer result is very close to this, with a DC gain of $k = 0.9960$ and an inverse scattering parameter,

$$
S(s) = \frac{s^2 + 2.653 \times 10^6 s + 3.520 \times 10^{12}}{s^2 + 8.886 \times 10^6 s + 3.946 \times 10^{12}}
$$

(8.4)

Example 3 is similar to the example 2. Though, in this case the desired equaliser cut frequency is $\omega_c = 2 \times 10^5 \text{rad/s}$.

Youla's result [1], has a DC gain of unity, and all pass function,

$$
a(s) = \frac{s - 0.927 \times 10^6}{s + 0.927 \times 10^6}
$$

(8.5)

The associated inverse scattering parameter $S(s)$ is,

$$
S(s) = a(s)S(s) = \frac{s^2(s - 0.927 \times 10^6)}{(s + 0.927 \times 10^6)(s^2 + 8.86 \times 10^6 s + 39.48 \times 10^{10})}
$$

(8.6)

The computer solution to the problem yields a DC gain of unity, with an all pass function,
\[ a(s) = \frac{s - 0.7395 \times 10^6}{s + 0.7395 \times 10^6} \] (8.7)

and an inverse scattering parameter numerator is \((s - 0.7395) (s^2 + 433.85 + 94.12 \times 10^3)\),

\[ S(s) = \frac{a(s)S_0(s)}{(s - 0.7395 \times 10^6)(s + 0.7395 \times 10^6)(s^2 + 8.886 \times 10^6 s + 39.48 \times 10^{10})} \] (8.8)

Unfortunately, the all pass function generated by EQDIO differs from that obtained by Youla [1], however in view of the other errors found in this example it is conceivable that a mathematic error has been made by Youla [1].

It is not possible to find any errors as Youla shows no working of the problems [1].

Fortunately all other results agree favourably with those derived by Youla [1] using Laurent series expansions.

In all three examples most of the computer results compare favourably with the Laurent series results derived by Youla [1] to three significant figures.

Finally it was decided to equalize the Sonar transducer impedance (8.1) found by curve fitting a rational function to the measurement data given by the manufacturer.

The response chosen for Equalization was a band pass version of the third order Bessel filter with a centre frequency of 1 rad/s, a bandwidth of approximately 1 rad/s and a terminal impedance of 1Ω.
In order to produce a useful curve fit the sonar transducer impedance was normalized such that the frequency range of interest, 50 KHz - 110 KHz would fall within the Bessel filter Bandwidth with a midband impedance of 1Ω. Setting the frequency of normalization to the Geometric mean, of 50 KHz and 110 KHz, the transducer impedance was normalized using the relationship

\[ S = 74.162 \text{ KHz} \quad S' \]

Where \( S' \) is the normalized frequency variable.

It is fortunate that scaling the sonar transducer by the geometric mean is sufficient to ensure that the frequency range of interest falls within the normalized bandpass Bessel-filter response. In general this will not occur therefore it may be necessary in some applications, to use a bandpass Bessel filter response with a non unit bandwidth and unit centre frequency to cover the desired frequency range of interest. Simple lowpass to bandpass frequency response conversion cannot be used to generate the desired bandwidth at unit centre frequency [5].

The program DEQDIO alters the centre frequency gain of the filter response in the same way as the DC gain of the low pass filter as discussed in section 7.3.

Using DEQDIO the back impedance of the sonar transducer equalizer was found,

\[ Z_2(s) = 140.1 \times 10^{12} s^7 + 22.99 \times 10^{15} s^6 + 31.28 \times 10^{21} s^5 + 14.39 \times 10^{27} s^4 \]
\[ + 10.62 \times 10^{33} s^3 + 1.874 \times 10^{39} s^2 + 133.15 \times 10^{42} s + 1.034 \times 10^{36} \]
\[ s^8 + 3.219 \times 10^{6} s^7 + 9.257 \times 10^{12} s^6 + 17.66 \times 10^{18} s^5 + 10.10 \times 10^{24} s^4 \]
\[ + 5.841 \times 10^{30} s^3 + 1.663 \times 10^{36} s^2 + 242.7 \times 10^{39} + 173.0 \times 10^{45} \]
The correctness of the equalizer is verified in the following section where the equalizer and load are realized as reactive networks terminated in resistors using DLREAL as discussed in section 7.3.
8.3 Load Impedance Realization Examples

Once the back impedance of the equalizer was found for the commercial sonar transducer of the preceding section the equalizer back impedance was realized using the techniques discussed in section 7.3 yielding the circuit shown in Fig 8.9.

Fig 8.9 Realization of Equalizer Back Impedance (8.9)

\[ Z_2(s) = 140.1 \times 10^{12} s^7 + 22.99 \times 10^{15} s^6 + 31.28 \times 10^{21} s^5 + 14.39 \times 10^{27} s^4 \\
+ 10.62 \times 10^{33} s^3 + 1.874 \times 10^{39} s^2 + 133.15 \times 10^{42} s + 1.034 \times 10^{36} \\
s^8 + 3.219 \times 10^{63} s^7 + 9.257 \times 10^{125} s^6 + 17.66 \times 10^{185} s^5 + 10.10 \times 10^{244} s^4 \\
+ 5.841 \times 10^{30} s^3 + 1.663 \times 10^{36} s^2 + 242.7 \times 10^{39} + 173.0 \times 10^{45} \]
In order to simulate the equalizer and load using the computer program LINSIM [6], the impedance function of the commercial sonar transducer (8.1),

\[ Z(s) = \frac{(3.958 \times 10^{-16})s^3 + (5.724 \times 10^{-12})s^2 + (3.324 \times 10^{-6})s + 1.693 \times 10^{-2}}{(1.288 \times 10^{-24})s^4 + (1.99 \times 10^{-16})s^3 + (1.177 \times 10^{-14})s^2 + (7.255 \times 10^{-11})s + 2.539 \times 10^{-9}} \]

The load realization program DLREAL was also used to realize the sonar transducer as a reactive network terminated with a resistor. The corresponding network is shown in Fig 8.10.

**Fig 8.10** Realization of Commercial Sonar Transducer Impedance Function
In order to verify that the circuit of Fig 8.10 is a valid representation of equation 8.1 the circuit was simulated using the linear interactive circuit simulation program LINSIM [6]. By connecting a 1A constant current source between the input terminals, as shown in Fig 8.11,

![Fig 8.11 Simulation of the Input Impedance of the Sonar Transducer Equivalent Circuit Given in Fig 8.10.](image)

The input impedance of Fig 8.10 is found. The voltage across the input terminals is numerically equal to the impedance. The frequency of the constant-current source was varied over the same frequency range as the measurement data giving the frequency varying impedance shown in Fig 8.12. The measurement data is also plotted in Fig 8.12 for comparison.
The frequency dependent impedance of Fig 8.12 matches the measurement data as well as the original equation 8.1 generated by the rational function curve fit in section 8.1.

Having verified that the load realization program LREAL, of section 7.3, will produce a valid equivalent circuit it was decided to measure the equalized voltage transfer, of the sonar transducer impedanced, from a voltage source with real internal resistance to the terminating resistor of the sonar transducer equivalent circuit. By exciting the equalizer input with a 1V constant voltage source the voltage across the terminating load resistor is numerically equal to the transfer function of the load and equalizer combination.
The gain and phase of the equalized sonar transducer load shown in Fig 8.13 are shown in Fig 8.14.

**Fig 8.13 Measurement of Equalized Sonar Transducer**

**Fig 8.14 Equalized Sonar Transducer Transfer Function**

a) Gain

![Diagram](image-url)
The gain varies by less than 1 dB over the frequency range of interest as shown in Fig 8.14 (a). The phase of Fig 8.14 (b) is nearly linear and corresponds to the expected phase of a 3rd order Bessel filter function [7].

These measurements verify the validity of each phase of the load equalization procedure as discussed in chapter 7.
REFERENCES - CHAPTER 8

[1] YOULA, D.C.
   "A New Theory of Broad-Band Matching."

[2] STRAZA INDUSTRIES
   "Stage Industries Type 8B45C Deep Submergence Acoustic Transducer."
   Straza Industries, 790 Greenfield Drive, El Cajon, California, 92021.

[3] HOPPER, M.J.
   "Harwel Scientific Subroutine Library - A Catalogue of Subroutines"

[4] HOPPER, M.J. (as for [3])

[5] SKILLING, H.
   "Electric Network."
   John Wiley and Sons, INC, 1974. pp

[6] EDWARD, L.N.M.
   "Liasion, A Linear Electrical Network Simulation and Optimization Program."

"A Bessel Rational Filter."

9.1

9. FUTURE PROPOSALS FOR OPTICAL AND ELECTRONIC DESIGN AND EXPERIMENTATION

During the course of optical and electronic design and experimentation discussed in the preceding chapter, proposals for new techniques of design and new experiments have occurred. Time limitations have prevented the realization of equipment, using new design techniques, and experimentation to verify the validity of the new ideas for application to the transmission of optical information. However some of these ideas have been developed to some extent at a theoretical level.

In this chapter, it is proposed to review the theory of new ideas intended to enhance the characteristics of optical communications systems, and new techniques of broadband circuit design based on the techniques discussed in chapter 7. These techniques could be tested at some future date.

To improve the SNR of detection, by coherent or correlative means as discussed in chapter 5, a local oscillator laser of great frequency stability is required. Such lasers are very expensive; their application in commercial situations is therefore quite unattractive. In section 9.1 a technique of differential detection is introduced which does not require the use of a highly stable local oscillator. With a minimum of sophisticated equipment, it may be a commercially attractive proposition.

In section 9.2, reduced carrier optical transmission systems are discussed. Using the reduced carrier, the receive local oscillator could be locked to the carrier such that the local oscillator tracks the frequency fluctuations of the transmit laser. Using this technique the
local oscillator laser need not be of high quality. To alter the local oscillator frequency an external frequency modulator may be used.

In section 9.3, a new technique of broadband amplifier design is introduced. Based upon linear models of the transistor, it offers a compromise between the design techniques discussed in chapter 6. It involves not only the input and output impedance terminations of the transistor but its overall transfer characteristics.

In some cases the loading impedances of the transistor may be non-PR due to the presence of negative resistance. The LDE technique of equalization is extended to application with non-PR loads, in section 9.4.

In order to extend the LDE techniques of equalization to cope with non-PR loads new LDE solution constraints are required. Some preliminary observations about these constraints are made in section 9.4.
9.1 THE DIFFERENTIAL OPTICAL DETECTOR

The frequency content of light emerging from a laser was discussed in section 3.1.4. Without stabilization a laser emits a signal in which several cavity modes are present. During its operation, the laser tends to move from one predominant cavity mode to another, emitting a different frequency at each mode. The frequency range, over which the movement occurs, defines the emission bandwidth $v_i$ of the laser \[1],\[2].

Associated with the laser emission bandwidth is a period of time $t_l$ which represents the maximum rate at which changes in amplitude or frequency can occur if the output is confined to a bandwidth of $V_L$. Whence

$$\Delta t_l = \frac{1}{\Delta v_i} \tag{9.1}$$

Associated with the time $\Delta t_l$, which is known as the coherence time, is a length $\Delta l_L$ in free space over which light can travel during the coherence time \[3],\[4],\[5].

$$\Delta l_L = C \Delta t_l = \frac{C}{\Delta v_i} \tag{9.2}$$

Where $l_L$ is the laser output coherence length.

Using a beam splitter, as shown in Fig 9.1, one may separate the laser output into two distinct beams. Using a second beam splitter (shown in Fig 9.1 also) the two beams may be recombined to form an interference pattern.
When separated, one lightbeam travels over a path of length $l_1$, the other over a path length of $l_2$. If the path length difference $\Delta l$ between the two beams given by,

$$\Delta l = |l_2 - l_1| \quad (9.3)$$

is less than the coherence length, that is,

$$|l_2 - l_1| < \Delta l_c \quad (9.4)$$

an interference pattern is observed on the screen [3],[5]. If the path length difference is greater than the coherence length,

$$|l_2 - l_1| > \Delta l_c \quad (9.5)$$

it is not possible to observe an interference pattern [3],[5]. In fact, an interference pattern is present but it is moving rapidly and cannot therefore be observed.
If the screen was replaced by a photo-receiver, the output of the photo diode would vary according to the path length difference of the interferometer.

If
\[ |l_2 - l_1| > \Delta l_L \]
a direct current would be generated by the photo-receiver [5].

If however,
\[ |l_2 - l_1| < \Delta l_L \]
a noise like signal would be observed across the electrical bandwidth of the photo-receiver [5].

If the photo-receiver output is integrated, the two input waveforms to the photo-receiver are correlated. The SPD of the photo-receiver forms the multiplier.

When the path length difference between the two beams is less than the coherence length, the integrator output will be large. When the path length difference exceeds the coherence length, the integrator output will be much lower [5]. In the case where the path length difference is less than the coherence length the two fluctuations in amplitude and frequency are correlated [5]. When the path length difference is greater than the coherence length the two beams are uncorrelated; or more accurately they are only partially correlated [5].

The reduction in correlation due to the path length difference may be exploited to reduce the effects of relative frequency fluctuations between the local oscillator and the transmit laser.
Consider the receiver configuration shown in Fig. 9.2.

The incident light is split into two beams using a beam splitter as shown. The local oscillator is mixed with one beam and the two are shone onto a photo-receiver. The second beam is mixed with the local oscillator signal such that the path length difference of the two local oscillator beams to the detectors exceeds the coherence length. Using a photo detector the second beam combination is detected. Each detector produces an electrical signal containing the modulation and frequency fluctuations of the input beam, and the frequency fluctuations of the local oscillator beam which are independent of each other.

The detector outputs are mixed and integrated. The fluctuations due to local oscillator noise will be uncorrelated. The fluctuations due to the input beam will be correlated, therefore the output of the receiver will be due only to the fluctuations in the input beam. These fluctuations are a combination of source-laser frequency instability and modulation. Consider now, a transmitter of the type shown in Fig 9.3.
The output signal of the laser is divided into two beams of orthogonal polarization. Using a modulator (discussed in section 4), each beam is phase modulated with the same signal. If the path length difference of each beam to its modulator differs by more than a coherence length, the fluctuations in each beam due to the laser will be uncorrelated. The fluctuations in each beam due to the modulation however, will be correlated.

The two beams are then combined and transmitted down an optical fibre to a receiver, similar to that of Fig 9.2. The beam splitter is replaced by a polarized beam splitter such that the two orthogonally polarized beams can be separated and mixed with the local oscillator signals.

The output of each detector consists of fluctuations due to modulation, the transmit and local oscillator lasers. However the fluctuations due to the lasers are uncorrelated and produce low outputs from the integrator. The modulation is however completely correlated and is easily detected at the output of the integrator.
Thus the dependence of the output on stability of the transmit and local oscillator lasers has been reduced. Provided information rates are significantly lower than fluctuations due to the lasers, highly stable lasers are not needed for such a transmission system.
9.2 Reduced Carrier Optical Transmission Systems

Another technique that may be employed to reduce the significance of transmit laser and local oscillator laser instability is to lock the local oscillator to the transmit laser. By transmitting a low level pilot carrier with the modulated signal, a reference signal will be available at the receiver to lock the local oscillator to the transmitter. The only fluctuations in amplitude, frequency or phase not tracked by the local oscillator will be those due to the modulation.

In radio systems the phase locked loop (PLL) is often used to lock the local oscillator to the signal transmitted and demodulate the incoming signal. Once the PLL is locked it tracks changes in the incoming signal at a rate determined by the loop gain [6].

A typical PLL for use in radio systems is shown in Fig 9.4 [6].

Fig 9.4 Block Diagram of Electrical Phase Locked Loop
The PLL consists of a phase comparator, a loop filter and a voltage controlled oscillator (VCO).

By using optical components corresponding to the electrical components of the PLL of Fig 9.4, an optical PLL can be built. Enloe and Redda [7] describe the contribution of an optical PLL intended for use at 1.52 μm optical wavelength. A block diagram of their PLL is shown in Fig 9.5.

**Fig 9.5 An Optical Phase Locked Loop**
With an appropriate choice of loop transfer function, the optical PLL will track changes in the frequency of the transmit laser, maintaining a constant frequency difference between the two.

If the rate of transmit laser frequency changes is different from the modulation frequency, the modulation may be recovered at the output of the photo detector.

Using filters, fluctuations due to transmit-laser frequency instability may be eliminated.

If changes in frequency due to transmit laser instability occur at the same rate as frequency changes due to the modulation, a carrier signal, containing only the transmit laser frequency information must be used to lock the optical PLL.

It is not possible to produce optical filters of sufficiently narrow bandwidth to isolate the carrier to lock the PLL. However if extra cavity laser modulation is used the carrier may be removed using a beam splitter. Using two polarizers it is possible to transmit the carrier and signal independently, if they are orthogonally polarized.

If the carrier and signal are transmitted down a polarization preserving optical fibre, the carrier and signal may be extracted again using two polarizers at the receiver.

The amplitude of the carrier transmitted must be sufficiently large to overcome system loss and polarizer loss and lock the optical PLL. Little advantage is gained if the carrier is of much greater amplitude than this.
When the optical PLL is locked to the carrier recovered at the receiver, the output of the PLL tunable laser will replicate the frequency instability of the transmit laser. If the tunable laser output is mixed with the incoming signal, homodyne or heterodyne detection can be performed without the presence of noise due to instability of the transmit-laser.

Fig 9.6 is a block diagram of a stabilized optical heterodyne detector.

**Fig 9.6 An Optical Heterodyne Detector Using PLL Local Oscillator**

[Diagram of optical heterodyne detector with labels for beamsplitter, mirror, laser, reduced pilot carrier, modulator, beamsplitter, polarizer, input signal, mirror, polarization preserving optic fibre, detected signal, input light, output light, error signal, variable bias, photo detector, low pass filter.]
9.3 AMPLIFIER DESIGN USING EQUALIZATION THEORY

Initially, the theory of equalization was introduced to improve the coupling of a broadband amplifier, designed using Kwok's technique [8], to the input and output devices and to improve the coupling between stages. As mentioned in chapter 5 Kwok's design technique has a number of failings which make it difficult to realize an amplifier which performs as predicted.

Using the theories of broadband equalization, theoretically based techniques of amplifier design can be found. These involve the matching of input and output to the reference impedance. In addition, it should be possible to equalize the transducer gain of the amplifier, to produce a broadband amplifier with optimum gain that can be connected through its matching circuits to any reference impedance circuit.

From measurements of the scattering parameters of a transistor in the common emitter mode, it is possible to calculate the equivalent two port impedance parameters, which are defined by the circuit given in Fig 9.7.

Fig 9.7 Impedance Parameter Circuit
The impedance parameter is readily described by the impedance parameter matrix equation [9],

\[
\begin{bmatrix}
V_1(s) \\
V_2(s)
\end{bmatrix} = 
\begin{bmatrix}
Z_{11}(s) & Z_{12}(s) \\
Z_{21}(s) & Z_{22}(s)
\end{bmatrix}
\begin{bmatrix}
I_1(s) \\
I_2(s)
\end{bmatrix}
\]  
(9.6)

Where \(V_1(s)\) is the frequency dependent input terminal voltage, \(V_2(s)\) is the frequency dependent output terminal voltage, \(I_1(s)\) is the frequency dependent output terminal currents in the direction shown, \(Z_{11}(s), Z_{22}(s), Z_{21}(s)\) and \(Z_{22}(s)\) are the frequency dependent impedance parameters.

If the output of the two port is terminated with a load impedance \(Z_L(s)\), the input impedance of the two port is given by [9].

\[
Z_{in}(s) = \frac{Z_L(s) Z_{11}(s) + Z_{11}(s) Z_{22}(s) - Z_{21}(s) Z_{12}(s)}{Z_L(s) + Z_{22}(s)}
\]  
(9.7)

Similarly the output impedance can be calculated when the input is terminated with a source impedance \(Z_S(s)\) [9].

\[
Z_{OUT}(s) = \frac{Z_S(s) Z_{22}(s) + Z_{11}(s) Z_{22}(s) - Z_{21}(s) Z_{12}(s)}{Z_S(s) + Z_{11}(s)}
\]  
(9.8)

In a similar manner one can define the terminated forward and reverse voltage gain [9],

\[
G_p(s) = \frac{Z_{21}(s) Z_L(s)}{(Z_L(s) + Z_{22}(s))(Z_S(s) + Z_{11}(s)) - Z_{12}(s) Z_{21}(s)}
\]  
(9.9)

and

\[
G_f(s) = \frac{Z_{12}(s) Z_S(s)}{(Z_L(s) + Z_{22}(s))(Z_S(s) + Z_{11}(s)) - Z_{12}(s) Z_{21}(s)}
\]  
(9.10)

Where \(G_f(s)\) and \(G_p(s)\) are the forward and reverse voltage gains, of the two port, respectively.
A large number of degrees of freedom, in design strategy, are available to produce an amplifier of the desired response. One could terminate the output terminal in the system reference impedance $Z_0$ and design an equalizer to match a voltage source, with internal impedance $Z_0$, to the input. The filter transfer function of the equalizer could be chosen to remove the frequency dependence of the forward transfer and improve the desired transfer function.

If for example, it was decided that the amplifier should exhibit a maximally flat delay, the equalizer transfer function would be of the form [10],

$$G(s) = \frac{G_B(s)}{G_F'(s)} \quad (9.11)$$

Where $G_B(s)$ is the Bessel filter transfer function and $G_F'(s)$ is $G_F(s)$ normalized to unit maximum gain.

The transfer to the resistive port of the amplifier load impedance, is equalized with a linear phase response.

A similar approach may be taken by equalizing the output for a given source impedance $Z_S(s)$.

Unfortunately associated with these approaches, is the risk of attempting to equalize the response to a non-bounded scattering parameter derived from (9.11). As it stands the Youla equalization theory cannot cope with such transfer functions [10]. Moreover the forward equalizer gain must change for each load equalized involving a new frequency dependence of $G_F(s)$. 
A more flexible technique of designing broadband amplifiers using the equalization approach, is to include the frequency dependence of amplifier forward and reverse transimpedances with the input impedances. In the way controlled mismatch of the equalizer output impedance and the input or output impedances of the amplifier may be exploited, to produce a constant forward power transfer with response \( G(s) \).

Consider the two port impedance network shown in Fig 9.8.

Fig 9.8 Two Port Network With Constant Forward and Reverse Transimpedance
The forward transimpedance is a constant $Z'_{21}(s) = g_f$, similarly the reverse transimpedance is also a constant $Z'_{12}(s) = g_r$. When circuit input and output impedances, $Z'_{12}(s)$ and $Z'_{22}(s)$ are as yet unknown. Regardless of the source or load impedances, it is desired to produce the same open circuit forward and reverse voltage gain as for the original transistor two port (Fig 9.7).

The open circuit input impedance, of the transistor, is $Z_{11}(s)$ and the open circuit output impedance is $Z_{22}(s)$. The open circuit voltage transfer functions of the two networks must be equal thus,

$$\frac{Z_{21}(s)}{Z_{11}(s)} = \frac{g_f}{Z'_{11}(s)} \quad \text{and} \quad \frac{Z_{12}(s)}{Z_{22}(s)} = \frac{g_r}{Z'_{22}(s)}$$

and so

$$Z'_{11}(s) = \frac{Z_{11}(s) g_f}{Z_{21}(s)} \quad (9.12)$$

and

$$Z'_{22}(s) = \frac{Z_{22}(s) g_r}{Z_{12}(s)} \quad (9.13)$$

Using computer based equalization techniques, the source and load impedances are matched to the modified input and output impedances of the open circuit two port.

This approach to amplifier design may be suitable for the design of broadband amplifiers with any desired frequency and phase response. Further theoretical and practical investigation of the technique, described above, is necessary before it is certain whether the technique will produce the amplifier desired.
Finally, a third approach to the problem of broadband amplifier design, using the Youla equalization technique, is possible.

Consider the terminated forward voltage gain of an impedance parameter two port network (9.9),

\[
G_f(s) = \frac{Z_{21}(s) Z_L(s)}{Z_L(s) + Z_{22}(s) (Z_s(s) + Z_{11}(s)) - Z_{12}(s) Z_{21}(s)}
\]

In terms of real frequency, \( j\omega \), \( G_f(j\omega) \) becomes,

\[
G_f(j\omega) = \frac{Z_{21}(j\omega) Z_L(j\omega)}{Z_L(j\omega) + Z_{22}(j\omega)(Z_s(j\omega) + Z_{11}(j\omega)) - Z_{12}(j\omega) Z_{21}(j\omega)}
\]

(9.14)

With a source impedance of zero, i.e. \( Z_s(j\omega) = 0 \) the transmission to the real part of the load is,

\[
T_f(j\omega) = G_f(j\omega) \left| \frac{Z_L(j\omega) + Z_L^*(j\omega)}{Z_s(j\omega) = 0} \right| = 2Z_L(j\omega)
\]

(9.15)

In order to apply Youla equalization to this situation, an equivalent impedance must be found such that the transmission to the real part of the load is given by (9.16). Thus

\[
\frac{Z_e(j\omega) + Z_e^*(j\omega)}{2Z_e(j\omega)} = \frac{Z_{21}(j\omega) (Z_L(j\omega) + Z_L^*(j\omega))}{(Z_L(j\omega) + Z_{22}(j\omega) Z_{11}(j\omega) - Z_{12}(j\omega) Z_{21}(j\omega)}
\]

(9.16)

Where \( Z_e(j\omega) \) is the unknown equivalent impedance.
Let \( Z_1(s) = \frac{N_1(s)}{D_1(s)} \) and \( Z_1(s) = \frac{N_1(s)}{D_1(s)} \)

\[
\begin{align*}
Z_{11}(s) &= \frac{N_{11}(s)}{D_{11}(s)} \quad Z_{12}(s) = \frac{N_{12}(s)}{D_{12}(s)} \\
Z_{21}(s) &= \frac{N_{21}(s)}{D_{21}(s)} \quad \text{and} \quad Z_{22}(s) = \frac{N_{22}(s)}{D_{22}(s)}
\end{align*}
\]

Using analytic continuation, and substituting each polynomial ratio for the impedances of (9.16), two diophantine equations are obtained.

\[
N_1(s)D_1(-s)+N_1(-s)D_1(s) = N_{21}(s)D_{11}(s)D_{22}(s)D_{12}(s)(N_1(s)D_1(-s)+N_1(-s)D_1(s))
\]

and

\[
D_1(-s)N_1(s) = D_1(-s)[N_1(s)D_{22}(s)+ N_{22}(s)D_1(s)]N_{11}(s)D_{12}(s)D_{21}(s) - N_{12}(s)N_{21}(s)D_{22}(s)D_{11}(s)D_{1}(s)]
\]  

(9.18)

Using equations (9.17) and (9.18), and the vector matrix techniques of solution given in section 7.2, \( N_1(s) \) and \( D_1(s) \) may be found for a given terminating impedance \( Z_1(s) \). \( Z_1(s) \) may be chosen to ease calculations or make \( Z_1(s) \) PR.

Having found \( N_1(s) \) and \( D_1(s) \), LDE equalization may be used to produce an amplifier with the desired power gain function.

It is clear that a number of variations on this technique can be found which may make calculations easier. Obviously considerable investigation of the proposal is required.

Using simple two port network theory, other techniques of building broadband amplifiers, by applying Youla equalization techniques, may be found.
9.4 Extensions to Equalization Theory

The theory of Broad-band equalization, discussed in section 7.2, is confined to the equalization of PR loads. Unfortunately, due to the gain and feedback combinations that occur in transistors, the impedances, to be equalized, in the design of a broadband transistor amplifier are not always PR. Over some regions of the transistor response the real part of the transistor input impedance can be negative.

In this section the Youla theory of equalization is re-examined. It is intended to find solution constraints which ensure that the equalizer back-impedance is PR, even when the load impedance $Z_L(s)$ is non-PR.

In his book on broadband circuit design, Chen [11] proposes a technique involving a circulator, for designing a broadband equalizer, when, $Z_L(s)$ is non-PR. Unfortunately it can be applied only with certain types of non-PR load impedance and is not general, it cannot be applied to all non-PR load types.

Here, it is intended to discuss some aspects of a new technique which can be applied to any non-PR load $Z_L(s)$.

Firstly the non-PR load is examined to determine which of the 3 PR conditions of section 7.1, it does obey and which conditions it violates.

Violations of the PR condition of a non-PR load are confined to violations of PR conditions 2 and 3. Its poles may be located in the RHP or on the imaginary axis of any multiplicity and with any residue value.
Moreover, the real part \( r_1(s) \), of the impedance \( Z_1(s) \), may become negative at any values on the real frequency \((j\omega)\) axis. The only PR condition that the non-PR load will obey is the condition that the coefficients of a series expansion of the numerator and denominator polynomials will be real. Therefore PR condition 1 is obeyed, a non PR load is real.

The equalizer itself must be PR, therefore, as in section 7.2, \( Z_2(s) \) must be PR.

Now, consider the scattering parameter \( S_{22}(s) \) for a non-PR load,

\[
S_{22}(s) = \frac{Z_2(s) - Z_1(-s)}{Z_2(s) + Z_1(s)} \tag{9.19}
\]

As in section 7.2, a more general scattering parameter \( S(s) \) may be introduced which obeys the relation,

\[
S(s) S(-s) = 1 - S(s) S(-s) \tag{9.20}
\]

Where \( S(s) \) is the forward transducer gain of the desired response and,

\[
S(s) = b(s) S_{22}(s) \tag{9.21}
\]

\( b(s) \) is an all pass function of the form,

\[
a(s) = \prod_{i} \frac{(s - b_i)}{(s + b_i)} \tag{9.22}
\]

Because \( Z_1(s) \) is no longer PR and \( Z_2(s) \) is unknown, \( b(s) \) cannot be determined at this stage, unlike its counterpart when \( Z_1(s) \) is PR.

Using synthetic division, the denominator of \( S(s) \) is divided into the numerator giving,
\[ b(s) - S(s) = 2r_1(s) \]
\[ \frac{Z_2(s)}{Z_1(s)} \]  

(9.23)

Where \( S(s) = a(s) S_0(s) \)  

(9.24)

And \( S_0(s) S_0(-s) = S(s)S(-s) = 1 - S(s) S(-s) \)  

(9.25)

\( A(s) \) is also unknown but is of the form,

\[ A(s) = \prod \frac{(s-a)}{s+a} \]  

(9.26)

By rearranging equation (9.23) an expression for \( Z_2(s) \) is found

\[ Z_2(s) = 2r_1(s)b(s) - Z_1(s) \]

\[ \frac{b(s) - S(s)}{Z_1(s)} \]  

(9.27)

If \( Z_2(s) \) is to be PR, it must obey the 3 PR conditions given in section 7.1.

Finally consider PR condition 3, \( r_e(Z_2(j\omega)) \geq 0 \) for all \( j\omega > 0 \).

Forming the real part \( r_e(Z_2(j\omega)) \)

\[ r_e Z_2(j\omega) = \frac{r_1(j\omega) (1 + |S(j\omega)|^2)}{|b(j\omega) - S(j\omega)|^2} \geq 0 \]  

(9.28)

The only violations of PR condition 3 that can occur are due to the zeros of \( r_e(j\omega) \). If \( Z_1(s) \) is PR, (9.28) is always obeyed. The only way in which (9.28) will be obeyed is if a function \( c(j\omega) \) is introduced which multiplies \( r_1(j\omega) \) producing even multiplicities of imaginary axis zeros in \( r_e(Z_2(j\omega)) \). In which case \( r_e Z_2(j\omega) \geq 0 \) always. Thus

\[ r_e Z_2(j\omega) = \frac{r_1(j\omega) C(j\omega) (1 + |S(j\omega)|^2)}{|b(j\omega) - S(j\omega)|^2} \]  

(9.29)
Therefore $C(s)$ contains imaginary axis zeros which must be of odd multiplicity. Unfortunately $|S(j\omega)|^2 \geq 0$ always, as is $|b(j\omega) - G(j\omega)|^2$, therefore $Z_1(s)$ must be modified to satisfy (9.29) by introducing $C(s)$.

If it is not possible to find a suitable method of modifying $Z_1(s)$, then non-PR loads violating PR condition 3 cannot be equalized directly by Youla methods. Obviously more investigation of Non-PR loads and methods of modifying their behaviour must be found.

Once $c(j\omega)$ is found, or if the load impedance $Z_L(s)$ obeys PR condition 2 diophantine equations may be derived. With the additional constraint that $Z_2(s)$ is PR, these could be solved for unknowns, $a(s)$, $b(s)$ and the unknown transducer gain $k$ where,

$$G(s) = k \frac{J(s)}{K(s)}$$

Using the methods of Load Synthesis discussed in section 7.3, an equalizer is then synthesised from $Z_2(s)$.
REFERENCES - CHAPTER 9

[1] BOYD, G.D. and GORDON, J.P.
"Confocal Resonator for Millimeter Through Optical Wavelength Masers."

"Fields and Waves in Communications Electronics".

"Coherent Light."

[4] RATNER, A.M.
"Spectral, Spatial and Temporal Properties of Lasers."

[5] HANBURY-BROWN, R.
"The Intensity Interferometer - its Applications to Astronomy."

[6] STREMLER, F.G.
"Introduction to Communication Systems."
   "Laser Phase Locked Loop."

[8] KWOK SIANG-PING.
   "UHF Broadband Amplifier Design."
   Motorola Application Note, AN-406, Motorola Inc.

[9] SKILLING, H.
   "Electric Networks."

[10] HA, T.T.
    "Solid-State Microwave Amplifier Design."

    "Theory and Design of Broadband Matching Networks."
10. **CONCLUSION**

In the nine preceding chapters, aspects of optical communications systems have been discussed in detail. The major emphasis is on the possibility of commercially applying some of the techniques discussed to reduce the cost of installing fibre optic communications systems. One must always be mindful of the commercial potential of any research conducted in the engineering field, as cost will dictate the extent to which any idea is used.

In chapter 1 and 2 the cost structures of trunk communications systems were discussed briefly. These are dependent on the domestic and commercial requirements within each country, and on the costs of competing technologies. A broad overview of the New Zealand telecommunications environment was given; the environment is unique to New Zealand. It dictates the overall design priorities of the New Zealand Post Office and therefore the areas in which most money is invested.

Within the New Zealand context, the development of the fibre optic communications network was discussed. In this context, the demand for long distance direct circuits, unique to New Zealand, has continued to move the Post Office towards the vast bandwidth capabilities of Optic fibre transmission systems. Interest in such fibre optic transmission systems has increased rapidly.

The Chapter 2, a historical view of optical communications was given. It highlights the advances that have been made since light was first used
as a means of communication. It also positions the present development and indicates the way ahead for optical communications. Optical communications system development is presently at the same stage as radio was just after the turn of the century.

The history of radio development over the past 80 years perhaps indicates the future path for the development of optical systems. Optimum methods of modulation and detection should become the modes of optical communications of the future. These will reduce the system costs as the distance between regenerators is increased and the number needed on any system reduced.

In chapter 2 the fibre optic equation was introduced. It expresses the relative effects of transmit power, optic fibre attenuation and noise in the SNR of detections. Using this equation with a margin of error and the power budget concept of system design, one can derive the maximum distance that can be achieved between regenerators for a given detection SNR or corresponding bit error rate.

The most effective means of increasing the spacing between regenerators is to alter the attenuation per unit length of optic fibre. It is exponentially related to the SNR and will therefore yield the most significant reductions. However other methods of SNR reduction include optimal detection techniques to improve the efficiency of detection and reduce the amount of noise power in the recovered signal. While such reductions are numerically smaller than reductions in cable attenuation, at very small values of attenuation, drastic increases in regenerator spacings can be made. 3dB increase in the SNR will result in a possible doubling of regenerator spacings. In this dissertation techniques of improving the detection efficiency have been examined.
Unfortunately, in order to apply some of the techniques, used in optimal radio systems, to optical systems a number of practical problems must be overcome. The most significant problem is one of laser frequency stability. Generally, a laser exhibits a very high $Q$, usually many orders of magnitude greater than a quartz crystal. Unfortunately the high $Q$ is of little importance. The figure of most importance is the absolute laser frequency stability, which must be of the same order as that of a crystal oscillator, if the laser is to be used in optimum heterodyne or homodyne methods of detection.

In chapter 3, the characteristics of laser light generation, frequency content and characteristics are discussed in some detail. In all types of laser the frequency content is determined by the frequency width over which stimulated emission will occur and its relationship to the dimensions of the optical cavity enclosing the active medium.

If the cavity length is reduced the number of modes that it can subtend, which fall within the bandwidth over which stimulated emission will occur, will fall. Unfortunately the laser gain will also fall, as the cavity will enclose a smaller population inversion, therefore the output power of the laser will fall.

The most effective approach seems to be to actively stabilize the laser using a feedback arrangement; this technique makes lasers very expensive.

If however frequency or phase shift techniques are used to modulate the light, one could also introduce signals into the modulator to stabilize the modulated light.
Uncertainty in optical frequency due to macro instability is not the only effect that one must reduce or tolerate. The second major effect causing frequency instability, or more correctly uncertainty, is due to the discrete nature of light. At low light levels the quantum mechanical uncertainty, between photon number and phase, becomes significant. As the optical power level falls the certainty in photon number rises. However the uncertainty in photon phase, and thus frequency, rises, broadening the frequency content of the received signal and reducing the SNR. When designing an optical system to receive very low light levels the phase uncertainty must be included in the SNR of detection.

Some of the more general aspects of quantum field theory are also introduced in chapter 3, as both a tutorial on the fundamental nature of light, and as a preliminary study to form an understanding of the concepts of the recently discovered phenomenon of the squeezed states of light. If the alien concepts of Dirac notation are removed or fully explained, it is possible to gain an elementary grasp of the physics involved in the quantum mechanical description of electro-magnetic fields.

Having characterised the nature of light generated by a Laser, methods of modulating such laser light are discussed in Chapter 4. Unlike the physical circuits involved in modulating the carrier of a radio signal, the modulation of light must be undertaken at the atomic level. By exploiting the optical properties of materials in electric, magnetic or acoustic fields it is possible to alter the amplitude, frequency, phase or polarization of the light in sympathy with changes in the field amplitude. Modulators manufactured from these materials may either be placed within the cavity, in which case the laser gain is exploited, or outside the cavity. Modulators placed within the optical cavity are subject to the frequency restrictions of the cavity dimensions. Modulators placed outside the cavity are largely independent of these restrictions.
Modulators exploiting the Electro-Optic or Magneto-Optic materials are usually expensive and complicated. The materials in which the effects are most pronounced are usually composed of expensive elements. In many cases they are unattractive for commercial applications.

Acousto-optic modulators are, on the other hand, simpler and use cheaper materials. With a suitable choice of material, very high modulation frequencies may be achieved using either surface acoustic waves or bulk waves. In chapter 4 a general overview of the bulk wave acousto-optic modulator was given. Using the quantum theory of light and sound the modulation process is discussed, emphasising the Bragg mode of deflection.

In the Bragg deflection mode, the angles of rays emerging from the acousto-optic modulator vary with instantaneous frequency. Such frequency dependent angles of variation impose constraints on the bandwidth over which the modulator may be used, as a functions of either receive aperture or spatial filter aperture position and size. Such aperture effects tend to restrict the bandwidth of application further than the inherent bandwidth associated with the modulator material itself. The maximum and minimum modulation bandwidth restrictions are quantized using simple ray optics, which give a general expression for bandwidth. Using more detailed geometric optics will yield a more accurate answer, however much analyses are dependent on the configuration of equipment. Each experimental configuration should be independently analysed, which is a time consuming process and of limited practical value in this experimental context.

The restrictions in modulation bandwidth imposed by the combination of frequency deflection angle and receive aperture may be overcome using passive optical components as discussed in chapter 6.
The acousto-optic technique of modulation may become the first method of generating FSK or PSK signals for commercial application. Acoustic methods of signal processing are already widely used in devices such as the Surface Acoustic Wave filter, the technology is well known.

Having characterized the transmitter of optical signals in Chapter 4, the characteristics of optical receivers were discussed in chapter 5. The basic behaviour of the vacuum-photo-diode were discussed, to give an overview of the basic detection process. From this analysis, the frequency dependence of the detector on electron transition time is derived. The electron transition time can also be viewed as a carrier lifetime, which fixes the electrical frequency response of semiconductor based photo receivers. The photo-multiplier tube is also introduced to illustrate the gain bandwidth trade-off found in optical detectors. Generally the higher the quantum efficiency of the detector, the narrower the bandwidth over which it can operate.

With the simple concepts involved in detection of light using a vacuum-tube-photo-diode, semiconductor photo detectors are briefly discussed. In commercial application, the silicon photo-diode is the most common detector type used; the SPD was also used in experimentation. It is therefore discussed in some detail. Using characteristic equations for the interaction of radiation and matter, the optical and electrical frequency responses of the SPD are characterized. The electrical bandwidth of the SPD is fixed by the combination of the average lifetime within the crystal and the time constant formed by the junction capacitance, extrinsic resistance, contact resistance and lead resistance, of the SPD. The optical frequency response is fixed by the frequency dependent reflectivity and absorbitivity of the semiconductors forming the diode.
Using the frequency dependent expressions, thus derived, an expression relating the electric field of the incident light to the detector output current is derived.

With the transfer function derived, the sources of noise in optical detection were discussed. SPD noise is due to a combination of thermally generated noise, intrinsic to the detector and optical noise, generated by the presence of light itself.

The thermal noise sources are similar to those found in any semiconductor junction diode due to linear resistance and leakage current.

The optical noise is due to the uncertainty in number and arrival times of the photons of light. The noise power is proportional to the incident optical power. The signal power is proportional to either the signal power or its square depending in the type of detection used. If a high-power local-oscillator is used, to coherently detect the signal, the optical-granular-noise will dominate the receiver noise, the quantum limit of detection will be reached and the SNR of envelope detected ASK coherently detected ASK, FSK and PSK are derived. Using an high powered local oscillator, quantum limited detection of each modulation type can be achieved. Therefore for a given transmit power, SNR improvements over ASK systems may be produced. PSK using a simple detector is the most efficient modulation type. It affords a 3dB gain over detection of the other modulation types discussed.

It is important to note that the process of mixing the two signals for coherent detection is somewhat similar to the addition of a local carrier prior to envelope detection of AM and modified AM signals. At low signal levels the SNR relation to signal power becomes highly non linear. The
amplitude of the signal is not significantly larger than the amplitude of
the signal multiplied by noise. This effect is due to the presence of
noise in both the local oscillator signal and information signal. Noise is
therefore multiplied by the signal as it multiplies the desired signals.

At low light levels, the contribution of phase uncertainty to the
noise power of detection, must also be included in the expressions for SNR.
As the intensity falls the number of photons fall. The uncertainty in
photon number is proportional to the square root of the number, therefore
the uncertainty also falls. Unfortunately due to the quantum mechanical
uncertainty between photon number an phase, the phase uncertainty of the
light beam and therefore its frequency, increases. Associated with the
phase uncertainty is a phase noise. Failure to include the quantum
mechanical phase uncertainty produces erroneous results in SNR
calculations. Especially when heterodyne and thermodyne detection systems
are compared.

The homodyne system of signal detection is merely a special case of
the heterodyne technique. It is a nonsense to suggest that an improvement
of 3dB, in detection SNR, can be acheived over an heterodyne method, using
optical homodyne detection.

In order to achieve further improvements in the SNR of detection in
optical systems, techniques of post detection processing were examined. At
low data speeds some improvement over the coherent detection schemes
discussed above can be achieved by correlating the output signals of two
such detectors, however at high speeds, the period over which the
correlation can be performed is very small; to achieve any improvement in
SNR at all, very high SNR 's of detection are required by each detector of
the correlator.
Another method of improving the detection SNR is offered by the concept of the squeezed state. Using techniques of non-linear optics, a state may be produced in which the uncertainty in photon number OP phase may be reduced below the quantum limit. The modulation is impressed upon light in the state and detected using a coherent detector. In this case the quantum mechanical noise is reduced below the traditional quantum limits, the detector SNR is thus improved.

Some experiments in this area have been reported, however the resultant squeezing is only just beyond what one would expect the experimental error to be. More research into this area is required before the technique could be used in commercial applications.

The SNR of detection, in coherent detection systems, is also reduced by misalignment of the local oscillator and signal beam. As the beams are misaligned the conversion efficiency of the SPD, as a mixer, falls.

To verify the SNR improvement proposals of chapter 5. Equipment must be built to process the signal in the manner proposed. Moreover, the signal levels at the output of a typical photo receiver are very small; they are of the same order as those obtained from a radio antenna. Before any form of signal processing can be attempted, the photoreceiver output must be amplified to a level suitable for mixing or digital level decision making. Amplification is also very necessary to ensure that additional noise of the processing circuitry has no significant effect on the SNR of detection. In chapter 6 the design of photo receivers, amplifiers, mixers and integrators was discussed.
The development of a photo receiver based on the Opto-electronics CD-10 SPD was discussed. The SPD is itself a broadband packaged in a case with appropriate thick film circuitry, to ensure its compatibility with a 50 Ω standard impedance system. The SPD was connected directly to 50 Ω stripline milled from epoxy insulated, double sided, circuit board. It was found that the stripline made in this way was quite suitable for the DC to 500 MHz frequency range of interest.

To protect each detector, a 3 terminal voltage regulator was used to supply the required bias voltage via radio-frequency-chokes and decoupling capacitors. These were present to prevent the entry of any high frequency signals into the 3 terminal regulator, which would cause it to fail. Each SPD and associated circuitry was mounted in on Eddystone diecast aluminium box, using insulated material to ensure that the circuitry ground was isolated from the metal of the box. In order to reduce the effects of differential atmospheric-noise signals, each box would be connected to a common earth point via a separate wire.

Once the photo receivers were built, they were tested using an Hitachi cooling fan as a modulator. Somewhat later, each was tested using the Acousto-Optic modulator.

Neither of the two photo receivers reached the specified frequency bandwidth. Unfortunately, to date, no reasons for this have been found.

In order to conduct some of the proposed experiments, temporary detectors, using the Motorola MRD 500 SPDs, were built. Using a simple bias circuit, mounted on strips of veroboard and connected directly to 50 Ω coaxial cable, the temporary detectors were most successful.
The lens was removed from one MRD 500 to expose the SPD crystal itself. Using this photo receiver, preliminary mixing experiments were conducted between beams selected from the ensemble produced by the AO modulator. Mixer products were observed between two beams emerging from a Mach-Zehnder Interferometer.

By selecting various beam combinations, from the AO modulator output beam ensemble, and adjusting the interferometer to optimize the mixer products, the temporary photo receiver was swept over a wide electrical frequency range from DC to 140 MHz. No significant reduction in photo-receiver sensitivity could be measured, at the high frequencies however the received signal strength was reduced by the frequency response of the AO modulator.

The Mach-Zehnder interferometer was used to align the optical beams to be mixed, because once properly adjusted, it ensures that the beams appear to emerge colinearly. Unfortunately, as the AO modulator input signal frequency is varied, the deflection angles of various beams in the output ensemble change. In order to mix two beams of different frequencies, the interferometer must be adjusted each time the frequency is changed. Using combinations of pre-AO-modulator and post-modulator lenses, the dependence of deflection angle, on modulation frequency, could be reduced. Unfortunately separation of carriers and sidebands of the zero and first orders of deflection was difficult. Further research into the reduction of deflection angle frequency dependence is required.

In order to amplify the photo receiver output signals, amplifiers capable of operating over very wide bandwidths are required.
Initially methods of mis-match design, using the Smith Chart, were investigated. Unfortunately Smith Chart techniques are generally applied to amplifiers operating over very small bandwidths; at high frequencies.

By applying the narrow band techniques over a range of spot frequencies it is possible to define input, and output, impedance: frequency curves which form tangents to the circles of constant gain at the spot frequencies. Unfortunately it is difficult to synthesise suitable circuits from such curves. Moreover, the spot frequency techniques requires the use of a unilateral transistor model. Such models are not sufficiently accurate for some transistors and configurations.

Using scattering parameter techniques and signal flow graph theory, it is possible to unilateralize the transistor and broaden the bandwidth over which it can be used. Such techniques are extremely complicated. These were abandoned in favour of the broadband design technique proposed by KWOK. [1]

Unfortunately, KWOKs technique has a number of flaws. It requires the use of a non-linear inductor of rather specific response. Such inductors are difficult to find. The technique also fails to consider the effects of input, output and interstage coupling, which must be roughly selected and then altered, using the energised circuit, largely on a trial and error basis.

Using kwok's [1] technique, an amplifier intended to operate over a 500 MHz bandwidth was designed. Correctly loaded, a one stage amplifier could produce a gain of 10 dB, well below the gain figure expected. Moreover, when connected to a second stage, the second stage loaded the first, such that the overall gain was again 10 dB.
It was believed that selection of suitable interstage coupling networks could overcome the gain reduction problem, thus extensive investigation of equalization was undertaken. It is discussed extensively in chapter 7.

Based upon the ideas of equalization, an entirely new technique of amplifier design was proposed in chapter 9. Unfortunately, time restrictions have not permitted the technique to be verified experimentally. It is therefore an area for future research.

If the amplifier is used to amplify digital signals its phase response must also be considered. Using the theory of equalization, discussed in chapter 7, it is possible to design an amplifier exhibiting a linear phase response, which should produce little pulse smearing.

Amplifiers with uncontrolled phase response cannot be used to their full -3dB bandwidth because the inter-symbol interference produced by pulse smearing must be reduced, for a given bandwidth, by reducing the data rate.

In order to achieve some gain for experimentation, amplifier gain modules based in the MWA 210 hybrid amplifier were built. These were designed according to the manufacturers specifications, on stripline milled from double sided epoxy insulated circuit board.

The power supply and BNC input and output sockets were mounted directly onto the board, which was, in turn, mounted on insulated stand-offs in the lid of an eddystone box. The circuit ground was again isolated from the metal of the box, which would be grounded via a separate wire.
Measurements of the amplifier gain were flat to beyond 500 MHz. Unfortunately the phase is completely uncontrolled.

Attention was subsequently focussed on the design and construction of integrators for long and short term correlation applications. By reversing the capacitor, of an analogue integrator, at regular intervals together with the input signal, amplifier drift errors can be removed.

In order to reduce the effects of amplifier leakage current, a second capacitor in series with the leakage path can be used to store the charge that leaks away from the capacitor. By switching this capacitor with the main integrator capacitor, the leakage charge can be added to the net integrator charge.

To further reduce the effects of leakage, reed relays were used in preference to MOS analogue switches. The use of a two-phase non-overlapping clock ensures that charge is not removed from the integrator capacitors, due to the brief short circuits, which occur during the reversal process if one clock, and its logical inverse are used.

The analogue integrator built, in the way described, exhibited stability of better than 10% over a 5 hour period. Moreover it was completely insensitive to amplifier offset voltage.

In integrator applications, the choice of amplifier frequency response is dictated by the proposed period of integration. If the integration period is short, the amplifier must respond rapidly, adding significant numbers half cycles of voltage to the total charge. It is possible to fix the amplifier frequency response, required for a particular application, by the error that may be tolerated in the accumulated total, due to an extra
half cycle of voltage. For a given integration period, the error is reduced by increasing the amplifier bandwidth.

In applications, where the integration period is for many hours, the required amplifier bandwidths are very small indeed, in which case operational amplifiers are quite suitable for the applications.

For much shorter integration periods, the amplifier must be made using high frequency transistors.

In order to perform some of the proposed post detector methods of signal processing, to improve the SNR of detection, a mixer or multiplier was required. Most commercial mixers are unsuitable for the task as they will not operate at D.C. For these reasons the differential mixer was studied.

Initially, an experimental mixer was built to gain some ideas about its performance characteristics. It was found that such a mixer should operate over approximately 500 MHz. During the informal tests the mixer was damaged.

Using MRF 901 transistors and current mirrors to stabilize the MRF901 collector currents, a second differential mixer was built. Unfortunately some reduction in the current mirror precision was required before the D.C. operating point, of the mixer, could be achieved. Each current mirror was by-passed with a resistor and capacitor, and isolated from the high frequencies, present in the mixer circuitry, with RF chokes.
Once the D.C. operating point was reached, the mixer response was measured. From the results, it is clear that some equalization of the two inputs, and the output, is required before it can be operated over the DC to 500 MHz desired frequency range.

Using the differential mixer as a building block, it is possible to build a full multiplier. Undesired signals are rejected by exploiting the common mode rejection of the differential pair. Unfortunately the full multiplier was not built.

In order to improve the input, interstage and output coupling of the amplifiers, mixers and integrators, the source and load impedance of each stage must be matched to the amplifiers input and output, over the desired frequency range. This is essentially a load equalization problem of which there are 3 phases.

The manufacturers data for high frequency transistors is normally given as scattering parameter measurements at a number of spot frequencies. In order to equalize the input and output impedance of the amplifier, the measured data must be described by a mathematical function, which fits the measurement data. Moreover, if a non linear element is used in the feedback circuit of an amplifier designed by kwoks method [1], the input and output impedance will be further modified. The modified input and output impedances must also be described by a mathematical function if they are to be equalized.

Once a mathematical description of the input and output impedances of the amplifier are found, the back impedance function of an equalizer must be found, to equalize the impedances of the amplifier to the desired source or load impedances. This is the second phase of load equalization.
Once the equalizer back impedance function is found, it must be realized by a reactive circuit terminated in a resistor. The load must therefore be synthesised. This is the third and final stage of equalization.

The three phases of load equalization were discussed in detail in chapter 7.

A variety of methods intended to characterize the manufacturers measurement data, with mathematical functions, were tried. Polynomial representations are unsuitable for the load characterization due to the nature of the equalization process. They cannot adequately describe poles located close to the real frequency axis. Therefore methods of fitting rational functions to the measurement data were sought.

Divided difference and orthogonal polynomial rational function data fitting techniques were unsuitable for the task. Divided differences produced rational functions which were the ratios of polynomials with complex coefficients. These could never represent a real impedance function.

The orthogonal polynomial techniques simply did not work at all.

The failure of established methods to adequately match the measurement data prompted investigation of rational function curve fitting using least-squares regression methods. Based upon general minimization routines, the first successful rational function curve fit was produced, using a product of zeros expressions, for the numerator and denominator polynomials.
Unfortunately it was not possible to impose the PR constraints on the rational functions produced, as required by the theory of broadband equalization. Therefore the product expression in the objective function was replaced by a sum of partial fractions.

Each element of the partial fraction sum has one of four regular forms, enabling one to enforce the PR conditions as a constraint of minimization. Using applications of the Marguardt and VA02AD methods of minimization, PR rational function curve fits to measurement data, could be produced.

The quality of the curve fit and rates of convergence can be altered by selecting different starting points and different difference weights.

Once the measurement data has been characterized, the back impedance of an equalizer must be found.

While perfectly general the Youla technique of equalization [2] cannot be readily applied to complicated load functions. It uses Laurent series expansions which produce sets of simultaneous nonlinear equations. Moreover, it is difficult to apply the Youla Laurent series approach to transfer function frequency responses, which do not have closed form. Explicit expressions for DC or mid-band gain cannot be found, therefore the Laurent series procedure must become iterative and very time consuming, if not completely intractible.

To write a computer program to perform Laurent series equalization is also very difficult, as the Laurent series form changes with the location and type of each transmission zero.
Based on Youla's original theory [2] a new method of load equalization has been found using the theory of LDE's. The technique is much more readily solved using a digital computer.

Each diophantine equation is represented as a vector-matrix product. The matrix contains the known polynomial coefficients and the vector contains the unknown. Using the pseudo-inverse, the unknown vector is found.

The vector matrix representation is also used to improve the accuracy of polynomial synthetic division, reducing the cumulative effects of error.

The LDE equalization technique can be used to find equalizers for loads of any complexity without changing the techniques of solution. Using the technique it is possible to predict the existence of common factors in the solution.

Unfortunately the LDE technique can produce trivial solutions. To overcome this, a bounding polynomial was introduced to force solutions away from the trivial solutions and generate a valid equalizer back impedance.

The computer based LDE equalization technique is applied to arbitrary loads in the program DEQDIO. A major tool used in this program is the Routh array, which is used to find imaginary axis zeros and their multiplicities. It is also used to verify the PR nature of the solution; on its own and in the Sturm PR test.
In order to ensure that the appropriate error tolerance is carried throughout the program to normalize polynomials, removing leading zeros, after each polynomial or matrix operation, a new estimate of the absolute error is generated from the errors of the contributing polynomials. If the error estimates are too low, the results become inaccurate. Polynomial coefficients, which are very small are still considered larger than zero, in which case some polynomials which are PR violate PR tests and equalization solutions cannot be found. Therefore reasonable values of finite precision error are needed.

Having found the equalizer back impedance, a circuit, to realize it, must be found. The only configuration appropriate for an equalizer is a reactive network which is terminated in a resistor.

In many cases, Cauer synthesis is sufficient to realize a reactive network terminated in a resistor. However, in some cases, the synthesis procedure terminates with neither a constant term nor a remainder containing poles or zeros, located on the imaginary axis.

Again in some cases, Brune synthesis may be a suitable method of continuing the realization procedure. Unfortunately in general, Brune synthesis involves the extraction of a series resistor, thus introducing a resistor in addition to the terminating resistor. In this case Brune Synthesis cannot be used to realize the network. However in a small number of cases, Brune synthesis does not produce an extra resistor, in such cases it is quite suitable to realize the network.

Due to the failure of the Brune method of synthesis, to realize a reactive equalizer terminated with a resistor, in the general case, another method of load realization was sought.
By implementing the method of Darlington synthesis using LDE's in vector matrix form, the load realization procedure can be continued. Each time it is applied the remainder rational function, polynomial orders, are reduced by two.

Unlike Guillemin's technique of cascade Darlington synthesis, the vector-matrix-LDE approach is readily included in a load synthesis program. The form of solution permitted in Darlington synthesis confines an infinite number of solutions to one unique solution for each transmission zero eliminated. Using the matrix pseudo inverse the components of each cascade section produced by Darlington's synthesis are found.

In both the Darlington and Brune Synthesis procedures, negative values of inductance and capacitance may be produced. Associated with a "T" network of inductors, such negative components can be produced using a transformer. Both procedures guarantee that the load can be made from ideal but nevertheless real, positive valued, components.

The Cauer and Darlington methods of load synthesis have been included in the load synthesis program DLREAL. Initially Cauer Synthesis is used to produce the load. If the synthesis procedure terminates prematurely with a frequency dependent remainder, synthesis is continued using LDE based Darlington synthesis.

In order to ensure that errors, due to finite precision, do not produce erroneous and over complicated networks. Estimates of the absolute polynomial error are made after each polynomial operation, based on the preceding error values.
The value of each component is associated with a label to fix its position within the resulting circuit.

Given any PR rational function, DLREAL will produce a reactive network realization of it, terminated in a resistor.

In order to test each phase, of the three phases of equalization, a number of examples were used.

The rational function curve fitting phase was tested using the frequency response measurements of either the impedance or admittance of 3 different sonar transducers. The curve fit errors were small and convergence quite rapid. Small deviations, due to experimental error in the measurement data, were ignored by the curve fitting process; which yielded a realistic curve fit in all cases. Further verification of the solution quality was given by synthesising each load using DLREAL. The circuits produced were similar to those given by the manufacturer.

To verify that DEQDIO would generate equalizers suitable to equalize a variety of load types, the examples considered by Youla [2] in his paper were equalized using DEQDIO. In all cases the results given by the computer matched those given by Youla without significant difference. The gain values back impedance coefficient values and, if present, all pass function values matched Youla's values to more than 4 decimal places.

A Bessel equalizer was also produced for the commercial sonar transducer load for which a rational function was found in section 7.1. The computer generated solution produced reasonable coefficient values for gain, back impedance coefficients and the coefficients of all pass function.
Using the load realization program DRREAL, a reactive network terminated in a resistor was produced for the equalizer back-impedance and the load. The load impedance, and equalized load voltage transfer function from the source to the load resistor were simulated using LINSIM. The load response matched the measurement data very closely, at many points it was not possible to measure the difference.

The equalized transfer from source, to the real part of the load, was flat to within ±2dB over the frequency range of interest, and the phase was almost linear, as expected, deviating by less than 5% from linearity. From these tests, it is clear that DLREAL is capable of synthesising a reactive load, terminated in a resistor, to match the response of the input rational function. Moreover equalization of the impedance function obtained from impedance measurement of the commercial transducer has been achieved, verifying the applicability of the three phases of equalization.

In many cases the network produced by Darlington Synthesis may be more complicated than is desired. However once the component values have been obtained, well known canonical transformations of the circuit configuration may lead to networks more suitable for construction.

In order to highlight any failings of the three phases of equalization, more examples must be considered. This is one area for further future research.

A number of the areas discussed in this dissertation have indicated areas in which future research could be conducted. In a number of cases, some thought has already been given to the direction of such research. Future proposals for research were discussed in chapter 9. New methods of
detection based on correlation or phase locked techniques are discussed. Such techniques are intended to reduce the need for lasers of great frequency stability. By tracking the fluctuations in frequency of the transmit laser, or indeed correlating the fluctuations out of the data, cheaper transmission systems may be produced. The laser complexity would be required only at the optical receiver.

Using polarization preserving fibres and polarizers at the fibre input and output points, a reduced carrier may be transmitted with the signal. The reduced carrier may be used to force the output of a tunable laser in an optical PLL to track fluctuations in frequency of the transmit laser due to frequency instability. The tunable laser can then be used as a local oscillator to coherently demodulate the incoming signal without the presence of laser frequency instability noise.

During the development of LDE equalization, load characterization and realization, methods of designing broadband amplifiers were found. Although it is clear that further research is required, it seems possible to include all frequency dependence of the amplifier forward transfer with the characterization of open circuit, or loaded input and output impedances. The amplifier design problem then becomes a load equalization problem which is readily solved using the 3 phases of equalization.

Unfortunately, it is evident that, when using some active devices and recasting the amplifier design problem as a load equalization problem, the load impedances will be non-PR.

In order to apply Youla-LDE equalization to non-PR loads, the equalization technique must be modified. In section 9.5, some preliminary
modifications and problems were discussed. With some further work it may be possible to produce a modified Youla-LDE technique to equalize any load.

It is evident from the discussion of the 9 preceding chapters that many problems must still be overcome before coherent optical detectors become a commercial reality.

Considerably more research into the design of optical and electronic design, is needed before many of the problems associated with coherent optical detection can be overcome.

This dissertation has investigated a number of problems in design and construction of coherent optical detectors. It has also discussed a number of possible solutions to these problems.

Based upon this work and that of others, the problems of coherent detection of optical signals will be overcome in the future. As a result coherent optical detection systems, and the gains they offer will eventually become a commercial reality.
Appendix I  Maximum Modulation Frequency of Amplitude Modulated Light

Appendix II  Minimum Modulation Frequency

Appendix III  USB Aperture Bandwidth Limit

Appendix IV  LSB Aperture Bandwidth Limit

Appendix V  Instruction Manual
              Acoustic Optic Modulator Driver

Appendix VI  Extruded Aluminium Heatsink

Appendix VII  Motorola MRD 500 MRD 510 PIN Silicon Photo Diode

Appendix VIII  Motorola MRF 901 NPN Silicon High-Frequency Transistor

Appendix IX  Reduction of Parallel Two Port Network Scattering Parameters

Appendix X  Reduction of Simplified Parallel Two Port Network Scattering Parameters

Appendix XI  Motorola MWA 210/MWA 220/MWA 230 Wideband Hybrid Amplifiers

Appendix XII  NEC PRA/PRB Reed Relays

Appendix XIII  Switched Capacitor Integrator Connections (Oscillator and Timing)

Appendix XIV  Optoelectronics CD 10/CD 20 Components Photodetectors
APPENDIX I

MAXIMUM MODULATION FREQUENCY OF AMPLITUDE MODULATED LIGHT
Appendix I

Fig I.1 Maximum Modulation Frequency of Amplitude Modulated Light

\[ U = \text{USB} \]

\[ L = \text{LSB} \]

\[ l = \frac{l}{d} \]

\[ w = \frac{w}{D} \]

\[ w = \frac{\tan(\theta_U - \theta_B) + \tan(\theta_B - \theta_L)}{1} \]
\[
\frac{w}{1} = \frac{\tan(\theta_U) - \tan(\theta_B) + \tan(\theta_B) - \tan(\theta_L)}{1 + \tan(\theta_U)\tan(\theta_B)} + \frac{\tan(\theta_B) - \tan(\theta_L)}{1 + \tan(\theta_L)\tan(\theta_B)}
\]

\[
\frac{w}{1} = \frac{(\tan(\theta_U) - \tan(\theta_B))(1 + \tan(\theta_L)\tan(\theta_B))}{1 + \tan(\theta_U)\tan(\theta_B)} + \frac{(\tan(\theta_B) - \tan(\theta_L))(1 + \tan(\theta_U)\tan(\theta_B))}{1 + \tan(\theta_U)\tan(\theta_B)}
\]

\[
\frac{w}{1} = \frac{((\tan(\theta_U) - \tan(\theta_L))(1 + \tan^2(\theta_B)))}{1 + \tan(\theta_U)\tan(\theta_B) + \tan(\theta_L)\tan(\theta_B) + \tan(\theta_U)\tan(\theta_L)\tan^2(\theta_B)}
\]

\[
w + wtan(\theta_B)(\tan(\theta_U) + \tan(\theta_L)) + wtan^2(\theta_B)\tan(\theta_U)\tan(\theta_L) = 1(\tan(\theta_B) + 1)(\tan(\theta_U) - \tan(\theta_L))
\]

\[
w + wtan(\theta_B) + \frac{2C}{C_s\omega_L\cos(\theta_B)} + \frac{2\tan(\theta_B)}{nC_s\omega_L\cos(\theta_B)} + \frac{wtan^2(\theta_B) C(\omega_c - \omega_m) + tan(\theta_B)}{nC_s\omega_L\cos(\theta_B)} + \frac{C(\omega_c + \omega_m) + tan(\theta_B)}{nC_s\omega_L\cos(\theta_B)} = 1(\tan^2(\theta_B) + 1) \frac{2C_m}{nC_s\omega_L\cos(\theta_B)}
\]

\[
w + \frac{2C\omega_c tan(\theta_B)}{nC_s\omega_L\cos(\theta_B)} + 2wtan^2(\theta_B) + \frac{C^2w(\omega_c^2 - \omega_m^2)tan^2(\theta_B)}{n^2C_s^2\omega_L^2\cos^2(\theta_B)}
\]

\[
+ \frac{2C\omega_c tan^3(\theta_B)}{nC_s\omega_L\cos(\theta_B)} + wtan^4(\theta_B) = \frac{2C_m(\tan^2(\theta_B) + 1)}{nC_s^2\omega_L\cos(\theta_B)}
\]
\[ \begin{align*} 
wn^2 v_s^2 \omega_L^2 \cos^2(\theta_B) + 2 C w v_s \omega_C \omega_L \sin(\theta_B) + 2 wn^2 v_s^2 \omega_L^2 \sin^2(\theta_B) \\
+ C^2 w \omega_c^2 \tan^2(\theta_B) - C^2 \omega_m^2 \tan^2(\theta_B) + 2 C w n v_s \omega_C \omega_L \tan^2(\theta_B) \sin(\theta_B) \\
+ wn^2 v_s^2 \omega_L^2 \tan^2(\theta_B) \sin^2(\theta_B) = 2 C ln v_s \omega_m \omega_L \omega_c (\tan^2(\theta_B) + 1) \cos(\theta_B) \\
C^2 w \tan^2(\theta_B) \omega_m^2 + [2 C ln v_s \omega_L (\tan^2(\theta_B) + 1) \cos(\theta_B)] \omega_m \\
- wn^2 v_s^2 \omega_L^2 \cos^2(\theta_B) - 2 C w n v_s \omega_C \omega_L \sin(\theta_B) - 2 wn^2 v_s^2 \omega_L^2 \sin^2(\theta_B) \\
- C^2 \omega_c^2 \tan^2(\theta_B) - 2 C w n v_s \omega_C \omega_L \tan^2(\theta_B) \sin(\theta_B) \\
- wn^2 v_s^2 \omega_L^2 \tan^2(\theta_B) \sin^2(\theta_B) \\
(C^2 w \tan^2(\theta_B)) \omega_m^2 + [2 C ln v_s \omega_L (\tan^2(\theta_B)) \cos(\theta_B)] \omega_m \\
- (1 + \tan^2(\theta_B)) wn^2 v_s^2 \omega_L^2 \sin^2(\theta_B) - (1 + \tan^2(\theta_B)) wn v_s \omega_C \omega_L \sin(\theta_B) \\
- wn^2 v_s^2 \omega_L^2 - C^2 \omega_c^2 \tan^2(\theta_B) = 0 
\end{align*} \]
APPENDIX II

MINIMUM MODULATION FREQUENCY
Using simple trigonometry,

\[ \gamma = \beta + \frac{\varphi}{2} \]

\[ \alpha = \beta - \frac{\varphi}{2} \]

Using the sine rule for triangles,

\[ W_1 = \frac{l \sin(\gamma/2)}{\sin(\gamma)} \]

\[ = \frac{l \sin(\gamma/2)}{\sin(\beta - \varphi/2)} \]
Similarly,

\[ W_2 = \frac{\sin(\psi/2)}{\sin(\alpha)} = \frac{\sin(\psi/2)}{\sin(\beta - \psi/2)} \]

Using the cosine rule, \( l' \) can be found in terms of \( W_1, W_2, \alpha'' \) and \( \gamma' \),

\[ l' = \frac{\sin^2(\gamma')(W_1 + W_2)^2 + W_1^2}{\sin^2(\gamma' + \alpha'')} - \frac{2\sin(\gamma')(W_1 + W_2)\cos(\alpha'')}{\sin(\gamma' - \alpha'')} \]

The unknowns of this equation are,

\[ x = \frac{\sin(\gamma')}{\sin(\gamma' - \alpha'')} \]

Solving the quadratic equation for \( x \),

\[ \frac{\sin(\gamma')}{\sin(\gamma' - \alpha'')} = \frac{(W_1 + W_2)\cos(\alpha'')}{(W_1 + W_2)^2 + W_1^2} + \frac{1}{(W_1 + W_2)^2 + W_1^2} \]

Let

\[ A = \frac{(W_1 + W_2)\cos(\alpha'') + (W_1 + W_2)^2(\cos(\alpha'') + 1') + W_1^2}{(W_1 + W_2)^2 + W_1^2} \]
In which case,

\[ A = \frac{\sin(Y')}{\sin(Y' + \alpha''')} \]

By re-arranging,

\[ \sin(Y') = A\sin(Y' - \alpha''') \]

Solving for \( Y' \),

\[ \sin(Y') = A\sin(Y')\cos(\alpha''') - A\cos(Y')\sin(\alpha''') \]

and so,

\[ (A\cos(\alpha''') - 1)\sin(Y') - A\cos(Y') = 0 \]

Which may be written as,

\[ r\cos(Y' + \xi) = r\cos(\xi)\cos(Y') - r\sin(\xi)\sin(Y') = 0 \]

where,

\[ r = A^2 + A^2\cos(\alpha''') - 2A\cos(\alpha''') + 1 \]

and,

\[ \xi = \cos^{-1}\left[ \frac{A}{A^2 + A^2\cos(\alpha''') + 2A\cos(\alpha''') + 1} \right] \]

The general solution for \( Y' \) is,

\[ Y' = \frac{\pi}{2} + 2N\pi - \xi \]

where \( N \) is an arbitrary integer.

But \( Y' \) is fixed by the photon-phonon interactions of the AO cell. thus,

\[ Y' = \tan^{-1}\left[ \frac{c(\omega_c + \omega_m)}{nV_s\omega_1\cos(\beta - \Psi/2)} + \tan(\beta - \Psi/2) \right] \]
Which may be rearranged to give \( \omega_m \)

\[
\omega = \frac{\pi}{2} - 2N\pi - \cos^{-1} \left[ \frac{A}{A^2 + A^2 \cos (a''') + 2A\cos(a''') + 1} \right] \\
- \tan(\theta_B - \Psi/2) \left[ \frac{n V_s \omega c}{\cos(\theta_B - \Psi/2)} - \omega_C \right]
\]

And so \( \omega \) is given in terms of the fundamental quantities of AO interaction.
APPENDIX III

USB APERTURE BANDWIDTH LIMIT
Using the relationship between incident and deflected angles derived in appendix II,

\[
\theta_c = \tan^{-1} \left[ \frac{\omega_{sc}c}{\omega_1 V_s \cos(\theta)} + \tan(\theta) \right]
\]

\[
\theta_{USB} = \tan^{-1} \left[ \frac{(\omega_{sc} + \omega_{sm})c}{\omega_1 V_s \cos(\theta)} + \tan(\theta) \right]
\]

The tangent of the difference between the USB angle of deflection and the angle of carrier deflection is given by,

\[
\tan(\theta_{USB} - \theta_c) = \frac{\tan(\theta_{USB}) - \tan(\theta_c)}{1 + \tan(\theta_{USB})\tan(\theta_c)}
\]

Using simple trigonometry,

\[
w = \frac{\tan(\theta_{USB} - \theta_c)}{1}
\]
Substituting for $\theta_c$ and $\theta_{usb}$, and applying the tangent identity,

\[
w = \frac{\omega_{sm} \cos(\theta) \text{cn} \omega_1 V \text{cs}(\theta)}{1 - n^2 \omega_1^2 \omega_2 - \left(\omega_{sc} + \omega_{sm}\right) \omega_2 \text{cs}^2 + \left(2 \omega_{sc} + \omega_{sm}\right) \text{cn} \omega_1 V \text{cs}(\theta) - \omega_{sm} \text{ci} \omega_1 V \text{ci}(\theta)}
\]

Which may be solved for $\omega_{sm}$,

\[
\omega_{sm} = \frac{\omega_{sc} \omega_1^2 \omega_2^2 + \omega_2^2 \text{cs}^2 + 2 \omega_{sc} \text{cn} \omega_1 V \text{cs}(\theta)}{1 \text{cn} \omega_1 V \text{cs}(\theta) - \omega_{sc} \text{cs}^2 - \omega_{sm} \text{ci} \omega_1 V \text{cs}(\theta)}
\]
APPENDIX IV

LSB APERTURE BANDWIDTH LIMIT
Appendix IV

Fig IV.1 LSB Aperture Bandwidth Limit

Using the relationship between incident and deflected angles derived in appendix II,

\[ \theta_c = \tan^{-1} \left[ \frac{\omega_{sc}}{\omega_1 V_s \cos(\theta)} + \tan(\theta) \right] \]

\[ \theta_{usb} = \tan^{-1} \left[ \frac{(\omega_{sc} - \omega_{sm})c}{\omega_1 V_s \cos(\theta)} + \tan(\theta) \right] \]

The tangent of the difference between the LSB angle of deflection and the angle of carrier deflection is given by,

\[ \tan(\theta_{lsb} - \theta_c) = \frac{\tan(\theta_{lsb}) - \tan(\theta_c)}{1 + \tan(\theta_{lsb})\tan(\theta_c)} \]

Using simple trigonometry,

\[ w = \frac{\tan(\theta_{lsb} - \theta_c)}{l} \]
Substituting for $\theta_c$ and $\theta_{lsb}$, and applying the tangent identity, we have:

$$w = \frac{\omega_{sm} cn_{\omega_1 V} \cos(\theta)}{1 - n^2 \omega_1^2 V_s^2 + (\omega_{sc} - \omega_{sm}) \omega_{sc}^2 c_2^2 + (2\omega_{sc} - \omega_{sm}) cn_{\omega_1 V} \sin(\theta)}$$

which may be solved for $\omega_{sm'}$:

$$\omega_{sm} = \frac{wn^2 \omega_1 \omega_s^2 + w_{sc} \omega_{sc}^2 + 2\omega_{sc} cn_{\omega_1 V} \sin(\theta)}{1cn_{\omega_1 V} \cos(\theta) + w_{sc} \omega_{sc}^2 + wc_{\omega_1 V} \sin(\theta)}$$
APPENDIX V

INSTRUCTION MANUAL
ACOUSTO - OPTIC MODULATOR DRIVER
INSTRUCTION MANUAL
ACOUSTO-OPTIC MODULATOR DRIVER
SERIES 230
(Analog Modulation)

PRECAUTIONS

NEVER OPERATE THE DRIVER WITHOUT PROPER COOLING. THE MOUNTING FACE TEMPERATURE MUST NOT EXCEED 70°C

NEVER OPERATE THE DRIVER INTO AN OPEN CIRCUITED OR SHORT CIRCUITED LOAD

THE VIDEO INPUT LEVEL MUST NOT EXCEED 4V PEAK TO PEAK (±2V WITH RESPECT TO GROUND).
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General</td>
</tr>
<tr>
<td>2</td>
<td>Analog Modulation</td>
</tr>
<tr>
<td>3</td>
<td>Installation and Adjustment</td>
</tr>
<tr>
<td>4</td>
<td>Operation</td>
</tr>
</tbody>
</table>

# List of Illustrations

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Installation, Series 230 Driver</td>
</tr>
<tr>
<td>2</td>
<td>Block Diagram</td>
</tr>
<tr>
<td>3</td>
<td>Typical Waveforms</td>
</tr>
<tr>
<td>4</td>
<td>Intensity vs. RF Envelope Voltage</td>
</tr>
<tr>
<td></td>
<td>Transfer Function</td>
</tr>
</tbody>
</table>
Figure 1 - Installation, Series 230 Driver

Thermal Compound

Video In

RF Out

+28VDC

Ground

Snap-In Plugs

Power Adjust

Bias Adjust

Mounting Flange

Heat Sink

TEMPERATURE MUST NOT EXCEED 70°C
1. General

The Series 230 Analog Driver is an RF power source specifically designed to operate with Isomet acousto-optic modulators such as the 1201, 1205, and 1206. The driver accepts an analog modulating signal at baseband video frequency and provides a double-sideband amplitude-modulated RF output to the acousto-optic modulator at a particular center frequency. Six different models of the Series 230 Drivers are available; these differ as to center frequency and maximum power output as shown in the following table:

### 1 Watt RF Output

<table>
<thead>
<tr>
<th>Model</th>
<th>Center Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>231-1</td>
<td>40 MHz</td>
</tr>
<tr>
<td>232-1</td>
<td>80 MHz</td>
</tr>
<tr>
<td>233-1</td>
<td>110 MHz</td>
</tr>
</tbody>
</table>

### 1.6 Watt RF Output

<table>
<thead>
<tr>
<th>Model</th>
<th>Center Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>231-2</td>
<td>40 MHz</td>
</tr>
<tr>
<td>232-2</td>
<td>80 MHz</td>
</tr>
<tr>
<td>233-2</td>
<td>110 MHz</td>
</tr>
</tbody>
</table>

Figure 2 is a block diagram of the Series 230 Driver. Center frequency of the driver is determined by the free-running quartz-crystal oscillator at 40, 80, or 110 MHz according to the model
Figure 2 - Block Diagram
used. This frequency is accurate to within ±0.005% and its stability is better than ±0.003%; the oscillator is not temperature stabilized.

A transistor r-f preamplifier isolates the crystal controlled oscillator from the load. It includes a gain control for adjusting the maximum r-f power output of the 230 Series Driver.

A high-frequency, diode ring modulator is used to double-sideband amplitude-modulate the r-f carrier with baseband video. The carrier input to the ring modulator is obtained from the preamp. The baseband video input, from an external 50Ω drive source, is the modulating input to the ring modulator. For the purpose of setting the average output power of the ring modulator, a biasing voltage is added to the video input. Normally the BIAS is adjusted for 50% deflection efficiency in the acousto-optic modulator with no video input. With this setting, a video input swing of 1 volt peak to peak (±0.5V with respect to ground) will result in 100% depth of modulation. The video input level must not exceed 4 volts peak to peak (±2V with respect to ground).

The amplitude-modulated rf from the ring modulator is amplified to the specified power level in the power amplifier stage. This amplifier is designed to operate at full rated power into a 50Ω load with 100% duty cycle.

Figure 3 illustrates the principal waveforms of the Series 230 Driver. The functions of the POWER ADJUST and BIAS controls are noted.

Conduction cooling of the driver from the mounting face to a heat sink or forced-air convection cooling is mandatory. The mounting face temperature must not exceed 70°C. Serious damage
Figure 3 - Typical Waveforms
to the amplifier may result if the temperature exceeds 70°C. Serious damage to the amplifier may also result if the RF Output connector is operated open-circuited or short-circuited.

A low impedance source of d-c power is required for operation of the 230 Series Driver. The required voltage is +28VDC at a current drain of 360 mA for 1 watt drivers and 500 mA for 1.6 watt drivers. The external power source should be regulated to ±2% and the power supply ripple voltage should be less than 25 mV for best results.

2. Analog Modulation

To intensity modulate a laser beam in an acousto-optic modulator requires that the input RF carrier voltage (power) be varied according to the video or baseband information. From the viewpoint of intensity modulation, the deflection efficiency equation is normalized as:

\[ i_1 = a \sin^2 (kE_{RF}) \]

Where \( i_1 \) is the instantaneous intensity in the first order diffracted beam and \( E_{RF} \) is the instantaneous RF envelope voltage across the matched transducer.

Figure 4 shows the intensity vs. RF envelope voltage transfer function of the acousto-optic modulator in normalized units with the typical waveforms superimposed. It will be noted that the driving RF waveform is a double-sideband amplitude-modulated carrier. If effect, the acousto-optic interaction demodulates the RF carrier, transforming the modulation envelope (baseband signal) into intensity variation of the first order diffracted laser beam.
Figure 4 - Intensity vs. RF Envelope Voltage Transfer Function
From the transfer function of figure 4, it can be seen that best linearity is obtained when the operating point of the modulator is set near the center of the $i$ vs. $E_{RF}$ curve, that is $i_{avg} = I_{SAT}/2$ and $e_{avg} = E_{SAT}/2$. With this setting and with the depth of modulation limited to 80%, the total distortion in the intensity modulated beam will be less than 5%. Depth of modulation is defined as:

$$\frac{E_{max} - E_{min}}{E_{max} + E_{min}} = \frac{i_p - i_v}{i_p + i_v}$$

In the 230 Series Modulator Driver, two controls are included RF POWER ADJ and BIAS ADJ. The RF POWER ADJ control sets the peak Driver output at saturated power. The BIAS control sets the average drive power at $E_{avg}$. Depth of modulation is controlled by the amplitude of the video signal from the external video source.

3. Installation and Adjustment

A. Install the Series 230 Driver on a heat sink as shown in figure 1. Use heat conducting compound between the Driver mounting face and the heat sink.

B. With no d-c power applied, connect the +28VDC line to the center terminal of the feed-thru terminal as shown in figure 1. DO NOT APPLY POWER.

C. Connect the RF output BNC jack to an acousto-optic modulator (or a 50Ω RF load, if it is desired to measure the modulator RF output power).

D. Adjustment of the RF output power is best done with the Series 230 Driver connected to the acousto-optic modulator. The Driver maximum output power is factory preset to the specified level (1 watt or 1.6 watts).
The optimum RF power level required for the modulator to produce maximum first order intensity will be different at various laser wavelengths. Applying RF Power in excess of this optimum level will cause a decrease in first order intensity (a false indication of insufficient RF power) and make accurate Bragg alignment difficult. It is therefore recommended that initial alignment be performed at a low RF power level.

1. Remove the BIAS ADJ and PWR ADJ snap-in plugs from the driver case (see Fig. 1).
2. With an insulated alignment tool or screwdriver:
   a. Rotate the recessed bias adjust potentiometer fully ccw.
   b. Rotate the Power ADJ potentiometer fully ccw, than cw approximately ¼ turn.
   c. Apply +28VDC to the driver.
   d. Observe the diffracted first-order output from the acousto-optic modulator and the undeflected zeroth order beam. Adjust the Bragg angle (rotate the modulator) to maximize first order beam intensity.
   e. After Bragg angle has been optimized, slowly increase the RF power (rotate PWR ADJ. cw) until maximum first order intensity is obtained. Record this intensity value \( I_{\text{SAT}} \).
   f. Rotate the BIAS ADJ control clockwise to reduce the first order beam intensity to \( I_{\text{SAT}}/2 \).
   g. Replace the snap-in plugs.

E. The driver is now ready for use as an analog modulator. Connect the external 50Ω drive source to the VIDEO INPUT jacks. Adjust the video level for minimum dis-
tortion of the intensity modulated signal. A video input level of 1V PP (±0.5V with respect to ground) will drive the modulator into saturation. The video input must not exceed 4V PP (±2V with respect to ground).

4. Operation

The Series 230 Modulator Driver is operated by applying +28VDC to the feedthrough terminal. There are no operating controls or adjustments.
The Series 230 Driver is designed to operate with Isomet acousto-optic modulators for the proportional (analog) control of laser-beam intensity. For this mode of operation, a double sideband amplitude modulated source of RF energy is required at a selected frequency and power level.

Contained in the 230 Driver are a free-running quartz-crystal carrier-frequency oscillator, a wideband balanced diode ring modulator, a bias level adjustment and a broadband Class-A amplifier. Under control of the low-level video (analog) input signal, the balanced ring modulator impresses the modulating video on the carrier. The resulting double-sideband AM signal is amplified in the broadband Class A amplifier. An externally accessible bias control permits adjustment of average output power from the driver. Depth of modulation is determined by the amplitude of the modulating input signal.

Efficient heat transfer from the driver requires that the mounting base be attached to an external heat sink not exceeding 70°C in temperature. Regulated DC power must be supplied by the user.

**SPECIFICATIONS**

<table>
<thead>
<tr>
<th>Specification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output Impedance</td>
<td>50Ω nominal</td>
</tr>
<tr>
<td>Load Mismatch VSWR</td>
<td>2:1 Max</td>
</tr>
<tr>
<td>Video Input Voltage</td>
<td>1V peak to peak for 100% depth of modulation, DC coupled</td>
</tr>
<tr>
<td>Input Impedance</td>
<td>50Ω source</td>
</tr>
<tr>
<td>Frequency Accuracy</td>
<td>±0.005%</td>
</tr>
<tr>
<td>Frequency Stability</td>
<td>±0.003%</td>
</tr>
<tr>
<td>DC Power Input</td>
<td>+28VDC regulated to ±0.25%</td>
</tr>
<tr>
<td>Temperature Range</td>
<td>0°C to 60°C ambient, temperature at mounting face must not exceed 70°C</td>
</tr>
<tr>
<td>Mounting Orientation:</td>
<td>Any</td>
</tr>
<tr>
<td>Dimensions</td>
<td>See Outline, Reverse side</td>
</tr>
</tbody>
</table>

**PERFORMANCE**

<table>
<thead>
<tr>
<th>Model</th>
<th>Center Frequency</th>
<th>Modulation BW</th>
<th>RF Drive Power</th>
<th>Use With Modulator</th>
</tr>
</thead>
<tbody>
<tr>
<td>231A-2</td>
<td>40MHz</td>
<td>20MHz</td>
<td>&gt;1.6W</td>
<td>1201E-1</td>
</tr>
<tr>
<td>232A-1</td>
<td>80MHz</td>
<td>35MHz</td>
<td>1W</td>
<td>1205C</td>
</tr>
<tr>
<td>233A-1</td>
<td>110MHz</td>
<td>50MHz</td>
<td>&gt;1W</td>
<td>1206C</td>
</tr>
<tr>
<td>RFA-108*</td>
<td>-</td>
<td>-</td>
<td>&gt;6W</td>
<td>1201E-2, 1207B</td>
</tr>
</tbody>
</table>

* Model RFA-108 RF amplifier may be used with 231A-2 to produce 8 watts of drive power for the 1201E-2 at 1.06μm or the 1207B at 10.6μm.

All specifications subject to change without notice.

5263 PORT ROYAL ROAD, P.O. BOX 1634, SPRINGFIELD, VA. 22151 (703) 321-8301 TELEX 899434
DIMENSIONS IN MM UNLESS OTHERWISE SPECIFIED. (ENGLISH EQUIVALENTS IN PARENTHESIS)
APPENDIX VI

EXTRUDED ALUMINIUM HEATSINK
EXTRUDED ALUMINIUM HEATSINK

Extruded heatsink of aluminium alloy.
The extrusion is supplied unpainted, in lengths of 1.5 m.
Weight: 4 kg per 1.5 m.

Dimensions in mm

Stud: 10-32UNF
Mounting base across the flats: 11.0 mm

Stud: ¼" x 28UNF
Mounting base across the flats: 14.0 mm

Stud: M8 x 125
Mounting base across the flats: 17.0 mm

Stud: ⅜" x 20UNF
Mounting base across the flats: 27.0 mm
APPENDIX VII

MOTOROLA MRD 500/MRD 510 PIN SILICON PHOTO DIODE
PIN SILICON PHOTO DIODE

... designed for application in laser detection, light demodulation, detection of visible and near infrared light-emitting diodes, shaft or position encoders, switching and logic circuits, or any design requiring, radiation sensitivity, ultra high-speed, and stable characteristics.

- Ultra Fast Response — (<1.0 ns Typ)
- High Sensitivity  
  MRD500 (1.2 μA/mW/cm² Min)  
  MRD510 (0.3 μA/mW/cm² Min)
- Available With Convex Lens (MRD500) or Flat Glass (MRD510) for Design Flexibility
- Popular TO-18 Type Package for Easy Handling and Mounting
- Sensitive Throughout Visible and Near Infrared Spectral Range for Wide Application
- Annular Passivated Structure for Stability and Reliability

MAXIMUM RATINGS (TA = 25°C unless otherwise noted)

<table>
<thead>
<tr>
<th>Rating</th>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse Voltage</td>
<td>VR</td>
<td>100</td>
<td>V</td>
</tr>
<tr>
<td>Total Device Dissipation @ TA</td>
<td>PD</td>
<td>100</td>
<td>mW/°C</td>
</tr>
<tr>
<td>Operating and Storage Junction</td>
<td>TJ_Tot</td>
<td>-65 to +200</td>
<td>°C</td>
</tr>
</tbody>
</table>

FIGURE 1 — TYPICAL OPERATING CIRCUIT
### STATIC ELECTRICAL CHARACTERISTICS (T_A = 25°C unless otherwise noted)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Fig. No.</th>
<th>Symbol</th>
<th>Min</th>
<th>Typ</th>
<th>Max</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark Current</td>
<td>4 and 5</td>
<td>I_D</td>
<td>-</td>
<td>-</td>
<td>2.0</td>
<td>nA</td>
</tr>
<tr>
<td>(V_R = 20 V, R_L = 1.0 megohm; Note 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_A = 25°C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_A = 100°C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Breakdown Voltage</td>
<td>-</td>
<td>BVR</td>
<td>100</td>
<td>200</td>
<td>-</td>
<td>Volts</td>
</tr>
<tr>
<td>(I_P = 10 μA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Voltage</td>
<td>-</td>
<td>V_F</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>Volts</td>
</tr>
<tr>
<td>(I_P = 50 mA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Series Resistance</td>
<td>-</td>
<td>R_s</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>ohms</td>
</tr>
<tr>
<td>(I_P = 50 mA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Capacitance</td>
<td>6</td>
<td>C_T</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>pF</td>
</tr>
<tr>
<td>(V_R = 20 V; f = 1.0 MHz)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### OPTICAL CHARACTERISTICS (T_A = 25°C)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Fig. No.</th>
<th>Symbol</th>
<th>Min</th>
<th>Typ</th>
<th>Max</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation Sensitivity</td>
<td>2 and 3</td>
<td>S_R</td>
<td>1.2</td>
<td>1.8</td>
<td>-</td>
<td>μA/mW/cm²</td>
</tr>
<tr>
<td>(V_R = 20 V, Note 1)</td>
<td>MRD500</td>
<td></td>
<td>0.3</td>
<td>0.42</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>MRD510</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity at 0.8 μm</td>
<td>-</td>
<td>S_L(0.8 μm)</td>
<td>6.6</td>
<td>-</td>
<td>-</td>
<td>μA/mW/cm²</td>
</tr>
<tr>
<td>(V_R = 20 V, Note 2)</td>
<td>MRD500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRD510</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Time</td>
<td>-</td>
<td>I(resp)</td>
<td>1.0</td>
<td>-</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>(V_R = 20 V, R_L = 50 ohms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength of Peak Spectral Response</td>
<td>7</td>
<td>λ_P</td>
<td>0.8</td>
<td>-</td>
<td>-</td>
<td>μm</td>
</tr>
</tbody>
</table>

**NOTES:**

1. Radiation Flux Density (I) equal to 5.0 mW/cm² emitted from a tungsten source at a color temperature of 2970 K.
2. Measured under dark conditions. (I = 0).
3. Radiation Flux Density (I) equal to 0.5 mW/cm² at 0.8 μm.
MRD500, MRD510 (continued)

**TYPICAL ELECTRICAL CHARACTERISTICS**

**FIGURE 2 - IRRADIATED VOLTAGE - CURRENT CHARACTERISTIC FOR MRD500**

- TUNGSTEN SOURCE TEMP = 2870 K
- I = 20 mW/cm²

**FIGURE 3 - IRRADIATED VOLTAGE - CURRENT CHARACTERISTIC FOR MRD510**

- TUNGSTEN SOURCE TEMP = 2870 K
- I = 20 mW/cm²

**FIGURE 4 - DARK CURRENT versus TEMPERATURE**

- V, = 20 V
- I = 0

**FIGURE 5 - DARK CURRENT versus REVERSE VOLTAGE**

- T = 25°C
- H = 0

**FIGURE 6 - CAPACITANCE versus VOLTAGE**

- f = 1 MHz

**FIGURE 7 - RELATIVE SPECTRAL RESPONSE**

- RELATIVE RESPONSE (U)
- λ, WAVELENGTH (µm)
OPTOELECTRONIC DEFINITIONS, CHARACTERISTICS, AND RATINGS

BV \text{R}  
Reverse Breakdown Voltage – The minimum dc reverse breakdown voltage at stated diode current and ambient temperature.

CT \text{F} 
Cathode Capacitance

H \text{r}  
Radiation Flux Density (Irradiance) \text{[mW/cm}^2\text{]} – The total incident radiation energy measured in power per unit area.

ID \text{r}  
Dark Current – The maximum reverse leakage current through the device measured under dark conditions, \text{(H=0)}, with a stated reverse voltage, load resistance, and ambient temperature.

PD \text{r}  
Power Dissipation

RS \text{s}  
Series Resistance – The maximum dynamic series resistance measured at stated forward current and ambient temperature.

SR  
Radiation Sensitivity \text{[\mu A/mW/cm}^2\text{]} – The ratio of photo-induced current to the incident radiant energy measured at the plane of the lens of the photo device under stated conditions of radiation flux density \text{(H)}, reverse voltage, load resistance, and ambient temperature.

TA \text{t}  
Ambient Temperature

TF \text{r}  
Junction Temperature

Tst \text{r}  
Storage Temperature

VF \text{r}  
Forward Voltage – The maximum forward voltage drop across the diode at stated diode current and ambient temperature.

VR \text{r}  
Reverse Voltage – The maximum allowable value of dc reverse voltage which can be applied to the device at the rated temperature.

\lambda_{\text{p}} \text{[um]}  
Wavelength of peak spectral response in micrometers.

OPTODEVICES

AN-460 – THEORY AND CHARACTERISTICS OF PHOTOTRANSISTORS

A brief history of the photoelectric effect is discussed, followed by a comprehensive analysis of the effect in bulk semiconductors, pn junctions and phototransistors. A model is presented for the phototransistor. Static and transient data for the MRD300 provide typical phototransistor characteristics. Appendices provide a discussion of the relationship of irradiation and illumination and define terms specifically related to phototransistors.

AN-508 APPLICATIONS OF PHOTOTRANSISTORS IN ELECTRO-OPTIC SYSTEMS

This note reviews phototransistor theory, characteristics and terminology, then discusses the design of electro-optic systems using device information and geometric considerations. It also includes several circuit designs that are suited to dc, low-frequency and high-frequency applications.
APPENDIX VIII

MOTOROLA MRF 901 NPN SILICON HIGH-FREQUENCY TRANSISTOR
The RF Line

NPN SILICON HIGH-FREQUENCY TRANSISTOR

... designed primarily for use in high-gain, low-noise small-signal amplifiers. Also usable in applications requiring fast switching times.

- High Current-Gain Bandwidth Product — $f_T = 4.5 \text{ GHz (Typ)}$ @ $I_C = 15 \text{ mA}$
- Low Noise Figure @ $f = 1.0 \text{ GHz}$: $NF = 2.0 \text{ dB (Typ)}$ and $2.5 \text{ dB (Max)}$
- High Power Gain — $G_{pe} = 10 \text{ dB (Min)}$ @ $f = 1.0 \text{ GHz}$
- Third Order Intercept = +23 dBm (Typ)

MAXIMUM RATINGS

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collector — Emitter Voltage</td>
<td>$V_{CEO}$</td>
<td>15</td>
<td>VDC</td>
</tr>
<tr>
<td>Collector — Base Voltage</td>
<td>$V_{CEO}$</td>
<td>25</td>
<td>VDC</td>
</tr>
<tr>
<td>Emitter — Base Voltage</td>
<td>$V_{CEO}$</td>
<td>3.0</td>
<td>VDC</td>
</tr>
<tr>
<td>Collector Current — Continuous</td>
<td>$I_C$</td>
<td>30</td>
<td>mA DC</td>
</tr>
<tr>
<td>Total Device Dissipation</td>
<td>$P_D$</td>
<td>0.375</td>
<td>Watt</td>
</tr>
<tr>
<td>$\Theta_JC + 25^\circ C$</td>
<td></td>
<td>3.3</td>
<td>mW/°C</td>
</tr>
<tr>
<td>Storage Temperature Range</td>
<td>$T_{Stg}$</td>
<td>150</td>
<td>°C</td>
</tr>
</tbody>
</table>

THERMAL CHARACTERISTICS

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Symbol</th>
<th>Max</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal Resistance, Junction to Ambient</td>
<td>$R_{JJA}$</td>
<td>300</td>
<td>°C/W</td>
</tr>
</tbody>
</table>

2.5 dB @ 1.0 GHz
HIGH FREQUENCY TRANSISTOR
NPN SILICON

MAXIMUM POWER DISSIPATION

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4.44</td>
</tr>
<tr>
<td>B</td>
<td>4.37</td>
</tr>
<tr>
<td>C</td>
<td>0.150</td>
</tr>
<tr>
<td>D</td>
<td>0.225</td>
</tr>
<tr>
<td>E</td>
<td>0.050</td>
</tr>
<tr>
<td>F</td>
<td>0.029</td>
</tr>
<tr>
<td>G</td>
<td>0.019</td>
</tr>
<tr>
<td>H</td>
<td>0.019</td>
</tr>
<tr>
<td>I</td>
<td>0.019</td>
</tr>
<tr>
<td>J</td>
<td>0.019</td>
</tr>
<tr>
<td>K</td>
<td>0.019</td>
</tr>
<tr>
<td>L</td>
<td>0.019</td>
</tr>
</tbody>
</table>

CASE 31F61

17-125
### ELECTRICAL CHARACTERISTICS (T_J = 25°C unless otherwise noted)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Symbol</th>
<th>Min</th>
<th>Typ</th>
<th>Max</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OFF CHARACTERISTICS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collector-Emitter Breakdown Voltage</td>
<td>BVCEO</td>
<td>15</td>
<td>-</td>
<td>-</td>
<td>Vds</td>
</tr>
<tr>
<td>Collector-Base Breakdown Voltage</td>
<td>BVCEO</td>
<td>25</td>
<td>-</td>
<td>-</td>
<td>Vds</td>
</tr>
<tr>
<td>Emitter-Base Breakdown Voltage</td>
<td>BVCEO</td>
<td>20</td>
<td>-</td>
<td>-</td>
<td>Vds</td>
</tr>
<tr>
<td>Collector Oufit Current</td>
<td>ICBO</td>
<td>-</td>
<td>-</td>
<td>50</td>
<td>mAdc</td>
</tr>
<tr>
<td><strong>ON CHARACTERISTICS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DC Current Gain</td>
<td>NFE</td>
<td>30</td>
<td>80</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td><strong>DYNAMIC CHARACTERISTICS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current-Gain-Bandwidth Product</td>
<td>IT</td>
<td>-</td>
<td>4.5</td>
<td>-</td>
<td>GHz</td>
</tr>
<tr>
<td>Collector-Base Capacitance</td>
<td>Ccb</td>
<td>-</td>
<td>0.4</td>
<td>1.0</td>
<td>pF</td>
</tr>
<tr>
<td>Noise Figure</td>
<td>NF</td>
<td>-</td>
<td>2.0</td>
<td>2.5</td>
<td>dB</td>
</tr>
<tr>
<td><strong>FUNCTIONAL TESTS</strong> (Figure 1)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common-Emitter Amplifier Power Gain</td>
<td>GmEE</td>
<td>10</td>
<td>12</td>
<td>-</td>
<td>dB</td>
</tr>
<tr>
<td>Third Order Intercept</td>
<td>-</td>
<td>-</td>
<td>+23</td>
<td>-</td>
<td>dBm</td>
</tr>
</tbody>
</table>

---

**FIGURE 1 - 1.0 GHz TEST CIRCUIT SCHEMATIC**

![Test Circuit Schematic](image)

**FIGURE 2 - MAXIMUM UNILATERAL GAIN versus FREQUENCY**

![Graph showing maximum unilateral gain versus frequency](image)
### Table I - S11

<table>
<thead>
<tr>
<th>Voltage (V)</th>
<th>200 MHz</th>
<th>500 MHz</th>
<th>1.0 GHz</th>
<th>2 GHz</th>
<th>3 GHz</th>
<th>5 GHz</th>
<th>10 GHz</th>
<th>20 GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table II - S21

<table>
<thead>
<tr>
<th>Voltage (V)</th>
<th>200 MHz</th>
<th>500 MHz</th>
<th>1.0 GHz</th>
<th>2 GHz</th>
<th>3 GHz</th>
<th>5 GHz</th>
<th>10 GHz</th>
<th>20 GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 Volt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 MHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

17-129
### TABLE III - S12

<table>
<thead>
<tr>
<th>FREQUENCY (MHz)</th>
<th>2.5G</th>
<th>5G</th>
<th>10G</th>
<th>20G</th>
<th>100G</th>
<th>150G</th>
<th>200G</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCC 1.0mA</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>2.5</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>5.0</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>10</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>15</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>20</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

### TABLE IV - S22

<table>
<thead>
<tr>
<th>FREQUENCY (MHz)</th>
<th>2.5G</th>
<th>5G</th>
<th>10G</th>
<th>20G</th>
<th>100G</th>
<th>150G</th>
<th>200G</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCC 1.0mA</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>2.5</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>5.0</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>10</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>15</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>20</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

17-129
APPENDIX IX

REDUCTION OF PARALLEL TWO PORT NETWORK SCATTERING PARAMETERS
CENTRE HAS 8 $\alpha \beta$ LOOPS + TWO $\alpha h$ LOOPS
\[ \frac{\alpha \beta}{1 - \alpha \beta} \]

\[ \frac{\alpha \beta}{1 - \alpha \beta} = \frac{\alpha \beta}{1 - \beta - \alpha \beta} = \frac{\alpha \beta}{1 - \beta - \alpha \beta} = \frac{\alpha \beta}{1 - 2\alpha \beta} \]

Remove 7 loops
\[
\frac{\alpha \beta}{1 - 7\alpha \beta} \quad \frac{A}{1 - 8\alpha \beta} \quad \text{Eliminating 8 loops}
\]
\[ S_{II}'(jw) = \frac{ah+\gamma \alpha + \delta \alpha}{1-8\alpha \beta - 2dh} \quad (b + \beta g + 2\beta gc) + c + \delta g + \delta gc \]

\[ S_{21}'(jw) = \frac{ah+\gamma \alpha + \delta \alpha}{1-8\alpha \beta - 2dh} \quad (d + dh + 2e\beta) + a + ah + \gamma e \]

\[ S_{12}'(jw) = \frac{1+c+\alpha \zeta + \epsilon \beta}{1-8\alpha \beta - 2dh} \quad (b + 2\beta g + 2\beta gc) + f + fc + \epsilon g + \epsilon gc \]

\[ S_{22}'(jw) = \frac{1+c+\alpha \zeta + \epsilon \beta}{1-8\alpha \beta - 2dh} \quad (d + dh + 2e\beta) + h + \delta e \]

\[ \gamma = \delta \]

\[ \epsilon = \xi \]

\[ \gamma = \frac{1+c}{1+2gc} \quad \epsilon = \frac{cf}{1+2gc} \quad \beta = \frac{b}{1+2gc} \quad \alpha = \epsilon \]

1. \[ 1-8\alpha \beta - 2dh = 1 - \frac{8eb}{1+2gc} - 2dh = \frac{1+2gc - 8eb - 2dh - 4cdgh}{(1+2gc)} \]

2. \[ \frac{(ah + 2\alpha \gamma)(1+2gc)(b+2/\beta g + 2\beta gc)}{1+2gc - 8eb - 2dh - 4cdgh} + c + \delta g + \delta gc \]

\[ a = S_{21}(jw) \quad b = S_{12}(jw) \quad c = S_{11}(jw) \quad d = S_{22}(jw) \]

\[ e = S_{21}'(jw) \quad f = S_{12}'(jw) \quad g = S_{11}'(jw) \quad h = S_{22}'(jw) \]
APPENDIX X

REDUCTION OF SIMPLIFIED PARALLEL TWO PORT NETWORK SCATTERING PARAMETERS
\[
\frac{fa}{(1-3ce)(1-3bd)} = \frac{fa}{(1-3ce)(1-3bd) - 4fa} = \frac{fa}{1-3ce - 3bd + 9bcde - 4fa}
\]

\[
\frac{fa}{(1-3ce)(1-3bd)} - \frac{a + ab - 2abd}{(1-3bd)(1-3ce)} \cdot \frac{1 + c}{1 - 4\alpha}
\]

\[
\frac{fa}{(1-3ce)(1-3bd)} + \frac{(a + ab - 2abd)}{(1-3bd)(1-3ce)} \cdot \frac{1 + c}{1 - 4\alpha}
\]

\[
\frac{f(1+b)}{1-3bd} + \frac{2f(1+b)}{1-3bd} \cdot \frac{\alpha}{1-4\alpha}
\]
\[
\begin{align*}
a & = S_{21}(j\omega) \\
b & = S_{11}(j\omega) \\
c & = S_{22}(j\omega) \\
d & = S_{11}'(j\omega) \\
e & = S_{22}'(j\omega) \\
f & = S_{12}'(j\omega)
\end{align*}
\]

\[
S_{11}''(j\omega) = A(j\omega) + \frac{S_{21}(j\omega)S_{11}(j\omega) - 2S_{21}(j\omega)S_{11}'(j\omega)}{2S_{12}'(j\omega)(1 + S_{11}(j\omega))(1 - 3S_{22}(j\omega)S_{22}'(j\omega))(1 - 4A(j\omega))}
\]

\[
S_{12}''(j\omega) = \frac{S_{12}'(j\omega)(1 + S_{11}(j\omega))}{1 - 3S_{11}(j\omega)S_{11}'(j\omega)} + \left(2S_{12}'(j\omega)(1 + S_{11}(j\omega))\right)\left(\frac{A(j\omega)}{1 - 4A(j\omega)}\right)
\]

\[
S_{21}''(j\omega) = \frac{(S_{21}(j\omega)S_{11}(j\omega) - 2S_{21}(j\omega)S_{11}'(j\omega))(1 + S_{22}(j\omega))}{(1 - 3S_{11}(j\omega)S_{11}'(j\omega))(1 - 3S_{22}(j\omega)S_{22}'(j\omega))(1 - 4A(j\omega))}
\]

\[
S_{22}''(j\omega) = S_{22}(j\omega) + \frac{A(j\omega)(1 + S_{22}(j\omega))}{1 - 4A(j\omega)}
\]
APPENDIX XI

MOTOROLA MWA 210/MWA 220/MWA 230 WIDEBAND HYBRID AMPLIFIERS
MWA210
MWA220
MWA230

WIDEBAND HYBRID AMPLIFIERS

- Single stage amplifiers designed for broadband linear applications up to 800 MHz.
- Low-Cost TO-39 Type Package
- Gain 10 dB Typ
- 50 Ω Input and Output Impedance
- Fully Cascadable for Any Gain
- Thin Film Construction
- Hermetic Package
- Guaranteed Performance from -25°C to +100°C

MAXIMUM RATINGS

<table>
<thead>
<tr>
<th>Rating</th>
<th>Symbol</th>
<th>MWA210</th>
<th>MWA220</th>
<th>MWA230</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF Input Power</td>
<td>P_m</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>mW</td>
</tr>
<tr>
<td>DC Supply Current</td>
<td>I_D</td>
<td>25</td>
<td>55</td>
<td>100</td>
<td>mA</td>
</tr>
<tr>
<td>Maximum Case Temperature</td>
<td>T_J</td>
<td>125</td>
<td>-5 to +200</td>
<td>°C</td>
<td></td>
</tr>
<tr>
<td>Storage Temperature Range</td>
<td>T_stag</td>
<td>-65 to +200</td>
<td>°C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OPERATING CONDITIONS

| Device Voltage         | V_D     | 1.75   | 3.2    | 4.4    | Vdc   |
| Device Current         | I_D     | 10     | 25     | 60     | mA    |
| Decoupling Impedance   | Z_D     | 1000   | 1000   | 330    | Ω     |

DC-600 MHz WIDEBAND GENERAL-PURPOSE HYBRID AMPLIFIERS

MOTOROLA RF DEVICE DATA

5-96
MWA210, MWA220, MWA230

FIGURE 10 - INPUT AND OUTPUT IMPEDANCE versus FREQUENCY MWA210

FIGURE 11 - INPUT AND OUTPUT IMPEDANCE versus FREQUENCY MWA220

FIGURE 12 - INPUT AND OUTPUT IMPEDANCE versus FREQUENCY MWA230

FIGURE 13 - 1.0 dB GAIN COMPRESSION versus FREQUENCY

MOTOROLA RF DEVICE DATA

5-99
FIGURE 14 - 1.0 dB GAIN COMPRESSION versus DEVICE CURRENT f = 600 MHz

FIGURE 15 - 1.0 dB GAIN COMPRESSION versus CASE TEMPERATURE f = 600 MHz

FIGURE 16 - NOISE FIGURE versus FREQUENCY

FIGURE 17 - REVERSE ISOLATION versus FREQUENCY

FIGURE 18 - SECOND HARMONIC OUTPUT versus FREQUENCY

FIGURE 19 - SECOND AND THIRD ORDER INTERCEPT MWA210

MOTOROLA RF DEVICE DATA
The MWA series hybrid amplifiers are designed for wideband general purpose applications in 50 Ω systems. Fully cascadable for any gain combination, operable at voltages as low as 3 Vdc, and external control of the low frequency corner make the MWA amplifiers extremely versatile gain blocks.

Basic Circuit Configuration

Figure 26 shows the basic internal circuit. It is important to note that the specified operating conditions of voltage, current, and external decoupling impedance must be applied to the units in order to achieve the published electrical characteristics.

Amplifier Applications

The circuit schematic for a simple amplifier design is shown in Figure 27. External to the MWA hybrid amplifier the only components required are:

- Decoupling elements – Bypass Capacitor
- Decoupling Impedance (resistor/inductor)
- DC Blocking Capacitors at the RF input and output.

External Decoupling Impedance

In all cases the external bias (decoupling elements) must present an impedance which is large compared to the 50 Ω load impedance to minimize RF gain reduction. The loss in gain due to the decoupling impedance is given by the equation:

\[
\text{Loss} = 20 \log \frac{Z_D}{Z_D + Z_0} \text{ dB}
\]

where \(Z_D\) = decoupling impedance in ohms. For example, if \(Z_D = 1\) kΩ, Loss = 0.214 dB.

Supply Voltage

The value of the external decoupling resistive impedance (RD) determines the supply voltage (+VCC) and is determined by the following equation:

\[
V_{CC} = R_D \times I_D + V_D
\]

where \(I_D\) and \(V_D\) are the device current and voltage stated in the data sheet. For example, for MWA110,

\[
I_D = 10 \text{ mA} \quad V_D = 2.9 \text{ V}
\]

and, if \(R_D = 330 \text{ Ω}\), then

\[
V_{CC} = 6.2 \text{ V}
\]

More commonly \(V_{CC}\) is predetermined and \(R_D\) may be calculated from:

\[
R_D = \frac{V_{CC} - V_D}{I_D}
\]

If an RF choke is used for decoupling, then the supply voltage (VCC) required is equal to the device voltage (Vp).

Low Frequency Response

The value of the blocking capacitors determines the low frequency response of the amplifier. The following expression is used to determine the blocking capacitor value to yield a desired 3 dB low frequency corner (fLFC):

\[
C_{Block} (\text{Farads}) = \frac{1}{100 \pi f_{LFC} (\text{Hz})}
\]

Bypass Capacitor

The reactive impedance of the bypass capacitor should be small compared to the impedance of the decoupling element at the lowest frequency of operation.
APPENDIX XII

NEC PRA/PRB REED RELAYS
The PR series reed relays have been designed to be compatible with dual in-line package IC devices in communication and general control circuit application design.

FEATURES
- Easy mounting on P.C. board
- Compatible with dual in-line IC devices
- Dust proof and gas proof

SPECIFICATIONS

<table>
<thead>
<tr>
<th>ITEMS</th>
<th>SPECIFICATIONS</th>
<th>SPECIFICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact Form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>1 Make Contact</td>
<td>2 Make Contacts</td>
</tr>
<tr>
<td>Voltage</td>
<td>10 VA max.</td>
<td>100 Vdc max.</td>
</tr>
<tr>
<td>Current</td>
<td>1 A max. carry only, 0.5 A max. switched</td>
<td></td>
</tr>
<tr>
<td>Initial Contact Resistance</td>
<td>150 milliohms max.</td>
<td></td>
</tr>
<tr>
<td>Life Expectancy</td>
<td>5 x 10⁶ operations at 100 Vdc, 100 mA.</td>
<td></td>
</tr>
<tr>
<td>Capacitance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Across open contacts</td>
<td>1.3 pF typical</td>
<td></td>
</tr>
<tr>
<td>Coil to contact</td>
<td>2.1 pF typical</td>
<td></td>
</tr>
<tr>
<td>Operate Time</td>
<td>0.6 ms max. incl. bounce at nominal voltage</td>
<td></td>
</tr>
<tr>
<td>Release Time</td>
<td>0.03 ms max. at nominal voltage, w/o diode</td>
<td></td>
</tr>
<tr>
<td>Breakdown Voltage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Across open contacts</td>
<td>250 Vdc max.</td>
<td></td>
</tr>
<tr>
<td>Coil to contacts</td>
<td>500 Vdc max.</td>
<td></td>
</tr>
<tr>
<td>Insulation Resistance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10¹⁰ ohms min. at 100 Vdc</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX XIII

SWITCHED CAPACITOR INTEGRATOR CONNECTIONS (OSCILLATOR AND TIMING)
APPENDIX XIV

OPTOELECTRONICS CD 10/CD 20 COMPONENTS PHOTODETECTORS
SERIES CD10
SERIES CD20
COMPONENT PHOTODETECTORS

DESCRIPTION

Series CD10 and CD20 Component Photodetectors are large area, high-speed photodiodes mounted in a special microwave circuit and housed in TO-3 packages. They are used to detect high frequency CW optical signals or short optical pulses. Each series consists of two models: the -01 model has a double convex lens in front of the photosensitive area and the -02 model has a high quality, 50/125 micron core/cladding optical fiber pigtail.

Series CD10 employs a silicon photodiode sensitive from 300-1100 nm and Series CD20, a germanium photodiode which may be used at longer wavelengths, up to 1800 nm.

These devices are designed for OEM applications in which other commercial photodiodes, having risetimes well above 0.5 nsec, are not fast enough. Their low cost permits systems consisting of many picosecond photodetectors to be constructed economically. The TO-3 packages are easily soldered and unsoldered to a PC board. A miniature PC board with bias voltage and signal output connectors is available as an option. This board facilitates connections to the package and converts it to an easy-to-use high-speed photodetector.
TEMPORAL RESPONSE

Figure 1

Typical response of a Series CD10 photodetector to a 70 psec wide, 900 nm wavelength, light pulse as monitored by a Tektronix 7904 oscilloscope with an S-4 sampling head.

SPECTRAL RESPONSE

Figure 2

ORDERING INFORMATION

All prices and specifications are subject to change without prior notice. For further information and technical assistance, contact

Opto-Electronics Inc.
Unit 9, 2538 Speers Road
Oakville, Ontario, Canada, L6L 5K9
Telephone: (416) 827-6214  Telex: 06-982392