95

Investigation into the feasibility of an Electro-Optic Rail Surface Monitor

by
Alexander F. Rawstorne

Dissertation submitted to the Department of Electrical and Electronic Engineering, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Science in Engineering.

Cape Town, May 1993

Declaration

I declare that this dissertation is my own unaided work. It is being submitted for the degree of Master of Science in Engineering at the University of Cape Town. It has not been submitted before for any degree or examination at this or any other university.

A. F. Rawstorne

May 1993.

Acknowledgments

I would like to thank the following for their assistance.

My Parents who have made this all possible.

Mr J R Greene who with his experience and knowledge has taught me a great deal about research. I would like to thank him especially for being my supervisor and for his patience.

Mr S Schrire for his patience, invaluable help with the circuit design, and most important, his humour.

Professor B J Downing for his advice.

Mr D Kenyon for construction of the optical test bed.

My gratitude also to Mr A Martin and Mr P Daniels for their practical assistance.

Many thanks to the Foundation of Research and Development (FRD) for their bursary.

Synopsis

Cracks and surface imperfections in railway track are important when considering its lifetime. This thesis investigates the feasibility of a new electro-optic technique which records surface irregularities on a track. It is more sensitive to crack detection than the traditional trailing wheel method and approximately one thousand times less expensive than the highly sensitive ultrasonic technique used by British Rail.

The electro-optic method consists of an optical projection–detection system with two mutually oblique axes. Light projected onto the rail is reflected and imaged onto a split photocell. A depression or rise in the reflective surface causes the axis of focus to shift and more light falls on one of the photocells. The difference in optical power between the two photocells is therefore a quantative measure of displacement and consequently rail surface profile.

The intention is to fix the device onto a bogey of a train so that measurements are taken as it moves over the track at speeds of upto $40ms^{-1}$. The relationship between vertical displacement and differential optical power is dependent on the shape and size of the spot. It is nonlinear for a circular spot and imperfections spanning less than $1mm$ vertically can be resolved. Vertical displacements greater than $5mm$ cannot be detected solely using the difference in optical power and the nett power falling on both photocells must be considered in addition.

An investigation into the reflective properties of rail lead to the optics being configured for specular reflection. The angles of incidence and reflection are optimally set at 35°. In addition, the use of an infra-red source is found to be an added advantage.

The system consists of a transmitter, receiver and the optics. An infra-red semiconductor laser is used in the transmitter as it is the only source with sufficient sterance (power per unit solid angle per unit area) to achieve a signal to noise ratio of at least $30dB$. It is intensity modulated with a $2MHz$ subcarrier which is demodulated asynchronously in the receiver. This eliminates ambient light variations decreasing the signal to noise ratio. The sum of optical power falling on both photocells is used for automatic light output control which dynamically adjusts the transmitter so that the sum of power on both photocells is constant.

A prototype of the system was built and tested in a simulated railway environment. An eight bit data acquisition module was used to accumulate data and verified that the electro-optic technique is feasible. The nonlinear differential voltage versus displacement characteristic is observed when the surface scanned is uniformly reflective. In general though, spot changes in reflectance mean that actual displacements cannot be accurately determined when moving over rail at $27ms^{-1}$. This is due to the automatic light output control being insufficiently fast to adapt to these rapid changes. The effect did not however prevent displacements of greater than $100\mu m$ from being consistently detected.

Correlation coefficients of data for different scans of a known surface were evaluated and are greater than 99%. This means that surface profile is measured consistently and only spot changes in reflectance give rise to errors.

Resolving $100\mu m$ cracks in rail demands a high bandwidth ($400kHz$) and sampling rate ($500kHz$). The high throughput of data to the storage mechanism, an AT computer's hard disc, therefore requires some means of data compression. Four techniques; threshold compression, zero, first and second order adaptive sampling were investigated and the least complex zero-order technique was evaluated with the test data. Data integrity was satisfactory for compression ratios of $1:3$ when there were imperfections every $4cm$ along the rail.

Contents

Declaration i

Acknowledgments ii

Synopsis iii

Table of Contents xiii

List of Tables xiv

List of Figures xix

1 Introduction 1

1.1 Analysis of the problem 2

1.1.1 Specifications 4

1.2 Current techniques used to monitor railway track imperfections 4

1.2.1 Trailing Wheel Technique 5

1.2.2 Ultrasonic technique of measuring track imperfections 6

1.2.3 Electro-optic technique of measuring imperfections 7

1.2.4 A comparison of the techniques 8

1.3 The electro-optic technique 9

1.3.1 A more detailed investigation 9

1.3.2 System feasibility 10

1.4 Summary of the remaining chapters 15

2 Design of an electro-optic track monitoring system **16**

2.1 Design goals 16

2.1.1 Optics 17

2.1.2 Receiver 18

2.1.3 Transmitter 19

2.2 Relationship between differential optical power and vertical displacement . 20

2.2.1 Rectangular spot 20

2.2.2 Circular spot 21

2.2.3 A summary of the optimum parameters 23

2.3 Reflective properties of railway line 24

2.3.1 Polar Diagrams and Reflectance of Rusty Steel 25

2.3.2 Reflectance 27

2.3.3 Reflectance of rusty steel as a function of optical wavelength 27

2.4 Conclusions 28

3 The receiver system **30**

3.1 Components of a receiver 30

3.2 The front-end amplifier 32

3.2.1 Resistive load termination 32

3.2.2 Tuned circuit coupling 33

3.2.3 Transimpedance front end 34

3.2.4 Noise performance of the different front ends 34

3.3 Design of the receiver circuitry 37

3.3.1 Circuitry of the front end amplifiers 37

3.3.2 The differential amplifier and summer circuitry 42

3.3.3 Configuring the optimum receiver 44

3.4 Noise considerations 45

3.4.1 Light source noise 45

3.4.2 Photodetector noise 46

3.4.3 Amplifier noise 46

3.4.4 Johnson noise 47

3.4.5 Total noise 47

3.4.6 Measurement of noise 48

3.4.7 Noise equivalent power 48

4 The Transmitter System **50**

4.1 Source 50

4.1.1 Requirements 51
4.1.2 LED versus Laser 52
4.1.3 Choice of a suitable source 54
4.2 Modulation 56
4.2.1 Requirements 56
4.2.2 Modulation Techniques 56
4.3 Laser driver 59
4.4 Automatic light output control 60
4.5 The transmitter circuitry 63
4.5.1 The temperature compensated laser driver 64
4.5.2 Modulation circuitry 64
4.5.3 Automatic light output control 65
4.6 Probability of error 65

5 System feasibility **68**

5.1 The electro-optic track monitoring system 69
5.1.1 The system 69
5.1.2 Focusing and configuring the system 70
5.1.3 A test setup 71
5.2 Design of a data capture unit 72
5.2.1 The analogue to digital converter module 73

5.2.2 Temporary storage module 75

5.2.3 PC interfacing and processing utilities 79

5.3 Performance test requirements 81

5.4 Static system tests 82

5.4.1 Procedure 82

5.4.2 Results 83

5.5 Dynamic system tests 85

5.5.1 Simulation of a train moving over railway track 85

5.5.2 Experimental procedure 87

5.5.3 Dynamic test results 89

5.6 Conclusions 96

6 Processing of information **97**

6.1 Correlation of data 98

6.1.1 Effect of averaging on correlation 100

6.1.2 Conclusions 100

6.2 Compression of data 100

6.2.1 Threshold data compression 102

6.2.2 Zero order adaptive sampling 102

6.2.3 First order adaptive sampling 103

6.2.4 Second order adaptive sampling 104

6.3 Application of the zero order technique to sampled data 104

6.4 Preliminary design of a zero order adaptive sampler 106

6.5 Summary 108

7 Conclusions and recommendations **110**

7.1 Receiver 111

7.2 Transmitter 111

7.3 System limitations and characteristics 112

7.4 Data processing 113

References **115**

A Calculation of vertical displacement for trailing wheel and electro-optic techniques **117**

A.1 Relationship between change in height and rotation angle for the trailing wheel technique 117

A.2 Derivation of change in power vs height relationship for an electro-optic track monitor 118

A.2.1 Rectangular Spot 120

A.2.2 Circular Spot 120

A.3 Mathcad listing for the circular spot relationship 122

B Optimum reflection angles for maximum change in power with height for rectangular and circular spots **125**

B.1 Rectangular Spot 126

B.2 Circular Spot 126

C Polar diagram measurement **129**

C.1 Measurement of Polar Diagrams 129

C.1.1 Aim 129

C.1.2 Method 129

C.1.3 Results 130

C.2 Measurement of specular reflectance 131

C.2.1 Aim 131

C.2.2 Method 132

C.2.3 Results 132

D Reflectance properties of rusty steel as a function of wavelength **134**

D.1 Aim 134

D.2 Method 134

D.3 Results 135

E Mathcad listings **137**

F Information relevant to design and operation of the receiver **144**

F.1 LRC Butterworth filter analysis 144

F.1.1 Sensitivity 146

F.2 Frequency Response of the receiver 147

F.2.1 Calibration of the source 147

F.2.2 Frequency response 148

F.3 Choice of suitable component values in a passive envelope detector 150

F.4 Measurement of the BPX 48 silicon photodetector responsivity and active area 152

F.4.1 Measurement of detector responsivity 152

F.4.2 Active Detector Area 153

G Measurement of noise and noise equivalent power in the receiver 156

G.1 Calculation of inverting amplifier input noise 156

G.1.1 Inverting amplifier input noise voltage 157

G.1.2 Second order filter 158

G.1.3 Total input noise 158

G.1.4 Noise due to the transimpedance and amplifier stage 158

G.2 Measurement of noise equivalent power 159

G.2.1 Source Calibration 159

G.2.2 Noise equivalent power measurement 160

H Circuit Diagrams 164

I Pascal Program Listings 171

I.1 S.PAS 171

I.2 BINASC.PAS 174

I.3 FIND.PAS 180

I.4 AVERAGE.PAS 184

I.5 ADAPTIVE.PAS 186

List of Tables

1.1 Specifications and design targets 4

2.1 Design goals for the optical system 20

2.2 Design criteria for circular and rectangular spots 23

5.1 States used to initiate sampling and transfer data to and from RAM 80

5.2 Thicknesses of aluminium used for imperfections 88

C.1 Reflectances for rusty steel 133

List of Figures

1.1 Trailing Wheel technique used to record track imperfections 5

1.2 Relationship between displaced height δh and angle θ 6

1.3 Ultrasonic scanning of railway line 7

1.4 Relationship between displaced height δh and angle θ a) Steady state b) imperfection present 8

1.5 Block diagram of an electro-optic track monitor 10

1.6 Potential problems in an electro-optic track monitoring system 11

1.7 The electro-optic system used in feasibility calculations 13

2.1 The optical system of an electro-optic track monitor 17

2.2 Normalised differential power as a function of the angle between vertical and receive axes 22

2.3 Normalised differential power vs height for a circular spot 23

2.4 Polar diagrams of specular and diffuse reflective surfaces. a)Specular b)Diffuse 24

2.5 Contrast between good and bad reflection for specular and diffuse cases . . 25

2.6 The three samples of rusty steel used to obtain the polar diagram and reflectance of steel track 26

2.7 Polar Diagram of surface 2 27

2.8 Reflection as a function of wavelength for rusty steel 28

3.1 Asynchronous technique used to reduce the low frequency noise 31

3.2 The resistive load termination receiver 33

3.3 A high impedance integrating front end receiver 33

3.4 The transimpedance front end 34

3.5 The transimpedance amplifier with low pass filter 38

3.6 Amplifier stages to realise a gain of 95dB 40

3.7 Asynchronous demodulation circuit 41

3.8 Frequency response of the receiver system 42

3.9 Circuit diagram of the differential amplifier 43

3.10 Circuit diagram of the summing amplifier 44

3.11 The noise sources in the electro-optic track monitor 45

4.1 Normalised differential power for spot radii 700, 350, 100 and 50μm for a BPX-48 photocell 51

4.2 Representation of the current pulse in the laser 57

4.3 Relative power distribution for first, second and third harmonics as a function of duty cycle x 58

4.4 Light output power vs forward current in an injection laser diode 62

4.5 Feedback system for automatic gain control 62

4.6 Circuit diagram of the transmitter 63

5.1 The electro-optic track monitor 69

5.2 The test setup 71

5.3 Principle used to focus the system for a 200 - 300μm spot 72

5.4 Analogue to digital converter specification 74

5.5 Analogue to digital converter module 75

5.6 Block diagram of the data capture unit 76

5.7 Circuit diagram of the temporary storage unit 77

5.8 Timing diagram of external sampling 78

5.9 A $10kHz$, $1.9V$ sine wave sampled at $500kHz$ 81

5.10 Apparatus for the static system tests 83

5.11 Differential voltage vs vertical displacement 84

5.12 Absolute differential voltage vs vertical displacement 84

5.13 Apparatus used to simulate moving railway track 86

5.14 A photograph of the apparatus 87

5.15 Noise of the system when scanning the stationary wheel 90

5.16 Differential voltage versus time : 1527rpm, 62.5kHz, 7000 samples 91

5.17 Differential voltage versus time : 1527rpm, 62.5kHz, 1500 samples, top graph taken 5 mins later 92

5.18 Differential voltage versus time : 1500rpm, 62.5kHz, 4000 samples 93

5.19 Differential voltage versus time : 1539rpm, 500kHz, 4000 samples 94

5.20 Differential voltage versus time : 1510rpm, 500kHz, 4000 samples 95

6.1 Cross-Correlation between two scans of the wheel surface 99

6.2 Autocorrelation vs cross-correlation between repetitive scans of the wheel surface 99

6.3 Differential power vs displacement for actual and zero order sampling : threshold = 10 105

6.4 Differential power vs displacement for actual and zero order sampling : threshold = 20 106

6.5 Distribution of data points for the zero order system in Fig 6.4 107

6.6 Block diagram of an adaptive sampler 108

A.1 Representation of the trailing wheel technique used to measure track imperfections 118

A.2 Representation of rectangular and circular spots in an electro-optic track monitoring system 119

A.3 Representation of the spot when moved by $\triangle I$ 121

A.4 Angles associated with the circular spot incident on the split photocell . . 122

C.1 Apparatus used to measure polar diagrams 130

C.2 Polar diagrams for different pieces of steel 131

C.3 Apparatus used to measure reflectance 132

D.1 Apparatus used to measure reflectance as a function of wavelength 135

D.2 Reflectance as a function of wavelength for rusty steel 136

F.1 Circuit diagram of a LRC second order Butterworth filter 145

F.2 Phase sensitivity as a function of frequency for for a second order Butterworth filter 147

F.3 Circuit used to measure the light output power of the LED as a function of frequency 148

F.4 Frequency response of LED 149

F.5 The apparatus used to measure receiver frequency response 149

F.6 Frequency response of the receiver 150

F.7 A passive envelope detector 150

F.8 Apparatus used to measure optical responsivity 153

F.9 Apparatus used to measure Detector Area 154

G.1 LRC filter and inverting amplifier stage in the receiver 157

G.2 Apparatus used to calibrate transmitter in NEP measurement 159

G.3 Relationship between power and distance 160

G.4 Apparatus used to calibrate the power curve 161

Chapter 1

Introduction

Railways are an important means of transport for people and goods. It is critical to maintain a high degree of reliability and efficiency to minimise the loss of goods and maximise passenger safety. Integrity of a railway system is dependent on the condition of the track and of the vehicles that use it. The latter is well understood and extensive procedures exist to test carriages and locomotives. Condition of the track is equally important and must be evaluated. A crack in a rail can lead to accidental loss of life and equipment. The presence of imperfections also increases wear and tear of vehicles on the track. The huge capital cost of replacing track periodically is not economically viable, and since it is impractical to remove sections of track, testing has to be performed on-site. Instrumentation to measure condition of rail is therefore required, and this thesis proposes and investigates the feasibility of a new electro-optic approach.

The design of an instrument to measure imperfections in track must emphasise high reliability at minimum cost. Extensive rail networks require periodic monitoring of large amounts of track. Consequently many scanners are needed to evaluate the whole network on a regular basis. Each monitor should be low cost and maintain a high level of efficiency. Costs of the scanner can also be reduced if existing equipment is used. This implies the use of routinely available rolling stock, rather than dedicated locomotives or carriages to carry the instruments. Mechanical wear and tear of the instrument should also be minimised to increase the replacement interval.

Three techniques are used to establish the condition of railway track. The first is labour intensive and involves observation of track by railway employees. It is unreliable, inefficient and impractical in terms of labour required. Consequently two automated approaches are favoured. The rolling wheel technique used in most countries including South Africa is the most common. It is however not as reliable as the more expensive ultrasonic method used in Britain. Lower reliability of the rolling wheel and high cost of the ultrasonic techniques make a new electro-optic approach proposed by Greene [10] a viable alternative. It is based on the triangulation of reflected light and is discussed with the two other automated approaches in the following section.

The aim of this chapter is to formalise a specification for an instrument capable of measuring the imperfections on a railway track and to evaluate the proposed method in the context of alternative possible approaches. The following section analyzes the problem by characterising and quantifying imperfections. This is followed by an analysis of the rolling wheel, ultrasonic and electro-optic approaches. The proposed electro-optic system is then discussed in detail together with its associated problems. Solutions and calculation of electrical requirements are presented to establish feasibility. The final section summarizes the contents of further chapters.

1.1 Analysis of the problem

A formal specification is required to achieve a set of design goals for a given problem. In order to specify a track monitoring system it is necessary to characterise and classify imperfections. Imperfections may be considered to be surface or subsurface. Subsurface irregularities are caused by metal fatigue and more commonly, poor welds. The following two irregularities differ in the length of track they span and constitute surface imperfections. The spinning motion of wheels when a locomotive pulls away causes heat and melts the surface of the track causing a cusp. Another example is a crack in a spot weld on the track. In general, both of these imperfections will span the width of the track.

Although it is desirable to acquire information on both surface and subsurface imperfections, the rolling wheel and electro-optic techniques only measure surface irregularities. These upper surface imperfections caused by wear and tear by trains or other mechanisms

are sufficient when considering the lifetime of a railway track. Imperfections are specified for length of track they span and their vertical impression or depression in the rail. Length of an imperfection is dependent on the mechanism that caused it. The cusp caused by spinning of a train's wheel constitutes the largest imperfection and may be as large as $10cm$. Cracks in spot welds give rise to the minimum resolvable imperfections and are typically a fraction of a millimetre in length. It is good practise to design conservatively and an interval ranging from $1m$ to $0.1mm$ is designed for. The maximum vertical extent of an imperfection is represented by a break in the track i.e. a depression of at least $10cm$. Minimum vertical extent on the other hand, is limited by system noise and a resolution of $10\mu m$ is assumed sufficient.

No mention has yet been made of the rate at which information is accumulated, and consequently the time taken to evaluate a piece of track. This is chiefly dependent on the response time of the apparatus used. Observation of track by railway employees is inherently the slowest technique. Since the aim is to fit instruments to existing coaches, it is desirable for acquisition to take place at maximum train speed and it was a stated requirement [10] that the system would operate at speeds of $40ms^{-1}$. At these speeds, there should be limited mechanical contact between device and track to limit wear and tear.

Equally important to identification of an imperfection is its location. This is required for repair and a high accuracy is not required. Location can be obtained by equispaced markers placed alongside the track, or by integrating the speed of the train. If average speed v_i in time interval $\triangle t_i$ is provided, the location l can be calculated by

$$l = \sum_{i=1}^{n} v_i \triangle t_i + d$$

where d is the starting reference. Accuracy can be arbitrarily increased by decreasing interval $\triangle t_i$. It is assumed that inputs are provided for calculation of position.

Having met the technical requirements above, it is important to keep system cost as low as possible. As mentioned previously, this can be achieved by making use of available equipment and instruments can be fixed to existing locomotives or coaches. Limiting mechanical contact with the track will also reduce replacement and servicing costs.

1.1.1 Specifications

The previous discussion establishes the requirements of an instrument to accumulate information about the surface of a railway track. This can be formalised by the specification in Table 1.1

Cost	Keep as low as possible - use available equipment and limit contact of instrument with track.
Type of Imperfection	Surface, and if possible internal track imperfections
Maximum length of imperfection	$1m$
Minimum length of imperfection	$0.1mm$
Width of imperfection	Spans the width of the track
Maximum vertical extent	$10cm$
Minimum vertical extent	$10\mu m$
Maximum train speed	$40ms^{-1}$
Location	Assume speed and reference inputs are provided

TABLE 1.1: SPECIFICATIONS AND DESIGN TARGETS

The requirements of an instrument used to scan the surface of a railway track have been presented. Now existing techniques and the electro-optic approach must be investigated in more detail.

1.2 Current techniques used to monitor railway track imperfections

Three techniques for determining track integrity are discussed in the following sections. The first two are in use, and the last is the electro-optic approach proposed. The first is low cost, but unreliable and not capable of detecting cracks. The second is highly reliable and moderately efficient, but extremely expensive. An electro-optic technique is then proposed which attains high reliability at low unit cost. The three systems are discussed and compared in terms of the specifications in §1.1.1.

1.2.1 Trailing Wheel Technique

The trailing wheel technique has been used in South Africa to record the condition of railway tracks. The system comprises a wheel riding on the rail fixed by an axle to the bogey of a moving train. The arrangement is shown in Fig 1.1 and movement of the axle through angle θ represents the stored information. Knowing the length L of the axle

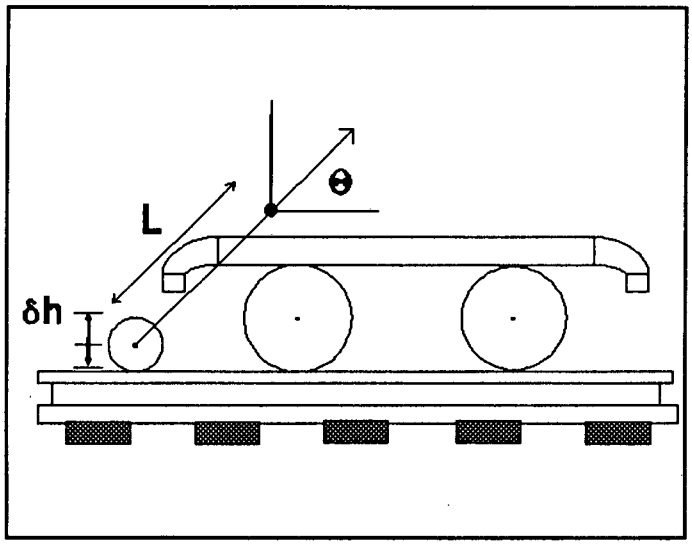


FIGURE 1.1: TRAILING WHEEL TECHNIQUE USED TO RECORD TRACK IMPERFECTIONS

and the difference in angle at two locations gives the vertical movement of the wheel. This displacement $\delta h \approx L \sin\theta \cos\theta$ is illustrated in Fig 1.2 (see App A). Although the system is fairly low cost, it is not capable of resolving small imperfections and only shallow depressions can be detected by this means. A major disadvantage of the technique is that large imperfections on the track may dent or smash the wheel. The faster the train moves, the more likely this is to happen. Other disadvantages include the effect of train vibration, inertia of the wheel that tends to 'smooth' irregularities, error associated with tilt of the bogey as it moves over an imperfection which leads to erroneous results.

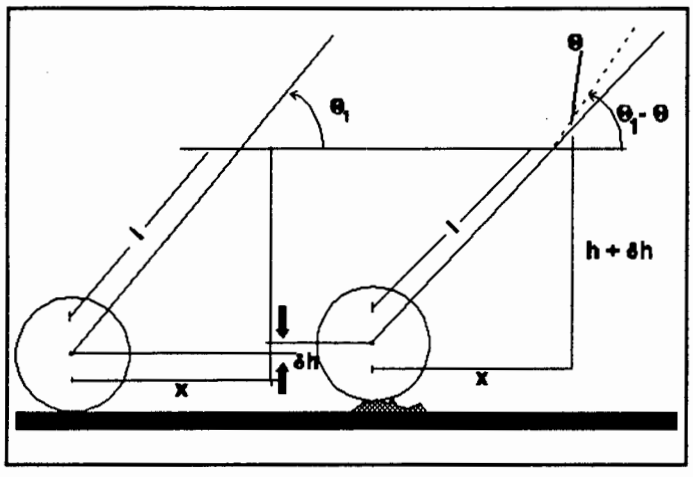


FIGURE 1.2: RELATIONSHIP BETWEEN DISPLACED HEIGHT δh AND ANGLE θ

1.2.2 Ultrasonic technique of measuring track imperfections

Ultrasonic imaging of railway track was developed by British Rail, and is presently the most reliable system for detecting surface and sub-surface imperfections [2]. It consists of three ultrasonic transducers scanning the rail at different angles. The arrangement is shown in Fig 1.3 and line can be scanned from a system moving at a maximum speed of $17.36ms^{-1}$. The principle is based on reflection of ultrasonic waves from discontinuities in the rail i.e. imperfections and was developed from research on the ultrasonic scanning of welds. The shoe–water–rail interface is used for acoustic matching. Each transducer views an identical piece of rail, the output of which is stored on a three-track, continuous tape. Presence of a crack constitutes a large reflection which triggers a threshold circuit activating the pick up head located further back in time than the recording head. Data is recorded with a position reference on another tape. This is later analyzed at a laboratory where the typical reflection 'signature' is isolated from imperfections. The method is very reliable giving high resolution information about the sub-surface condition of rail. The technique is however extremely costly and mechanically cumbersome requiring a specially designed carriage.

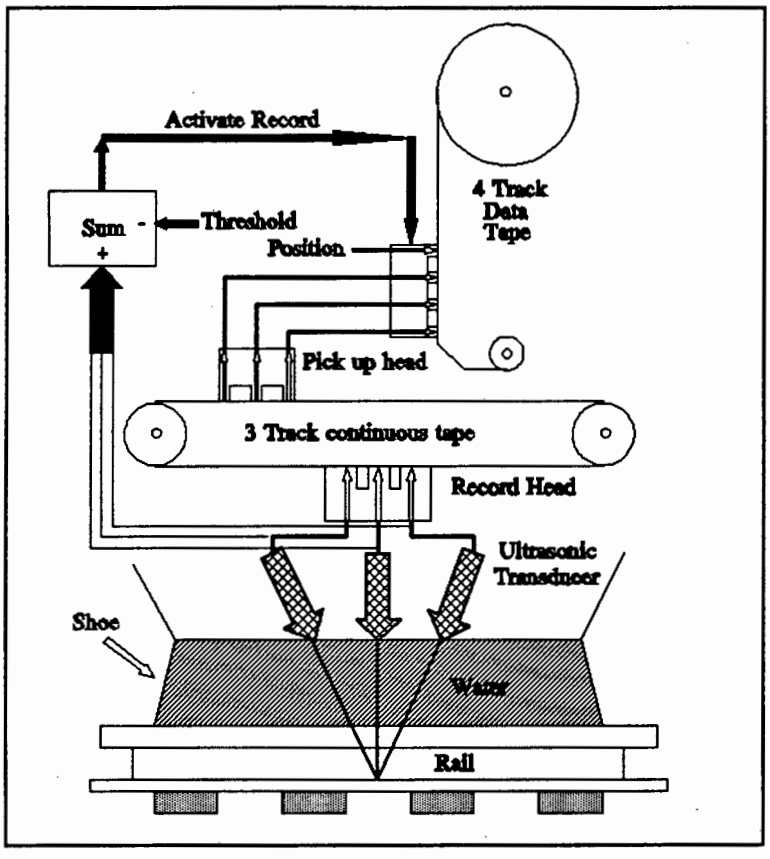


FIGURE 1.3: ULTRASONIC SCANNING OF RAILWAY LINE

1.2.3 Electro-optic technique of measuring imperfections

A new technique using triangulation of light reflected off the track is proposed to overcome the disadvantages of the last two methods. It should have higher resolution and be less costly than its ultrasonic counterpart. The principle is shown in Fig 1.4 and only surface imperfections are resolved. The transmitted spot on the rail surface is reflected and focused by the receive lens onto a split photocell (Fig 1.4a). Fig 1.4b illustrates how a depression δh is detected. An equal amount of light falls on both photocells in the absence of a bump. A depression causes the position of the image to shift laterally and more light falls on one of the detectors. This assumes that the defocusing effect of the depression is negligible, which it is for large image distances ($l \geq 2f$ where f is the focal distance of the lens). The relationship between change in power $\triangle P$ with height $\triangle h$ is given in Equation 1.1 for a rectangular spot (see App. A for derivation). A similar formula is worked out

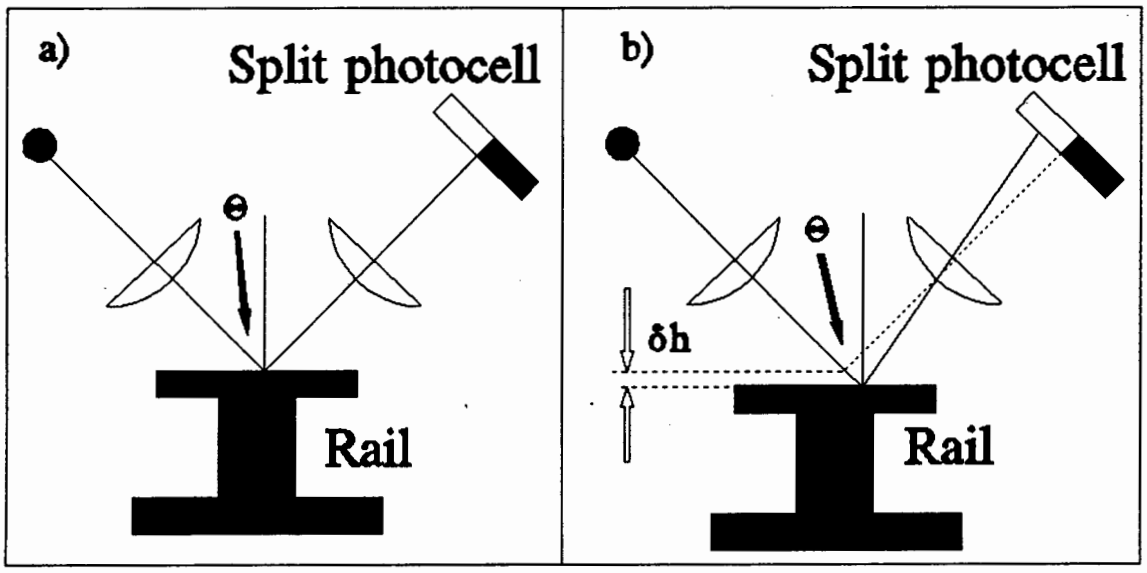


FIGURE 1.4: RELATIONSHIP BETWEEN DISPLACED HEIGHT δh AND ANGLE θ A) STEADY STATE B) IMPERFECTION PRESENT

in the next chapter for circular spots.

$$\triangle P = \frac{4P\delta h \sin\theta}{x} \tag{1.1}$$

P is the power falling on both photocells and θ the angle between transmit and vertical axes. Nothing moves physically on the rail surface in this method, reducing the probability of damage to the transducer, and consequent measurement error. It is also low cost and data acquisition can take place on a train moving at $40ms^{-1}$. Disadvantages of the method are that vibrations and tilt of the bogey lead to error. The possibility of using more than one transponder (i.e. redundancy) as in the ultrasonic technique could overcome this and also improve reliability.

1.2.4 A comparison of the techniques

The ultrasonic and electro-optic approaches both meet the specifications of Table 1.1. The trailing wheel technique fails to resolve small imperfections and cannot operate at high train speeds which makes it unsatisfactory. Although the ultrasonic technique has

the advantage that it detects subsurface imperfections, and has a high acquisition rate, a niche may exist for the simpler and lower cost electro-optic alternative. The absence of a transducer-water-rail interface is an additional asset when considering wear and tear. These advantages are felt to warrant investigation of the electro-optic technique in this thesis.

1.3 The electro-optic technique

The electro-optic technique has been presented and it is necessary to fully characterise the components that make up a system. In the light of this characterisation, calculations and limitations of the technique are investigated to verify feasibility.

1.3.1 A more detailed investigation

An electro-optic track monitoring system consists of a transducer, data storage and processor. A basic system overview is given in Fig 1.5. It consists of two parts, the digital storage and processor unit, and the optical system with transducer and signal conditioning electronics. The optical system is fixed to the bogey of the train and has a $10cm$ clearance between it and the track. It generates an analogue voltage which represents deviations in the track and is connected by shielded cable to a personal computer (PC) storage and processing unit. This digitises, compresses (preprocesses) and stores the information with the location of the train. Analysis of track imperfections takes place in two phases. Firstly there is the measurement and storage stage which captures information and its location. On completion of accumulating data for a track, it is transferred to a non-realtime environment for processing. This enables detection and characterisation of imperfections on a piece of track. If the data throughput is sufficiently low, the system may be real time with on-site processing. A compromise may be to have online detection of large imperfections and offline analysis of data at a remote laboratory.

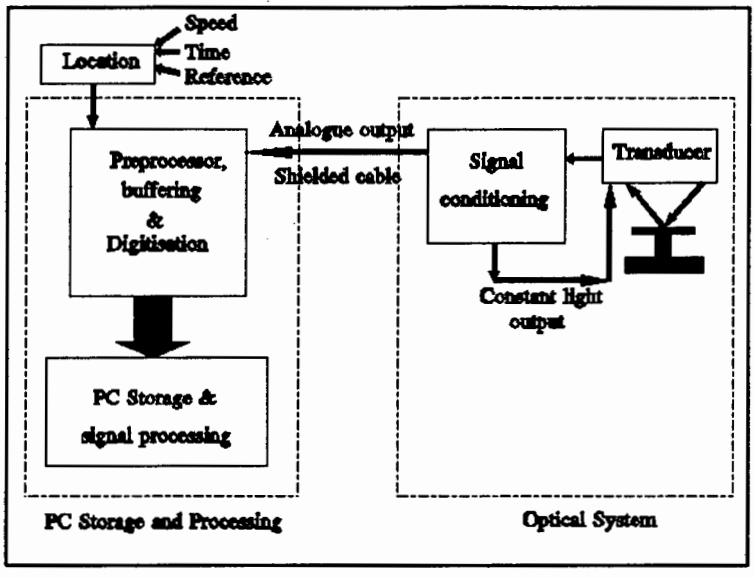


FIGURE 1.5: BLOCK DIAGRAM OF AN ELECTRO-OPTIC TRACK MONITOR

1.3.2 System feasibility

It is important to establish feasibility of the electro-optic technique before proceeding with design. Accurate measurements and detailed design are required to prove conclusively that the system is viable. At this stage approximations have to be made. There are essentially two areas which have to be investigated. Firstly the problems inherent in the technique have to be identified and solutions proposed. Secondly, the electronic requirements have to be established to determine practicality.

Possible design problems

A number of possible problems and possible sources of error must be considered. The most important are:

1. Any vibration of the bogey is measured as a track imperfection.
2. If the cross-section of the track is curved or twisted, movement of the focus point over this cross-section will contribute measurement error.

3. Tilt of the bogey contributes to measurement error.

4. Change in radius at different points on the bogey wheel circumference are measured as imperfections.

These four scenarios are illustrated in Fig 1.6. The first is overcome with high pass

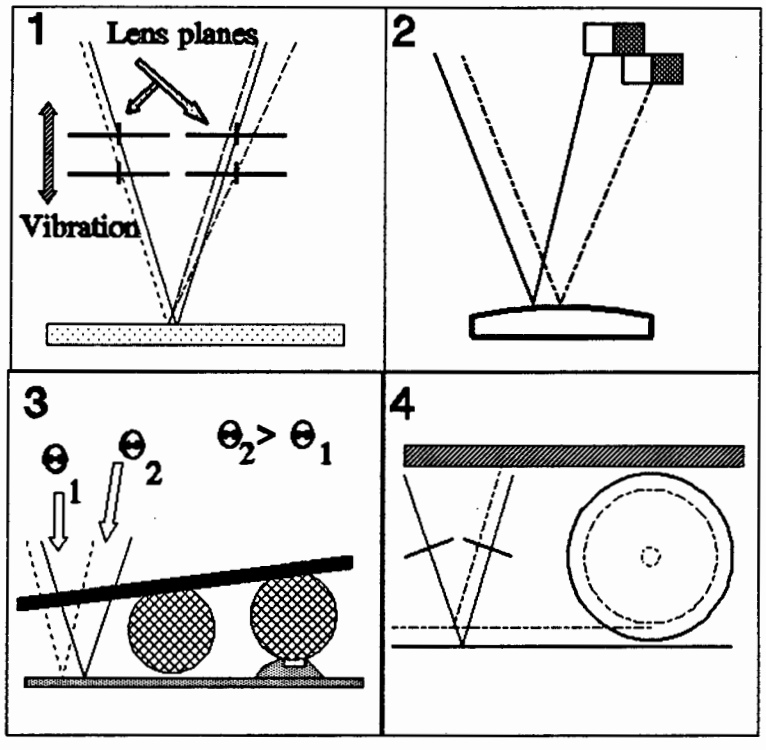


FIGURE 1.6: POTENTIAL PROBLEMS IN AN ELECTRO-OPTIC TRACK MONITORING SYSTEM

filtering and/or non real-time signal processing. Mechanical vibration noise statistics are not known accurately but are not expected to have frequency components greater than $20Hz$ for a train moving at $40ms^{-1}$. For lower speeds, this upper frequency limit has increasing influence on maximum resolvable distance since the train covers less distance in the period of vibration. A differential technique can be used to remove vibration artifact in this scenario and involves using two systems, one to measure vibration and the other to measure the rail imperfections. The difference between the two is the desired signal.

Problem 2 is insignificant as the track is almost flat. If measurement shows that this is in fact a problem, techniques used to fix the spot at a position in the track cross-section will have to be investigated.

Tilt of the bogey is not expected to contribute large measurement error and is proportional to wheel separation. This error is minimised by placing the instrument halfway between the wheels of the bogey.

Changes in wheel radius will also contribute to error. As this is periodic, filtering can be used to remove it.

An increase in temperature results in thermal expansion of the wheels. Halliday and Resnick [11] give the coefficient of linear expansion for steel as $11 ppm/°C$. Change of temperature during measurement of track imperfections is assumed to be less than $5°C$. The change in wheel radius for this temperature range is insignificant and therefore thermal expansion is ignored.

Electronic requirements

This section contains an approximate calculation to verify the feasibility of obtaining a reflected signal from the track. Individual system components have not yet been investigated, and approximations have to be made in calculations. Perhaps the most important electronic requirement is availability of a suitably powerful source. The optical system in Figure 1.7 is assumed and the on-axis pointance of the Lambertian transmitter of power X watts is :

$$I_s = \frac{X}{\pi}$$

Power captured by the $2cm$ transmit lens is calculated from source pointance (I_s), lens area (A) and distance between the lens and transmitter (r).

$$\begin{aligned} P_{TxL} &\approx I_s \times \frac{A}{r^2} \\ &\approx 0.01X \end{aligned}$$

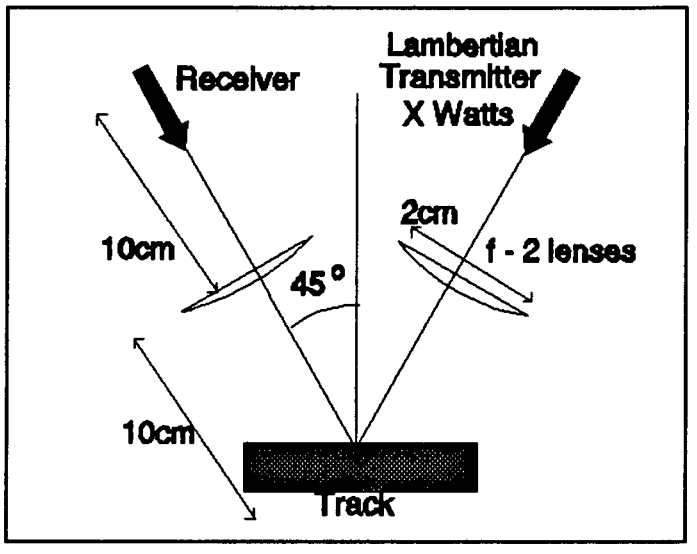


FIGURE 1.7: THE ELECTRO-OPTIC SYSTEM USED IN FEASIBILITY CALCULATIONS

An 8% transmission loss associated with the lens is due to absorption and refractive index mismatch. The power transmitted to the lens surface is thus :

$$\begin{aligned} P_{surface} &\approx 0.01X \times 0.92 \\ &\approx 0.0092X \end{aligned}$$

The worst case radiation profile of the reflective surface is Lambertian. Peak on-axis pointance is therefore

$$\begin{aligned} I_r &\approx \frac{P_{surface}}{\pi} \\ &\approx 0.003X \end{aligned}$$

The receive lens is oriented at 45° to the vertical axis. Pointance is relatively constant over the 2cm diameter receive lens. Power captured by the receive lens is calculated from pointance, lens area and source–lens distance.

$$\begin{aligned} P_{RxL} &= \frac{I_{r45^\circ} \times A}{r^2} \\ &= \frac{(0.003X \cos 45)(\pi 0.01^2)}{0.1^2} \\ &= 63 \times 10^{-6}X \end{aligned}$$

Assuming a lens transmission loss of 8%, power incident on the photocells is

$$P_{Photo} = 58 \times 10^{-6}X$$

Signal current in each photocell is half P_{Photo} multiplied by silicon responsivity.

$$\begin{aligned} I &= 0.5 \times 0.5 \times 58 \times 10^{-6} X \\ &= 14.5 \times 10^{-6} X \end{aligned}$$

Signal power is the product of voltage and current developed in a terminating resistance R. The choice of R is dependent on bandwidth of the system and is limited by shunt capacitance of the photodiode ($20pF$ max). System bandwidth can be estimated from minimum span of an imperfection and maximum train speed. A train moving at $40ms^{-1}$ sees $0.1mm$ every $2.5\mu s$ or at a rate of 400000 per second. The bandwidth can therefore be assumed to be $400kHz$. The resistance R can be evaluated for bandwidth and shunt capacitance.

$$\begin{aligned} R &\leq \frac{1}{2\pi BC} \\ &\leq 19000\Omega \end{aligned}$$

Signal power is the product of this resistance and square of the signal current.

$$\begin{aligned} P &= \left(14.5 \times 10^{-6} X\right)^2 \times 19000 \\ &= 3.99 \times 10^{-6} X^2 \end{aligned}$$

For a $1mW$ source, the signal power is therefore $4nW$.

In order to calculate dynamic range, or signal to noise ratio, noise power must be estimated. It is made up of shot and Johnson noise [21]. Johnson noise is dependent on temperature and bandwidth. Horowitz [17] gives the formula for Johnson noise. For a temperature of $20°C$ and $400kHz$ bandwidth the noise power P_j may be calculated

$$\begin{aligned} P_j &= 4kTB \\ &= 6.47 \times 10^{-15} W \end{aligned}$$

where k is Boltzman's constant. Shot noise is dependent on load resistance, bandwidth and signal current. Horowitz [17] gives the equation for shot noise and the respective parameters are substituted.

$$\begin{aligned} P_s &= 3.2 \times 10^{-16} IBR \\ &= 3.65 \times 10^{-7} W \end{aligned}$$

Total noise power is the sum of the uncorrelated noise sources.

$$\begin{aligned} P_n &= 3.65 \times 10^{-7} + 6.47 \times 10^{-15} \\ &= 6.51 \times 10^{-15} W \end{aligned}$$

The signal to noise ratio is therefore $30dB$ which is sufficient to justify further investigations.

1.4 Summary of the remaining chapters

The second chapter gives a detailed description of the aims, design goals and optical requirements of the system. This includes an analysis of rusty rail reflective properties which are of interest in choosing specular or diffuse detection techniques. Having obtained a more concise specification for the system, chapter 3 gives a detailed description of the receiver, for example its design and noise specifications. Design of the receiver precedes that of the transmitter as receiver noise defines the signal power required. Full analysis of the receiver must take place before the noise power can be calculated theoretically. Chapter 4 follows with a similar description of the transmitter. Chapter 5 examines system feasibility by establishing dynamic and static performance of the receiver and transmitter in a simulated railway environment. A description of the data acquisition is given with the relevant results which establish system feasibility. The next chapter evaluates correlation between data obtained previously and discusses techniques used in data compression. Application of one technique is applied to data and its performance observed. A theoretical design of this zero-order adaptive sampling is then given. The last chapter discusses conclusions of the thesis and gives recommendations.

Chapter 2

Design of an electro-optic track monitoring system

This chapter discusses the design parameters in an electro-optic track monitor. These are dependent on the optical system, electronic requirements and environment in which the system operates. The requirements of the two optical projection systems with mutually oblique axes are discussed first. This is followed by an investigation of the receiver and transmitter. The change in received differential optical power with vertical displacement is dependent on three parameters; optimal choice of these is important to achieve an effective system. Lastly, the reflective properties of rusty rail are measured. These include obtaining the polar diagram and reflection vs optical wavelength characteristic.

2.1 Design goals

In this section, design parameters are derived which satisfy the specifications of the previous chapter. It is assumed that vibration can be adequately filtered and the differential system discussed in §1.1.3 is not considered. The system is divided into three parts :

- Optics
- Receiver
- Transmitter

2.1.1 Optics

In essence the optics comprises two projection systems with mutually oblique axes. A source projects a focused spot of light onto the track and this in turn is focused onto a split or dual photodetector. Due to the mutual obliqueness of the optical axes, local variations in height translate into variations in differential optical power falling on the two photocells. The optical system is illustrated in Fig 2.1. The parameters 'a', 'b', 'd' and θ have to be chosen.

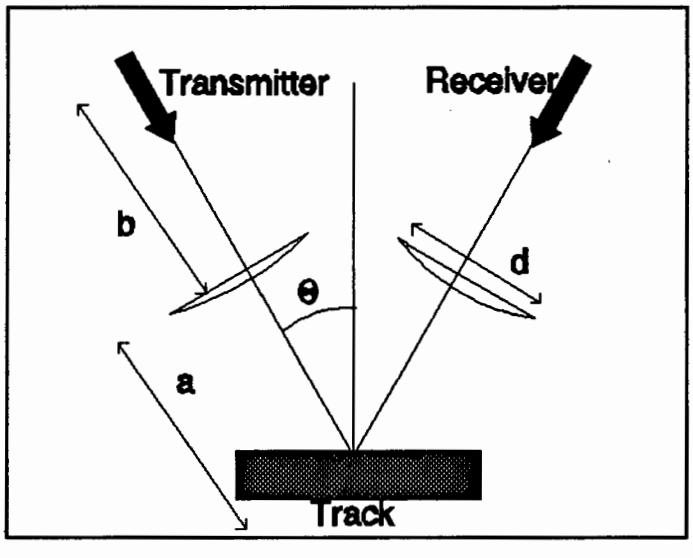


FIGURE 2.1: THE OPTICAL SYSTEM OF AN ELECTRO-OPTIC TRACK MONITOR

To minimise the likelihood of damage to the unit as a result of objects such as stones on the track, a clearance of at least $50mm$ is required. A clearance of $150mm$ is probably the most that should be considered. Spacing greater than this would unnecessarily reduce

returned signal without any significant improvement in system robustness. A nominal clearance of 'a' = $100mm$ is assumed sufficient in the present study.

It is assumed that the light source area is smaller than $100\mu m$ by $100\mu m$, which is less than the required resolution. This means image magnification is uncritical and is initially assumed as 1 : 1. Consequently 'b' is equal to 'a' and is $10mm$.

The actual size of the projected optical image is determined by the (predominantly spherical) aberration of the lens [22]. Spot size of a simple optimally-configured equi-bi-convex lens is given by $\frac{0.07f^3}{d^2}$ where d is lens diameter and f is the focal length [9]. Assuming a standard $50mm$ focal length, $f/2$ lens, the spot size is $0.014mm$. Since this is inadequate in terms of the target specification it is necessary to resort to a compound lens system. The simplest and most economical solution is to use a compound camera lens, even the crudest of which (at, say, 10 line pairs per mm) has a resolution performance far in excess of that required.

The choice of θ involves a tradeoff between depth resolution and collected optical flux from a diffuse surface. Initially, an angle of $\theta = 20°$ is a reasonable compromise, subject to more detailed investigation.

2.1.2 Receiver

The receiver must be designed to achieve the information bandwidth. This is calculated from horizontal resolution ($0.1mm$) and maximum train speed ($40ms^{-1}$) and is $400kHz$. The photodetector must therefore have sufficiently low capacitance to be fast enough to achieve the $30dB$ signal to noise ratio.

It is proposed that differential voltage will be digitised and stored. A sampling rate of $800kHz$ is required to satisfy the $400kHz$ bandwidth. This can be achieved with low cost eight bit analogue to digital converters. The quantisation is assumed sufficient to achieve the resolution desired with the $30dB$ signal to noise ratio at this data rate.

Split photocells have finite width and detector separations which influence vertical resolution. If the separation is larger than the $100\mu m$ diameter spot there will be no change

in differential voltage for small displacements. A separation of order of $100\mu m$ or less is assumed sufficient. The width of the photocells restricts the range of displacements that can be measured with differential voltage. A width of $1mm$ is deemed sufficient, yielding, at 20° separation of the optical axes and $100\mu m$ spot size, a depth range of $\pm 3mm$.

2.1.3 Transmitter

Optical power and emitting area of the source are important in the design of the transmitter. Emitting area has to be sufficiently small to achieve the $0.1mm$ resolution. The optical power also has to be large enough to achieve the $30dB$ of signal to noise ratio when reflected off the worst-case Lambertian surface.

The size of the projected image of the source is determined by the specified resolution as $0.1mm$. It is assumed that the imperfections to be detected span the full width of the track. This means the source's image may take the form of a circular patch of light.

The source may be either a LED or semiconductor laser. Lasers are able to radiate from a smaller emitting area than LEDs. This means that they can achieve greater power density (W/m^2)in the spot. Lasers are therefore capable of higher signal to noise ratio than LEDs. They are also inexpensively available as a replacement component for compact disc players in which their capability to produce high power density in a focused spot is exploited. The effective source area is typically about $10\mu m$ by $1\mu m$. Since the laser is somewhat more expensive and fragile than the LED, the discussion of which to use will be deferred at this stage, until more detailed calculations have been made.

Design goals for the optics, transmitter and receiver are summarised as follows :

Optics	Clearance between track an optics $\geq 50mm$, point to point focusing, use of standard $50mm$, $f/2$ camera lenses, angle $\theta \approx 20°$
Receiver	$400kHz$ information bandwidth, 8 bit quantisation of differential voltage, photocell width $\approx 1mm$, photocell separation $\approx 0.1mm$
Transmitter	Source emitting area diameter $> 0.1mm$, circular or line spot, LED or laser source

TABLE 2.1: DESIGN GOALS FOR THE OPTICAL SYSTEM

2.2 Relationship between differential optical power and vertical displacement

The relationship between differential optical power and vertical displacement must be understood in order to design an optimum system. It is different for circular and rectangular spots and it is necessary to consider each.

2.2.1 Rectangular spot

A change of $\triangle P$ in differential optical power is related to the change in height $\triangle h$ for a rectangular spot by

$$\triangle P = \frac{4P \sin\theta}{x} \triangle h \tag{2.1}$$

The relationship is derived in Appendix A and assumes a uniformly illuminated spot. P is the source power, x the width of the spot and θ the angle between vertical and optical axes. Equation 2.1 represents a linear change in differential optical power with height.

The smaller the width, x, of the spot, the higher the vertical resolution for small displacements. The minimum width is limited by photocell separation and is optimally $100\mu m$. The maximum width is limited by horizontal resolution ($100\mu m$) and is less than $200\mu m$. A width of $100\mu m$ is assumed optimal.

Increasing the incident optical power, P, increases the vertical resolution. This can be achieved by using a more powerful source, or by making use of the specular reflective

properties of the railway track. In the interests of efficiency and cost effectiveness the latter is optimised. Specular reflection and a $5mW$ source is assumed to provide adequate resolution for the $30dB$ signal to noise ratio.

The last parameter is choice of θ, the angle between optical and vertical axes. Movement of the spot on the photocell is largest for angles above 45°. However, reflected power is greatest for the worst case Lambertian reflective surface if θ is small. The angle at which there is maximum change in differential power is calculated in Appendix B as 35°. This assumes the worst case Lambertian reflective surface.

2.2.2 Circular spot

The relationship between differential power and height for a circular spot is calculated in Appendix A. The nonlinear function is given by :

$$\triangle P = P\left(1 - \frac{2}{\pi}\left(\arccos\frac{2\triangle h \sin\theta}{r} - 0.5\sin\left(2\arccos\frac{2\triangle h\cos\theta}{r}\right)\right)\right) \tag{2.2}$$

P is optical power, $\triangle h$ change in height, θ the angle between vertical and optical axes, and r is the radius of the spot.

An increase in optical power, P, increases the resolution of small vertical displacements. As with the rectangular spot, specular reflection and a $5mW$ source are assumed to achieve the $10\mu m$ vertical resolution and $30dB$ signal to noise ratio.

Increasing spot radius means that larger imperfections can be measured. This increased dynamic range is at the expense of the horizontal resolution. Alternatively, if the spot r is made smaller, large vertical displacements cannot be measured, only detected. In addition, if diameter is too small, the spot can fall in between the two photocells. It is initially assumed that a spot diameter of $100\mu m$ is sufficient.

An optimum angle θ exists that gives maximum differential power for a Lambertian reflective surface. Since the relationship between differential power and height is nonlinear, this angle varies for different displacements. Figure 2.2 shows the effect for vertical displacements 0.9, 0.5, 0.3 and 0.1 of the spot radius as angle θ is varied from 0° to 90°. As displacement increases, so too does differential power. The most important trend however

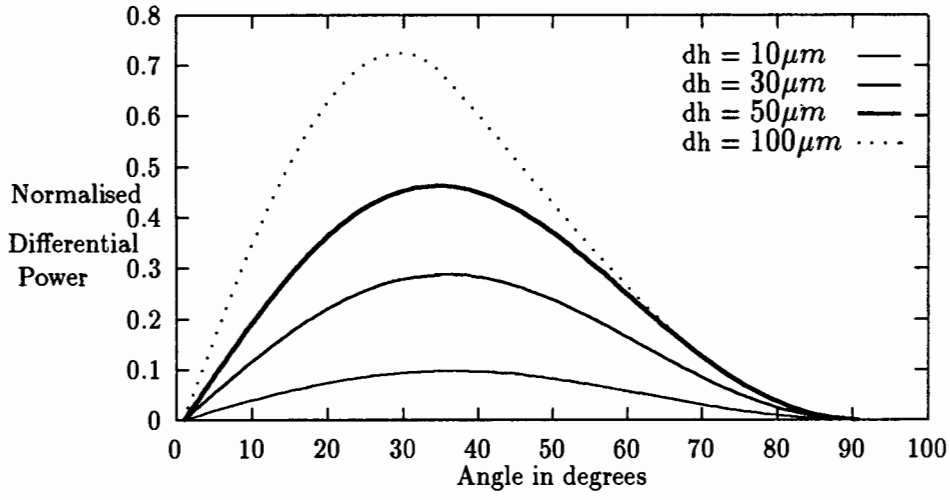


FIGURE 2.2: NORMALISED DIFFERENTIAL POWER AS A FUNCTION OF THE ANGLE BETWEEN VERTICAL AND RECEIVE AXES

is that peak power occurs at angles between 20 and 40°. An angle θ falling within this range is therefore suitable. It should preferably be between 20 and 30° where the peak responses are sharpest.

The implications of the nonlinear relationship between differential power and displacement must be investigated. Perfectly rectangular spots are a misnomer and emitting areas are not geometrically well defined. In reality, the spot will be a combination of the rectangular and circular spot characteristics.

Normalised differential power of Equation 2.2 is plotted in Fig 2.3. The characteristic is linear for small vertical displacements (60% of spot radius). Differential power does not increase significantly after this and eventually reaches saturation. The spot then falls exclusively on one photocell.

Differential power is directly related to the difference in fractional area of the uniformly illuminated spot that falls on the photocells. The spot is divided into two segments by the photocell. As the area of one segment increases, the other decreases and results in an increase in differential area. The area in a segment of a circle decreases non-linearly with the size of the segment, and the contribution of the smaller segment becomes less and less significant. This is essentially when 80% of the spot falls on one photocell and it

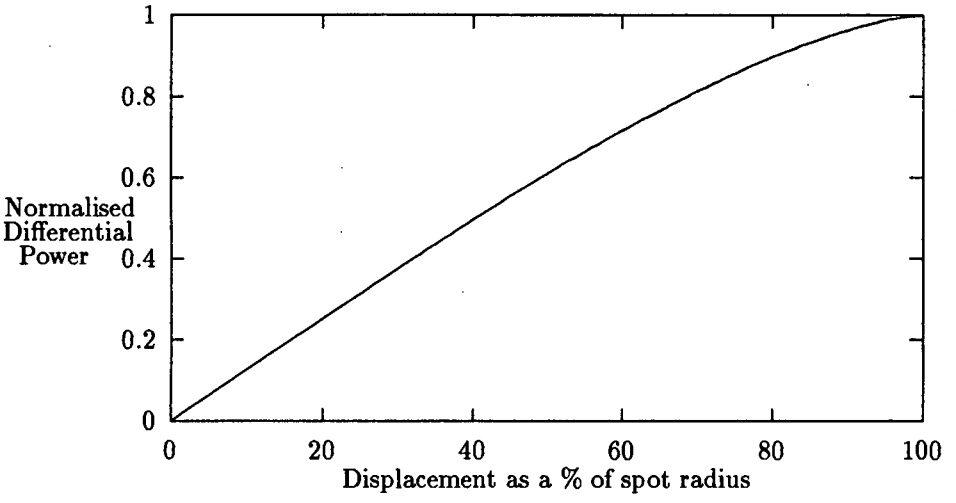


FIGURE 2.3: NORMALISED DIFFERENTIAL POWER VS HEIGHT FOR A CIRCULAR SPOT

gives rise to the nonlinearity in Fig 2.3.

An eight-bit, linear, analogue to digital converter was used to digitise the differential voltage. The nonlinear characteristic above implies that resolution varies with displacement. For imperfections less than 60% of the spot radius, maximum resolution is achieved. Greater displacements result in increasingly less resolution and their amplitude is not known as accurately. This is assumed to be acceptable.

2.2.3 A summary of the optimum parameters

The optimal choice of the parameters in Equations 2.2 and 2.1 are summarised for rectangular and circular spots in Table 2.2.

Parameter	Rectangular	Circular
Spot Size	$100 \times 100\mu m$	$100\mu m$ diameter
θ	35°	25°
P	significant power for $30dB$ $\frac{S}{N}$	significant power for $30dB$ $\frac{S}{N}$

TABLE 2.2: DESIGN CRITERIA FOR CIRCULAR AND RECTANGULAR SPOTS

The choice of θ depends on the reflective properties of the railway track and it is necessary to discuss these.

2.3 Reflective properties of railway line

The reflective properties of rail are required to calculate source power and obtain an optimum optical system. This includes knowledge of the polar diagram of the reflective surface and the power captured by the receive lens (reflectance). These depend on the type of track surface which is either rusty or polished. The surface is thus diffuse or specular.

Specular surfaces reflect light in one direction only and an example is a mirror. The angle of highest reflectance is equal to the angle of incidence. Surfaces which scatter light equally in all directions like matt white paper are diffuse. To understand the relevance of specular or diffuse reflection consider the polar diagrams in Fig 2.4. Figure 2.4a) is the

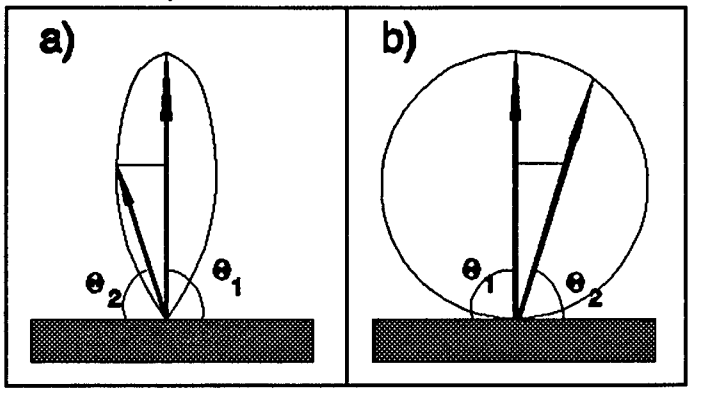


FIGURE 2.4: POLAR DIAGRAMS OF SPECULAR AND DIFFUSE REFLECTIVE SURFACES. A)SPECULAR B)DIFFUSE

polar plot of polished steel and Fig 2.4b), a diffuse reflector.

Real surfaces are neither perfectly diffuse nor perfectly specular. Part of the light is reflected and part scattered. If Figure 2.4a) was perfectly specular, the polar plot would have been an impulse with 0° beamwidth.

The optics can be arranged to take advantage of either diffuse or specular reflection. Light from the transmitter is reflected and then captured by the receiver. If the angle of incidence, θ_1, is not equal to the angle of reflectance, θ_2, reflection is diffuse. If $\theta_1 = \theta_2$ then reflection is specular.

Railway tracks are in general highly reflective and are thus predominantly specular. However, unused track is rusty and significantly diffuse. If the diffuse mechanism is used ($\theta_1 \neq \theta_2$), there will be poor return on shiny track and good reflection on rusty track. If specular reflection is used ($\theta_1 = \theta_2$), the opposite occurs, good reflection for shiny, and bad reflection for rusty track. Reflected power is less for diffuse reflection off a shiny surface than it is for specular reflection off a diffuse surface. The contrast is shown in Figure 2.5 and as a result the optics are arranged for specular in preference to diffuse reflection.

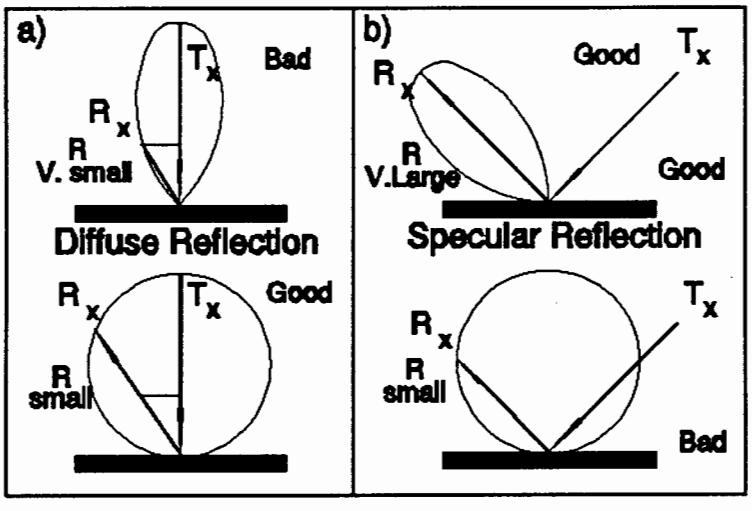


FIGURE 2.5: CONTRAST BETWEEN GOOD AND BAD REFLECTION FOR SPECULAR AND DIFFUSE CASES

2.3.1 Polar Diagrams and Reflectance of Rusty Steel

The reflectance and polar diagram of rusty rail is required to calculate signal to noise ratio. These were measured (Appendix C) for three different pieces of rusty steel. The three samples are visually quite different and are illustrated in Fig 2.6.

FIGURE 2.6: THE THREE SAMPLES OF RUSTY STEEL USED TO OBTAIN THE POLAR DIAGRAM AND REFLECTANCE OF STEEL TRACK

Polar Diagrams

Four measurements, one for surfaces 1 and 2, and two for surface 3 were taken. The polar diagram of surface 2 is given in Fig 2.7. Although the rusty surfaces are visually quite different, the polar diagrams are very similar. Three out of four graphs give a $\cos^2\theta$ distribution. This indicates that rusty steel is slightly directional or specular. It is assumed that rusty rail has a $\cos^2\theta$ distribution.

2.3.2 Reflectance

Standard, $50mm$, $f/1.7$ camera lenses were used to measure reflectance of the three surfaces (Appendix C). The ratio of reflected to incident power is the reflectance. Reflectance was measured twice for each surface and the worst case reflectance is 1.62×10^{-2} (Table B.1). All measurements are consistent, and the T-piece in Fig 2.6 has a slightly smaller

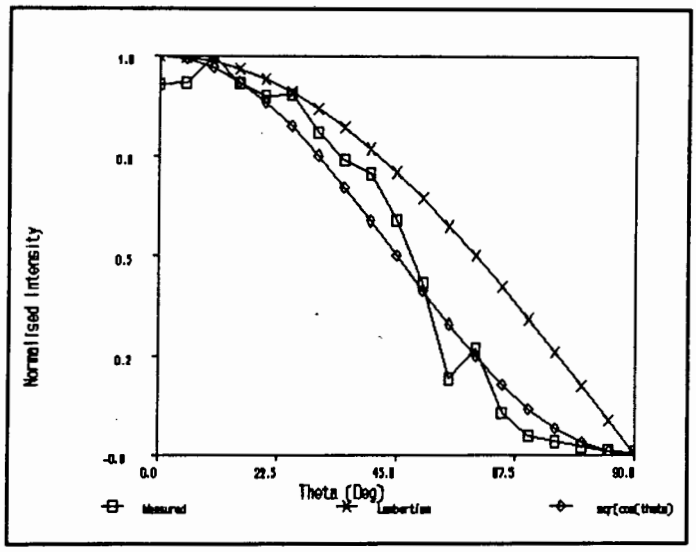


FIGURE 2.7: POLAR DIAGRAM OF SURFACE 2

reflectance. A conservative value of 1×10^{-3} is assumed to be the worst case reflectance.

2.3.3 Reflectance of rusty steel as a function of optical wavelength

It is important to see if a variation in optical wavelength changes the received optical power. If it does, the appropriate source 'colour' will be used. Two parameters, the sensitivity of the photocell and reflectance, change when subject to different optical wavelengths.

The reflectivity was measured for rusty steel (Appendix D) at infra-red ($910nm$), red ($675nm$), orange ($610nm$), yellow ($565nm$) and green ($510nm$) wavelengths. These were then plotted as a function of optical wavelength and are shown in Fig 2.8. The reflectance is best for infra-red and decreases with wavelength. The reflectances of red, orange, yellow and green relative to infra-red are 0.67, 0.66, 0.51 and 0.49 respectively.

Silicon photocells are most efficient in the infra-red. Photocells such as the proposed BPX-48 split photocell have peak responsivity at $850nm$. The relative responsivities of the infra-red, red, orange, yellow and green are 1, 0.67, 0.56, 0.47 and 0.35 respectively. In comparison with the measured values above, the responsivity is generally better than

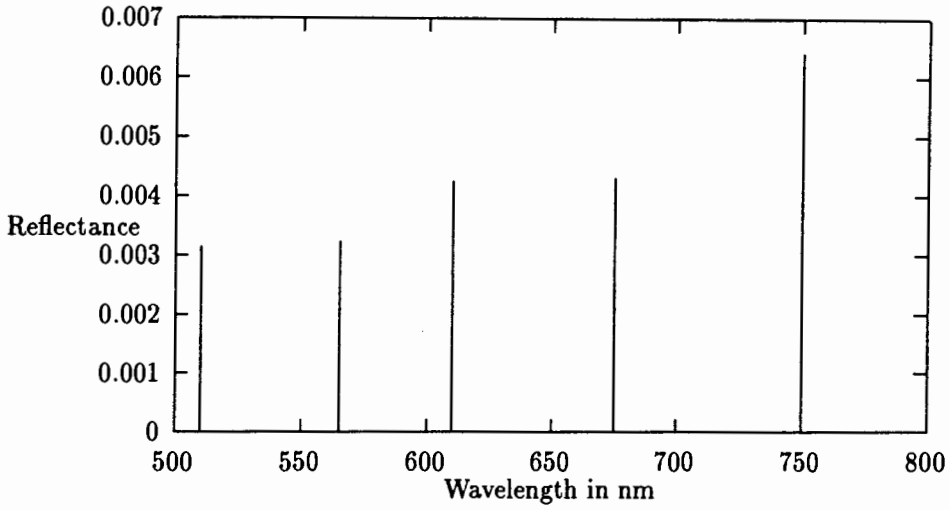


FIGURE 2.8: REFLECTION AS A FUNCTION OF WAVELENGTH FOR RUSTY STEEL

expected. This means that the apparent difference in reflectance is due to the changes in responsivity.

The results above indicate that a near-infra-red source must be used. Infra-red lasers and red, or infra-red LEDs are therefore the most suitable.

2.4 Conclusions

Characterisation and choice of optimal parameters in an electro-optic track monitor were discussed. These considered the design goals of the optics, transmitter and receiver. The relationship between differential optical power and height, and reflective properties of rusty steel were also discussed. The requirements of the system are summarised as follows :

- Proximity to track $\geq 50mm$
- Use of standard $50mm$, $f/1.7$ camera lenses
- The angle between the optical and vertical axis should lie between 20° and 40°

- $400kHz$ information bandwidth
- Use of a silicon split photocell with a detector width $\approx 1mm$ and separation $\approx 0.1mm$
- Use of an infra-red ($\approx 850nm$) source
- Emitter area $< 0.1mm \times 0.1mm$
- Optics arranged to make use of specular reflection

Chapter 3

The receiver system

This chapter describes the design of the receiver electronics. An important design goal is achieving the maximum signal to noise ratio. This involves minimising the noise in the receiver.

Variations in ambient light introduce additional noise and two approaches are discussed which minimise it. This is followed by an investigation of the three possible configurations for front end amplification. Of these, only one is suitable and it is incorporated into the receiver. Each of the modular components is tested and the system noise is then measured. A prediction of Noise Equivalent Power based on characteristics of the individual noise sources is then checked by actual measurements.

3.1 Components of a receiver

The receiver output is the differential sum of two transimpedance amplifiers with additional gain. Components have an electrical bandwidth of $400kHz$, and the transimpedance front end amplifiers are optimised for minimum noise. In addition it is assumed that the system is powered from $\pm 5V$ supplies.

The system is subject to high and low frequency noise sources which decrease signal to noise ratio in the receiver. The technique which eliminates the low frequency noise like drift, pink ($\frac{1}{f}$) noise and ambient light variations is considered. It consists of modifying the front end stages and shifting the receiver bandwidth upward in frequency.

The optical source is intensity modulated, reflected off the track, and then demodulated synchronously or asynchronously. The track imperfections appear as amplitude modulation on the subcarrier. As a result, the low frequency noise can be high pass filtered before demodulation, which leads to a much larger signal to noise ratio.

Synchronous demodulation has the advantage that it is a coherent detection system which constitutes a matched filter with respect to the signal. It is however impractical at the high carrier frequency ($\geq 1MHz$) associated with the large receiver bandwidth ($400kHz$)[23].

Asynchronous detection is therefore used in this thesis and involves non-coherent demodulation. The technique [1] is illustrated in Figure 3.1. A $2MHz$ subcarrier is amplitude

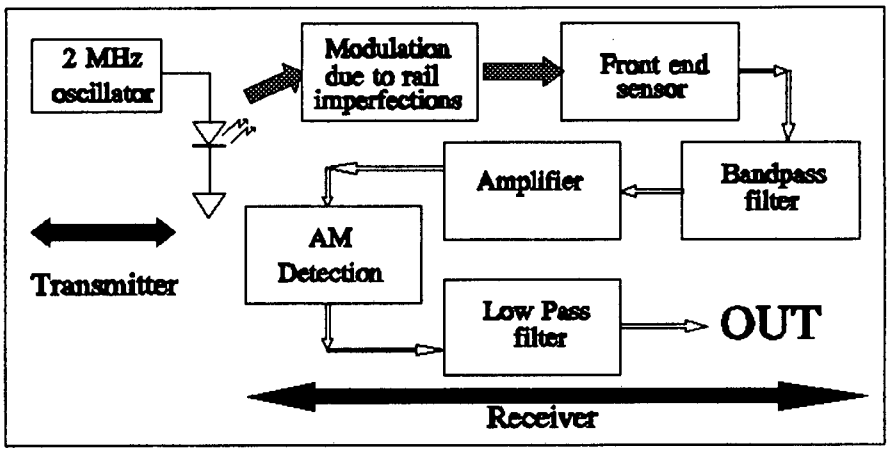


FIGURE 3.1: ASYNCHRONOUS TECHNIQUE USED TO REDUCE THE LOW FREQUENCY NOISE

modulated by rail imperfections. The front end sensor converts this intensity modulated light to the appropriate electrical signal. Bandpass filtering is then used to remove the low frequency noise and is followed by additional amplification. The AM detector is then used to demodulate the signal, and the low pass filter to remove the post detection ripple above the information bandwidth.

A subcarrier frequency of $2MHz$ with an information bandwidth between 1.6 – $2.4MHz$ is used. The lower the subcarrier frequency, the more the post detection ripple. Alternatively, the higher the frequency, the more complicated the detector and amplifier circuitry. A frequency of $2MHz$ was chosen as a reasonable compromise between performance and circuit complexity.

3.2 The front-end amplifier

The front-end amplifier consists of the transducer (photodiode) and front end amplifier. There are three alternative designs; the resistive termination, the tuned circuit and transimpedance approaches. These are compared in terms of cost, complexity and signal to noise ratio. The aim is to achieve the maximum signal to noise ratio by choosing the design with the least noise in the receiver bandwidth.

The requirement for a monolithic split detector constrains the choice of photodetectors. Avalanche detectors are not readily available in this form, entail considerable additional circuit complexity to establish stable bias conditions, and are expensive. In any event they offer only marginal performance improvement in this situation (moderate bandwidth requirement and only partial exclusion of ambient illumination). The use of a PIN detector with its reduced junction capacitance would also be marginally advantageous, but a split PIN detector was not available at reasonable cost.

3.2.1 Resistive load termination

The resistive load termination receiver is shown in Fig 3.2 [21]. Current in the photodiode is converted to a voltage and then amplified. The load resistance is limited by shunt input capacitance of the preamp and photodiode. Photodiode capacitance is nominally $20pF$ and Bipolar or FET preamp capacitance is $8pF$ which limits R_L to 2370Ω. This reduces the signal to noise ratio significantly [21] and consequently the technique is not suitable in this application.

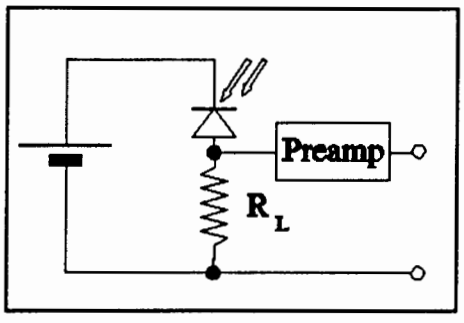


FIGURE 3.2: THE RESISTIVE LOAD TERMINATION RECEIVER

3.2.2 Tuned circuit coupling

This receiver exploits the high tuned circuit impedance at resonance to overcome the effect of the depletion capacitance [8] [21]. The circuit is shown in Figure 3.3 and consists of a parallel LC circuit tuned to the subcarrier frequency (f_0), and FET preamplifier.

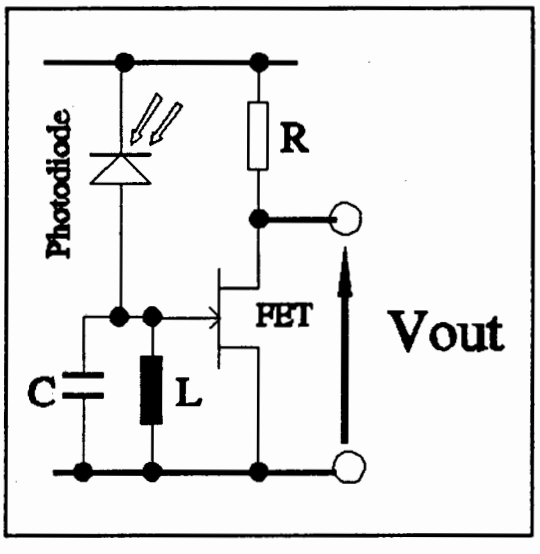


FIGURE 3.3: A HIGH IMPEDANCE INTEGRATING FRONT END RECEIVER

Shunt depletion capacitance of the photodiode and FET can now be tuned out, and load impedance R_L calculated from [8] :

$$R_L = \frac{Q}{2\pi f_0 C} \tag{3.1}$$

The load resistance is directly proportional to Q and the technique finds application in narrow band systems such as distance measurement.

3.2.3 Transimpedance front end

The transresistance amplifier converts the current in the photodiode to a voltage. Figure 3.4 illustrates the circuit which exploits the low input impedance (due to feedback)

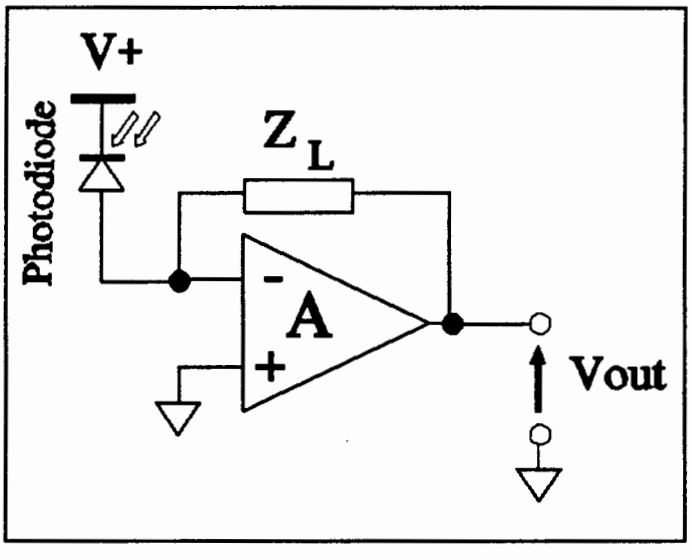


FIGURE 3.4: THE TRANSIMPEDANCE FRONT END

and high transimpedance gain ($-Z_L$). It is the most widely used front end stage and is implemented with discrete transistors at high frequencies and opamps at low frequencies [21].

3.2.4 Noise performance of the different front ends

The resistive termination has already been ruled out. In order to choose between the high and transimpedance front ends, their noise performance is compared by calculating Noise Equivalent Power (NEP). Equation 3.2 [21] [9] gives the signal to noise ratio of a

photodetector receiver which is used to calculate NEP

$$\frac{S}{N} = \frac{(\rho W_{sig})^2 R_L}{2qB(\rho(W_{sig} + W_{back}) + I_d) + {i_j}^2 R_L + {i_a}^2 R_L} \tag{3.2}$$

where ρ is photodetector responsivity; W_{sig}, signal power; R_L, load resistance; q, electronic charge; B, bandwidth; W_{back}, background power; I_d, dark current; i_j, Johnson noise current and i_a amplifier noise current.

The NEP is defined as the noise power at which signal to noise ratio is unity. To calculate NEP it is assumed that signal and background power does not contribute to the noise power. The tuned circuit and transimpedance NEP is calculated for $f_0 = 2MHz$ and $B = 800kHz$.

Tuned circuit :

The Q required for a bandwidth of $800kHz$ is :

$$\begin{aligned} Q &= \frac{f_0}{B} \\ &= \frac{2 \times 10^6}{1 \times 10^6} \\ &= 2 \end{aligned}$$

Assuming $C_{min} = 25pF$ (Diode capacitance + strays) and using Equation 3.1 the load resistance is :

$$\begin{aligned} R_L &= \frac{2}{2 \times \pi \times 10^6 \times 2 \times 25 \times 10^{-12}} \\ &= 6370\Omega \end{aligned}$$

Johnson noise current [17] is

$$\begin{aligned} i_j &= \sqrt{\frac{4kTB}{R_L}} \\ &= \sqrt{\frac{16 \times 10^{-21} \times 10^6}{6370}} \\ &= 1.6 \times 10^{-9} A \end{aligned}$$

Silicon FET noise current [8][21] is

$$i_a = \sqrt{\frac{2kTB \times 0.7}{R_L \times g_m}}$$

For a typical FET [8], the transconductance g_m is $1mS$.

$$\begin{aligned} i_a &= \sqrt{\frac{8 \times 10^{-21} \times 10^6 \times 0.7}{6370 \times 10^{-3}}} \\ &= 35 \times 10^{-9} A \end{aligned}$$

The dark current (I_d) is maximally $200nA$ for a BPX-48 photocell and the shot noise due to dark current is

$$\begin{aligned} i_d &= \sqrt{2qBI_d} \\ &= 253 \times 10^{-12} A \end{aligned}$$

To calculate NEP, $\frac{S}{N} = 1$. If X is the optical power incident on one photocell then using equation 3.2

$$\begin{aligned} (0.5X)^2 R_L &= 1.6 \times 10^{-13} X R_L + \left(2.5 \times 10^{-18} + 876 \times 10^{-18} + 64 \times 10^{-21}\right) R_L \\ 0 &= X^2 - 6.4 \times 10^{-13} X - 3.51 \times 10^{-15} \\ \rightsquigarrow X &= 59 \times 10^{-9} \\ \rightsquigarrow NEP &= \frac{X}{\sqrt{1 \times 10^6}} \\ &= 59 \times 10^{-12} \frac{W}{\sqrt{Hz}} \end{aligned}$$

A listing of the Mathcad calculation is given in Appendix E.

Transimpedance amplifier : opamp configuration

The individual terms in Equation 3.2 are substituted such that

$$\begin{aligned} JohnsonNoisePower &= 4kTB \\ &= 16 \times 10^{-15} W \end{aligned}$$

$$\begin{aligned} AmplifierNoisePower &= {i_{na}}^2 R_L B \\ &= 3.6 \times 10^{-14} W \end{aligned}$$

$$
\begin{aligned}
ShotNoisePower &= 2qBR_L\left(\rho\left(W_{sig}+W_{back}\right)+I_d\right) \\
&= 2.327\times 10^{-15}W
\end{aligned}
$$

Once again, the optical power when $\frac{S}{N}=1$ is X and using equation 3.2

$$
\begin{aligned}
0 &= 5.386X^2-6.302\times 10^{-9}X-7.065\times 10^{-14} \\
\rightsquigarrow X &= 3.622\times 10^{-9}W \\
\rightsquigarrow NEP &= \frac{X}{\sqrt{B}} \\
&= 2.56\times 10^{-12}\frac{W}{\sqrt{Hz}}
\end{aligned}
$$

Optimum front end

Johnson and amplifier noise dominate the shot noise in both cases. The Noise Equivalent Power is less for the transimpedance configuration and it is therefore used in preference to the tuned circuit arrangement.

3.3 Design of the receiver circuitry

The receiver circuit consists of two front-end stages which are inputs to differential and summing amplifiers. Each front-end photocell amplifier is made up of a transimpedance stage, amplification stages and a passive asynchronous demodulator.

3.3.1 Circuitry of the front end amplifiers

There are two front end stages in the receiver and each is made up of :

Transimpendance amplifiers

The transimpedance amplifier is illustrated in Fig 3.5. An operational amplifier is used

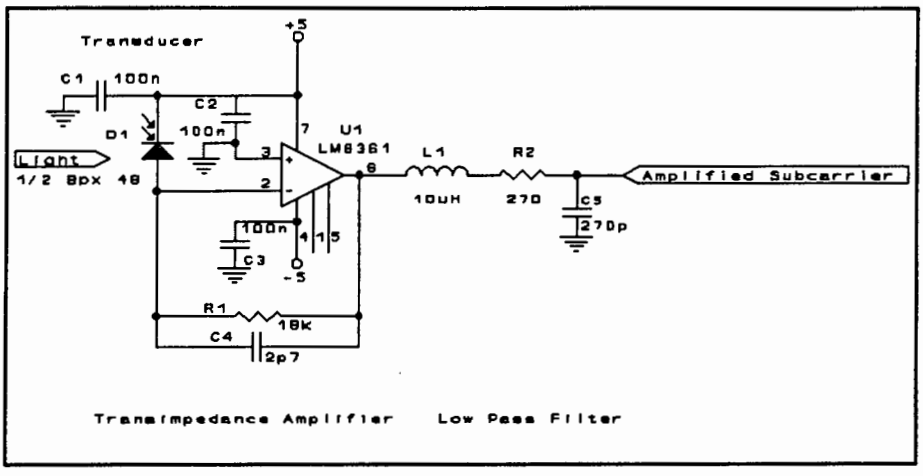


FIGURE 3.5: THE TRANSIMPEDANCE AMPLIFIER WITH LOW PASS FILTER

in preference to the transistor configuration although the latter is marginally superior in terms of noise performance [21].

The choice of a suitable opamp depends chiefly on the rise time of the output voltage, and to a lesser extent, on the minimum noise voltage and current. The rise time t_R is calculated from [14]

$$t_R = \pi\sqrt{\frac{R_f \times C}{2\pi GBW}}$$

Capacitance C is the sum of feedback, amplifier input and photodiode shunt capacitance which is maximally $30pF$. The feedback resistance (R_f) is $18K\Omega$ giving a rise time of $163ns$ which is sufficient for the $2MHz$ subcarrier and $1MHz$ bandwidth.

The feedback capacitance of $2.7pF$ is used to bandlimit the system to $3.27MHz$. This is larger than the $2.4MHz$ required, allowing for some reduction of bandwidth due to the additional LRC low pass filtering. A larger feedback resistance and smaller capacitance ($22K\Omega$ and $2.2pF$) is not suitable as the system is unstable. This is due to the depletion capacitance of the photodiode which forms a relaxation oscillator with the feedback resistance.

The Butterworth LRC second order low pass filter bandlimits the output of the transimpedance amplifier to $3MHz$. Calculation of component values and analysis of phase and amplitude response of the circuit is given in Appendix F. Low Q ferrite in-line inductors can be used as Q is dominated by the much larger series resistance. Phase response of the filters is not critical, and accurate component matching is not required for identical phase. Consequently, components with 5% tolerances are acceptable.

Additional amplification

Two amplification stages are provided for optimum demodulation of the subcarrier signals. The passive demodulators require a positive subcarrier voltage of at least $0.6V$ to turn on. It is therefore essential to provide enough amplification so that when the spot falls on only one photocell, the output amplifiers are nearly saturated.

At this stage, the source power is unknown and the approach is to provide enough amplification for $100mV_{rms}$ of noise. Shot noise is assumed insignificant and the maximum signal to noise ratio is therefore $28dB$. This is calculated from the ratio of the maximum allowable RMS voltage before amplifier saturation ($3.5V$) and the noise voltage.

The output noise voltage of the transimpedance amplifier is $47\mu V$ (see following section) which is amplified by 2130 to give $100\mu V$ of noise. Two uncompensated LM6365 opamps are used to realise this as shown in Figure 3.6. They have a minimum and maximum gain of 25 and 133 in the $1 - 3MHz$ bandwidth and are bandlimited from $880kHz$ to $14MHz$.

The low frequency breakpoint of the amplifiers is set at $880kHz$. In addition to filtering out drift, ambient light variation and pink noise, the a.c. coupling also prevents d.c. offset of the previous stage from being amplified. The noise bandwidth is also reduced for the 'white' Johnson and shot noises. The LM6365 and LM6361 can source and sink currents of $25mA$ and it is advantageous to make input resistances as small as possible for minimum noise. For a maximum input voltage of $3.5V$, this corresponds to a minimum resistance of 180Ω. A suitable compromise is therefore 220Ω with a $820pF$ capacitor.

The high frequency breakpoint is set to $14MHz$ by the feedback resistance and capac-

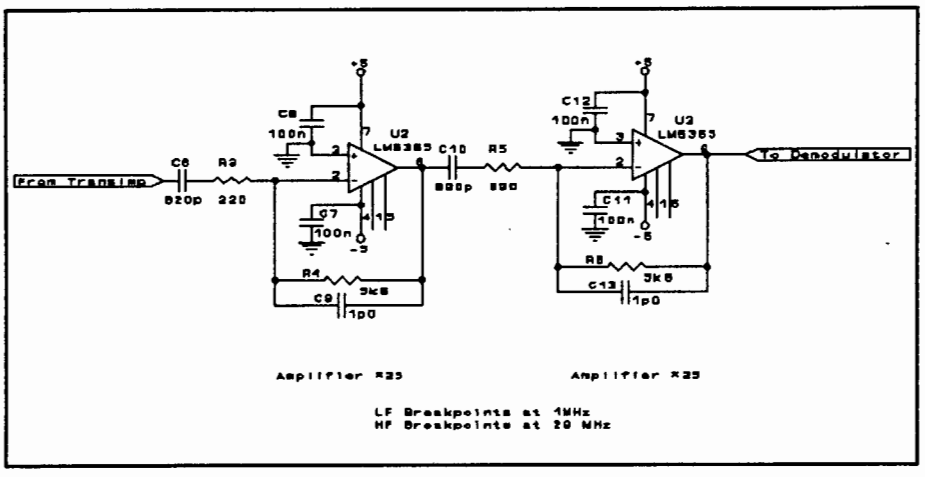


FIGURE 3.6: AMPLIFIER STAGES TO REALISE A GAIN OF 95DB

itance. It is set sufficiently high in frequency to prevent any significant attenuation in the $1 - 3MHz$ bandwidth. As the LM6365 is only partially compensated and requires a minimum gain of 25, the feedback capacitance must not be too large for the desired gain. Feedback of signals near the natural frequency at gains of less than 25 due to the low feedback impedance result in oscillation. A feedback resistance of $5.6K\Omega$ and capacitance of $1pF$ gives a gain of 25 and a breakpoint of $14MHz$.

Amplitude demodulation

The $2MHz$ subcarrier is amplitude modulated by imperfections on the railway track, and the passive asynchronous detector in Figure 3.7 is used for demodulation. It consists of an envelope detector which rectifies and smooths the signal, and a Butterworth LRC low pass filter.

The envelope detector consists of a diode and RC circuit. The diode half wave rectifies the subcarrier and the RC time constant is chosen to optimally recover the modulation. Appendix F shows how the values R and C are chosen. The diode is oriented so the summer output voltage (see next section) is positive. The use of active demodulators was not considered as at this frequency they are more complex, but have the advantage that they do not require a turn-on voltage.

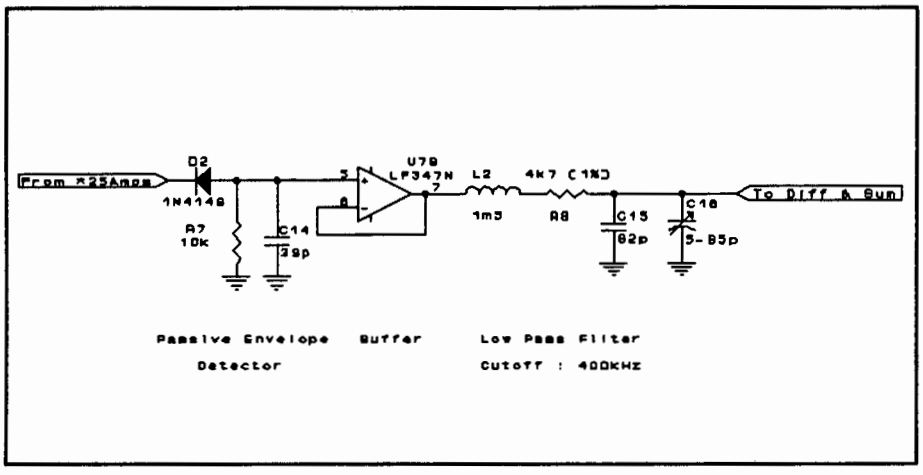


FIGURE 3.7: ASYNCHRONOUS DEMODULATION CIRCUIT

The Butterworth filter is used to remove the ripple after demodulation. Appendix F gives the frequency response of the filter and explains the choice of component values. The cutoff is at $400kHz$ and the variable capacitor is used to phase match the two amplifier stages before the differential summer. Sensitivity of phase with frequency is most significant near the cutoff frequency (Appendix F) and 1% resistors are used to minimise variation. The Q of the circuit is dominated by the larger series resistance and consequently Low Q ferrite in-line inductors are sufficient.

The load impedance of the envelope detector has to be greater than $10K\Omega$ so that the RC time constant is unchanged. Consequently, it cannot be directly coupled to the low pass filter stage and an opamp follower provides the high input, low output impedance.

Frequency response

A calibrated source was used to measure the frequency response of the transimpedance front end. The optical subcarrier frequency was varied from $600kHz$ – $10MHz$ and the output voltage measured. Figure 3.8 shows the response and the $3dB$ points at $1.1MHz$ and $3MHz$ can be observed. In addition, the rolloffs that flank either side of the passband are second order ($12dB$/octave). These results are as predicted and indicate the system is operating correctly.

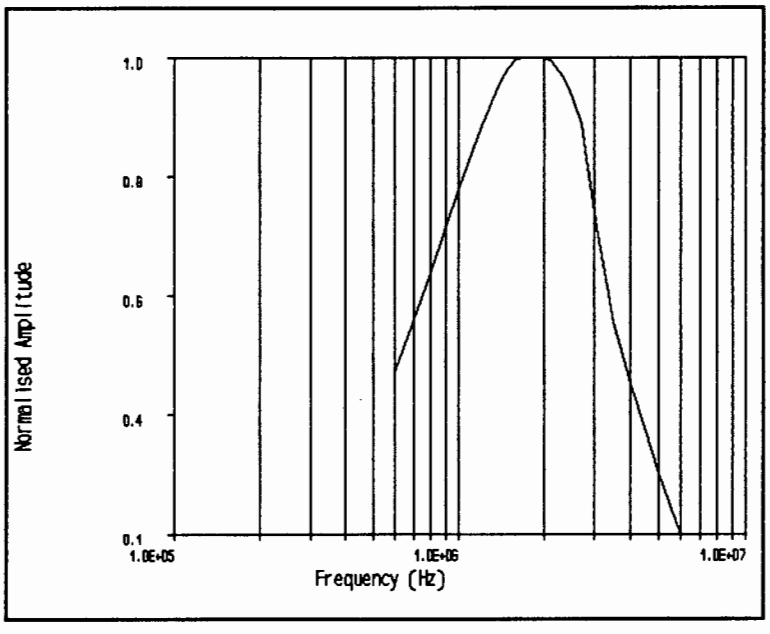


FIGURE 3.8: FREQUENCY RESPONSE OF THE RECEIVER SYSTEM

3.3.2 The differential amplifier and summer circuitry

The demodulated outputs of the two front end amplification stages are subtracted and added. The differential signal is used to determine track imperfections and the summer output is used to maintain constant reflected optical power. Summer output is also used to record the larger track imperfections ($\geq 3mm$).

Differential amplifier

The differential amplifier is shown in Fig 3.9 [17]. Inputs are connected to the demodulated outputs of the two front end stages and in order to achieve a high common mode rejection ratio, and accurate phase matching, 1% resistors are used. The 50Ω trimmer potentiometer is used to minimise the common mode signal.

The common mode rejection ratio (CMRR) of a differential amplifier is a measure of how well the common mode signal is rejected. Ideally $CMRR = \infty$ but is finite for real devices and is defined as the ratio of differential to common mode gain. In general CMRR

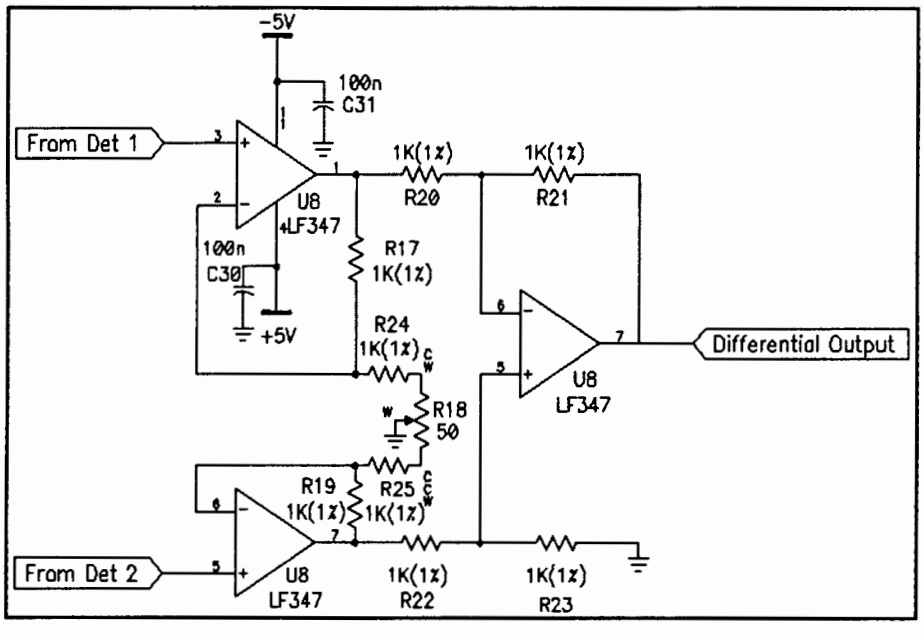


FIGURE 3.9: CIRCUIT DIAGRAM OF THE DIFFERENTIAL AMPLIFIER

of the circuit in Fig 3.9 decreases with frequency and its performance at $400kHz$ is worst. The measured CMRR at $400kHz$ is $36dB$ which is less than the 8 bit quantisation of the output signal.

Summing amplifier

The summing amplifier in Figure 3.10 adds the outputs of the two demodulators. The potentiometers prescale the inputs identically and prevent the output amplifier from becoming saturated.

A large summer input impedance is required to prevent loading the demodulator output stages. The $100K\Omega$ potentiometers are sufficient and prescaled outputs are buffered before the inverting summer.

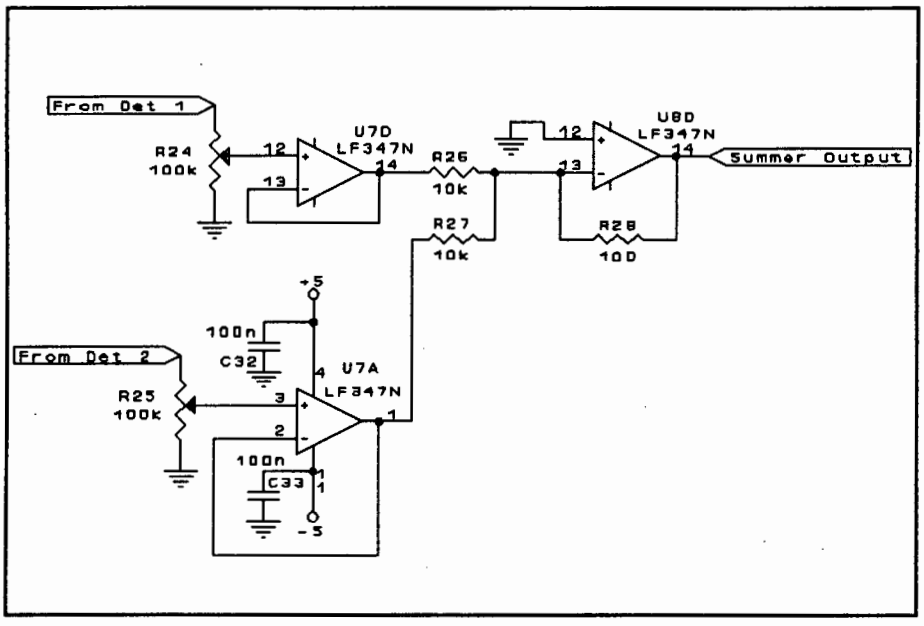


FIGURE 3.10: CIRCUIT DIAGRAM OF THE SUMMING AMPLIFIER

3.3.3 Configuring the optimum receiver

The receiver system must be fine-tuned for optimum results in the performance tests that follow. This involves tuning the post-demodulation low pass filters for identical phase, and minimising the differential amplifier's common mode signal by adjusting the $2K\Omega$ potentiometer.

The common mode rejection ratio is least when a $400kHz$ common mode signal is applied to the differential amplifier. Consequently a common mode $400kHz$ sine wave was applied, and the $2K\Omega$ trimmer potentiometer adjusted for minimum output.

Phase tuning of the LRC filters is completed in two parts. The first sets the cutoff of one filter to $400kHz$, and the second tunes the other to the identical phase. A $400kHz$ sine wave is applied to both low pass filters and Q of one filter is increased by adding a much smaller (270Ω) parallel resistance. The tunable capacitor is then varied for maximum output after which the parallel resistance is removed. The unadjusted capacitance is then varied for minimum differential voltage.

3.4 Noise considerations

It is important to verify that receiver noise is less than or equal to that calculated. Four types of noise, namely : light source, shot, thermal and amplifier noise are associated with the front-end transimpedance amplifiers and are illustrated in Fig 3.11.

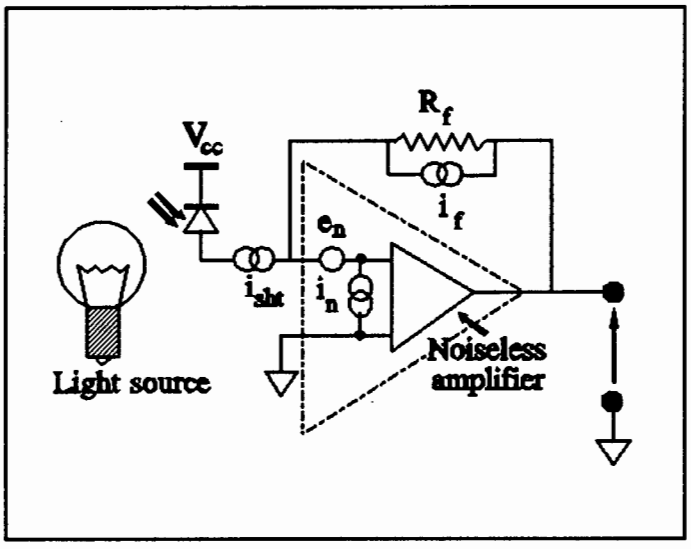


FIGURE 3.11: THE NOISE SOURCES IN THE ELECTRO-OPTIC TRACK MONITOR

3.4.1 Light source noise

The two sources of noise in a lightwave transmitter are quantum and macroscopic variations [4]. Quantum noise is taken into account by the receiver shot noise and macroscopic variations arise due to fluctuations in drive voltage which result in a change of brightness.

Two ways to minimise macroscopic variations are to use a well regulated, heavily decoupled supply and secondly, to use feedback, which dynamically adjusts current in the light source. The feedback technique consists of an auxiliary fully illuminated-photodiode which controls the source current thus maintaining a constant light output. As the variations are low frequency (a fraction of a Hz), the technique does not require fast feedback control.

Minimisation of macroscopic variations is a misnomer though, as the changes in rail reflectance are significantly greater. However, if feedback is used to compensate the changes in surface reflectance, the effect of the macroscopic variations will be eliminated.

3.4.2 Photodetector noise

Photodetector noise consists of shot, carrier generation-recombination and Johnson noise of the bulk resistance associated with the reverse-biased junction. The two latter noises are insignificant and shot noise originating in the pn junction of the photodiode dominates. It has a white spectrum and is composed of signal and background power, and dark current. The shot noise current i_{sht} is given by :

$$i_{sht} = \sqrt{2qB\left(\rho\left(W_{sig} + W_{back}\right) + I_d\right)}$$

The responsivity ρ was measured as $0.573\frac{A}{W}$ in Appendix F. Worst case dark current is $200nA$ (BPX-48 specifications) and background power is assumed to be zero. Assuming the optical signal produces a $2.47V_{rms}$ sine wave at the output of the both amplifiers immediately before demodulation, this corresponds to $424pW$ of optical power. In a $3MHz$ bandwidth, worst case shot noise is therefore $630pA$. In addition, for zero incident signal, shot noise current is maximally $438pA$.

3.4.3 Amplifier noise

Amplifier noise is represented by a series voltage source and shunt current source at the input of a noiseless amplifier. This representation is the notation used by the National Semiconductor Corporation and is equivalent to using two current and voltage sources [13].

The bipolar LM 6361 opamp noise voltage arises from Johnson noise in the base spreading resistance (r_{bb}), shot noise in the intrinsic emitter resistance (r_e) and $\frac{1}{f}$ shot noise from

current through r_{bb}. The vector sum of these voltages gives the net noise voltage $e_n = 17\mu V$ in the $1 - 3MHz$ bandwidth. The $\frac{1}{f}$ noise is insignificant and only the white shot and Johnson sources contribute.

Current noise in the opamp is caused by shot noise in the base current of the input transistors, and $\frac{1}{f}$ noise in r_{bb}. Ignoring the $\frac{1}{f}$ noise, the noise current is $1.73nA$ in the $1 - 3MHz$ bandwidth.

3.4.4 Johnson noise

The last source of noise is white Johnson Noise in the feedback resistance R_f. It is represented as a current i_j given by :

$$i_j = \sqrt{\frac{4kTB}{R_f}}$$

The Johnson noise is therefore $1.64nA$.

3.4.5 Total noise

The individual noise sources must now be combined to calculate the worst case output noise. Total current flowing through feedback resistance R_f is the vector sum of the uncorrelated noise currents :

$$\begin{aligned} {i_{nse}}^2 &= i_{sht}^2 + i_n^2 + i_j^2 \\ &= 3.95 \times 10^{-19} + 2.99 \times 10^{-18} + 2.69 \times 10^{-18} \\ &= 6.08 \times 10^{-18} A^2 \end{aligned}$$

The shot noise current is therefore $2.46nA$ and for zero incident signal, it is $2.39nA$.

The output noise voltage is the quadrature vector sum of the uncorrelated noise sources :

$$\begin{aligned} e_{out}^2 &= e_n^2 + i_{nsig}^2 R_f^2 \\ &= \left(17 \times 10^{-6}\right)^2 + \left(2.39 \times 10^{-9} \times 18000\right)^2 \\ &= 2.14 \times 10^{-9} V^2 \end{aligned}$$

The output noise voltage is therefore $47\mu V$.

Noise of subsequent stages is not significant provided their noise figure is sufficiently low [7] [13]. A small noise figure is achieved by making amplifier input resistance as low as possible and by minimising noise voltage. The opamps can drive a minimum load of 200Ω at full output voltage and input noise of the next stage is $e_n = 14\mu V$ (Appendix G). When added to the $47\mu V$ of the first stage, the vector sum gives a total noise voltage of $49\mu V$, $2\mu V$ larger than before. Noise of successive stages is therefore insignificant compared to the much larger amplified noise of the transimpedance stage.

The voltage and current noise for zero optical signal are $46\mu V$ and $2.39nA$ respectively. The shot noise associated with an optical input signal is therefore insignificant and measurement of noise is performed in the absence of an optical signal.

3.4.6 Measurement of noise

Two experiments need to be performed to ensure the noise performance is as predicted. The first measures the RMS voltage at the output of the last amplifier stage which should be $46 \times 10^{-6} \left(\frac{5.6}{0.22}\right)^2 = 30mV$. Secondly, a relative comparison of the last two amplifier stages should show amplification and no nett increase of noise.

The RMS noise at the output of the last amplifier was measured as $20mV$ which is less than the $30mV$ predicted. The spectrum analyzer was used to compare inverting amplifier input noise with output noise and showed that there was no additional noise due to the amplifier. Appendix G gives the experimental procedure used in the measurements.

3.4.7 Noise equivalent power

Noise equivalent power (NEP) of a detector is the signal power per root Hertz for which signal to noise ratio is unity. The NEP was calculated as $2.6 \times 10^{-12} \frac{W}{\sqrt{Hz}}$(Appendix E). Appendix G gives the procedure for measuring NEP with the result of $3.5 \times 10^{-12} \frac{W}{\sqrt{Hz}}$. The slightly larger measured value is acceptable and the difference is due to measurement

error.

Chapter 4

The Transmitter System

The transmitter is designed so that taking into account the receiver characteristics, a signal to noise ratio of $38dB$ can be achieved for a worst-case rusty rail. This is the maximum achievable in the receiver and is calculated from the Noise Equivalent Power and largest optical signal for which there is no amplifier saturation ($20Log(\frac{2.47}{0.031})$).

The transmitter is discussed in terms of the optical source, modulation technique, source driver and the mechanism for automatic light output control. This is followed by the circuit diagram of the transmitter which is explained in detail. The last section calculates the probability of a $100\mu m$ imperfection being missed when the transmitter power is adjusted for maximum signal to noise ratio and gives a measure of reliability of the system.

Design of the transmitter involves choice of an optical source, modulation technique, source driver and an automatic light output control stage.

4.1 Source

Choice of an optical source is important to achieve the maximum signal to noise ratio.

4.1.1 Requirements

A suitable source must have sufficient response time to be modulated at $2MHz$ and be powerful enough to image a spot on the photocells which gives a $38dB$ signal to noise ratio. The response time must therefore be less than $500ns$. Optical power incident on the photocell is dependent on the emitting area of the source, its power and its beamwidth.

The emitting area of the source is imaged on the photocells and has direct influence on the differential power versus vertical displacement relationship. Curves for spots of radii 700, 350, 100 and $50\mu m$ were calculated (App. A) for the BPX-48 photocell and are shown in Fig 4.1. The analysis assumes that the optical power is constant, and that there is no

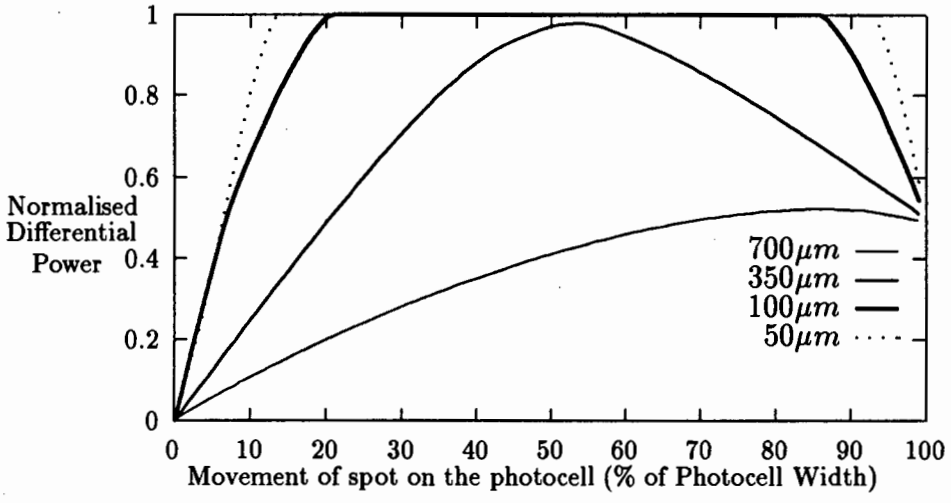


FIGURE 4.1: NORMALISED DIFFERENTIAL POWER FOR SPOT RADII 700, 350, 100 AND $50\mu m$ FOR A BPX-48 PHOTOCELL

magnification or spherical abberation.

The responses are almost linear before they approach the saturation region when the spot falls on one of the photocells. The smaller spots ($\leq 350\mu m$) are characterised by a sharp discontinuity near saturation and the larger spots by a reduction in maximum power. In addition, the range over which vertical displacement is unsaturated is bigger for the larger spots. This means that vertical displacement can be measured over the full range of each of the split photocells for the large spots.

In light of the analysis above, the horizontal resolution requirement is relaxed and spot diameter will be between 100 and $300\mu m$. This gives a suitably linear characteristic, exploits full differential power dynamic range, and achieves sufficient spatial resolution between samples.

Beamwidth represents the optical power distribution of the source and is ideally less than 10° (angle subtended by the lenses). This means that most of the power is concentrated in the lens for highest efficiency.

The optical power required is calculated from the source power necessary to produce maximum signal in the receiver during worst case reflection. The maximum unsaturated voltage in receiver is $2.5V$ and corresponds to $370nW$ of optical power on the photocell. According to Table C1 in Appendix C rusty steel has least reflectance and is 1.6×10^{-3}. A worst case of 10^{-3} is assumed which implies that $370\mu W$of incident power is required.

The worst case optical power requirements to produce $370\mu W$ of incident radiation is calculated assuming a Lambertian source. A source of X Watts is imaged on the rail with a $30mm$ $f/1.7$ lens with 8% transmission loss and the power incident on the rail is

$$\begin{aligned} P_{inc} &= \frac{X}{\pi}\frac{\pi 0.015^2}{0.1^2} \times 0.92 \\ &= 0.021 \times X \end{aligned}$$

which means that X is $19mW$.

4.1.2 LED versus Laser

Light emitting diodes (LEDs) and injection laser diodes (ILDs) are the two alternative sources considered. LEDs are spontaneous emitters and cheap, high irradiance diodes are easily available. Injection laser diodes are stimulated emitters with high temporal and spatial coherence (coherence lengths of order $1m$). The characteristics of ILDs and LEDs are now compared in terms of [19] power, speed, linearity, thermal behaviour, reliability and cost.

Power

The optical output power of LEDs and ILDs is typically of the same order. The laser is however capable of concentrating its power into a small spot unlike the LED. Sterance is a measure of the ability of the source being able to concentrate its power into a small area. Specifically this is the power radiated per unit solid angle from unit area. In general, lasers have a higher sterance than LEDs. This is an important advantage when small spot size ($\leq 1mm$ in diameter) is required.

Speed

The response time of LEDs decreases with increasing radiance and although circuitry can be used to speed up the devices, they are inherently slower than laser diodes. Fastest response times with lasers are achieved when they are biased slightly below their threshold, avoiding the charge time associated with their input capacitance.

Linearity

Linearity in light output with modulating signal is important and LEDs are linear. Lasers on the other hand are non-linear, threshold devices and drive current is not directly proportional to the output intensity. Below threshold they are spontaneous (LED) emitters and above threshold, light output is significantly increased due to lasing. However, above threshold their incremental light output is linearly related to current.

Effect of temperature

The intensity of lasers and LEDs is not constant if temperature is changed. LEDs have a negative (-1%/$^\circ C$) temperature coefficient and the laser threshold changes 1%/$^\circ C$ which can prevent lasing if compensation is not provided.

Reliability

The mean time between failure is largest for light emitting diodes and in general they are more reliable than ILDs (typically 10^5hrs vs 10^4hrs [19]).

Cost

The cost of easily available high irradiance LEDs is significantly less than that of the equivalent laser ($R2$ vs $R100$). On the other hand, higher radiance etched well Burrus and edge emitting LEDs are comparable in cost to lasers. Laser drivers are also more complex and expensive than those of LEDs.

4.1.3 Choice of a suitable source

The performance of three sources; a Siemens SFH487P $20mW$ infra-red LED, a Telefunken $4cd$ visible red LED and an infra-red $5mW$ semiconductor laser diode were compared.

Siemens IR LED

The Siemens LED has a back reflector which makes the effective emitter diameter greater than $1mm$. As this cannot be corrected optically, an optical mask was constructed from two razor blades separated by $200\mu m$ using a feeler gauge. This results in an effective line source.

The power reduction associated with the line source was calculated assuming the source radiates uniformly. For the $5mm$ LED, the active area is $19.6mm^2$ and for the line source it is $0.3mm^2$. This implies that the line source is $\frac{1}{65}$ of the source power. The line source power is therefore $308\mu W$ which is too low and this was verified by measurement.

Telefunken red LED

This LED has the same problem as the Siemens diode in that its source area is too large. By following the procedure above, a $300\mu m$ line source can be made but is inadequate in terms of the power required.

IR laser diode

The last alternative is the BLD101U infra-red $5mW$ laser diode. Laser diodes are generally not Lambertian and their polar diagrams are non symmetric. The far field polar diagrams are characterised by the transverse and lateral Full Width to Half Maximum amplitudes (FWHMs) which are typically 8.5° and 38° [12].

Calculation of the power incident on the rail is non trivial if the polar diagrams are not symmetric. A worst case estimate is to assume the laser has a symmetric profile with a radiation angle of the largest FWHM, which is 38°. In addition, the profile is represented by a $cos^{14}\theta$ distribution which has the same FWHM. The on axis pointance I_o is calculated from

$$\begin{aligned} I_o &= \frac{X(14+1)}{2\pi} \\ &= 2.39X \end{aligned}$$

where X is the source power. Assuming the lens is the same as above, the power incident on the rail is

$$\begin{aligned} W_{lens} &= I_o \frac{\pi 0.015^2}{0.1^2} \times 0.92 \\ &= 0.155X \end{aligned}$$

For a $5mW$ laser this is $777\mu W$ which is double the $390\mu W$ requirement.

The analysis so far assumes that the laser emitter is small enough for a $200\mu m$ diameter spot. Different lateral and transverse intensity profiles suggest that the emitting area is rectangular. This diode has only one transverse mode and the transverse width is a few tenths of a micrometer [12]. Lateral lengths are of order of a few hundred micrometers and if the laser is oriented on its side, it meets the source requirements of the transmitter. The BLD101U laser is therefore the source used in this thesis.

4.2 Modulation

4.2.1 Requirements

The techniques for modulating a laser at $2MHz$ must be evaluated so that the most efficient tradeoff between power output and circuit complexity can be found. The higher the modulation index, the greater the effective light output power. A modulation depth exceeding 80% [12] is readily possible, and will be a target value.

4.2.2 Modulation Techniques

Intensity modulation of the optical carrier is achieved by varying the laser current with the modulating signal. This signal need not be a sine wave and can be in the form of repetitive pulses. The harmonics associated with the pulses are attenuated by the low pass filtering in the receiver so its effective output is a sine wave. Analogue and digital modulation techniques will therefore be investigated.

Analogue modulation is achieved by varying the laser current sinusoidally at $2MHz$. This can be done directly or by transformer coupling via a tuned circuit [18]. The analogue techniques involve more complicated hardware in general and are subject to the nonlinear light output versus light output and temperature dependence characteristic discussed earlier.

Digital modulation is achieved by periodically turning the laser fully on and off at $2MHz$. If the peak optical power is increased with a decrease in duty cycle so that average power is kept constant, the harmonic energy distribution can be optimised by choice of the appropriate duty cycle. High modulation indexes are guaranteed and the light output is less affected by the nonlinear light output with drive current characteristic .

The simplicity and adequately low harmonic energy (3% for square wave modulation) associated with the technique make digital modulation suitable. It is necessary to calculate the optimum duty cycle which is a tradeoff between harmonic distortion, and cost and

complexity of the circuitry. The peak current in the pulse is assumed to increase proportionally with a decrease in duty cycle and a representation of the waveform is shown in Fig 4.2. The power associated with this current pulse is calculated assuming a 1Ω

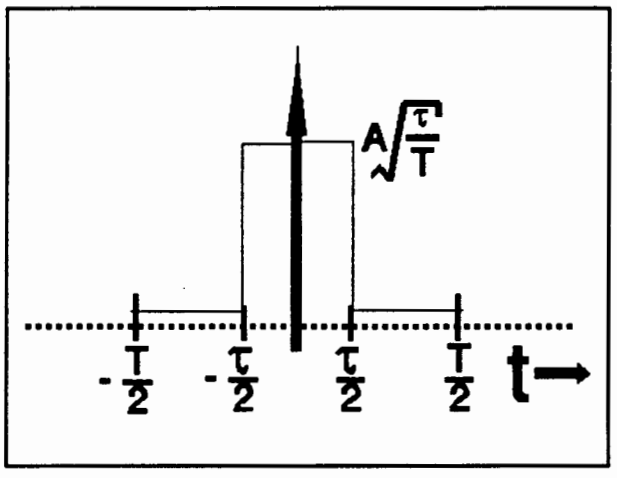


FIGURE 4.2: REPRESENTATION OF THE CURRENT PULSE IN THE LASER

resistance and is

$$\begin{aligned} P &= \frac{1}{T}\int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \left(A\sqrt{\frac{T}{\tau}}\right) dt \\ &= \frac{1}{T}\frac{T}{\tau}A^2\tau \\ &= A^2 Joules \end{aligned}$$

This means that the power in the pulse is independent of duty cycle and it is constant.

The distribution of power in the harmonics is a function of the duty cycle and the Fourier coefficient of the n'th harmonic is calculated from [7]

$$F_n(w_o) = A\sqrt{\frac{\tau}{T}} sinc \frac{w_o n\tau}{2}$$

where $sinc(X) \equiv \frac{\sin X}{X}$. Pulse power is the sum of the squares of the Fourier coefficients (Parseval's theorem [7])

$$P = \sum_{\tau=-\infty}^{\infty} A^2 \frac{\tau}{T} sinc^2 \frac{w_o n\tau}{2}$$

and a relative measure of power contained in the n'th harmonic is thus

$$| F_n |^2 = x sinc^2(n\pi x) \tag{4.1}$$

where $x = \frac{\tau}{T}$ is the duty cycle.

The function in Equation 4.1 is plotted for first, second and third harmonics as a function of duty cycle in Fig 4.3.

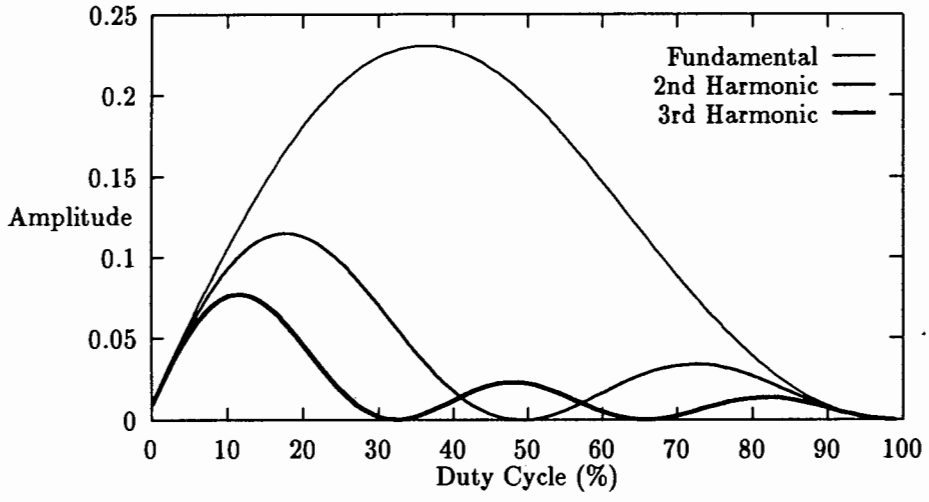


FIGURE 4.3: RELATIVE POWER DISTRIBUTION FOR FIRST, SECOND AND THIRD HARMONICS AS A FUNCTION OF DUTY CYCLE x

The duty cycle, x, for which there is maximum fundamental can be calculated by solving $\frac{d}{dx}F_1 = 0$ which simplifies to

$$\begin{aligned} \tan(\pi x) &= 2\pi x \\ x &= \frac{67}{180} \\ &= 0.37 \end{aligned}$$

This is observed in Fig 4.3 with the smaller second and third harmonics contributing 4% harmonic distortion. It is assumed that only the second and third harmonic components need be considered due to the receiver filtering.

Additional circuitry must be added to achieve a 37% duty cycle. However, comparison with a 50% duty cycle indicates that the performance improvement is insignificant. Apart

from the slightly less (3% versus 4%) second and third harmonic distortion, the marginally less (13%) fundamental power can be compensated by slightly increasing laser current. As a result, a duty cycle between 37 and 50% is acceptable.

Modulation index is the ratio of the average to peak amplitude of the modulating signal. In this context it is a measure of how well the dynamic range of the intensity is exploited. A modulation index of 100% is achieved by turning the laser fully on and off. This depends on the modulation circuitry and in practise, a value close to 100% is acceptable.

4.3 Laser driver

The laser driver establishes a modulating current in the laser diode. As laser diodes are sensitive to temperature, aging and nonlinearities in the light output versus drive current characteristic, the driver must compensate for these.

In order to facilitate dynamic control of light output, a monitor photodiode is integrated onto the laser. The photodiode current is a linear representation of the optical output power and can therefore be used to linearize the light output. In essence, a feedback loop is set up to dynamically adjust the current in the laser so that the light output is independent of changes in temperature and aging.

Digital modulation is used. Analogue laser drivers are not considered since their circuitry is more complex. Digital drivers basically consist of a biased laser diode in a feedback loop with the monitor photodiode. Modulation is coupled into the current path of the laser and is not compensated by the slower feedback loop. As a result, the laser diode is modulated while the slower feedback loop maintains constant average light output.

Modulation is coupled into the feedback circuit directly or by shunting current away from the laser diode. Direct drive of the laser is accomplished with a series transistor which switches the current on or off. The shunt configuration [19] consists of a transistor connected in parallel with the laser diode. By switching it on, the current is shorted away from the laser, and when off, all the current flows through the laser.

The direct or series driver is able to achieve high modulation indexes and the light output is in phase with the modulation. On the other hand, the light output of the shunt driver is out of phase with the modulation and has slightly lower modulation index as the laser is not turned fully off. Shunt modulation circuitry however has the advantage that it can be added to existing laser driver circuitry. It is this advantage that makes the shunt driver the most suitable.

Both shunt and series drivers use transistors to divert or switch the laser current. The FET or bipolar transistors must be suitably fast and capable of facilitating high modulation indexes. Bipolar transistors are significantly faster than their FET counterparts but require large (mA) turn on currents and have a positive temperature coefficient. FETs, although slower, require no (pA) input current, have small on resistance, and a negative temperature coefficient. On resistance is related to FET speed and the faster the switching speed, the larger the on resistance.

A comparison was made between a bipolar 2N2222 transistor and a power MOSFET VN2222 shunt driver for modulation at $2MHz$ and a laser diode current of $100mA$. Both transistors are fast enough, and the on voltage of the bipolar transistor is $0.2V$ as opposed to 0.75 for the FET. This means that the modulation index is somewhat less for the FET.

4.4 Automatic light output control

As the scanning system moves over the rail the reflectance changes and total power on the photocells varies. This is undesirable as changes in reflectance are interpreted as imperfections. A mechanism for keeping the power on the photocells constant is therefore necessary and is called automatic light output control (ALOC). It is achieved by dynamically changing the source power to cater for an increase or decrease in reflected radiation.

The main reason for introducing ALOC is to adjust to the local conditions of the railway track. It is assumed that once the system has adjusted itself to the rail surface there will be insignificant changes in reflectance. The feedback system's time constant is therefore large in relation to the modulation period, and is of order of milliseconds.

Before explaining the feedback system, the dynamic range required by the source for constant light output must be calculated. This is computed from the reflectances of the best and worst case surfaces. Best and worst case reflectances were measured as 0.6 and 1.62×10^{-3} (see App C) and are assumed to be conservatively 1 and 10^{-3}. This means the dynamic range is $28dB$ (1 : 617) which corresponds to light output powers varying from $5\mu W$ to $3mW$.

The sum of optical power falling on both photocells in the receiver is a direct measurement of the power in the incident spot. There are three possible ways to facilitate automatic light output control :

- Vary the gain of the receiver amplifiers.
- Use optical filters to provide different degrees of attenuation.
- Increase or decrease the average laser current.

The first requires dynamic adjustment of the amplifier feedback resistances. As this includes the transimpedance amplifier, the receiver NEP will change which is undesirable. The use of variable optical attenuator filters is mechanically cumbersome and is therefore also not suitable. The best solution is to dynamically change the transmitted power by varying the average laser current.

The laser light output has infinite dynamic range but the characteristic is a nonlinear function of the drive current as shown in Fig 4.4. The sharp transition between spontaneous and stimulated emission is clearly visible at the lasing threshold and is characterised by a rapid increase in optical output power with a small change in current. As the dynamic range requirement requires the system to operate in this region, the feedback system must compensate for this nonlinearity.

The most basic feedback system is illustrated in Figure 4.5 and u is the setpoint, y the output signal, e the error, A the amplification and g the plant model. The relationship between output y, and input u is

$$\frac{y}{u} = \frac{Ag}{Ag + 1}$$

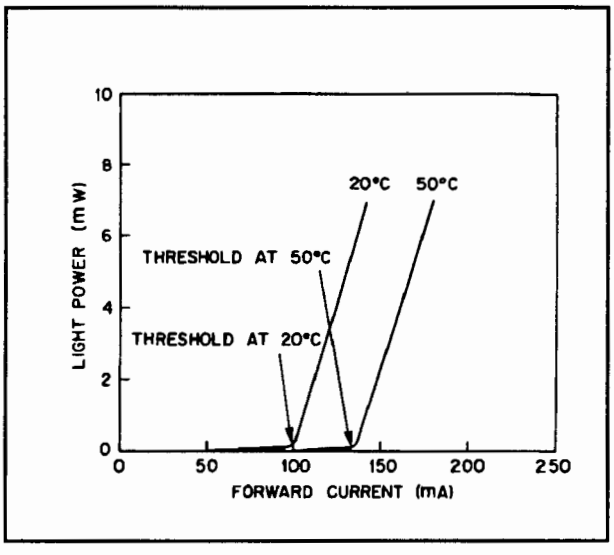


FIGURE 4.4: LIGHT OUTPUT POWER VS FORWARD CURRENT IN AN INJECTION LASER DIODE

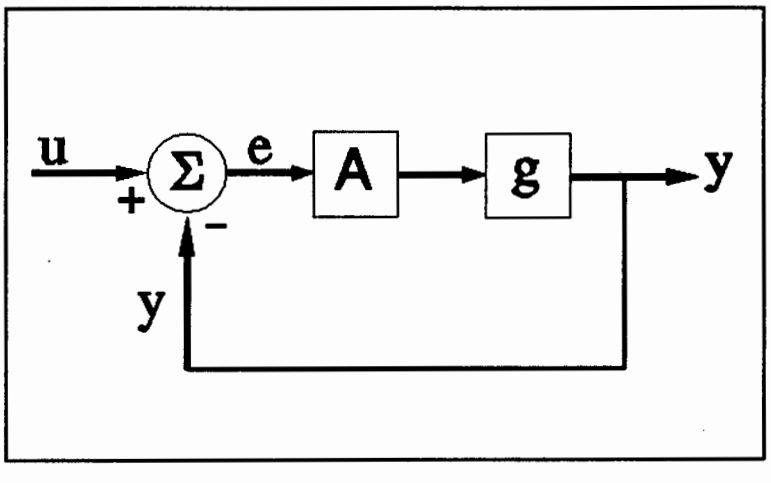


FIGURE 4.5: FEEDBACK SYSTEM FOR AUTOMATIC GAIN CONTROL

and therefore A should be made as large as possible so $y \approx u$.

The feedback must not compensate for changes in intensity due to the modulation and in addition, it does not have to be fast as rapid changes in reflectance are not expected. In general, a rail being scanned is assumed to have constant reflectance throughout its length. A time constant of order $100ms$ is assumed to be sufficient for the feedback loop.

4.5 The transmitter circuitry

This section explains the design and operation of the transmitter circuitry which is illustrated in Fig4.6. In order to explain its operation it is divided into three parts; the

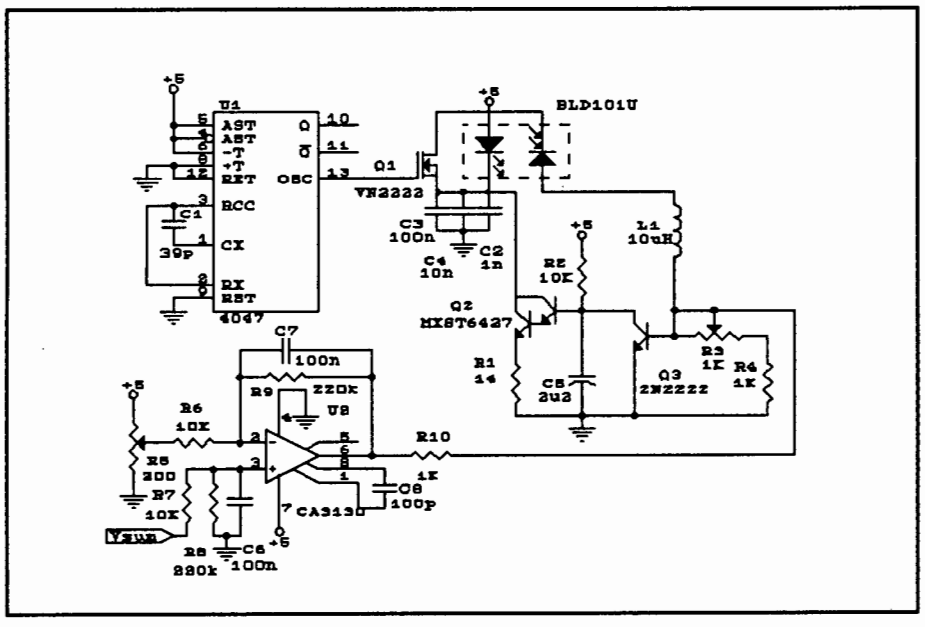


FIGURE 4.6: CIRCUIT DIAGRAM OF THE TRANSMITTER

temperature compensated laser driver, the modulation circuit and the automatic light output control.

4.5.1 The temperature compensated laser driver

The temperature compensated laser driver excludes the CD4047, VN2222 FET and CA3130 opamp. Initially no current flows in the laser and monitor diode of the BLD101U. The BOF01 transistor is off allowing the $2.2\mu F$ capacitor (C5) to charge. When base voltage of the Darlington transistor is above $1.2V$, a current flows in the 14Ω resistor and laser. Current in the monitor diode increases with optical output power until the base voltage of the BOF01 is $0.6V$ which turns it on. The capacitor (C5) is then discharged and a steady state light output is reached. Output power may be increased by decreasing the effective resistance of the $1K\Omega$ potentiometer (R3) and current in the laser is maximally $80mA$. The ceramic capacitors (C2 - C4) connected to the laser cathode prevent short-duration transient current spikes in the diode.

The feedback reduces temperature dependence of the circuit which is now dependent on the base emitter voltage of the transistors. Base emitter voltage drops $2.1mV/°C$ and the current in the laser is slightly lower for an increase in temperature.

4.5.2 Modulation circuitry

The modulation circuit consists of the VN2222A FET and the CD4047 astable. A $2MHz$, 45% duty cycle square wave from the CD4047 turns the VN2222A power mosfet on and off. The feedback circuit discussed above does not compensate for the modulation and its time constant ($22ms$) is set by the charge time of the $2.2\mu F$ capacitor (C5) and resistance R2.

A FET was chosen in preference to a bipolar transistor to prevent loading of the CD4047 and its on resistance is 7.5Ω. The voltage across the FET when on is therefore $0.6V$ when sinking $80mA$ and the minimum source–gate voltage ($2.5V$ max) is satisfied by the $5V$ output of the CD4047 monostable. Rise and fall times are limited by the gate capacitance of the FET which is $60pF$. The output source and sink current (I_{OH} and I_{OL}) of the CD4047 is $0.36mA$ which means the rise (fall) time ($\frac{C}{I_{OH}}$) is $417ns$.

Modulation index of the circuit above was measured and is approximately 95%.

4.5.3 Automatic light output control

Light output power may be varied by changing the current in the $1K\Omega$ potentiometer (R3) manually or electrically. Automatic light output control is achieved by varying the current in R3 with the CA3130 operational amplifier. The difference between the receiver summer voltage and a setpoint constitute the error voltage. The sense is such that when the sum is larger than the setpoint, the output error signal is positive which discharges the $2.2\mu F$ capacitor (C5) turning the laser off. The 200Ω setpoint potentiometer (R5) is adjusted for maximum unsaturated voltage when the spot falls on one photocell only. It was however not possible to get a nonsaturated output as the setpoint is too sensitive, and the pre demodulation signal in the receiver is slightly clipped.

Once again the feedback time constant was made slow ($140ms$) and the gain was set to 22. The CA3130 operational amplifier was used as it's output is able to swing within $50mV$ of the supplies allowing the highest possible current drive.

4.6 Probability of error

The aim of this section is to establish the reliability of the displacement measuring system. This is done by calculating the probability of a $100\mu m$ imperfection being missed.

The following assumptions are made about the system :

- Input noise is Gaussian and bandlimited to $2MHz$. The RMS noise at the input of each envelope detector is $30mV$.
- Input noise does not increase significantly with optical power i.e. contribution of shot noise is insignificant.
- The noise sources in each of the photodetector amplifier stages are statistically independent.
- A $100\mu m$ imperfection results in a $2V$ signal at one input and $0.5V$ at the other input of the differential amplifier in the receiver (see Chapter 5).

- The localised changes in height and reflectance of the rail surface give rise to peak deviations of $160mV$ from the mean differential voltage (see Chapter 5). The detection threshold of a $100\mu m$ imperfection is therefore $160mV$.
- The probability is evaluated for only one observation of the imperfection and the effect of averaging is not considered.

The probability density function of the noise at the output of the system must be calculated to compute probability. The output random variable, Z, with mean M_z, is the difference between two statistically independent random variables, X and Y, with means M_x and M_y. The signal and additive Gaussian noise entering the envelope detectors is transformed into the random variables X and Y. X and Y thus have identical probability functions but different means. The probability density function of X is as follows [20] :

$$Pdf_x(X) = \frac{X}{\sigma} \times e^{-\frac{X^2+M_x^2}{2\sigma}} \times I_0\left(\frac{XM_x}{\sigma}\right)$$

σ is the RMS input noise voltage ($30mV$) and I_0 is the modified Bessel function of zero order. The probability that X lies between $V1$ and $V2$ is therefore :

$$P_x(V2 > X > V1) = \int_{V1}^{V2} Pdf_x(X)dX \tag{4.2}$$

Equation 4.2 cannot be evaluated analytically and has to computed numerically.

The output probability density function of the receiver is calculated from the statistically independent variables X and Y and is [7] :

$$Pdf_z(X,Y) = Pdf_x(X) \times Pdf_y(Y) \tag{4.3}$$

The aim is to express Pdf_z in terms of one random variable Z and by making the substitution

$$X = Z + Y$$

the probability of Z being less than V_T is calculated from [16]

$$P(Z < V_T) = \int_{-\infty}^{V_T} \int_{0}^{\infty} Pdf_x(Z+Y) \times Pdf_y(Y)dYdZ$$

This expression is difficult to compute and a simplification is to impose constraints on both X and Y i.e. use Equation 4.3. The mean voltages M_x and M_y are 2 and $0.5V$

respectively and RMS noise, σ, is $30mV$. It is assumed that the worst case detection threshold of a $100\mu m$ imperfection is $160mV$ for X and Y. The probability of a $100\mu m$ imperfection being missed is

$$P = P_x(1.8 < X < 2.2)P_y(Y > 1.8) + P_x(X < 0.5)P_y(0.3 < Y < 0.7) + P_x(X < 0.5)P_y(Y > 1.8) \quad (4.4)$$

This is more specifically the probability that

- X is within $0.2V$ of its mean and Y is greater than $1.8V$ i.e. Y is in error.
- X is less than $0.5V$ and Y is within $0.2V$ of its mean i.e. X is in error.
- X is less than $0.5V$ and Y is greater than $1.8V$ i.e. both X and Y are in error.

Each of the probabilities was evaluated in Mathcad using Equation 4.2 (see Appendix E) and Equation 4.4 simplifies to

$$\begin{aligned} P &= 0.75 \times 6.6 \times 10^{-14} + 2.6 \times 10^{-18} \times 0.76 + 6.6 \times 10^{-14} \times 2.6 \times 10^{-18} \\ &= 5 \times 10^{-14} \end{aligned}$$

The probability of missing a $100\mu m$ imperfection is therefore negligible and the system is reliable.

Chapter 5

System feasibility

This chapter establishes feasibility of the electro-optic track monitor. The receiver and transmitter developed in the last two chapters are integrated and feasibility is determined by performing a series of tests. Not only do the results indicate whether specifications are met, but they also give a measure of system performance.

Before discussing the system tests it is necessary to integrate the transmitter and receiver modules into a test bed. This involves configuring the optimum optical system and calibrating the system for the tests.

Static and dynamic tests are used to evaluate system performance. The static tests show that the system is capable of measuring vertical impressions and depressions on the surface of the track. They also give the relationship between vertical displacement and change in differential voltage, and the minimum resolvable displacement.

Dynamic tests are performed to show the system is capable of measuring imperfections when moving over rail. Imperfections of known height and width are placed on the moving surface and are measured with the optical system. Detection of an imperfection and accuracy of measurement of its height is compared with actual values and the ability of detecting fine cracks of a known width is also measured.

The dynamic tests require high speed data acquisition unlike the static measurements

which are made with a voltmeter. Available hardware is not capable of the $800kHz$ sampling requirement and therefore a data capture unit is designed and built. Information is accumulated in on board RAM and is transferred at the end of sampling to an IBM personal computer. The data can then be stored and processed with available software.

5.1 The electro-optic track monitoring system

The transmitter and receiver discussed earlier are integrated into a system which is capable of measuring the surface profile of a railway track. This section discusses the system, focusing of the receiver and transmitter, and establishes an optical test bed from which measurements are made.

5.1.1 The system

The electro-optic track monitoring system is shown in Figure 5.1. The receiver and

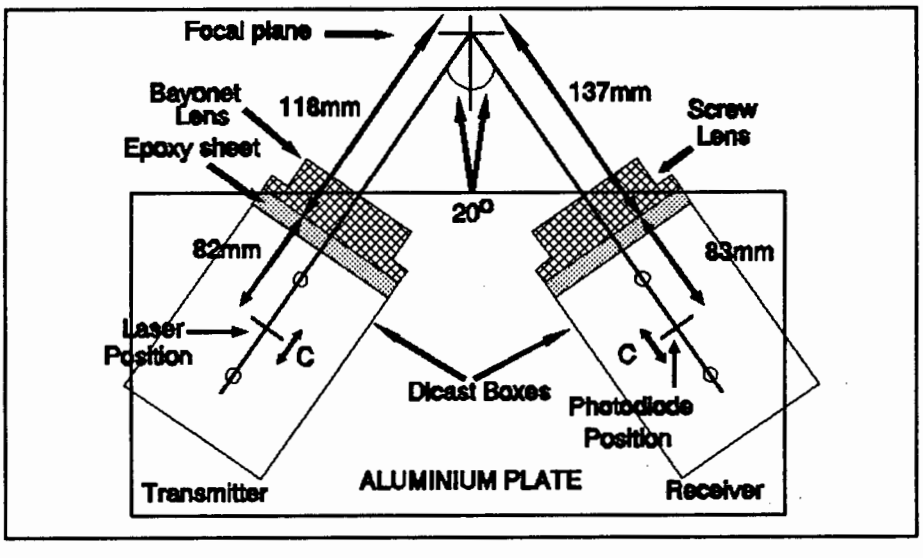


FIGURE 5.1: THE ELECTRO-OPTIC TRACK MONITOR

transmitter boxes are mounted rigidly on a $5mm$ thick aluminium plate and the dicast

boxes are used for electrical shielding. Standard camera ($f = 55mm$, $50mm$) lenses with f-number of 1.7 - 22 are mounted on $10mm$ thick epoxy blocks with trued up surfaces. The clearance between the reflective surface and the unit is $8cm$ and the angle between receiver and transmitter axes is 40^o for point–to–point focusing.

The laser diode and split photocell were positioned on axis with the lenses and no proviso was made for adjustments in the lens plane. Focus is adjusted coarsely by moving the image points along the c axes (Fig 5.1), and finely, by changing the adjustable lens focus. The laser and photodiode are mounted on epoxy angle pieces which sit in a groove giving movement along the c axes.

The transmitter consists of a semiconductor laser mounted in a brass ferrel with two selfoc gradient index lenses. The selfoc lenses are used for point to point focusing and the laser spot is imaged on the face of the outer selfoc lens.

The receiver electronics including the photocell is mounted on a printed circuit board to minimise the noise and avoid instability. This is clamped to the epoxy angle piece so the photocell is on axis with the centre of the $50mm$ bayonet lens.

5.1.2 Focusing and configuring the system

Power was applied to the transmitter, and the spot imaged on a piece of white paper was viewed with an infra-red camera. Although the laser is infra-red, a red spot was observed with the naked eye. By moving the paper backwards and forwards the spot size increases as expected. The coarse focus (along c in Fig 5.1) was then adjusted for the approximate smallest spot size.

A point source was placed at the receiver image point and the coarse focus adjusted so the image of the spot on the photocell was smallest. The receiver was then turned on without the light output control connected and the amplifier output voltages measured in the absence of a reflective surface. These were zero which proves that there is no electrical coupling between receiver and transmitter.

A highly reflective surface saturates the amplifiers and a rusty sheet only just saturates

the system. Light output control is introduced by feeding the summing output of the receiver back to the transmitter. A shiny reflective surface was positioned so the spot fell onto only one photocell. The transmitter setpoint was then adjusted for maximum unsaturated signal.

The transmitter setpoint is very sensitive and the system cannot be adjusted so there is no saturation. The sine waves are slightly clipped when the spot falls exclusively on one photocell and there is no saturation when light falls equally on both photocells. The linear range of vertical displacement with change in differential output voltage is therefore reduced as a result. Different reflective surfaces were used to verify that the problem is due to setpoint tracking and the degree of saturation is consistent for different reflective surfaces.

5.1.3 A test setup

The optical system above was fixed to an optical bench as shown in Figure 5.2 and the

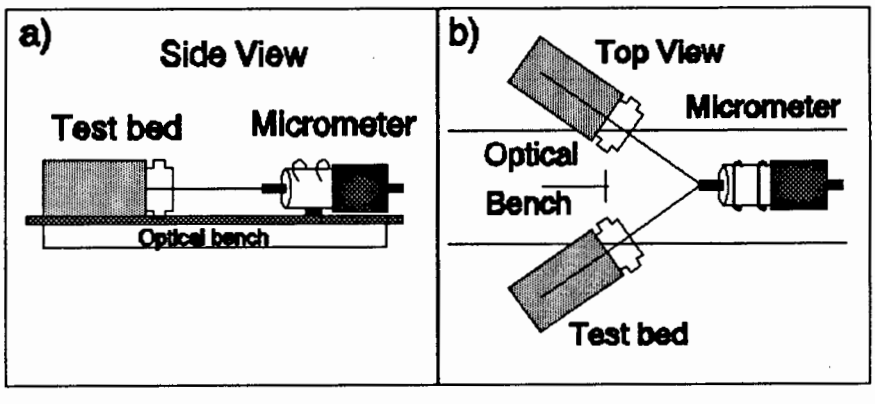


FIGURE 5.2: THE TEST SETUP

face of a micrometer screwgauge is the reflective surface. Displacements resulting in a change in differential voltage are obtained by varying the gauge.

A test was performed to make sure that the spot is adequately small and consists of moving a $200\mu m$ diameter wire across the reflective surface (see Fig 5.3). The focusing of the system was adjusted so that there was an appropriate change in differential power.

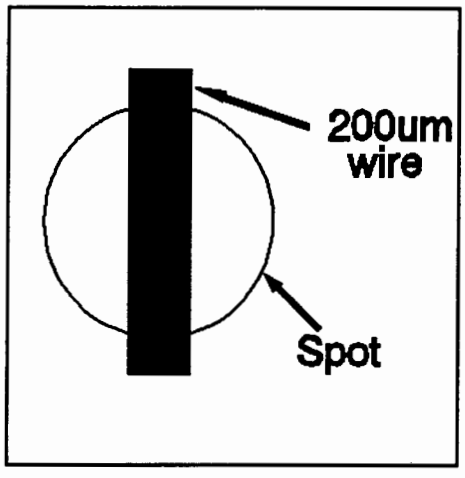


FIGURE 5.3: PRINCIPLE USED TO FOCUS THE SYSTEM FOR A $200 - 300\mu m$ SPOT

5.2 Design of a data capture unit

Data capture is required to accumulate information about the surface being scanned with the electro-optic track monitoring system. In order to establish what is required it is necessary to investigate the static and dynamic system tests that evaluate system performance. Static tests do not require high speed acquisition and a digital voltmeter or oscilloscope is adequate.

Dynamic tests involve measurement of differential voltage when moving over a surface at $40ms^{-1}$ which necessitates high speed data acquisition. In order to specify requirements of the equipment it is necessary to know the acquisition rate and quantisation. The bandwidth of the measurement system is $400kHz$ which implies at least an $800kHz$ sampling rate to avoid aliasing. The greater the sample rate though, the more the data accumulated, and $400kHz$ was chosen as a suitable compromise. This corresponds to a spatial sampling interval of $100\mu m$ when train speed is $40ms^{-1}$ which is small enough to avoid spatial aliasing. Eight bit quantisation is sufficient as system noise is of the order of the least significant bit.

Eight bit, $400kHz$ data loggers are not easily available and Eagle Electric's PC-30D acquisition card is only capable of a $200kHz$ throughput. As a result, an 8 bit data logger with a sampling rate of up to $400kHz$ and with a P.C. interface was designed and built.

The data acquisition unit consists of

- An analogue to digital converter module
- A temporary storage module
- An IBM PC interface and PC software

The analogue to digital converter is responsible for acquisition of data at the sample frequency ($\leq 400kHz$). It is self contained and is compatible with requirements of the digital data compression discussed in the next chapter.

A temporary storage module is required as the IBM PC is not capable of the $400kHz$ throughput of the A/D converter. It is a stand alone state machine which captures data from the A/D converter at $400kHz$ and then transfers it at a slower rate to the PC. Unlike the analogue to digital converter, the state machine is under direct control of the PC.

An IBM PC is the centralised processing and control centre used to initiate sampling, analyze and store data. Data in the temporary storage is downloaded via a parallel interface card and supervisory program which stores it on disc. Available signal processing and statistical software is then used to evaluate measurements.

5.2.1 The analogue to digital converter module

The analogue to digital converter acquires 8 bit data at a selectable sampling frequency of up to $400kHz$. A full specification including the inputs and outputs is given in Fig 5.4. Analogue input voltage is sampled at the frequency selected and output on the data bus $D0$ – $D7$. The clock output signal is used to establish synchronisation and indicates when sampling takes place. Finite conversion time of the analogue to digital converter is represented by the end of conversion signal. In addition to specifications already given, the analogue input voltage range is $\pm 5V$.

Figure 5.5 illustrates the analogue to digital converter circuit diagram. The $1MHz$ crystal oscillator is used in conjunction with the 74HC4040 binary ripple counter to set the

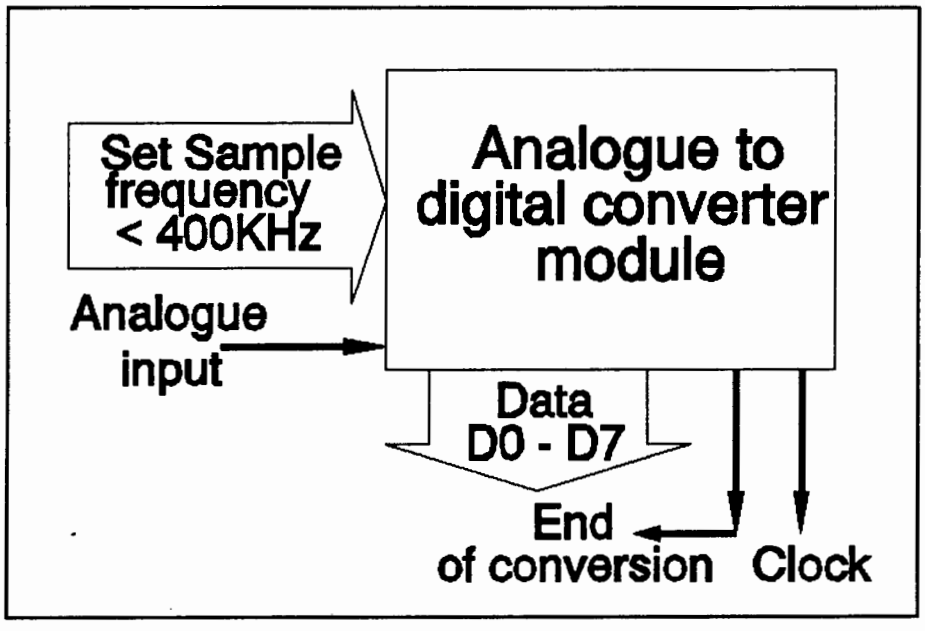


FIGURE 5.4: ANALOGUE TO DIGITAL CONVERTER SPECIFICATION

sampling frequency. A crystal oscillator is used in preference to an LC or RC oscillator as it is simpler and has high stability which is required to establish an accurate time base. The three inverter configuration is used so that sufficient propagation delay guarantees oscillation.

Sampling frequency is a modulo 2 multiple of $1MHz$ and can be one of 12 frequencies between $500kHz$ and $244Hz$. It is jumper selectable and the TTL 74121 monostable (U3) produces a $35ns$ falling edge pulse which initiates a conversion in the A/D converter. The other 74121 (U6) signals the end of a conversion and is low for $105ns$, the duration of a conversion.

National Semiconductor's ADC0820 and Motorola's MC10319 both meet the sampling and quantisation requirements already discussed. The ADC0820 is cheaper but requires an expensive sample and hold circuit and therefore the Motorola MC10319 eight bit flash converter was chosen. Two LM336's provide the positive and negative $\pm 1V$ references for the converter's resistor ladder and the opamp – transistor buffers source the high input current demanded by the resistor ladder. The analogue input voltage ($\pm 3.5V$) is reduced by the potential divider to $\pm 1V$ as required by the converter.

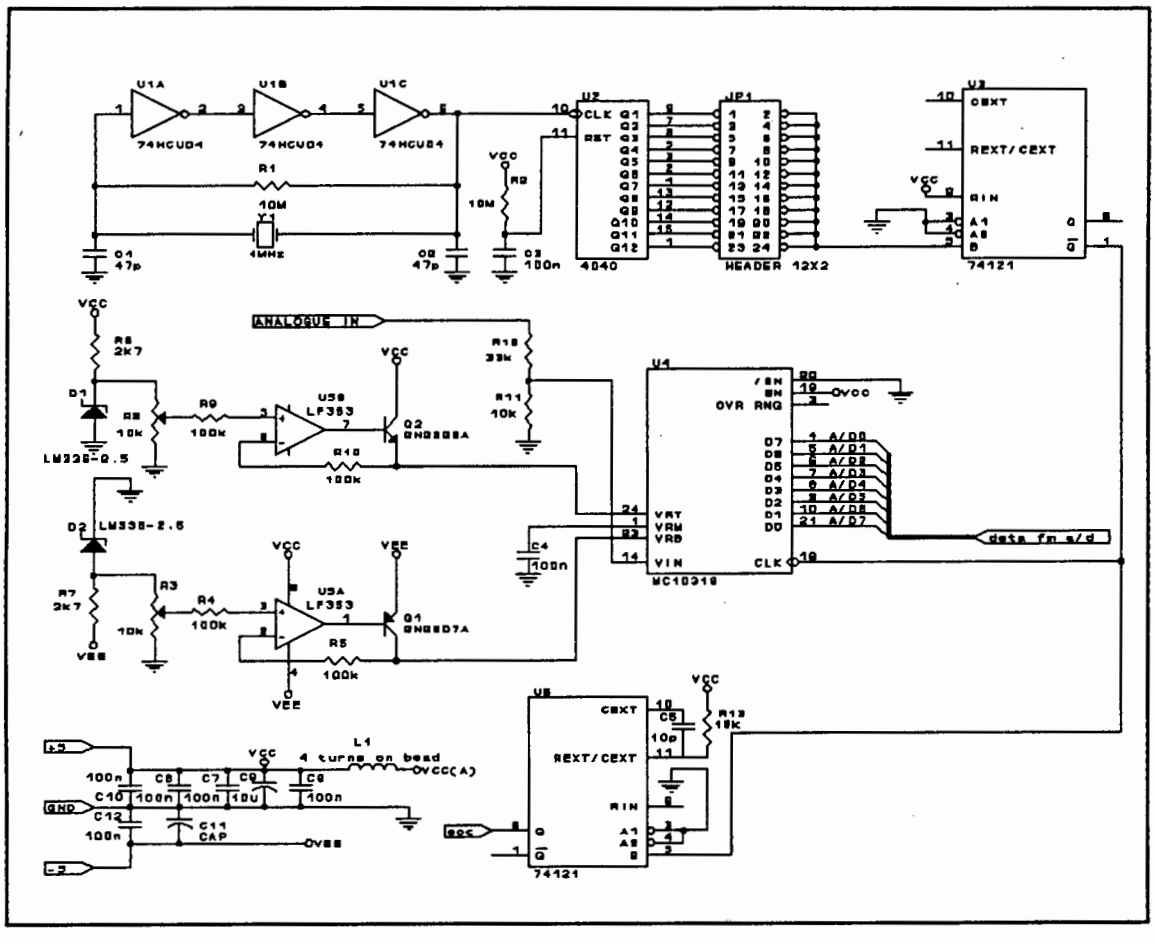


FIGURE 5.5: ANALOGUE TO DIGITAL CONVERTER MODULE

A printed circuit board of Fig 5.5 was made using the Tango PCB software. It was then assembled and tested. Tests revealed the digital outputs correlate to the DC input voltages and therefore that the circuit works.

5.2.2 Temporary storage module

The temporary storage unit accumulates 32 kilobytes of data at the sampling frequency which can then be read at a slower rate by the PC. Figure 5.6 illustrates the block diagram of the system with the associated control signals.

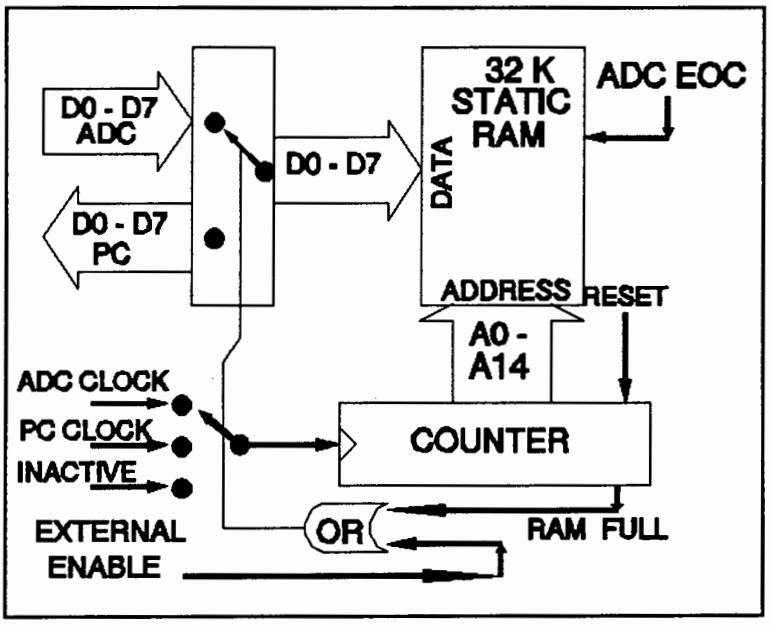


FIGURE 5.6: BLOCK DIAGRAM OF THE DATA CAPTURE UNIT

The External Enable signal is activated and the RAM data bus is switched through to the analogue to digital converter. Sampling occurs at the selected frequency and data is stored after activation of the ADC End of Conversion (EOC) signal. When the address counter indicates the RAM is full, the ADC EOC, Clock and ADC data bus are disabled. The Reset signal initiated by the PC is used to reset the address counter.

The circuit diagram of the temporary storage unit is given in Figure 5.7. The GAL was programmed using ABEL software (Appendix H) and implements the combinatorial logic.

There are three inputs from the ADC module which are the data bus, clock and end of conversion signals. The PC interface consists of two eight bit ports – a data and control port. The control port is made up of the External Enable (EE), reset (RS), PC clock (PClk) and write (WRI) signals, and the data port, of the data bus.

The RAM data bus is either accessed via an omnidirectional buffer (U1) from the ADC, or from the PC data port via a bidirectional, octal tristate buffer (U3). The tristate outputs cannot be enabled simultaneously and either the ADC or PC has access to the data bus. This is controlled by the External Enable (EE) signal which disables PC access when high. In addition, the 74HC245 (U3) is bidirectional and the DIR input allows data

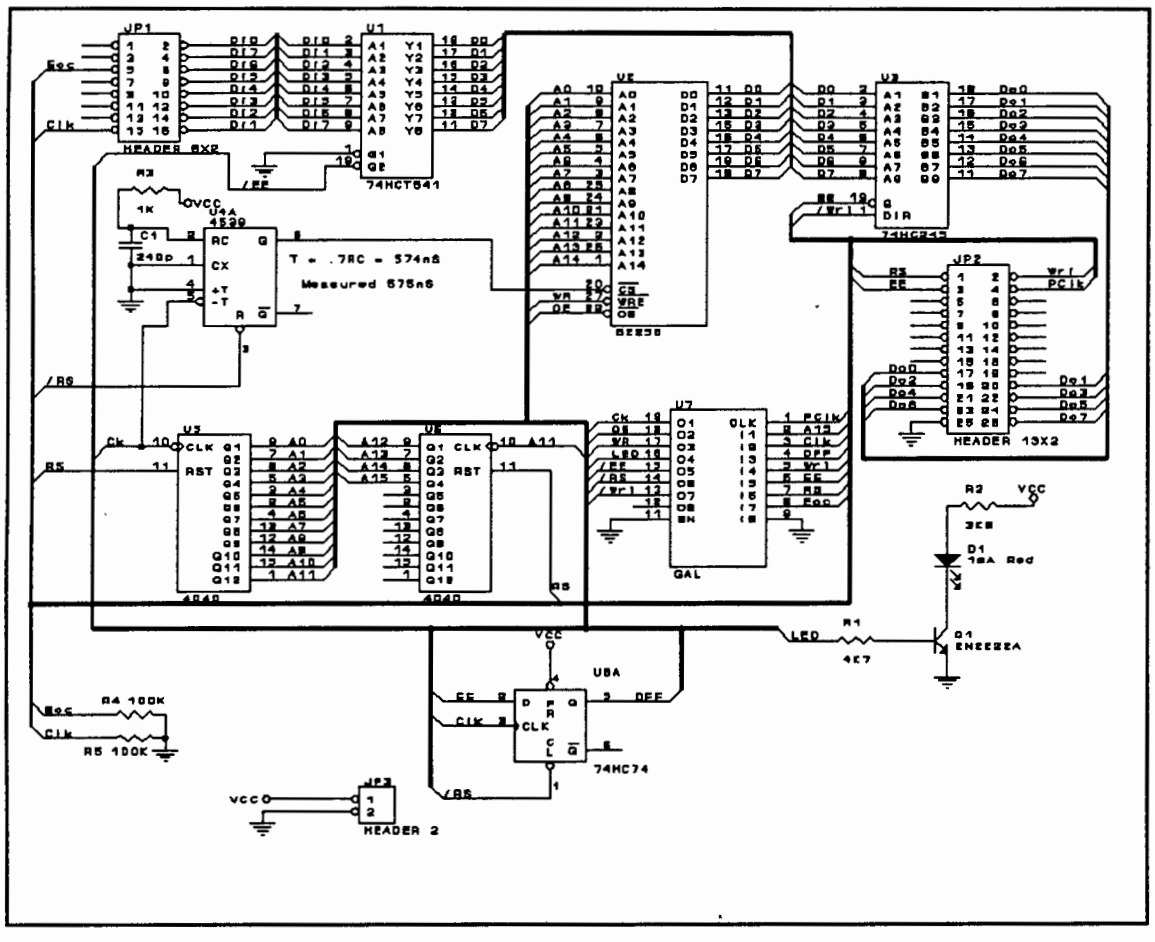


FIGURE 5.7: CIRCUIT DIAGRAM OF THE TEMPORARY STORAGE UNIT

to be written or read from the RAM by the PC when EE is low.

In order to write data into the RAM, the !WRE and !CS signals must be low. The write signal of the RAM is therefore active (low) when EE and EOC are true, or when the PC initiates a data write (EE low and WRI high). The RAM read can only be initiated by the PC, or when !WRE is high and EE is low.

RAM read and write signals are disabled (!CS low) during an address update. A falling edge on the Ck signal produces a $570ns$ pulse in the CD4538 (U4A) during which the address is updated. This is sufficient delay for the $190ns$ propagation delay of the 74HC4040 ripple counters and address settling time ($50ns$) of the RAM.

Addressing of the RAM is sequential and is performed by two 74HC4040 ripple counters (U5 and U6). A PC reset signal (RS) resets the counters to address 0 which is incremented with the Ck strobe. The Ck signal is either activated by the PC clock or external A/D clock and in addition, the ADC clock is only active provided the RAM is not full (A15 low) and DFF is high.

The red, $1mA$ LED indicates when the RAM is being filled by the external A/D converter. Specifically this is when the RAM is not full (A15 low) and EE is high. Storage of samples is initiated by making EE high. To prevent the address being incremented before the first sample is stored, the DFF flag of the 74HC74 D-type flipflop is activated only after this occurs. The initially low DFF is active after EE is high and after there is a rising edge on the Clk signal. The address is therefore only updated if DFF is active.

In summary, the timing diagram of external sampling is shown in Figure 5.8.

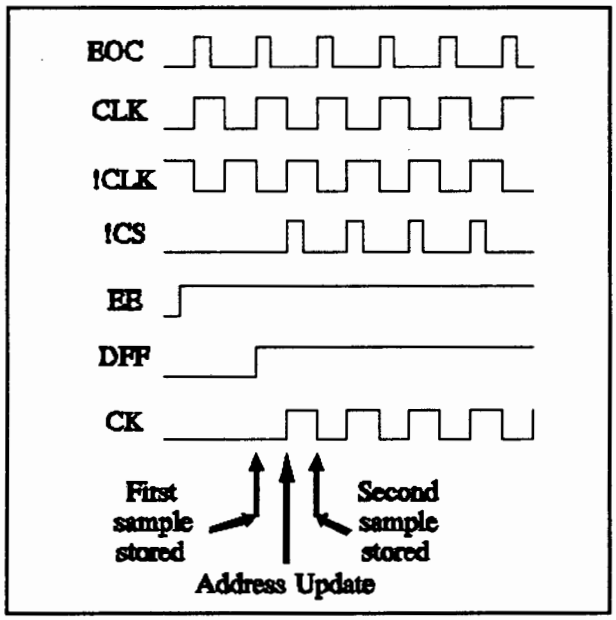


FIGURE 5.8: TIMING DIAGRAM OF EXTERNAL SAMPLING

A printed circuit board was made of the circuit in Fig 5.7 with the Tango PCB and Route (student version) software. The system was then tested in conjunction with the PC interface and software.

5.2.3 PC interfacing and processing utilities

This section is divided into the digital interfacing hardware, software written in Turbo Pascal and tests that were performed to verify the data capture is fully functional.

Hardware

The digital interface provides two eight bit IO mapped addresses in the PC which represent the data and control registers of the temporary storage module. Writing to the RAM from the PC is not necessary and use is made of a PC plug in card with two output ports and one input port. It is addressable as three consecutive IO locations.

The data bus and control port of the temporary storage unit are connected by a $1.5m$ ribbon cable and DB25 connector to the card. The control register is represented by a binary number and each bit is as follows :

7	6	5	4	WRI	EE	RS	PClk

Software

The states that are initiated by the PC to read from temporary storage, write to temporary storage and perform external sampling are given in Table 5.2.3.

The program s.exe (App. I) is used in conjunction with hardware to sample the analogue input. After the temporary storage has been filled, the data is transferred to the PC where it is stored as a binary file on disc. Sampling is initiated by a key press and terminated when the temporary storage is full or when another key is pressed – which ever happens first. The entire RAM contents is then transferred into P.C. memory before it is stored as a block of 32768 bytes in a binary data file, data.dat. A binary file is used as it stores the data most efficiently i.e. the ASCII equivalent is 128000 bytes.

State	Cycle	WRI	EE	RS	PClk
Reading from Memory	1	Off	Off	On	Off
	2	Off	Off	Off	Off
	Repeat	Read	Data	from	Port
	4	Off	Off	Off	On
	5	Off	Off	Off	Off
	Until	32K	samples	are	read
Writing to Memory	1	On	Off	On	Off
	Repeat	write	data	to	port
	3	Off	Off	Off	Off
	4	On	Off	Off	On
	5	On	Off	Off	Off
	Until	all	data	written	
External Sampling	1	Off	Off	On	Off
	2	Off	On	Off	Off
	Repeat				
	Until	Red	LED	is	Off
	5	Off	Off	On	Off

TABLE 5.1: STATES USED TO INITIATE SAMPLING AND TRANSFER DATA TO AND FROM RAM

Testing

The digital interface card was tested by setting various bits on the output port and rereading them with the input port. The s.exe program was then used to test the entire data capture system.

The first test involved initiating sampling and checking the time taken (i.e. the on time of the red LED) to fill the temporary storage (32768 bytes) at a sampling frequency of $244Hz$ was as calculated. The next test involved sampling d.c. voltages which were varied from $3V$ to $-3V$. The results indicated the system is linear and digital values correlated with those calculated.

The last test involved sampling a $10kHz$, $1.9V_{op}$ sine wave at $500kHz$, the maximum sampling rate. The binary values are plotted for two periods in Fig 5.9 and the sine wave is observed. Fifty sample points are counted for each cycle as expected, and the negative d.c. offset voltage observed was measured with a digital voltmeter. Peak deviation is 70

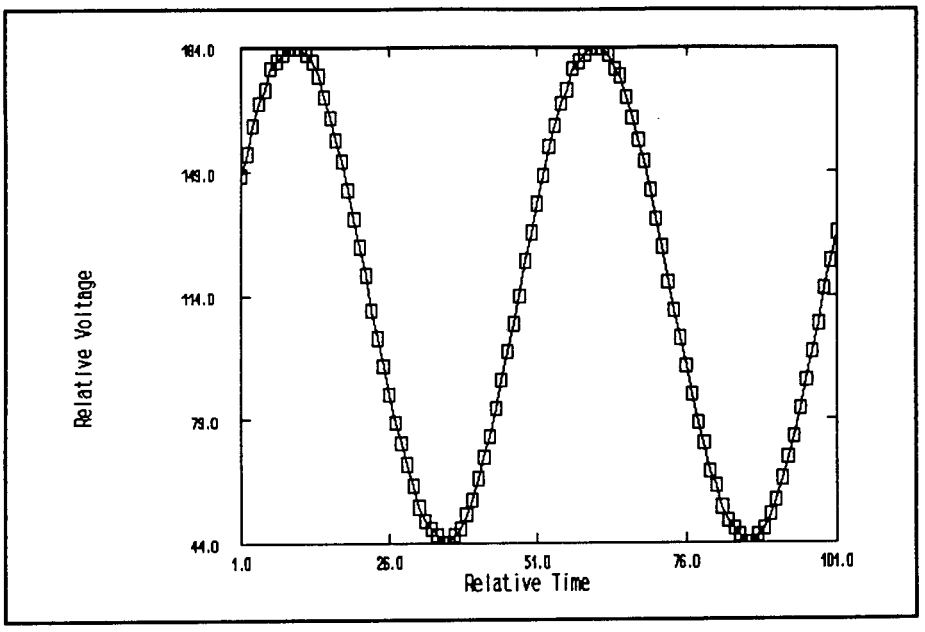


FIGURE 5.9: A $10kHz$, $1.9V$ SINE WAVE SAMPLED AT $500kHz$

units, or $\frac{70}{128} \times 3.5 = 1.91V$ which agrees with the measured value.

5.3 Performance test requirements

The performance of the optical test bed must be evaluated to see if the electro-optic technique is feasible. The most important requirement is to verify that the specifications of §1.1.1 are met and if they are not, the limitations are recorded and compromises discussed. The test results also indicate where modifications must be made and a systematic set of static and dynamic tests are used to measure the system's steady state and transient performance.

Static tests measure the relationship between vertical displacement and differential voltage. This is compared with the relationship calculated earlier. The region over which the differential voltage increases linearly with displacement is useful when comparing the sizes of imperfections.

Dynamic tests are used to verify that the system has sufficient bandwidth and that the sampling rate is adequate to resolve $100\mu m$ imperfections. It was not possible to take measurements on a railway track and therefore one was simulated. The essence of the tests is accumulation of differential voltage samples when the system moves over a known surface at $28ms^{-1}$. Imperfections of known length and height are placed on the reflective surface and are then measured dynamically.

In addition to verifying that the technique is feasible, the tests characterise the system performance. The steady state measurement of differential voltage versus height is required to resolve the approximate height of imperfections. Dynamic test data is also used to evaluate data compression techniques and to establish if repetitive sampling will average the noise out.

5.4 Static system tests

The static system tests measure the relationship between differential voltage and vertical displacement. This enables the linear region and vertical resolution in μm per digital unit to be determined.

5.4.1 Procedure

The apparatus in Figure 5.10 is used and differential voltage is measured with a digital voltmeter as the screwgauge is varied. The micrometer is adjusted so that there is zero differential voltage when the spot is on the centre of the screwgauge. When the micrometer moves forward, the spot moves in the b direction, and when it moves backwards, it moves in the a direction.

The screwgauge is adjusted so the spot falls just off the outer side of the photocells so the summer and differential outputs are zero. Differential voltage is then recorded for fixed intervals as the spot moves over both photocells. The results are then plotted in the ASEASYAS spreadsheet and analyzed.

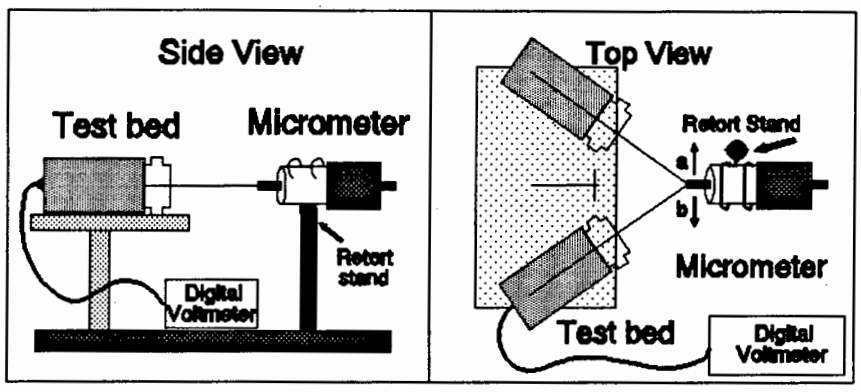


FIGURE 5.10: APPARATUS FOR THE STATIC SYSTEM TESTS

5.4.2 Results

Differential voltage is plotted against vertical displacement in Figure 5.11. Figure 5.12 illustrates the region in which vertical displacement can be resolved. A number of conclusions can be made from Figure 5.11 :

- Differential voltage can be used to detect imperfections of upto $\pm 5mm$ in extent
- The range over which vertical displacement can be resolved is approximately $800\mu m$
- An estimate can be made of the spot size using the relationship

The distance from zero to maximum differential voltage represents movement of the spot by its radius. Assuming the magnification is 1, this translates to a movement of $119\mu m$ on the reflective surface, and thus the spot diameter is $\approx 238\mu m$. Saturation of the amplifier which underestimates spot size is assumed to be compensated by the finite detector separation ($90\mu m$).

The saturation voltages in Fig 5.11 are different and are $2.1V$ and $3.2V$ respectively. A distinction between positive ($2.1V$) and negative ($3.2V$) saturation is made and they are characterised by PS and NS. It is not possible to avoid amplifier saturation when the spot falls predominantly on one photocell and as a result the subcarrier sine wave is distorted. NS results in slightly different saturated waveforms compared to PS with the result that PS has lower differential voltage.

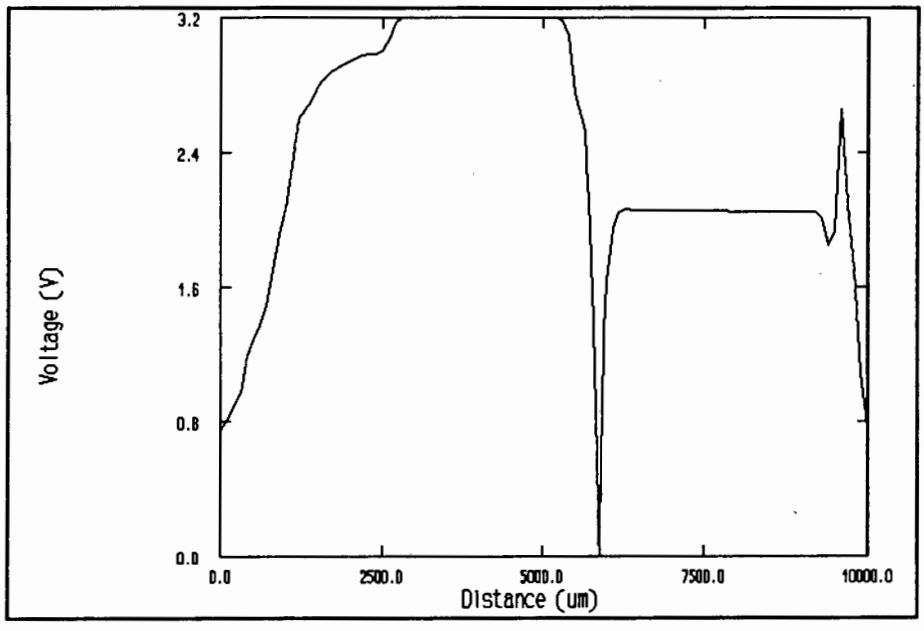


FIGURE 5.11: DIFFERENTIAL VOLTAGE VS VERTICAL DISPLACEMENT

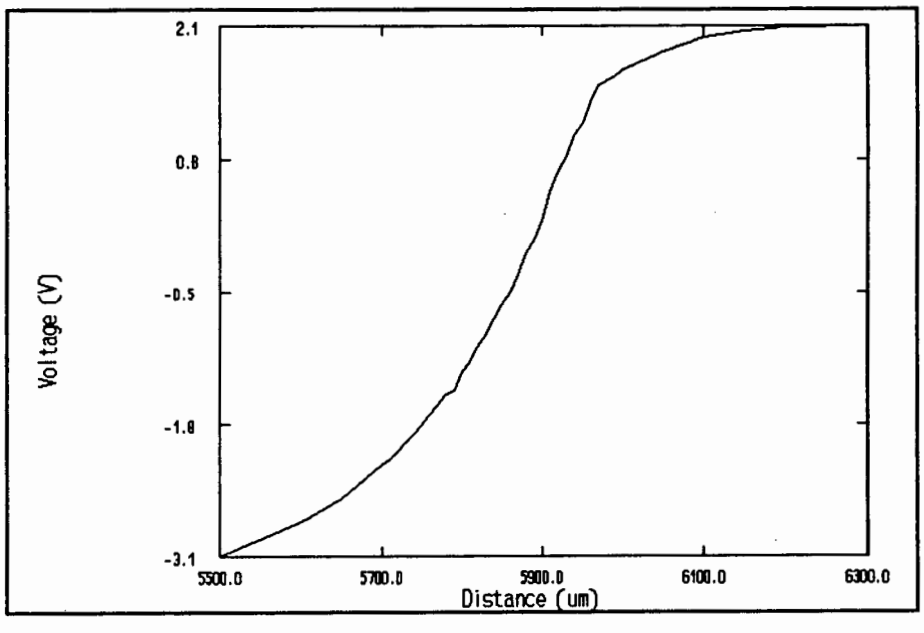


FIGURE 5.12: ABSOLUTE DIFFERENTIAL VOLTAGE VS VERTICAL DISPLACEMENT

The $800\mu m$ interval during which vertical displacement can be resolved is illustrated in Figure 5.12. Vertical displacement is linearly related to voltage for $150\mu m$ and in this region $4\mu m$ displacements can be resolved. On entering the non-linear region, the resolution is decreased until satutration when the magnitudes of imperfections can no longer be distinguished.

5.5 Dynamic system tests

Dynamic system tests determine transient performance of the system and imperfections must be detected when moving over rail at $40ms^{-1}$. The tests include ensuring the sampling rate and bandwidth are sufficient which in turn leads to a prediction of the system's performance. The effect of varying the sampling rate, changing reflectance ..etc are all important parameters to consider when evaluating the system. As the tests cannot be performed in a real railway environment, one is simulated.

5.5.1 Simulation of a train moving over railway track

The apparatus used in the dynamic tests which simulates a train moving over railway track is illustrated in Figures 5.13 and 5.14.It consists of a spinning steel wheel which represents the moving rail. The shaft onto which the wheel is fastened is coupled by fan belt to a d.c. motor whose speed is controlled by a variac.

The surface of the wheel is highly reflective and accurately represents the surface of a well used piece of rail. Automatic light output control will however compensate any minor differences in reflectance.

A train is subject to vibration damped by its inertia which cannot be accurately modelled in this system. The intention however is to minimise this effect and therefore the optical scanning system is fixed rigidly to the steel bench. It is noted that the bench itself has its own mechanical resonances which are minimised by low pass filtering and making sure the wheel runs true.

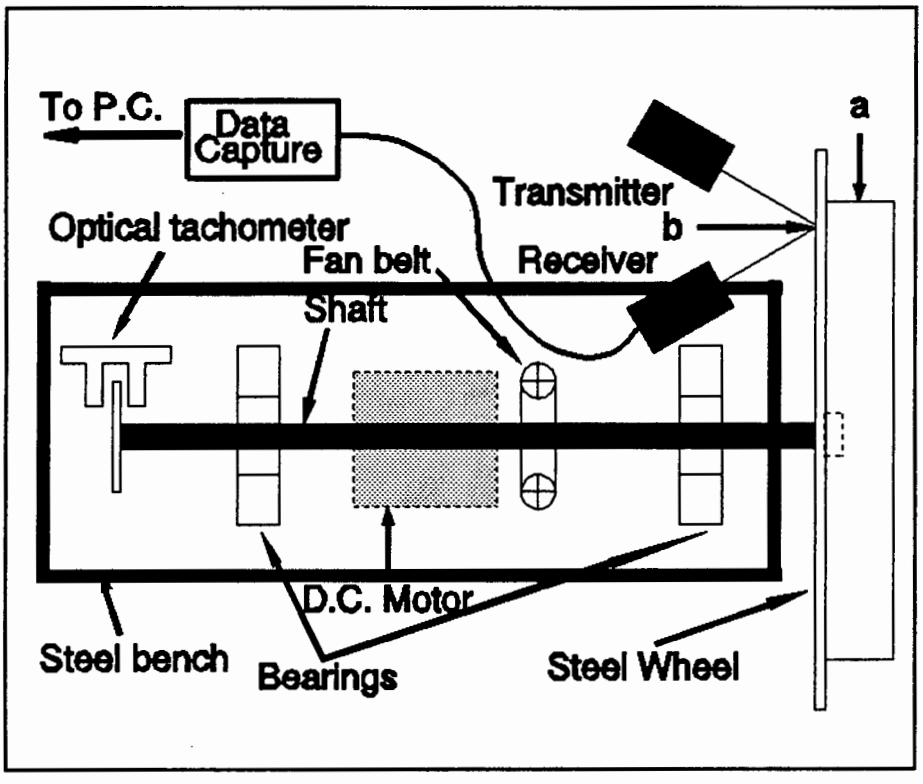


FIGURE 5.13: APPARATUS USED TO SIMULATE MOVING RAILWAY TRACK

There are two moving surfaces on the steel wheel which can be scanned and are illustrated as 'a' and 'b' in Figure 5.13. As it is not possible to establish good mechanical coupling between the optics and the metal bench in the 'a' direction, scanning is performed in the 'b' direction.

Angular speed of the shaft in rpm is recorded with an optical tachometer. The tangential speed ($\frac{dx}{dt}$) which represents the velocity of the train moving over the track is related to angular speed ($\frac{dw}{dt}$) by

$$\frac{dx}{dt} = r\frac{dw}{dt}$$

where r is the radius from the point being scanned. The maximum angular speed of the dc motor is $1500rpm$ and a radius of $25cm$ is required to achieve a velocity of $40ms^{-1}$. It is not possible to achieve good mechanical coupling at this radius and the best compromise is a radius of $17.5cm$. This corresponds to a velocity of $27.5ms^{-1}$ and 1 sample every $55\mu m$ for a $500kHz$ sampling rate.

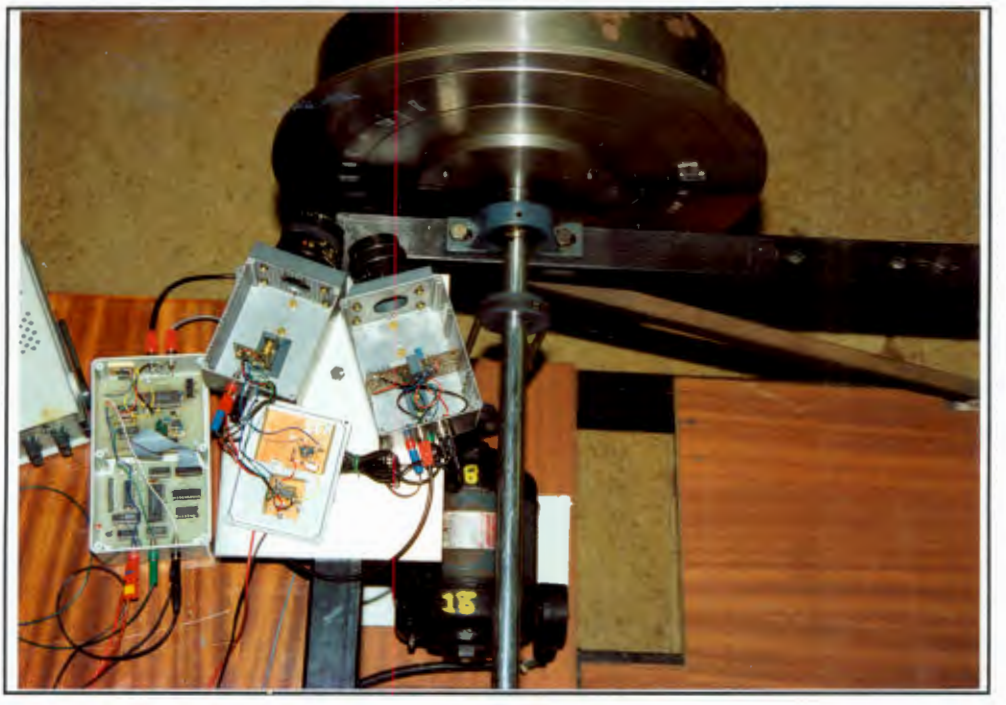


FIGURE 5.14: A PHOTOGRAPH OF THE APPARATUS

The shaft was loosened from its bearings and adjusted for zero differential voltage. The wheel was then rotated slowly which revealed a non uniform surface profile when differential voltage was observed on an oscilloscope. This non uniform profile was verified with a profilometer and the surface has a worst case depression of $50\mu m$.

5.5.2 Experimental procedure

The data capture unit is connected to the differential output voltage and this in turn is connected to an 8086 Bondwell computer. The s.exe program initiates sampling, downloads the 32 Kilobyte epoch and stores the information in a binary file on disk.

The binary file is incompatible with P.C. Datamaster (Durham Technologies) which is used to plot data, filter data, compute correlations ..etc and the program Binasc.exe (App I) converts the data to the correct format. As Datamaster runs under a command line interpreter (pseudo DOS), it is possible to create the appropriate data files with

Binasc.exe in the Datamaster environment.

The dynamic tests are divided into three parts :

- Measurement of the noise
- Measurement of large imperfections : detection and resolving their depth or height
- Measurement of narrow imperfections : just detection

Measurement of electrical noise

The steel wheel is stationary and epochs of 32768 samples of noise associated with the differential voltage are accumulated. This is repeated for various differential voltages and the noise statistics are then compared.

Measurement of large imperfections

Nine rolled aluminium strips of thicknesses varying from $1 - 0.1mm$ are super glued to the steel wheel and represent the imperfections. They are spaced approximately $2cm$ apart and are $0.5 - 1cm$ in width. Super glue is used as it has a thin bonding film of $20 - 30\mu m$.

A profilometer was used to measure the actual thickness of the imperfections and these (A – I) are shown in Table 5.5.2

A	B	C	D	E	F	G	H	I
118	228	300	304	800	420	800	506	1100

TABLE 5.2: THICKNESSES OF ALUMINIUM USED FOR IMPERFECTIONS

The wheel is positioned so that differential voltage is negative when there is no imperfection. This means that the system has maximum linear dynamic range available to it. The wheel is then rotated at maximum speed ($1500rpm$) and samples of the surface are

taken at $62.5kHz$ (3000 samples per revolution). Data is then analyzed to establish if imperfections are detected and secondly, if their height can be resolved.

Twenty five data files were accumulated and then analyzed with Datamaster.

Measurement of cracks

It is not possible to make cracks on the wheel and they were simulated. To detect cracks deeper than $5mm$ both differential and summer voltage must be observed. The change in differential voltage is however of most interest and cracks with depths between 0.1 and $1mm$ were created to observe this.

Three cracks were simulated and each comprises the gap between two strips of aluminium. A feeler gauge was used to set the crack widths to $100\mu m$, $200\mu m$ and $300\mu m$. The aluminium sheets on either side of the cracks varied from 0.1 - 1mm in thickness and they were superglued to the surface of the wheel.

The steel wheel was rotated at $1500rpm$ and 5 data files of the surface profile accumulated at a sampling rate of $500kHz$. Data were then analyzed in Datamaster to see if the cracks were consistently detected.

5.5.3 Dynamic test results

Measurement of noise

Three thousand points sampled at $62.5kHz$ from one of five data files are illustrated in Figure 5.15. The graph represents an undisplaced point on the wheel surface. The mean is 116 digital units which is close to zero differential voltage ($\approx$ 128 units) and peak to peak deviation 330mV (12units).

The graph is consistent with the other data files and the noise statistics do not change when measured on $200\mu m$ and $300\mu m$ imperfections. The noise associated with the sat-

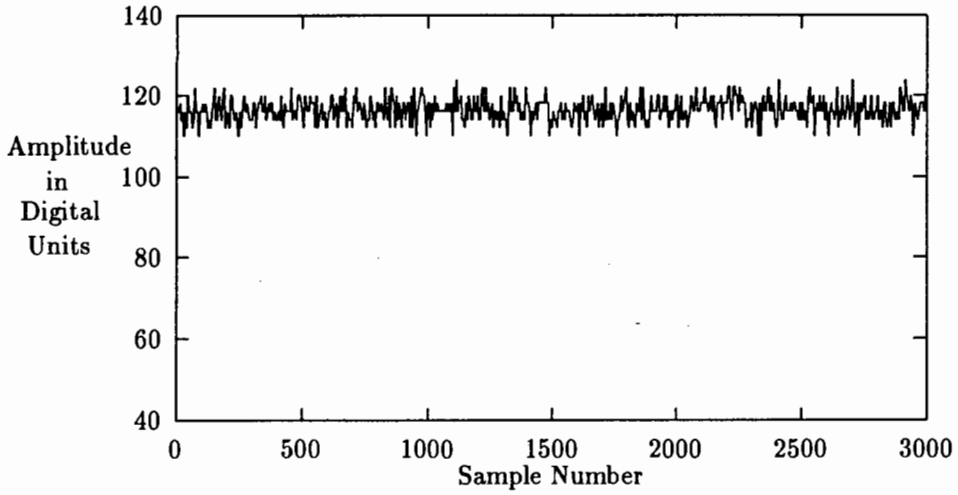


FIGURE 5.15: NOISE OF THE SYSTEM WHEN SCANNING THE STATIONARY WHEEL

uration of the amplifiers when measuring greater than $400\mu m$ imperfections was also measured and is less than that stated above.

Measurement of large imperfections

Twenty five data files were obtained at sampling rates of 500 and $62.5kHz$. Only two were sampled at $500kHz$ as the imperfections on the wheel do not require the resolution of a high sampling rate. The lower sampling rate means that in addition to reduced spatial resolution, the surface of the wheel is scanned several times in one epoch of 32768 samples.

Tangential speed was calculated from the angular speed and was $27.9ms^{-1}$ which corresponds to 2440 samples per revolution. Seven thousand samples from one of the data files is illustrated in Figure 5.16. The three groups of nine imperfections are consistently observed although the waveform is significantly noisier than that measured in the last section. There are three possible reasons for this :

- An increase in electrical noise
- Changes in the wheel profile

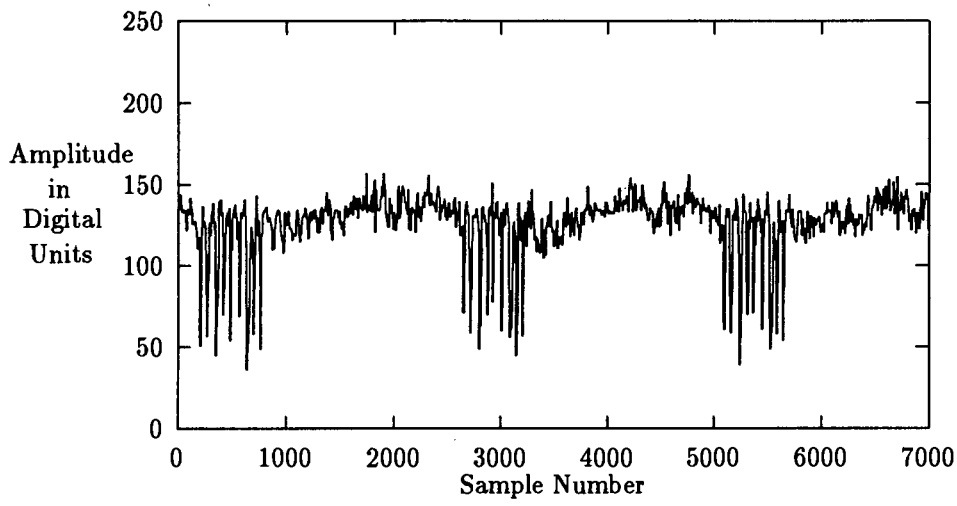


FIGURE 5.16: DIFFERENTIAL VOLTAGE VERSUS TIME : 1527RPM, 62.5KHZ, 7000 SAMPLES

- Localised changes in reflectance on the wheel surface

The first is unlikely as there is no reason for an increase in electrical noise. The small changes in signal amplitudes in the amplifiers due to localised changes in reflectance do not account for such a large increase in the noise.

Changes in the wheel profile were observed earlier and account for changes of up to $50\mu m$ in the wheel profile. As the peak amplitude of the noise is of order $100\mu m$ in Figure 5.16, wheel profile changes are not alone responsible for the noise observed.

Localised changes in reflectance on the wheel surface are a problem if they are not compensated for by the automatic light output control. The automatic light output control was designed for gradual changes in reflectance (time constant of the feedback loop of order $10Hz$) and therefore the rapid changes observed in Figure 5.16 will not be compensated for.

As the wheel profile and localised changes in reflectance do not change in time, the random noise may be evaluated by comparing successive revolutions of the wheel. This is achieved by physically overlaying the two sets of data or by computing the maximum correlation coefficient. The latter is discussed in the following chapter, and the similarity of data is

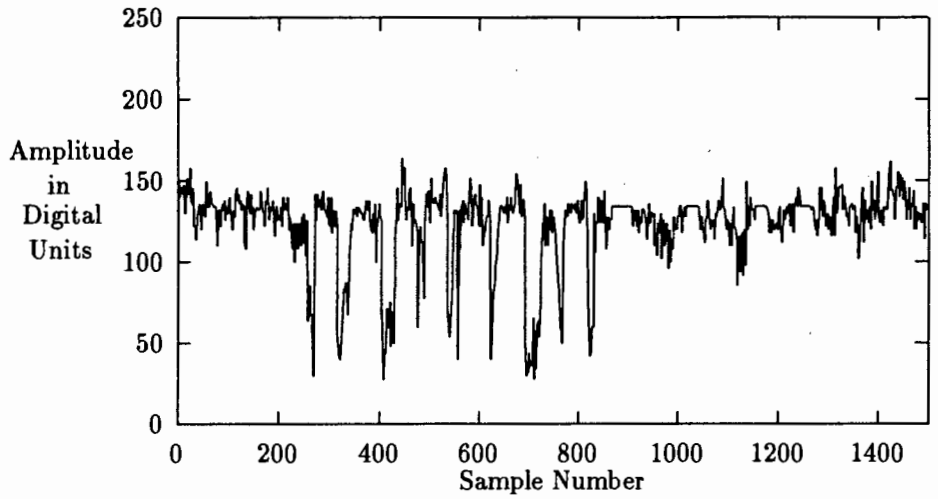


FIGURE 5.17: DIFFERENTIAL VOLTAGE VERSUS TIME : 1527RPM, 62.5KHZ, 1500 SAMPLES, TOP GRAPH TAKEN 5 MINS LATER

shown by overlaying the two graphs in Figure 5.17. The high correlation of data between the two graphs illustrates that the random noise is minimal and the noisy appearance of Figure 5.16 is in fact due to changes in reflectance and wheel profile.

The width of the imperfections in Figure 5.16 are as measured with a vernier. Tolerance on each measurement is $\pm 450\mu m$ since there is one sample every $450\mu m$. This tolerance is also satisfied when the other data files are considered which means that the system is able to resolve widths of $0.4cm$ and more at the current sampling rate.

The amplitude of the imperfections in Figure 5.16 do not correlate to those in Table 5.5.2. For example, imperfection A is of order of B, F and H. There are three possible reasons for this :

- Amplifier saturation
- The nonlinear relationship between differential power and displacement
- Localised changes in reflectance and surface profile

All imperfections in Figures 5.16 and 5.17 are of the order of the saturation voltage (25

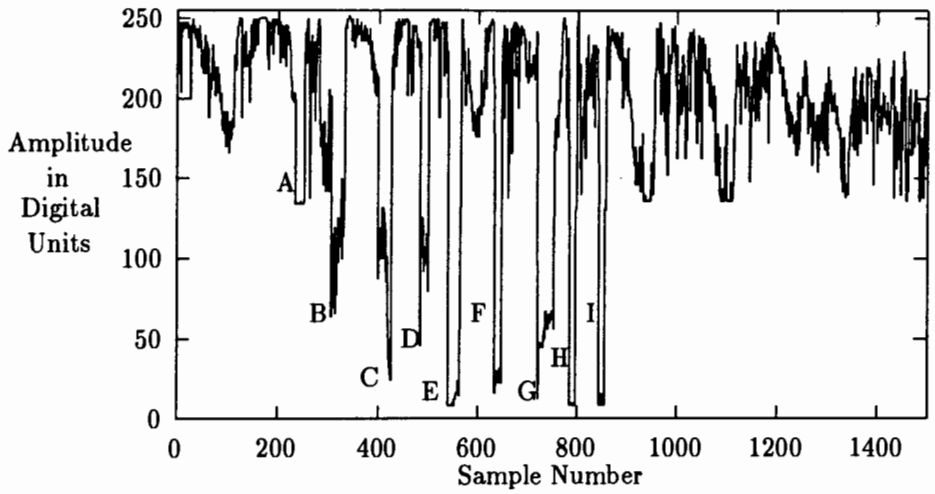


FIGURE 5.18: DIFFERENTIAL VOLTAGE VERSUS TIME : 1500RPM, 62.5KHZ, 4000 SAMPLES

units). This means the differential voltage will not increase for the thicker imperfections as the amplifiers are saturated.

Secondly, the relative amplitude of imperfections cannot be established by direct comparison. The nonlinear relationship between differential voltage and height must be used to calculate actual height.

Lastly, the changes in surface reflectance dominate the variations in differential voltage. A particular example is the less reflective imperfection F which has a lower differential voltage than imperfections A, B and C in Figures 5.16 and 5.17. This is verified by disconnecting the automatic light output control and measuring reflectance of F with respect to A, B and C while the wheel is stationary.

It is important to investigate the ability of resolving the amplitude of imperfections when their reflectances are similar. Each of the imperfections was polished so the reflectances were roughly equal. The differential voltage was then biased positively to give maximum dynamic range and the wheel rotated at 1500rpm. The resulting graph of the nine imperfections is shown in Figure 5.18. Imperfection A($100\mu m$) is almost half of B($228\mu m$) as expected. The non-linear region for the larger imperfections is increasingly significant for the larger imperfections and this is observed. Differential voltage for C($300\mu m$) is

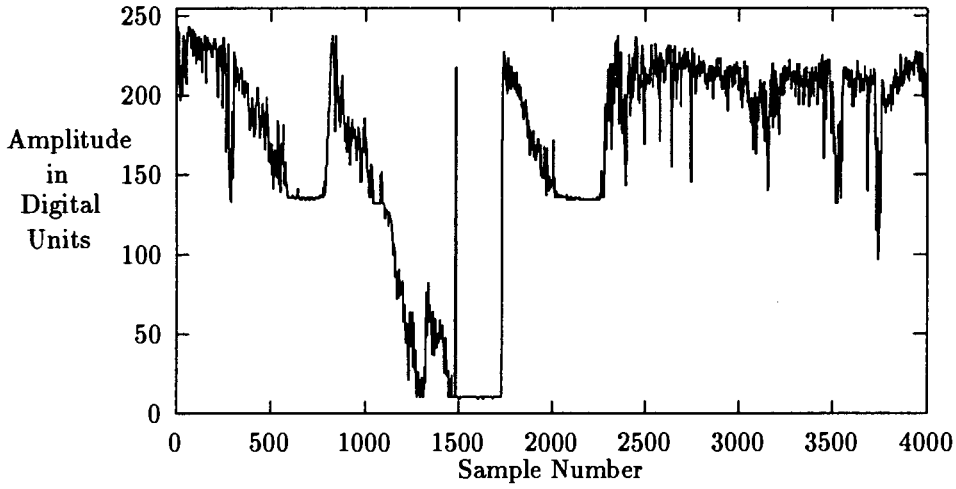


FIGURE 5.19: DIFFERENTIAL VOLTAGE VERSUS TIME : 1539RPM, 500KHZ, 4000 SAMPLES

not three times that of A, and D($304\mu m$) is within the noise tolerance of C. Imperfection E($800\mu m$) is clearly saturated and F($400\mu m$) is greater than C and D, and is just entering the saturation region. Imperfections G, H and I are all saturated as expected.

The results above indicate that differential voltage can be used to resolve amplitude of imperfections. However, this can only be achieved with the current system if the imperfections have equal reflectance and if amplifiers are not saturated.

Detection of cracks

The $140\mu m$ crack between two $510\mu m$ sheets is observed in the middle of Figure 5.19. Sampling frequency was increased to $500kHz$ (one sample every $56\mu m$) and the crack is observed consistently in the five data files. Width of the crack is estimated from the number of numerical values above 150 units. There are three successive data points in the crack which satisfy this condition (amplitudes 166, 212 and 196). Since the interval between samples is $56\mu m$, the crack width is between 112 and $168\mu m$ which correlates to the $140\mu m$ measured with a feeler gauge.

Two more cracks were added with widths of $220\mu m$ and $800\mu m$. All three are observed

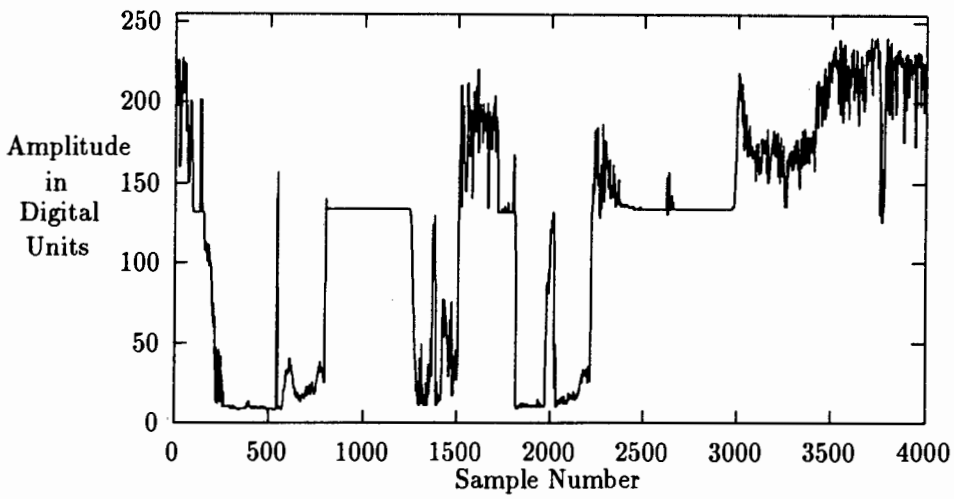


FIGURE 5.20: DIFFERENTIAL VOLTAGE VERSUS TIME : 1510RPM, 500KHZ, 4000 SAMPLES

in Figure 5.20. Five data files were accumulated and the increasing widths can be seen. An estimate of the width of the $800\mu m$ crack can be made from the graph as width of the aluminium sheet on the left of the crack is $4mm$. The crack is $\approx \frac{1}{4}$ of this i.e. 1mm, which ties up with the actual value.

Widths of the other two cracks were estimated from data by counting the values below 150 in the middle of a crack. These corresponded to widths of $280\mu m$, and $840\mu m$ which are within the error tolerance of $\pm 56\mu m$.

Depth of the 220 and $800\mu m$ imperfections is less than for the $100\mu m$ imperfection. This is consistent in all the data files and is due to the less reflective and slightly higher wheel surface. The areas either side of the aluminium imperfections not including the cracks represent the approximate crack depths.

Measurements above only include cracks less than $1mm$ in depth. Detection of a crack greater than $5mm$ will mean that the reflected signal does not fall on the split photocells i.e. zero differential and summer voltage. This means that differential voltage alone cannot be used to detect cracks and that summer voltage must also be recorded.

The experiments above do however reveal that cracks of widths $140\mu m$, and depths less

than 1mm can be resolved and the system therefore has adequate bandwidth and horizontal resolution.

5.6 Conclusions

The static and dynamic tests conclusively indicate the electro-optic technique is feasible for measuring track imperfections. More specifically the static tests indicate :

- Imperfections of upto $\pm 5mm$ in vertical extent can be detected
- The relationship between differential voltage and displacement is non-linear
- The differential power vs displacement graph must be used to relate measured differential voltage to displacement
- Vertical displacements larger than $100\mu m$ cannot be resolved
- Differential voltage is linear for displacements less than $\pm 50\mu m$ only
- Spot size is approximately $250\mu m$

The dynamic tests show the system is capable of the bandwidth designed for, and more specifically :

- Vertical imperfections of $100\mu m$ to $1mm$ can definitely be detected
- Results are consistent and reliably indicate the surface profile
- Amplitude information can be established if the surface is uniformly reflective and amplifiers are not saturated
- $4\mu m$ - $400\mu m$ amplitudes can be read from the differential voltage versus displacement graph
- Cracks $125\mu m \longrightarrow 800\mu m$ wide with depths of $100\mu m$ - $3mm$ can be consistently and reliably detected
- Summer output must be used to detect crack depths greater than $3mm$

Chapter 6

Processing of information

This chapter calculates the correlation between samples of the same surface at different time intervals, and investigates data compression techniques. The correlation between two epochs which represent the surface of the steel wheel is evaluated. Secondly, the techniques used to compress data and therefore to reduce its overall throughput are discussed.

The last chapter gave an informal approach to the similarity of different epochs of data for a section of the rotating steel wheel surface. A more conclusive test is to compute the maximum correlation coefficient of the two sets of data. This leads to a prediction of the amount of random noise in the data and indicates if it is necessary to introduce some sort of noise cancelling by using two time spaced transducers.

Associated with a high sampling rate is the large storage overhead of the data. It is therefore necessary to investigate compression techniques which reduce the amount stored data and maintain its integrity. Four techniques; threshold compression, zero, first and second order adaptive sampling are considered. Of these, the simplest zero-order approach is applied to data and compression ratios are evaluated. A preliminary design of the zero-order system is then presented.

The last section summarises the results and discusses their limitations and implications.

6.1 Correlation of data

Correlation is a measure of similarity of two epochs of data, and if they are identical, the correlation coefficient is 1. The amount by which the coefficient is less than one represents the randomness of the data.

One of the data files in the last chapter was divided into epochs of one wheel revolution and the cross-correlation between two of these was evaluated. The first step is to determine the number of samples per revolution of the wheel and a rough estimate is made from wheel radius and angular speed (2440 samples in this case).

A more accurate method is to count the number of samples between successive scans of a known imperfection. Find.exe (App I) does this and it displays the samples that exceed a threshold. A hysteresis parameter is specified so that only the first of a successive number of samples exceeding the threshold is displayed. In this way, one data point is printed for every imperfection and by calculating the difference between the sample numbers of the first and 'tenth' imperfections, the number of samples per revolution is evaluated. This leads to values between 2456 and 2470, with most being concentrated about 2467 samples per revolution.

Three epochs of 2467 data points were used to evaluate cross-correlations and correspond to the second, fifth and eleventh periods in a data file of 14 revolutions. One of the cross-correlations is shown in Figure 6.1 and the maximum correlation coefficient is 0.99. This reinforces the similarity observed between the two superimposed epochs in Chapter 5 and means that the noisy appearance of the differential voltage vs time plots is due to changes in reflectance and surface profile.

Autocorrelation of one of the epochs is plotted against the cross-correlation of Figure 6.1 in Figure 6.2. If the epochs in the cross-correlation were 100% correlated, Figure 6.2 would be a straight line. The high linearity of the graph once again illustrates the consistent similarity of epochs. Correlation between epochs of the surface taken at least 5 minutes apart were also computed and resulted in similar coefficients.

The results above indicate that data is highly correlated and the random noise is minimal. Secondly, the data is stationary, and noise statistics do not change noticeably with time.

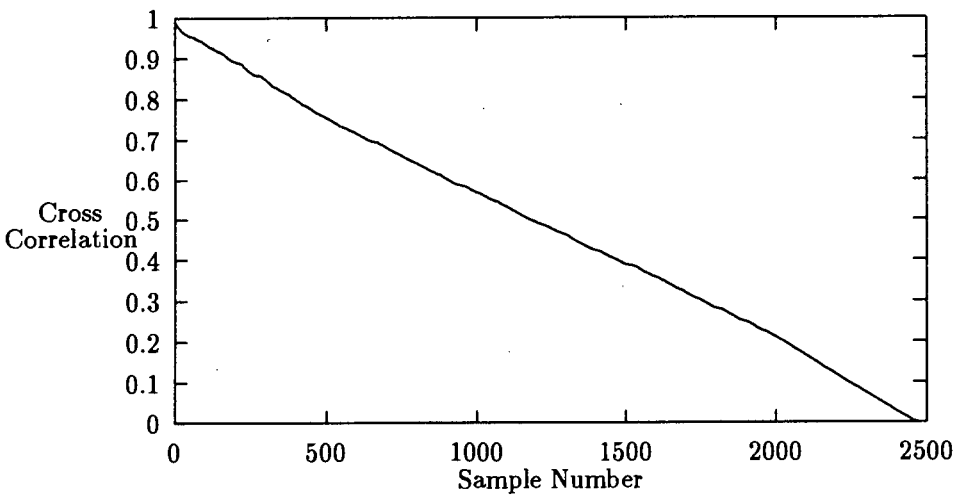


FIGURE 6.1: CROSS-CORRELATION BETWEEN TWO SCANS OF THE WHEEL SURFACE

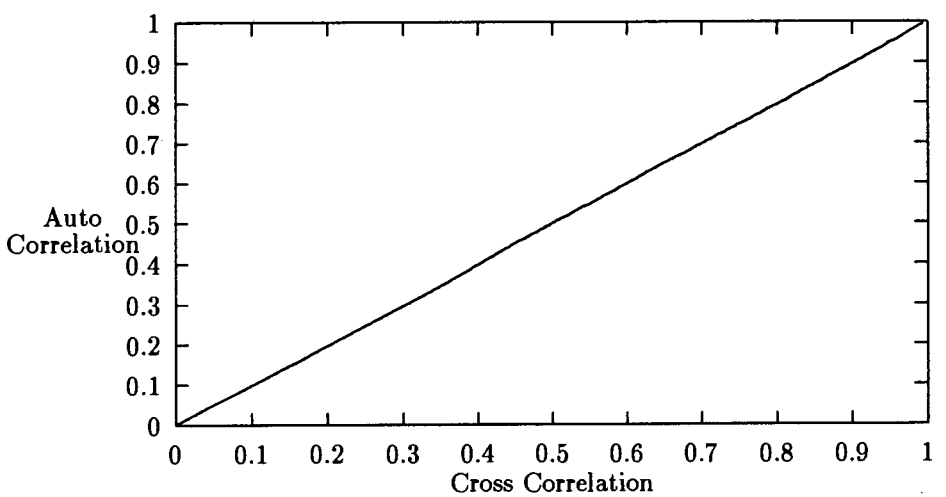


FIGURE 6.2: AUTOCORRELATION VS CROSS-CORRELATION BETWEEN REPETITIVE SCANS OF THE WHEEL SURFACE

6.1.1 Effect of averaging on correlation

It is necessary to investigate how time averaging of two or more epochs reduces noise. This is achieved by correlating averaged with unaveraged data.

Average.exe (App I) was used to average two epochs of 2467 samples and this was correlated to another independent epoch. The maximum correlation coefficients varied between 0.9993 and 0.9997 and reduction in noise is insignificant.

6.1.2 Conclusions

The results above indicate that the randomness of sampled data is in accordance with the signal to noise ratio designed for. In addition, averaging the output of two time spaced transducers will not significantly improve the signal to noise ratio. Time averaging with an extra transducer is therefore not necessary.

6.2 Compression of data

The high data throughput and storage requirements associated with $500kHz$ sampling cannot be achieved with standard equipment. A mechanism which reduces the throughput to the mass storage device is required and is achieved with data compression.

Magnetic tape or disk will be used for long term mass storage of data. Disk systems are more favourable as their higher throughput outweighs the significantly larger storage capacity of a tape. A PC AT with fixed disk storage capacity of 20 – 625Mbytes will be used as the mass storage device. This is the most cost effective choice as dedicated hardware and software is not required.

Before discussing the compression techniques, the implications of a PC AT's maximum throughput and hard disk capacity are discussed. Throughput of fixed disks varies marginally with capacity and depends mostly on cost which increases with disk size.

Presently [1], a P.C. AT with 105Mbyte hard drive is the most cost effective system.

Maximum throughput of the PC AT is used to calculate the compression required to support a $500kHz$ sampling frequency. Throughput is calculated from the data–to–memory and memory–to–disk times. For a P.C. AT, data–to–memory time is smallest if transfer is done with Direct Memory Access (DMA) which is maximally 200KBytes per second [6]. Faster times cannot be guaranteed due to higher priority system interrupts which are serviced in preference to the DMA request. Memory–to–disk time is also performed by DMA but is limited by the physical disk properties viz. access time, latency and data transfer rate. The equation relating these is given by :

$$T_{DSK} = T_{Latency} + T_{Access} + \frac{T_{Transfer}}{b}$$

where T_{DSK} is the time taken to transfer b bytes to disk. For a typical 105MByte fixed disk [15], latency is $18ms$, access is $277\mu sec$ and transfer rate is 2Mbyte/sec. DMA is fastest if data is transferred in 64KByte blocks which implies a $T_{DSK} = 50ms$. Nett throughput is calculated by adding data–to–memory and memory–to–disk times. This is $370ms$ for 64KByte blocks and implies a maximum throughput of 170KBytes/sec.

Data compression must therefore reduce the 500KByte/sec throughput to 170Kbyte/sec. The digitised differential voltage will in general be constant except for noise and only intermittent, fast changing intervals associated with an imperfection are of interest. It is therefore advantageous to sample at a high rate during these transient intervals and slowly during the quiescent intervals.

Another implication of data compression associated with non uniform sampling is that time information must be stored with each sample. Four data compression techniques are investigated and they are the threshold, zero, first and second order methods discussed by Cohen [3]. The underlying philosophy of all techniques is to sample at $500kHz$ and to reject or accept data depending on the compression algorithm.

[1] Dec 1992

6.2.1 Threshold data compression

The threshold technique involves sampling at $500kHz$ and accepting data if it is below (or above) a threshold. Thus for the duration of an imperfection, data is sampled at $500kHz$ giving maximum resolution. The mathematical rule is store if

$$X_n > R_1 \tag{6.1}$$

where R_1 is the threshold, and X_n the n'th $500kHz$ sample.

The technique has the advantage that in general, the threshold will be exceeded successively for duration of the imperfection and only the starting time of the imperfection need be stored. As positive and negative going transitions of differential voltage are of interest, an additional threshold R_2 is introduced. The modification to Equation 6.1 is don't store if

$$R_1 > X_n > R_2$$

where R_1 is the upper and R_2 the lower threshold.

The technique finds application in systems with high signal to noise ratio where the threshold is well defined. It can be implemented easily in hardware and the compression ratio depends on the threshold and interval between imperfections.

The disadvantage of this technique is that it is sensitive to drift of the differential voltage. For example, vibration will introduce low frequency variation in differential voltage. Say that the differential voltage is close to the threshold R_2 and a positive going imperfection is passed. The increase in differential voltage may not be large enough to exceed the R_1 threshold and consequently the imperfection is not stored.

6.2.2 Zero order adaptive sampling

Also known as the voltage triggered method, this technique stores data if the difference between differential voltage of the current, and last sample stored, exceeds a threshold. This is more formally, store if

$$\mid x(t_i + \tau_i) - x(t_i) \mid > R_0$$

Assume that at time t_i a sample $x(t_i)$ was stored. The next sample that is stored is interval τ_i later when the difference between it and the last sample stored, exceeds the threshold R_0. The absolute value indicates that the rule applies to either increasing or decreasing data.

The amount of data stored can be reduced by increasing threshold R_0 at the expense of losing resolution. Storage of data is intermittent and therefore each of the data points is stored with its associated time information.

The technique has the advantage that it is simple to implement in hardware and is insensitive to low frequency vibration. It is also capable of resolving transient data at $500kHz$ which prevents aliasing and performs better than the threshold method in lower signal to noise environments [3]. It however has the disadvantage that with each sample, time information must be stored. This overhead reduces overall compression which must be taken into account when setting threshold R_0.

6.2.3 First order adaptive sampling

Also known as the two points projection method, this technique compares slope of the data last stored to the current slope. If the absolute value of the slope of the last data point stored is different to the current slope by more than a threshold value R_1, data is stored. Thus data is stored if

$$\mid \dot{x}(t_i + \tau_i) - \dot{x}(t_i) \mid > R_1$$

The time derivative (slope) of data stored at t_i, $\dot{x}(t_i)$, is compared to the current slope $\dot{x}(t_i + \tau_i)$ an interval τ_i later. If the difference between the two slopes exceeds R_1, then data is stored. The time derivatives are estimated by the formula :

$$\dot{x}(nT_s) \approx \frac{x\left(nT_s\right) - x\left(\left(n-1\right)T_s\right)}{T_s}$$

T_s represents the minimum sampling interval ($\frac{1}{500\times10^3}$) and n, the nth sample. In the presence of significant additive noise, Cohen [3] suggests a time averaged estimate of the slope :

$$\dot{x}(nT_s) \approx \frac{\frac{1}{M}\sum_{j=0}^{M-1} x\left(\left(n+j\right)T_s\right) - \frac{1}{M}\sum_{j=0}^{M-1} x\left(\left(n-j\right)T_s\right)}{\left(M-1\right)T_s} \tag{6.2}$$

where $2M - 1$ samples are used to smooth data and obtain an estimate of the slope.

The method represents transitional slopes significantly better than the first two techniques and is insensitive to low frequency variation or drift associated with vibration. It operates more efficiently than the threshold technique in low signal to noise environments and has the disadvantage that time information must be stored with each sample. This is taken into account when selecting the threshold R_1. The technique is also significantly more difficult to implement in hardware than the last two methods.

6.2.4 Second order adaptive sampling

This technique examines the slope before and after the current sample. If the absolute value of the difference between these slopes is greater than a threshold R_2 then the sample is stored. Also known as the second differences method, this technique is represented by the following rule and the sample is stored if :

$$| \dot{x}(t_i^-) - \dot{x}(t_i^+) | > R_2$$

The time derivatives are estimated by approximations of averages of $2M - 1$ data points and are given by Cohen [3].

The advantage of the method is it accurately represents slopes in a highly transitional signal. It is insensitive to low frequency drift associated with vibration and operates as efficiently as the first order method in noisy environments. Once again, time information is associated with each sample which is taken into account when threshold R_2 is chosen. The major disadvantage of the technique is it is difficult to implement in hardware.

6.3 Application of the zero order technique to sampled data

Choice of a suitable adaptive sampling technique should ideally be based on data obtained in an actual railway environment. As this is not available, the zero order method was

chosen as it is insensitive to vibration and is easier to implement than the first and second order methods at the high data rate ($500kHz$).

Adaptive.exe (App I) implements zero order adaptive sampling and it stores compressed data with the associated time information in an output file. The total number of samples and the number in the compressed file are displayed and the compression ratio is calculated from these. It was assumed that associated with each data byte, there are two time bytes. The reason for this assumption is given in the next section.

Figures 6.3 and 6.4 illustrate actual and compressed signal representations for 2000 data points sampled at $62.5kHz$. The threshold was set to 10 units out of 255 and data was reduced from 2000 samples to 665. Imperfections are well represented and the similarity

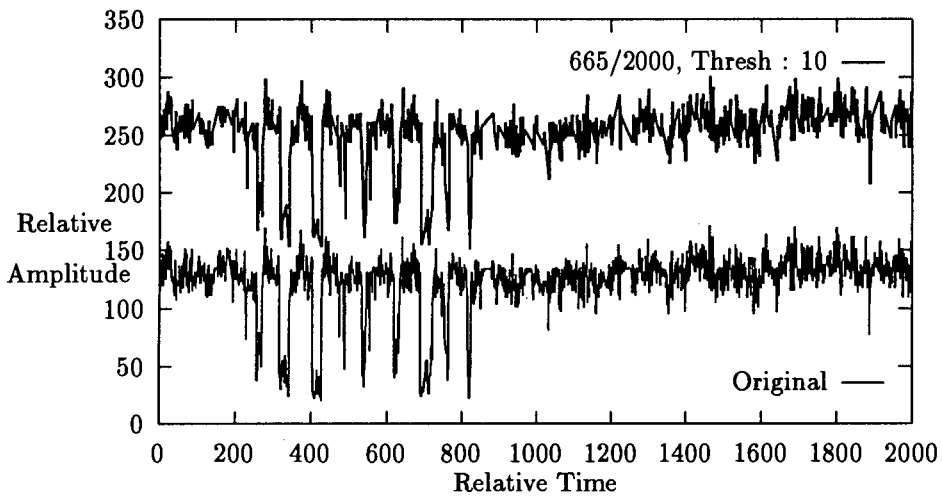


FIGURE 6.3: DIFFERENTIAL POWER VS DISPLACEMENT FOR ACTUAL AND ZERO ORDER SAMPLING : THRESHOLD = 10

between actual and compressed data is observed. If the time information is added to the data, the effective number of samples is $665 \times 3 = 1995$ i.e. there is not much compression. Although storage requirements aren't reduced, throughput is, by a factor of 3. For example, if the three bytes associated with a sample are transferred simultaneously, the data rate is one third of the original rate.

Figure 6.4 shows the compression of data when the threshold is 20. Imperfections are observed easily and the lower amplitude variations due to the surface profile are no longer

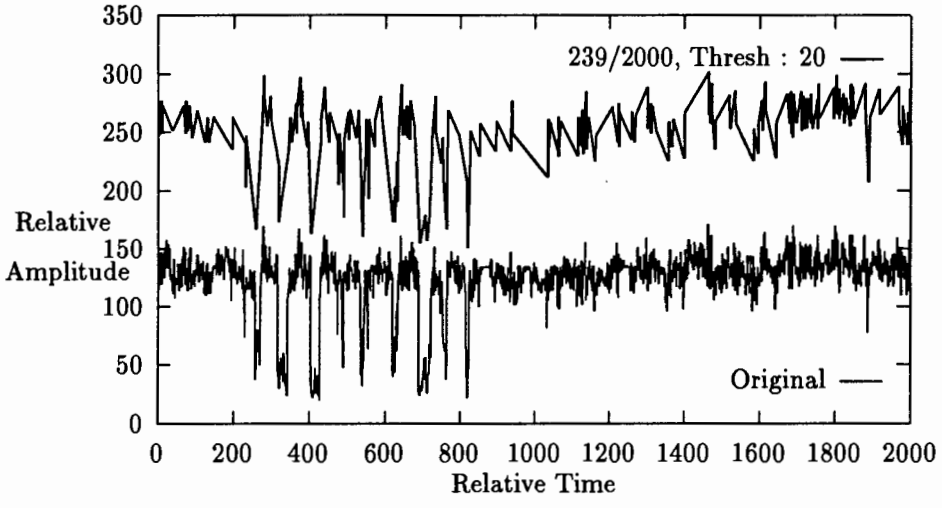


FIGURE 6.4: DIFFERENTIAL POWER VS DISPLACEMENT FOR ACTUAL AND ZERO ORDER SAMPLING : THRESHOLD = 20

clearly distinguishable. The number of samples stored is reduced from 2000 to 239 and including time information, this is a compression ratio of 1 : 3.

Most samples are stored in transient intervals of a data file. The distribution of stored points for the data in Figure 6.4 is illustrated in Figure 6.5 and was generated with Adaptive.exe.

The data in a railway environment is expected to have relatively few imperfections which are not clustered as in data above. In general, the imperfections are expected to be widely spaced apart. This means that the compression ratios above will improve as the majority of samples stored are due to transient intervals.

6.4 Preliminary design of a zero order adaptive sampler

The analysis so far has revealed that zero order adaptive sampling is simple and adequate. The threshold technique is sensitive to low frequency vibration and the first and second order techniques are unnecessarily complex. The high sample rate of $500kHz$ demands that compression computations be performed in less than $2\mu s$. Consequently, the more

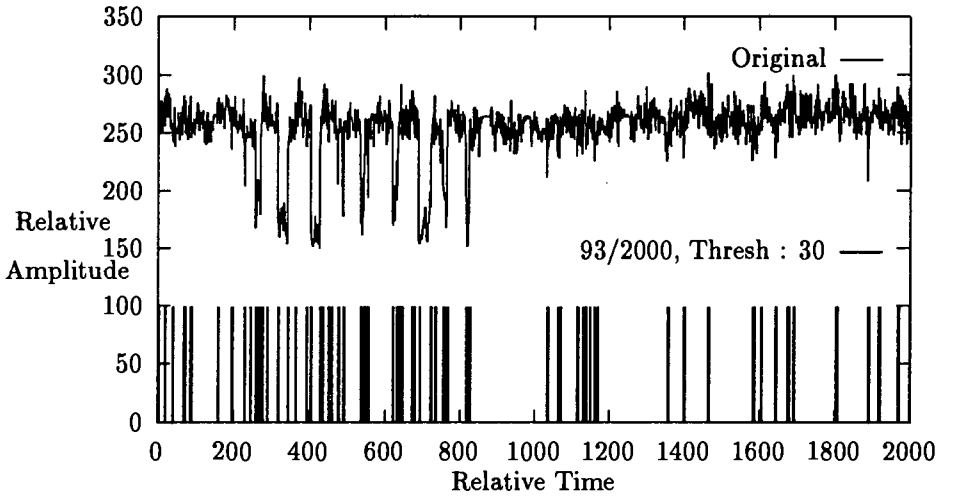


FIGURE 6.5: DISTRIBUTION OF DATA POINTS FOR THE ZERO ORDER SYSTEM IN FIG 6.4

mathematically intensive first and second order methods cannot be implemented efficiently with available low cost microprocessors and require complex dedicated hardware. The less complex zero order technique can easily implemented with hardwired logic and is therefore the best alternative.

A zero order sampler is represented by the block diagram in Fig 6.6. The system is divided into three modules, the analogue to digital converter, the arithmetic module and the RAM module.

The A/D module was developed in chapter 5 and its function is to provide 8 bit samples at $500kHz$ with an End of Conversion pulse. Sampled data is then fed to the arithmetic unit where it is subtracted from the last value stored. The absolute value of the difference is compared to a threshold, and if greater, activates the store signal. The store signal is then used to update the last sample register and to store the sample in RAM with the time information. The RAM is divided into two blocks which are alternatively filled by the A/D and read by the PC. As data is stored in one block, data in the other block is read by the PC.

The arithmetic module has been designed and tested and its circuit diagram is included in Appendix H. The RAM module has not been built and fully designed. Time information is

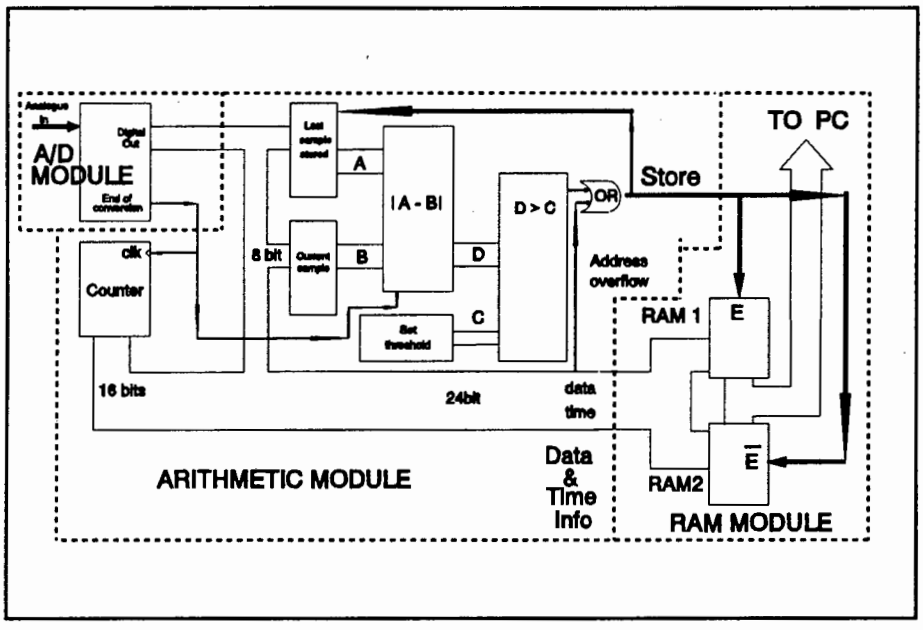


FIGURE 6.6: BLOCK DIAGRAM OF AN ADAPTIVE SAMPLER

maintained by the counter and with each datum, two bytes from the counter are stored. To ensure that synchronisation is not lost, a counter value of 0000 initiates a store command. Each RAM is able to store 21845 data points (1 data, 2 time bytes) and when full, data is transferred to the PC via DMA while the other RAM is filled.

6.5 Summary

This chapter investigated the correlation of data in Chapter 5 and discussed data compression techniques. Correlation of data obtained for the rotating wheel surface in Chapter 5 was of order of 99.3%. This means that errors in the system are not due to random noise and result from changes in surface reflectance.

Four compression techniques were investigated to reduce throughput and data storage requirements. Of these, the zero order technique was the simplest and easiest to implement. It was applied to data and resolution is traded off for compression. The analysis was applied to a data file of 2000 points which included the nine imperfections (Chapter

5) and a compression ratio of 1 in 3 suitably represented data. A preliminary hardware design of the zero order system was then presented.

Chapter 7

Conclusions and recommendations

The electro-optic technique for measuring imperfections on railway track is feasible. A summary of how the specifications are satisfied is presented and the system limitations are discussed. Recommendations based on these results can then be made.

The extent to which specifications in Table 1.1 are fulfilled is summarised as follows :

- Cost and maintenance of the system is low in contrast with other techniques.
- Vibration of the train is manifested as low frequency drift in the output of system. The ability of the technique to detect long imperfections which appear as low frequency signals in the output is therefore limited by this vibration. In this system, maximum length of an imperfection is assumed to be less than $1m$.
- The minimum resolvable length of an imperfection was tested to be $0.15mm$. This assumes a sampling rate of $500kHz$ at a train speed of $40ms^{-1}$ and less.
- Vertical resolution is a nonlinear function of vertical displacement. The highest resolution is achieved for small displacements and is typically $4\mu m$ in this system.

The limitations of the system are discussed in terms of the data acquisition hardware and the processing of the raw data. The latter discusses the techniques used to reduce

throughput of data and the optimum choice of a particular one. Data acquisition is more easily explained in terms of the

- Receiver
- Transmitter
- System characteristics

7.1 Receiver

The following conclusions and recommendations are made about the receiver :

- Asynchronous as opposed to direct detection increases the signal to noise ratio to $38dB$.
- The envelope detector used to asynchronously demodulate the incoming signals should be replaced with an active circuit which doesn't require a turn on voltage.
- Negative and positive supplies of the opamps should be increased from $\pm 5V$ to increase dynamic range.

7.2 Transmitter

The following conclusions and recommendations are made about the transmitter :

- A semiconductor laser is the most cost effective source to use.
- A $5mW$ laser has sufficient sterance to achieve a $38dB$ signal to noise ratio in the receiver for a worst case rail reflectance of 1×10^{-3}.
- Highest reflectance off rusty rail is achieved in the infra-red where silicon photocells are most sensitive and lasers are most efficient.

- Automatic light output control circuitry is used to keep the signal to noise ratio constant when the system is subject to differing rail reflectances.
- The light output control cannot adapt to rapid changes ($< 1s$) in reflectance and is used to adjust to long term variations.
- The light output control circuitry does not function correctly when light falls predominantly on one photocell and receiver amplifiers saturate. The problem is believed to be due to the very sensitive setpoint which cannot be adjusted accurately. A better feedback system should therefore be designed to eliminate the problem.

7.3 System limitations and characteristics

The system comprising the receiver and transmitter was evaluated and recommendations are made based on the results. The evaluation and recommendations are summarised as follows :

- Measurements were performed to evaluate the system. These could not be performed on a real railway track and one was simulated.
- The measurements should be repeated in a real railway environment.
- Test results showed that imperfections of $50\mu m$ can be seen by observing the raw data without additional processing.
- The output of the displacement measuring system is a non linear function of displacement.
- Vertical displacements of less than $50\mu m$ are represented by a linear change in voltage with displacement and have a resolution of $4\mu m$.
- Large imperfections cannot be resolved as accurately as small ones and resolution decreases with an increase in imperfection size.
- Spot changes in reflectance are not compensated by the automatic light output control.

- Automatic light output control must be implemented so it is sufficiently rapid ($2\mu s$) to compensate for spot changes in reflectance.
- Changes in reflectance presently prevent the output of the system being mapped onto absolute vertical displacement.
- Dynamic range for change in reflectance between best and worst case surfaces is maximally $30dB$(1:1000).
- System bandwidth has been tested for sample rates of $500kHz$ and train speeds of $27ms^{-1}$.
- The system must be tested for train speeds of $40ms^{-1}$.
- Imperfections are represented by differential and summing voltages.
- Imperfections of less than $1mm$ can be resolved in height, and can be detected if they are less than $5mm$ by looking at the differential output voltage.
- Imperfections greater than $5mm$ are detected by the summing amplifier.
- Tight mechanical coupling between the system and the bogey ensures that the effect of vibration is minimised.

7.4 Data processing

Before investigating data processing, the correlation of data was evaluated for a known surface. This lead to the following results and conclusions :

- The system has a high data integrity which is verified by the 0.993 correlation coefficient.
- Averaging of data from two separate transducers will not significantly improve the signal to noise ratio of the system.

Data processing is essential to reduce the throughput of raw data to the permanent storage which is a PC AT's hard disk. This involves an investigation into various data compression techniques which lead to the following :

- Data compression is required to reduce throughput and storage in the system.
- Four techniques were considered of which only one, zero-order adaptive sampling, was chosen.
- Zero-order adaptive sampling is the simplest, easiest to implement and adequately represents the track surface. It was evaluated with the raw data obtained from the tests in Chapter 5.
- First and second order techniques must be applied to data and their hardware implications investigated in more detail.
- The zero-order system requires a threshold which determines the amount of compression. The appropriate hardware should therefore be built and tested in a railway system so the appropriate threshold can be found.
- In general, compression is traded off for amplitude resolution and the less the data stored, the lower the resolution.

References

[1] Ambient Light rejection improves optical recognition sensor performance. *Dataweek*, 10 May 1991. South Africa.

[2] J. Bell, January 1992. Professor, Dept. Electronic Eng, University of Cape Town, Private communication.

[3] Arnon Cohen. *Biomedical Signal Processing Volume 1 : Time and Frequency Domain Analysis.* CRC Press, Florida, USA, 1st edition, 1986.

[4] Robert Z. Csizmadia. Digitisation of the output of a Gyrotheodolite. Master's thesis, University of Cape Town, 1991. Chapter 4.

[5] G. Daryanani. *Principles of Active Network Synthesis and Design.* Wiley & Sons, New York, USA, second edition, 1976. Chapter 5.

[6] Lewis C. Eggebrecht. *Interfacing to the IBM PC.* SAMS, New York, USA, 2nd edition, 1990.

[7] F.G.Stremler. *Introduction to Communication Systems.* Addison Wesley, New York, USA, second edition, 1982. Chapter 4.

[8] John Greene. An Electro-Optic in-shore survey aid. Master's thesis, University of Cape Town, 1980.

[9] John R. Greene. Optics course, 1990.

[10] J.R. Greene, June 1991. Senior Lecturer, Dept. Electronic Eng, University of Cape Town, Private communication.

[11] D. Halliday and R. Resnick. *Fundamentals of Physics.* Wiley & Sons, New York, USA, second extended edition, 1978. Chapter 19, p349.

[12] J.P.Wittke & I.Ladany H.Kressel, M.Ettenberg. *Semiconductor Devices for Optical Communication : Lasers Diodes and LEDs for Fibre Optic Communication.* Springer & Velag, Berlin, Germany, 2nd updated edition, 1982. Chapter 2, Topics in Applied Physics.

[13] H.W.Ott. *Noise reduction techniques in Electronic Systems.* John Wiley & Sons, New York, USA, 1975. Chapter 9.

[14] Eric Miller. *Introduction to practical Fiber Optics.* 1989. National Semiconductor Application Note, AN-244, p614-615.

[15] Scott Miller. *Upgrading and Repairing PC's.* Que Corporation, Indiana, USA, 1st edition, 1988.

[16] Athanasios Papoulis. *Probability, Random Variables and Stochastic Processes.* Mc Graw & Hill, New York, USA, second international student edition, 1976.

[17] P.Horowitz and W.Hill. *The Art of Electronics.* Cambridge University Press, Cambridge, USA, first edition, 1980. ISBN 0 521 23151 5.

[18] Plessey. The MA200 Tellurometer Technical reference, 1988.

[19] Jr P.W.Shumate & M.D.DiDomenico. *Semiconductor Devices for Optical Communication : Lightwave Transmitters.* Springer & Velag, Berlin, Germany, 2nd updated edition, 1982. Chapter 5, Topics in Applied Physics.

[20] MI Skolnik. *Introduction to Radar Systems.* Mc Graw & Hill, Tokyo, Japan, 1962. Chapter 1, p26.

[21] R.G. Smith and S.D. Personick. *Receiver Design for Optical Fibre Communication Systems.* Springer-Velag, Berlin, Germany, 2nd updated edition, 1982. Chapter 4, Topics in applied physics, Semiconductor Devices for Optical Communication.

[22] John Strong. *Concepts of Classical Optics.* W.H.Freeman & Co., San Francisco, USA, 1958. Chapter 14.6, p315.

[23] Sean Thain. *Electro-Optic Surface Monitor.* 1991. undergraduate thesis.

Appendix A

Calculation of vertical displacement for trailing wheel and electro-optic techniques

This appendix includes

- Derivation of the change in height with rotation angle in the trailing wheel technique.
- Derivation of the differential optical power vs vertical displacement in the electro-optic technique.
- A Mathcad listing of a worksheet to display the electro-optic technique's results.

A.1 Relationship between change in height and rotation angle for the trailing wheel technique

The relationship is worked out using Fig A.1.

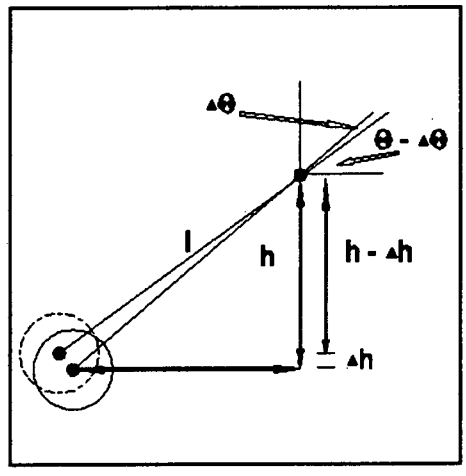


FIGURE A.1: REPRESENTATION OF THE TRAILING WHEEL TECHNIQUE USED TO MEASURE TRACK IMPERFECTIONS

$$\begin{aligned} h &= l\sin\theta \\ h - \triangle h &= l\sin(\theta - \triangle\theta) \\ \triangle h &= l\sin\theta - l\sin(\theta - \triangle\theta) \\ \triangle h &= l\sin\theta - l\sin\theta\cos\triangle\theta + l\sin\triangle\theta\cos\theta \end{aligned}$$

but $\cos\triangle\theta \approx 1$ for $\triangle\theta$ small

$$\begin{aligned} \triangle h &= l\sin\theta - l\sin\theta + l\sin\triangle\theta\cos\theta \\ \triangle h &= l\sin\triangle\theta\cos\theta \end{aligned}$$

A.2 Derivation of change in power vs height relationship for an electro-optic track monitor

Equations for this relationship are developed from Fig A.2

$$\angle AFC = 180° - \angle DFA - \angle CFB$$

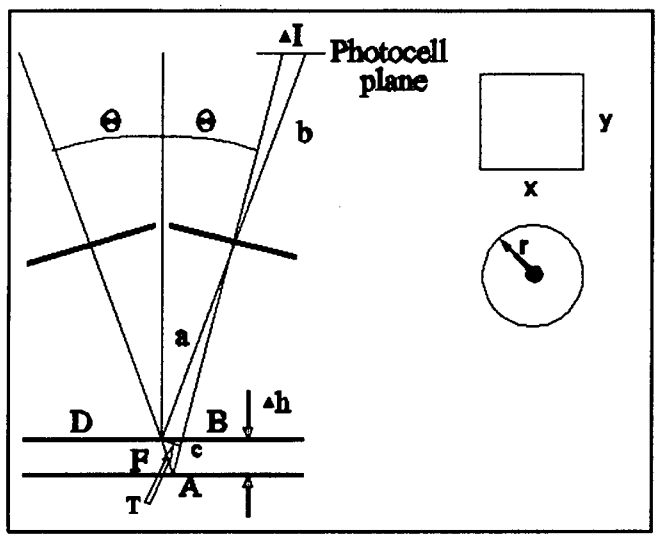


FIGURE A.2: REPRESENTATION OF RECTANGULAR AND CIRCULAR SPOTS IN AN ELECTRO-OPTIC TRACK MONITORING SYSTEM

$$
\begin{aligned}
\angle DFA &= 90 + \theta \\
\angle CFB &= \theta \\
\angle AFC &= 180 - (90 + 2\theta) \\
&= 90 - 2\theta \\
T &= FA\cos(90 - 2\theta) \\
&= FA(\cos 90 \cos 2\theta + \sin 90 \sin 2\theta) \\
&= FA \sin 2\theta
\end{aligned}
$$

but

$$
\begin{aligned}
FA &= \triangle h \sec\theta = \frac{\triangle h}{\cos\theta} \\
T &= 2\triangle h \sin\theta \\
\triangle I &= 2\frac{b}{a}\triangle h \sin\theta
\end{aligned}
$$

Call $\frac{b}{a}$ magnification M

$$\triangle I = 2M\triangle h \sin\theta$$

A.2.1 Rectangular Spot

P = power in the spot
Spot dimension x in the photocell plane = Mx
Spot dimension y in the photocell plane = My
Area of the spot $A_{spt} = M^2xy$
Assuming uniform power distribution, Areance = $\frac{P}{M^2xy}$
Movement $\triangle I$ of the spot :
Area of spot on photocell 1 = $(Mx + \triangle I)My$
Area of spot on photocell 2 = $(Mx - \triangle I)My$
Power on photocell 1 = $(Mx + \triangle I)My\frac{P}{M^2xy}$
Power on photocell 2 = $(Mx - \triangle I)My\frac{P}{M^2xy}$

$$\begin{aligned} NettDifferentialPower &= PwrPhoto1 - PwrPhoto2 \\ &= \frac{2\triangle IMyP}{M^2xy} \\ &= 4\frac{M^2\triangle h \sin\theta yP}{M^2xy} \\ &= \frac{4P\triangle h \sin\theta}{x} \end{aligned}$$

A.2.2 Circular Spot

P = Power in the spot
Spot radius r in the photocell plane = Mr
$\rightsquigarrow$ Area of the spot in photocell plane = πM^2r^2
$Areance = \frac{P}{\pi M^2r^2}$ (Assuming uniform power distribution)

Movement $\triangle I$ of the spot (refer to Fig A.3)

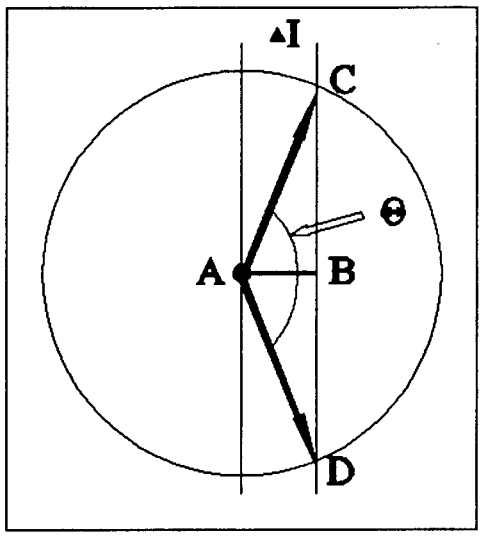


FIGURE A.3: REPRESENTATION OF THE SPOT WHEN MOVED BY $\triangle I$

Area of sector $= \pi M^2 r^2 \frac{\theta}{2\pi}$
but $\frac{\theta}{2} = \arccos \frac{\triangle I}{Mr}$
$\rightsquigarrow \theta = 2 \arccos \frac{\triangle I}{Mr}$

$$\begin{aligned} AreaOfSector &= \pi M^2 r^2 \frac{2 \arccos \frac{\triangle}{Mr}}{2\pi} \\ &= M^2 r^2 \arccos \frac{\triangle I}{Mr} \end{aligned}$$

$$\begin{aligned} AreaOf\triangle ACD &= 0.5 AD.AC \sin\theta \\ &= 0.5 M^2 r^2 \sin\left(2 \arccos \frac{\triangle I}{Mr}\right) \end{aligned}$$

$\rightsquigarrow$ Area of segment 1 $= M^2 r^2 \arccos \frac{\triangle I}{Mr} - 0.5 M^2 r^2 \sin\left(2 \arccos \frac{\triangle I}{Mr}\right)$
$\rightsquigarrow$ Area of segment 2 $= \pi M^2 r^2 - M^2 r^2 \arccos \frac{\triangle I}{Mr} - 0.5 M^2 r^2 \sin\left(2 \arccos \frac{\triangle I}{Mr}\right)$
$\rightsquigarrow$ Power of seg 1 $= \frac{P}{\pi}\left(\arccos \frac{\triangle I}{Mr} - 0.5 \sin\left(2 \arccos \frac{\triangle I}{Mr}\right)\right)$
$\rightsquigarrow$ Power of seg 2 = P - $\frac{P}{\pi}\left(\arccos \frac{\triangle I}{Mr} - 0.5 \sin\left(2 \arccos \frac{\triangle I}{Mr}\right)\right)$

$\rightsquigarrow$ Diff Pwr = P - $\frac{2P}{\pi}\left(\arccos \frac{\triangle I}{Mr} - 0.5 \sin\left(2 \arccos \frac{\triangle I}{Mr}\right)\right)$

$$\rightsquigarrow DP = P\left(1 - \frac{2}{\pi}\left(\arccos\frac{M2\Delta h\sin\theta}{Mr} - 0.5\sin\left(2\arccos\frac{2M\Delta h\sin\theta}{Mr}\right)\right)\right)$$
$$\rightsquigarrow DP = P\left(1 - \frac{2}{\pi}\left(\arccos\frac{2\Delta h\sin\theta}{r} - 0.5\sin\left(2\arccos\frac{2\Delta h\sin\theta}{r}\right)\right)\right)$$

The differential power is non linear with respect to height Δh. Magnification does not affect the differential power as increase in spot size is compensated by an increase in displacement.

A.3 Mathcad listing for the circular spot relationship

The Mathcad listing included is used to evaluate differential power vs displacement for a circular spot. The photodetector width and separation are taken into account and the parameters are defined in Fig A.4.

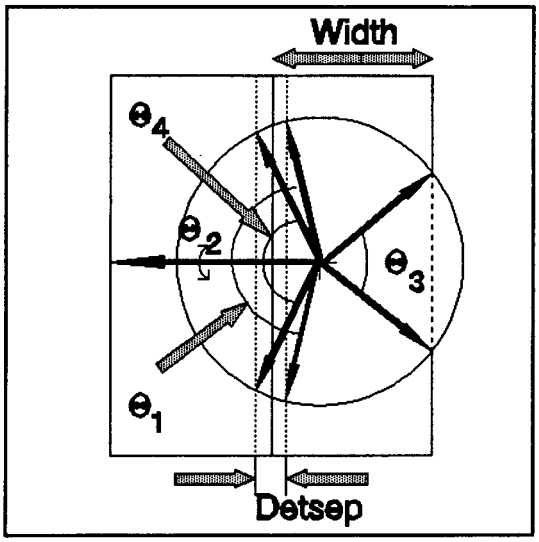


FIGURE A.4: ANGLES ASSOCIATED WITH THE CIRCULAR SPOT INCIDENT ON THE SPLIT PHOTOCELL

Relationship between change in power with height for a circular spot
on a splitphotocell. Photocell separation is DetSep.

File : Circ23.mcd

Parameters :

Width := .7 — Width of 1 photocell in mm

r := .7, .65 .. .05 — Radius of spot in mm

x := 0, .05·Width .. .99·Width — Movement of spot in mm from the centre

Pwr := 1 — Normalised power

DetSep := 0.09 — Detector separation in mm

$$DetSep := \frac{DetSep}{2}$$

Calculations :

$$Areance(r) := \frac{Pwr}{\pi \cdot r^2}$$

Areance(r)
0.65
0.753
0.884
1.052
1.273
1.572
1.989
2.698
3.537
5.093
7.958
14.147
31.831
127.324

$$\beta1(x,r) := if\left[((x - DetSep) < -r), 0, 2\cdot acos\left[\frac{DetSep - x}{r}\right]\right]$$

$$\beta1(x,r) := if\left[(x - DetSep) > 0, 2\cdot acos\left[\frac{x - DetSep}{r}\right], \beta1(x,r)\right]$$

$$\beta2(x,r) := if\left[((x + DetSep) > r), 0, 2\cdot acos\left[\frac{x + DetSep}{r}\right]\right]$$

$$\beta3(x,r) := if\left[(x + r) < Width, 0, 2\cdot acos\left[\frac{Width - x}{r}\right]\right]$$

$$\beta3(x,r) := if\left[(Width - x) < 0, \frac{\pi}{2}, \beta3(x,r)\right]$$

$$\beta4(x,r) := if\left[(r - x) < Width, 0, 2\cdot acos\left[\frac{Width + x}{r}\right]\right]$$

Area of triangles

$$AreaTx1(x,r) := 0.5\cdot r^2\cdot sin(\beta1(x,r))$$

$$AreaTx2(x,r) := 0.5\cdot r^2\cdot sin(\beta2(x,r))$$

$$AreaTx3(x,r) := 0.5\cdot r^2\cdot sin(\beta3(x,r))$$

$$AreaTx4(x,r) := 0.5\cdot r^2\cdot sin(\beta4(x,r))$$

Area of sectors

$$ASect1(x,r) := \pi \cdot r^2 \cdot \left[\frac{\theta 1(x,r)}{2 \cdot \pi}\right]$$

$$ASect2(x,r) := \pi \cdot r^2 \cdot \left[\frac{\theta 2(x,r)}{2 \cdot \pi}\right]$$

$$ASect3(x,r) := \pi \cdot r^2 \cdot \left[\frac{\theta 3(x,r)}{2 \cdot \pi}\right]$$

$$ASect4(x,r) := \pi \cdot r^2 \cdot \left[\frac{\theta 4(x,r)}{2 \cdot \pi}\right]$$

Area of Segments

$$Area1(x,r) := ASect1(x,r) - AreaTx1(x,r)$$

$$Area2(x,r) := ASect2(x,r) - AreaTx2(x,r)$$

$$Area3(x,r) := ASect3(x,r) - AreaTx3(x,r)$$

$$Area4(x,r) := ASect4(x,r) - AreaTx4(x,r)$$

Active Photocell Areas

$$a(x) := DetSep - x$$

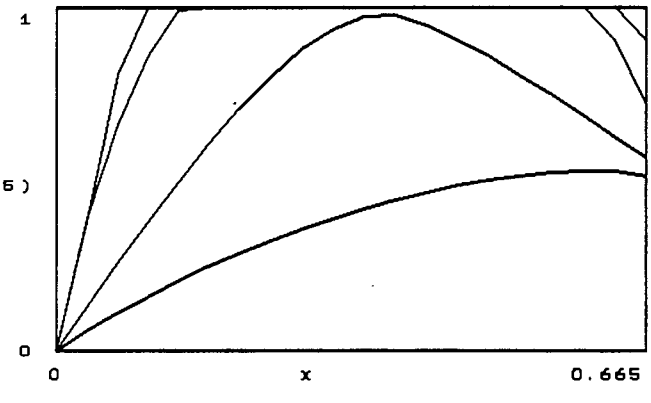
$$A1(x,r) := if\left[a(x) > 0, Area1(x,r) - Area3(x,r), \pi \cdot r^2 - Area1(x,r) - Area3(x,r)\right]$$

$$A2(x,r) := Area2(x,r) - Area4(x,r)$$

Differential Power

$$D(x,r) := Areance(r) \cdot (A1(x,r) - A2(x,r))$$

$D(x,0.7), D(x,0.35), D(x,0.1), D(x,0.05)$



Appendix B

Optimum reflection angles for maximum change in power with height for rectangular and circular spots

The angle θ dictates the differential power for a displacement $\triangle h$ on railway track. Change in power depends on absolute power and the angle at which the displacement is viewed. For small θ, the power reflected off a worst case surface ($\cos^2\theta$ distribution) is a maximum while the displacement resolution is a minimum. Alternatively, for $\theta = 90°$, resolution is largest with no reflected power. This appendix evaluates the optimum angle θ for largest change in power with vertical displacement for circular and rectangular spots. The optimum angle for circular reflection varies with displacement measured. A Mathcad listing which displays the relationship is included.

B.1 Rectangular Spot

$$\triangle P = \left(\frac{4P\cos^2\theta\sin\theta\triangle h}{x}\right)$$

To maximise $\triangle P$, $cos^2\theta\sin\theta$ must be maximised

$$\begin{aligned}
\rightsquigarrow 0 &= \frac{d}{d\theta}(cos^2\theta\sin\theta) \\
&= \cos^3\theta - 2\cos\theta\sin^2\theta \\
&= \cos^2\theta - 2\sin^2\theta = 0 \\
&= 1 - 3\sin^2\theta \\
\rightsquigarrow \theta &= \arcsin\frac{1}{\sqrt{3}}
\end{aligned}$$

B.2 Circular Spot

$$\triangle P = P\cos^2\theta\left(1 - \frac{2}{\pi}\left(\arccos\frac{2\triangle h\sin\theta}{r} - 0.5\sin\left(2\arccos\frac{2\triangle h\sin\theta}{r}\right)\right)\right) \qquad \text{(B.1)}$$

Analytically, the maxima are found by differentiation w.r.t. θ

$$\begin{aligned}
\frac{d\triangle P}{d\theta} &= -2P\cos\theta\sin\theta\left(1 - \frac{2}{\pi}\left(\arccos\frac{2\triangle hsin\theta}{r} - 0.5\sin\left(2\arccos\frac{2\triangle hsin\theta}{r}\right)\right)\right) \\
&- \frac{2}{\pi}P\cos^2\theta\left(\frac{-2\triangle h\cos\theta}{r\sqrt{1-\left(\frac{2\triangle hsin\theta}{r}\right)^2}} - 0.5\cos\left(2\arccos\frac{2\triangle hsin\theta}{r}\right)2\frac{2\triangle h\cos\theta}{r\sqrt{1-\left(\frac{2\triangle hsin\theta}{r}\right)^2}}\right)
\end{aligned}$$

$$\begin{aligned}
\rightsquigarrow 0 &= \sin\theta\left(1 - \frac{2}{\pi}\left(\arccos\frac{2\triangle hsin\theta}{r} - 0.5\sin\left(2\arccos\frac{2\triangle hsin\theta}{r}\right)\right)\right) \\
&+ \frac{4\cos\theta}{\pi}\cos\left(\arccos\frac{2\triangle hsin(\theta)}{r}\right)\frac{\triangle h\cos\theta}{\sqrt{r^2-(2\triangle h\sin\theta)^2}}
\end{aligned}$$

$$- \frac{2\triangle h \cos^2\theta}{\pi\sqrt{r^2 - (2\triangle h \sin\theta)^2}} \tag{B.2}$$

The roots of equation B.2 are difficult to evaluate analytically. Mathcad was used to plot equation B.1 as a function of θ ($0 - 90°$). These plots were then obtained for displacements of 10, 30, 50 and 90/maximum varying in the range of 20° to 40°. The Mathcad listing is as follows :

Optimum reflection angle for an Electro-Optic track monitoring system
Circular Spot
File : RefAng.mcd

Parameters :

$r := .1$ Radius of spot in mm

$h := 0.01, 0.02 .. 0.1$ Change in height

$P := 1$ Normalised Power

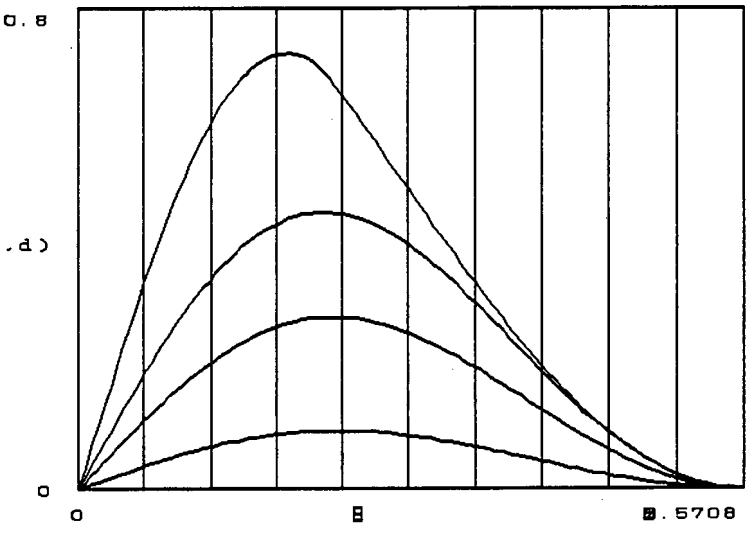
$\theta := 0, .01 \cdot \frac{\pi}{2} .. \frac{\pi}{2}$ Angle from receiver axis to the vertical

$$\tau(\theta, h) := \operatorname{acos}\left[\frac{2 \cdot h \cdot \sin(\theta)}{r}\right]$$

$$D(\theta, h) := P \cdot (\cos(\theta))^2 \cdot \left[1 - \frac{2}{\pi} \cdot (\tau(\theta, h) - 0.5 \cdot \sin(2 \cdot \tau(\theta, h)))\right]$$

D is differential power

$a := 0.01$ $b := 0.03$ $c := 0.05$ $d := 0.09$



WRITEPRN(d1dat) := D(θ,a)

WRITEPRN(d2dat) := D(θ,b)

WRITEPRN(d3dat) := D(θ,c)

WRITEPRN(d4dat) := D(θ,d)

Appendix C

Polar diagram measurement

This appendix explains how polar diagrams were measured. Four polar diagrams, one for polished steel (specular) and the other three for rusty surfaces are included. The reflectances when the transmitter and receiver are orientated for specular reflection are also given for the various surfaces.

C.1 Measurement of Polar Diagrams

C.1.1 Aim

The aim of this experiment is to measure the polar diagram of a reflective surface and to measure the specular reflectance of the surface through a standard receive lens.

C.1.2 Method

The experiment is divided into two parts, measurement of the polar diagram and specular reflectance through the receive lens. The polar diagram is calculated from measurements

made using the apparatus in FigC.1. The laser and reflective surface is fixed to a

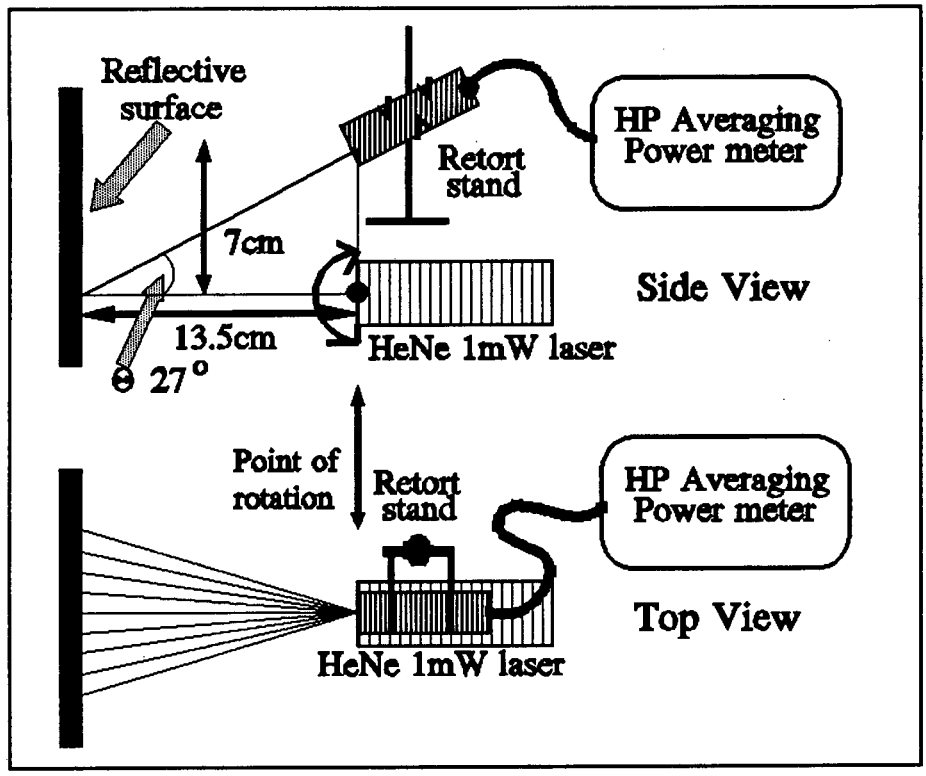


FIGURE C.1: APPARATUS USED TO MEASURE POLAR DIAGRAMS

table rotated through angle θ at the point indicated in Fig C.1. Ambient and absolute power is recorded for each angle θ and the results are plotted. The assumption that the polar diagram is symmetric in the vertical plane is valid. Polar diagrams are plotted for normalised powers.

C.1.3 Results

The results for θ vs Power are plotted for i) polished steel, ii) a rusty T piece, iii) a rusty rectangular piece and iv) a rusty sheet of metal. The next part of the experiment involves measurement of specular reflectance off rusty steel.

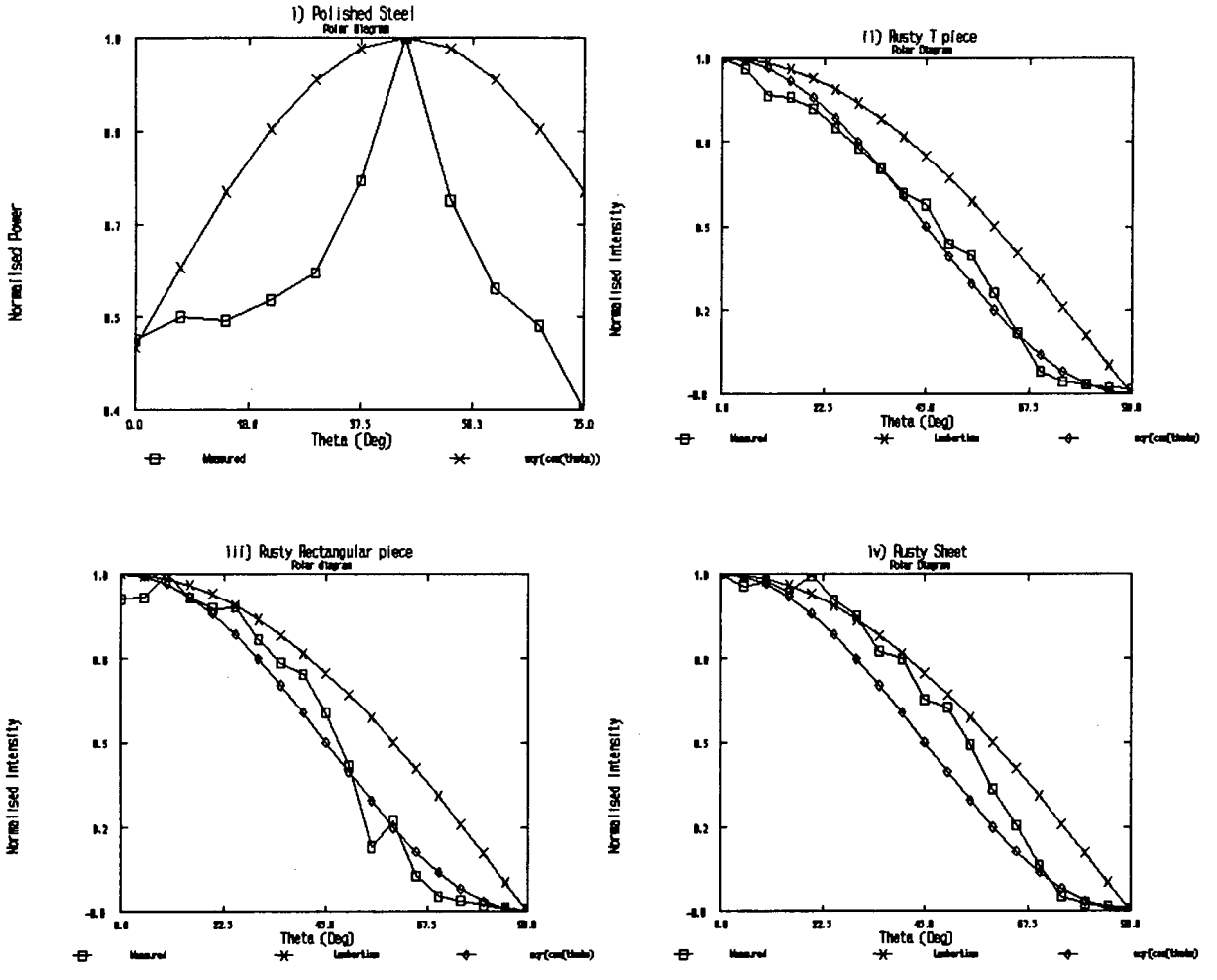


FIGURE C.2: POLAR DIAGRAMS FOR DIFFERENT PIECES OF STEEL

C.2 Measurement of specular reflectance

The specular reflectance is important for power budget calculation.

C.2.1 Aim

To measure reflectance of rusty steel through the receive lens.

C.2.2 Method

Figure C.3 shows the apparatus used in the experiment. The reflectance was calculated

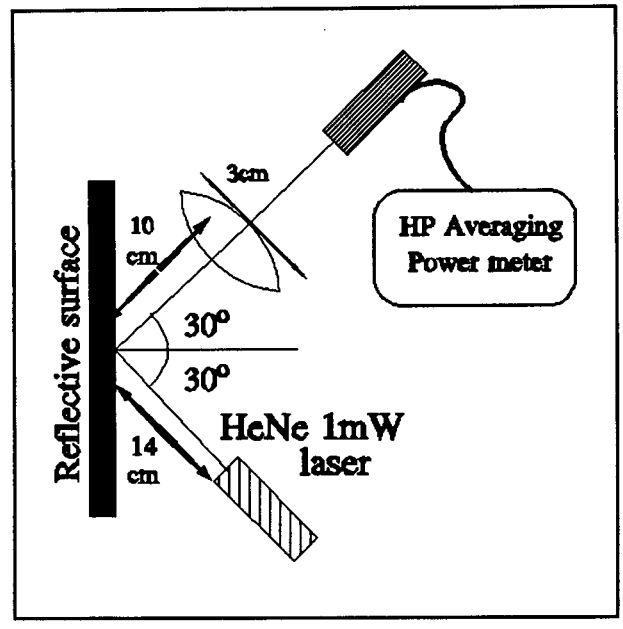


FIGURE C.3: APPARATUS USED TO MEASURE REFLECTANCE

as the ratio of reflected to transmitted power. Transmitted power was measured by placing the power meter in front of the beam and recording the reflected power. Reflected power was measured by the power meter placed at twice the focal point of the lens. The lens used has f-number 2, diameter $35mm$ and was used in initial experiments with the receiver. The angle θ was selected from the results of Appendix B $\Leftrightarrow$ it was chosen for maximum reflection.

C.2.3 Results

The results for sheets 1 to 3 (§2.3) are :

TABLE C.1: REFLECTANCES FOR RUSTY STEEL

	T	Rect	Sheet	Units
Pamb	1560	1502	1513	nW
Preflected	4.28	3.96	3.24	μW
Pincident	625	625	625	μW
Reflectance	4.35E-3	3.94E-3	2.77E-3	
Pamb	1513	1499	1493	nW
Pref	2.67	2.66	2.5	μW
Reflectance	1.86E-3	1.86E-3	1.62E-3	
Pamb	1532	1513	1469	nW
Pref	2.7	2.94	2.89	μm
Reflectance	1.87E-3	2.29E-3	2.27E-3	

Appendix D

Reflectance properties of rusty steel as a function of wavelength

D.1 Aim

To measure reflectance of rusty steel as a function of optical wavelength.

D.2 Method

The apparatus is given in Fig D.1. Infra-red, red, orange, yellow and green LEDs were used and their colour was used to estimate wavelength. LED surface lenses were ground off to decrease spot size at the focus. The worst case reflection is off rusty sheet and this was used as the reflector.

The small difference in optical wavelengths and large detector area of the power meter means that changes in focal length with wavelength is negligible. Reflected power was measured at $20mm$ from the reflective surface. The distance from the reflective surface and the point at which the LED was focused was kept the same in each experiment. On

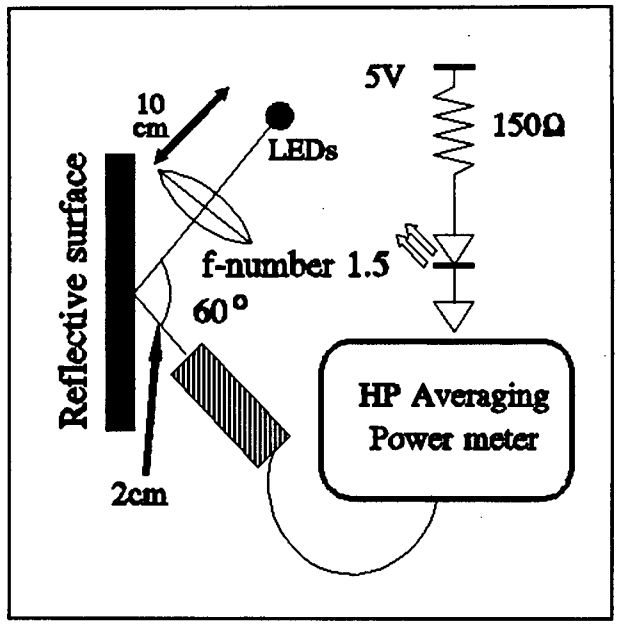


FIGURE D.1: APPARATUS USED TO MEASURE REFLECTANCE AS A FUNCTION OF WAVELENGTH

axis power was measured and reflectance calculated from the ratio of reflected to incident power.

D.3 Results

The results are shown in Figure D.2

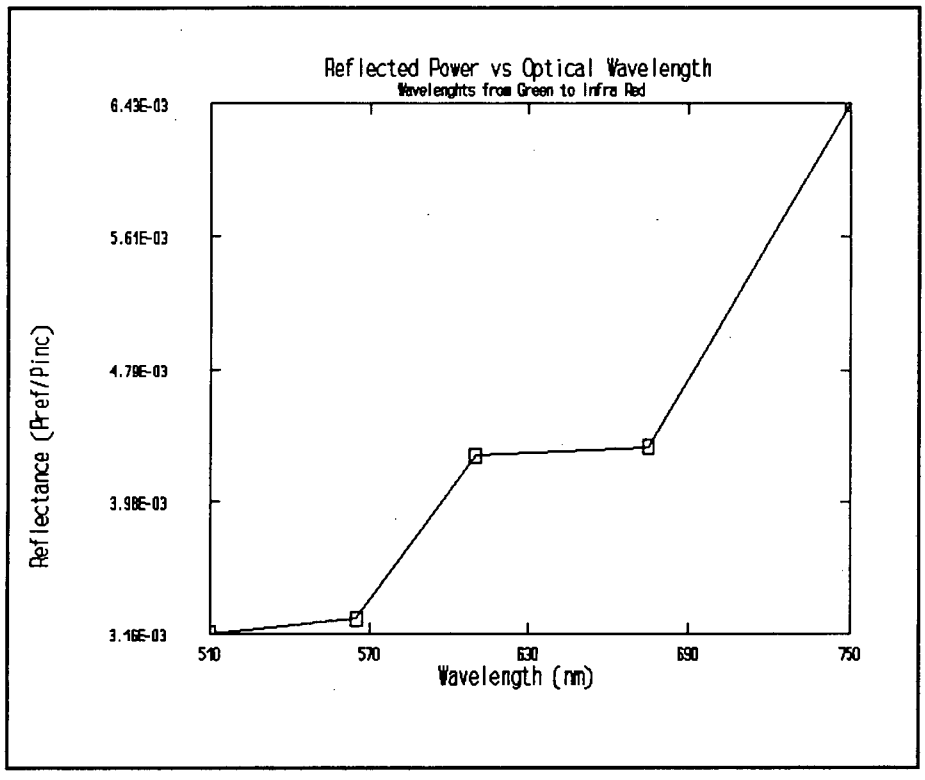


FIGURE D.2: REFLECTANCE AS A FUNCTION OF WAVELENGTH FOR RUSTY STEEL

Appendix E

Mathcad listings

This appendix contains the listings of the following Mathcad files :

Fetpwr.mcd	Calculates the signal to noise ratio and Noise Equivalent Power of a FET input end detector.
Power.mcd	Calculates the signal to noise ratio of an operational amplifier based transimpedance detector. Specifications of the LM6361 opamp are used
Pwrbug.mcd	Calculates the power budget for the electro-optic track monitor. A Lambertian source is assumed and specifications for the LM6361 opamp are used
Lpfphse.mcd	Calculates and displays the phase response of an LRC filter. The phase sensitivity is also plotted against frequency.
Probability.mcd	Calculates the probability of an event at the output of the envelope detectors.

Calculation of NEP for the integrating - high impedance front end amplifier.

Constants :

$f0 := 2 \cdot 10^{6}$ Center frequency

$BndWdth := 1 \cdot 10^{6}$ Electrical BW

$k := 1.38 \cdot 10^{-23}$ Boltzmann's const

$T := 293$ Absolute temperature

$\eta := 0.7$ Fet noise factor - 0.7 for silicon

$gm := 1 \cdot 10^{-3}$ Fet transconductance minimally 1mMho

$e := 1.6 \cdot 10^{-19}$ Electronic Charge

$Id := 200 \cdot 10^{-9}$ Worst case dark current for BPX48 diode

$Cmin := 25 \cdot 10^{-12}$ Diode Capacitance + strays

$\rho := 0.5$ Photodiode sensitivity

Calculations :

$Q := \frac{f0}{BndWdth}$ $Q = 2$ Tuned Circuit Q

$R := \frac{Q}{2 \cdot \pi \cdot f0 \cdot Cmin}$ $R = 6.366 \cdot 10^{3}$ Load Resistance

$JohnsonNoise := 4 \cdot k \cdot T \cdot BndWdth$ $JohnsonNoise = 1.617 \cdot 10^{-14}$

$FETNoise := 2 \cdot k \cdot T \cdot BndWdth \cdot \frac{\eta}{gm}$ $FETNoise = 5.661 \cdot 10^{-12}$

$ShotNoise := 2 \cdot e \cdot BndWdth \cdot Id \cdot R$ $ShotNoise = 0$

Shot Noise is very small, background is assumed negligeable. Shot noise consists of dark current term only.

Calculation of NEP

Coefficients of the Quadratic Equation :

$a := \rho^{2} \cdot R$ $b := -2 \cdot e \cdot BndWdth \cdot R \cdot \rho$

$c := -(FETNoise + ShotNoise + JohnsonNoise)$

$Pwr := \frac{-b + \sqrt{b^{2} - 4 \cdot a \cdot c}}{2 \cdot a}$ $Pwr = 6.973 \cdot 10^{-8}$

$NEP := \frac{Pwr}{\sqrt{BndWdth}}$ $NEP = 6.973 \cdot 10^{-11}$

File : Power.mcd

Calculation of NEP : Theoretical, based on Opamp and diode specs

Constants

$\eta := 0.8$ Spectral Efficiency

$e := 1.6 \cdot 10^{-19}$ Electron Charge

$k := 1.38 \cdot 10^{-23}$ Boltzman's Const

$T := 293$ Absolute temperature

$R1 := 18000$ Feedback Resistence

$B := 2 \cdot 10^{6}$ Bandwidth

$Id := 200 \cdot 10^{-9}$ Dark current

$h := 6.63 \cdot 10^{-34}$ Plank's Constant

$\lambda := 850 \cdot 10^{-9}$ Optical wavelength

$Ina := 10^{-12}$ Amplifier noise current

$c := 3 \cdot 10^{8}$ Speed of Light

$Pback := 0$ Background Power

Calculations

$f := \frac{c}{\lambda}$ $f = 3.529 \cdot 10^{14}$ Frequency of Light

Coefficients of the quadratic equation

$a := \left[\frac{\eta \cdot e}{h \cdot f}\right]^2 \cdot R1$ $a = 8.356 \cdot 10^{3}$

$b := \frac{2 \cdot e \cdot R1 \cdot B \cdot \eta \cdot e \cdot -1}{h \cdot f}$ $b = -6.302 \cdot 10^{-9}$

$c := -\left[2 \cdot e \cdot R1 \cdot B \cdot \left[Id + \frac{\eta \cdot e \cdot Pback}{h \cdot f}\right] + 4 \cdot k \cdot T \cdot B + Ina^2 \cdot R1 \cdot B\right]$

$c = -7.065 \cdot 10^{-14}$

$Power := \frac{-b + \sqrt{b^2 - 4 \cdot a \cdot c}}{2 \cdot a}$

$Power = 3.622 \cdot 10^{-9}$

$NEP := \frac{Power}{\sqrt{B}}$ $NEP = 2.561 \cdot 10^{-12}$

$ThrmNse := 4 \cdot k \cdot T \cdot B$

$AmpNse := Ina^2 \cdot R1 \cdot B$

$ShotNse := 2 \cdot e \cdot B \cdot R1 \cdot \left[Id + \frac{\eta \cdot e \cdot (Pback + Power)}{h \cdot f}\right]$

$ThrmNse = 3.235 \cdot 10^{-14}$ $AmpNse = 3.6 \cdot 10^{-14}$ $ShotNse = 2.327 \cdot 10^{-15}$

i.e. The system is thermal and amplifier noise limited !!!

File : Pwrbug.mcd

Power budget for specular reflection off rusty steel : Lambertian Transmitte

Units :

$$Deg := \frac{\pi}{180}$$

Transmitter Parameters

$PTx := 1 \cdot 10^{-3}$
TxLensDiam := .025
TxLensLoss := .92
TxFocalLength := .05
SceLensDist := .1
TxCosPower := 1

Receiver Parameters

RxAngleToVert := 15·Deg
RxLensDiam := .025
RxLensLoss := .92
Responsivity := 0.573
GainTr := 18000
GainAmp1 := 25
GainAmp2 := 25
RxFocalLength := .05
LensDetDist := .1
Rl := 18000
$Ina := 1 \cdot 10^{-12}$
$B := 2 \cdot 10^{6}$
$Id := 200 \cdot 10^{-9}$

Pback := 0

Reflective Surface Parameters

$Reflectance := 3.42 \cdot 10^{-3}$

RfCosPower := 2

Constants

$k := 1.38 \cdot 10^{-23}$
T := 293
$e := 1.6 \cdot 10^{-19}$

$$TxPointance := \frac{PTx}{\pi} \qquad TxPointance = 3.183 \cdot 10^{-4}$$

$$\theta max := \operatorname{atan}\left[\frac{TxLensDiam}{2 \cdot SceLensDist}\right] \qquad \theta max = 0.124$$

$$PowerTxLens := 2 \cdot \pi \cdot TxPointance \cdot \int_0^{\theta max} \cos(\theta)^{TxCosPower} \cdot \sin(\theta)\, d\theta$$

$PowerTxLens = 1.538 \cdot 10^{-5}$ PowerToSurface := PowerTxLens ·TxLensLoss

$$PowerToSurface := 1.415 \cdot 10^{-5}$$

$$PwrCapturedByRx := Reflectance \cdot PowerToSurface$$

$$PwrCapturedByRx = 4.841 \cdot 10^{-8}$$

$$PowerToDet := PwrCapturedByRx \cdot RxLensLoss$$

$$PowerOn1Det := \frac{PowerToDet}{2}$$

$$Sig := (Responsivity \cdot PowerOn1Det)^2 \cdot Rl$$

$$Nse := 2 \cdot e \cdot Rl \cdot B \cdot (Id + Responsivity \cdot (PowerOn1Det + Pback)) + 4 \cdot k \cdot T \cdot B$$

$$Nse := Nse + Ina^2 \cdot Rl \cdot B$$

$$Nse = 7.08 \cdot 10^{-14}$$

$$SignalToNoise := \frac{Sig}{Nse} \qquad SignalToNoise = 41.388$$

Sensitivity of phase as a function of frequency for a low pass filter.

File : Lpfphse.mcd

$flo := 100000$

$fhi := 1000000$

$f := flo, 1.2 \cdot flo \ldots fhi$

$L := 1.5 \cdot 10^{-3}$

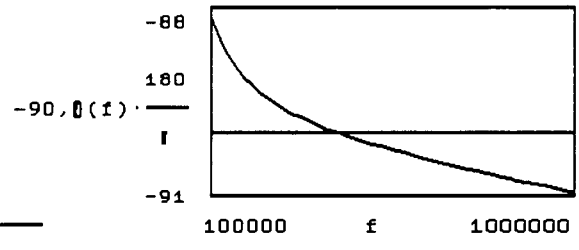
$R := 470000$

$C := 100 \cdot 10^{-12}$

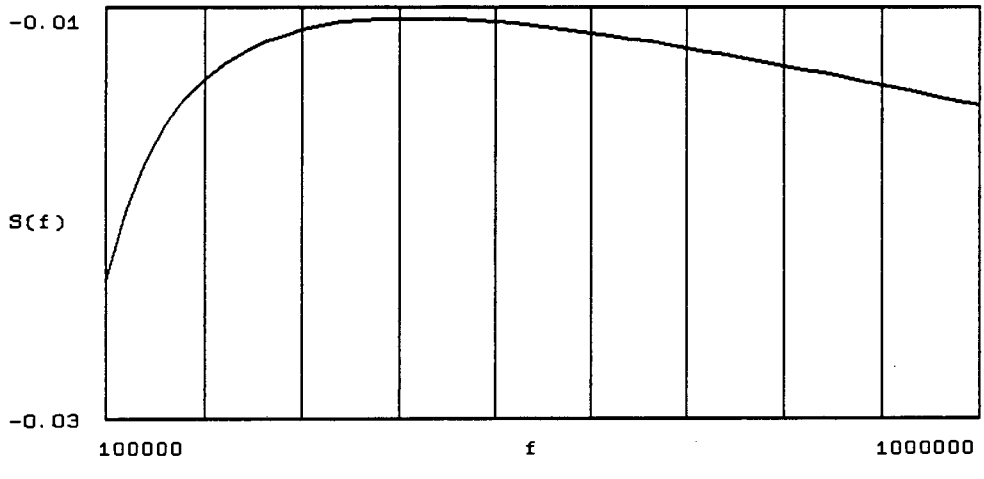
$w(f) := 2 \cdot \pi \cdot f$

Phase sensitivity with respect to radian frequency

$$\theta(f) := \frac{-\pi}{2} - \operatorname{atan}\left[\frac{w(f)^2 \cdot L \cdot C - 1}{w(f) \cdot R \cdot C}\right]$$



$$S(f) := \frac{w(f)}{\theta(f)} \cdot \frac{R \cdot C \cdot \left[w(f)^2 \cdot L \cdot C + 1\right]^2}{\left[w(f)^2 \cdot L \cdot C - 1\right]^2 + (w(f) \cdot R \cdot C)^2}$$



File : Probability.mcd
Calculates the relevant probabilities associated with the computation of the probaility of a 100um imperfection being missed

2V random variable

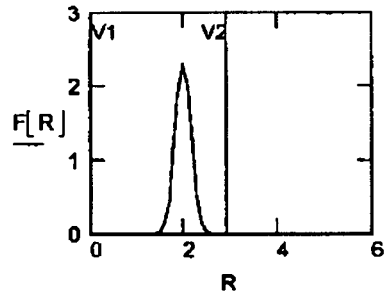
Limits V1 := 0.05 V2 := 2.9

Sine amplitude and RMS deviation A = 2.0 σ := 0.03

$$F(R) := \frac{R}{\sigma}\cdot\exp\left[\frac{-\left[R^2+A^2\right]}{(2\cdot\sigma)}\right]\cdot I0\left[\frac{(R\cdot A)}{\sigma}\right] \qquad R := 0.5, 0.6 .. 4$$

$$P := \int_{V1}^{V2} F(R)\, dR$$

P = 0.9999998



Mean

$$M := \int_{0}^{3} R\cdot F(R)\, dR$$

M = 2.0075141

0.5V random variable

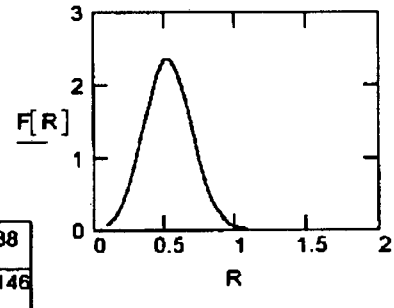
Limits V1 := .3 V2 := .7

Sine amplitude and RMS deviation A := 0.50 σ = 0.03

$$F(R) := \frac{R}{\sigma}\cdot\exp\left[\frac{-\left[R^2+A^2\right]}{(2\cdot\sigma)}\right]\cdot I0\left[\frac{(R\cdot A)}{\sigma}\right] \qquad R := 0.1, 0.15 .. 1.5$$

$$P := \int_{V1}^{V2} F(R)\, dR$$

P = 0.7599521



Mean R := 4, 5 .. 6

$$M := \int_{0}^{3} R\cdot F(R)\, dR$$

M = 0.5311737

F(R)
$1.4004197\cdot10^{-88}$
$1.9436461\cdot10^{-146}$
$8.8243836\cdot10^{-219}$

Appendix F

Information relevant to design and operation of the receiver

This appendix includes :

- Analysis of the low pass second order LRC Butterworth filter.
- Measurement of frequency response of the receiver.
- Choice of component values in a passive envelope detector.
- Measurement of active photocell area and responsivity for a BPX-48 silicon photocell.

F.1 LRC Butterworth filter analysis

The LRC low pass filter is illustrated in Fig F.1. The transfer function is calculated from :

$$V_o = \frac{\frac{1}{sC}}{R + sL + \frac{1}{sC}} V_i$$

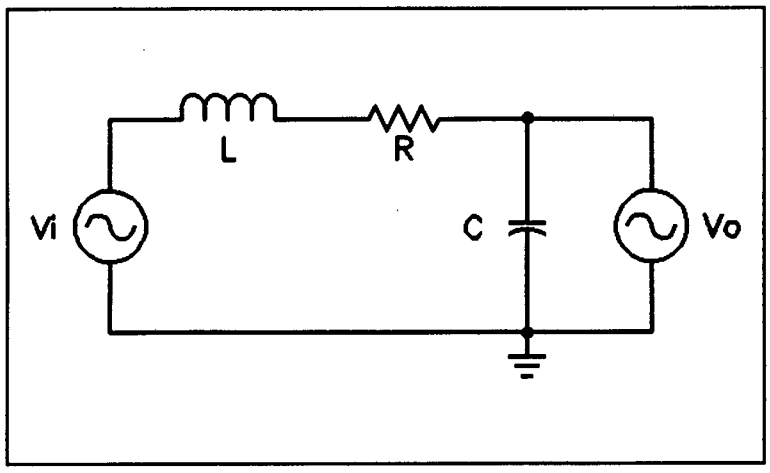


FIGURE F.1: CIRCUIT DIAGRAM OF A LRC SECOND ORDER BUTTERWORTH FILTER

$$
\begin{aligned}
&= \frac{1}{s^2LC + sRC + 1} V_i \\
\frac{V_o}{V_i} &= \frac{1}{LC} \frac{1}{s^2 + sRC + 1} \\
\rightsquigarrow \omega_c &= \frac{1}{\sqrt{LC}} \\
\rightsquigarrow Q &= \frac{1}{R}\sqrt{\frac{L}{C}}
\end{aligned}
$$

For the Butterworth case :

$$
\begin{aligned}
Q &= \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} &= \frac{1}{R}\sqrt{\frac{L}{C}}
\end{aligned}
$$

$2\pi\omega_c$ is the $3dB$ cutoff frequency at which phase is 90°
Additional constraints are :

- Input impedance > Source impedance at resonance, $Z_{sce} < R$
- Load impedance > Output impedance at resonance, $Z_L > \sqrt{\frac{L}{C}}$

F.1.1 Sensitivity

Sensitivity is a measure of how response changes if a component value is changed. Phase response is of interest in the low pass filter after envelope detection in the receiver circuit. The differential process requires phase of the two input signals be identical. Consequently if phase is sensitive to changes in component values, there must be proviso for tuning, or higher tolerance components must be used. Daryanani [5] gives equation for sensitivity of value a to component b as :

$$S_b^a = \frac{b}{a}\frac{da}{db}$$

In this case, change in phase with resonant frequency is of interest which leads to :

$$\begin{aligned}
\phi &= -arctan\frac{wRC}{1-w^2LC} \\
S_w^\phi &= \frac{w}{\phi}\frac{d\phi}{dw} \\
\frac{d\phi}{dw} &= \frac{d\phi}{dx} \times \frac{dx}{dw}
\end{aligned}$$

where

$$x = \frac{wRC}{1-w^2LC}$$

which leads to

$$\begin{aligned}
\frac{d\phi}{dx} &= \frac{1}{1+x^2} \\
&= \frac{(1-w^2LC)^2}{(1-w^2LC)^2+(wRC)^2} \\
\frac{dx}{dw} &= \frac{RC}{1-w^2LC} - \frac{wRC}{(1-w^2LC)^2} \times -2wLC \\
&= \frac{RC - w^2RLC^2 + 2w^2RLC^2}{(1-w^2LC)^2} \\
&= \frac{RC\,(1+w^2LC)}{(1-w^2LC)^2} \\
\frac{d\phi}{dw} &= \frac{RC\,(1+w^2LC)}{(1-w^2LC)^2+(wRC)^2}\frac{w}{-\arctan\frac{wRC}{1-w^2LC}}
\end{aligned}$$

The phase response and sensitivity was plotted in Mathcad as a function of frequency (see Appendix E for listing). The phase sensitivity as a function of frequency is illustrated in

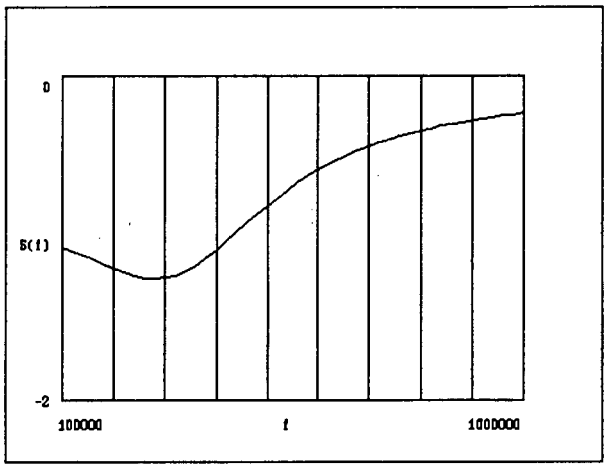


FIGURE F.2: PHASE SENSITIVITY AS A FUNCTION OF FREQUENCY FOR FOR A SECOND ORDER BUTTERWORTH FILTER

FigF.2 and is most sensitive at the resonant frequency ($400kHz$) and therefore tuning should be provided.

F.2 Frequency Response of the receiver

In order to measure the frequency response of the receiver the change in light output with subcarrier frequency for the source must be measured first.

F.2.1 Calibration of the source

The circuit in Figure F.3 is used to measure the light output of the LED source as a function of frequency.

The 47Ω resistors are used to terminate and match the synthesizer and spectrum analyzer. The low pass RC filter is used to filter transients on the supply. Assuming the sum of analyzer capacitance, stray and diode capacitance is less than $35pF$, the receiver diode has a $97MHz$ bandwidth. The HP signal synthesizer was used to subcarrier modulate

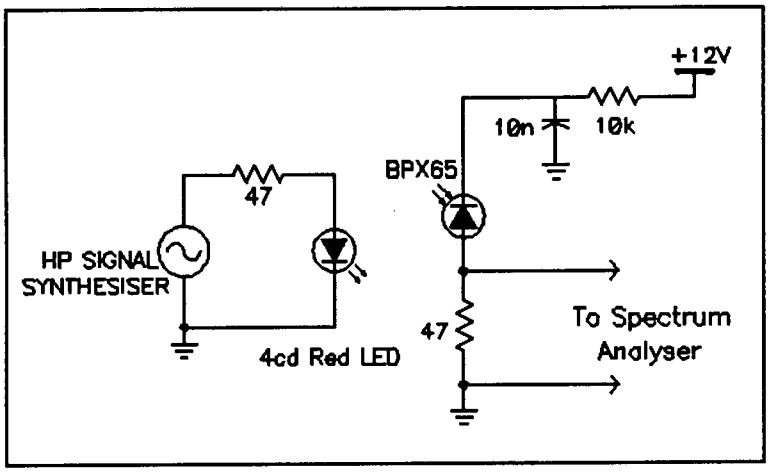


FIGURE F.3: CIRCUIT USED TO MEASURE THE LIGHT OUTPUT POWER OF THE LED AS A FUNCTION OF FREQUENCY

the LED source from $500kHz$ - $13MHz$ and the spectrum analyzer was used to record the equivalent amplitude on the detector.

FigF.4 shows the frequency response of the hi-brite LED.

F.2.2 Frequency response

The next stage is to measure frequency response of one of the front end amplifier/demodulator stages taking the normalised function of FigF.4 into account. The apparatus and receiver circuit are illustrated in Figure F.5 Output voltage V_o was recorded as a LED frequency was varied. The results were then corrected using the graph in Figure F.4. Figure F.6 illustrates the frequency response of the receiver.

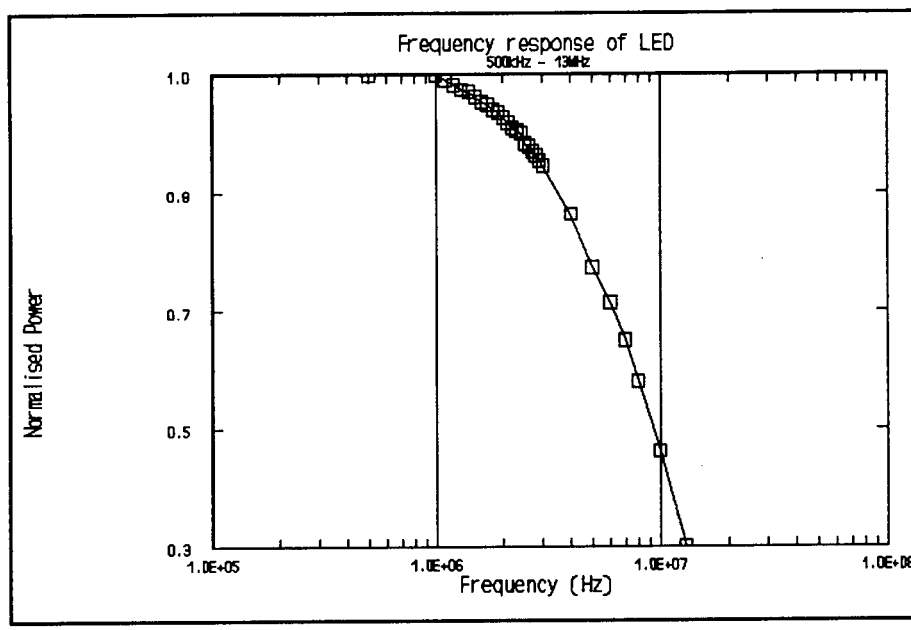


FIGURE F.4: FREQUENCY RESPONSE OF LED

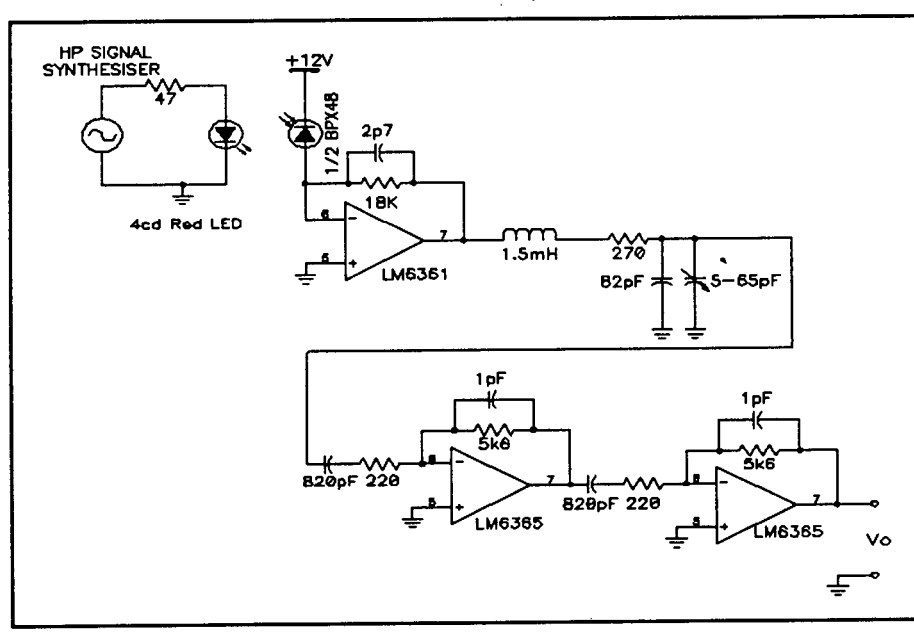


FIGURE F.5: THE APPARATUS USED TO MEASURE RECEIVER FREQUENCY RESPONSE

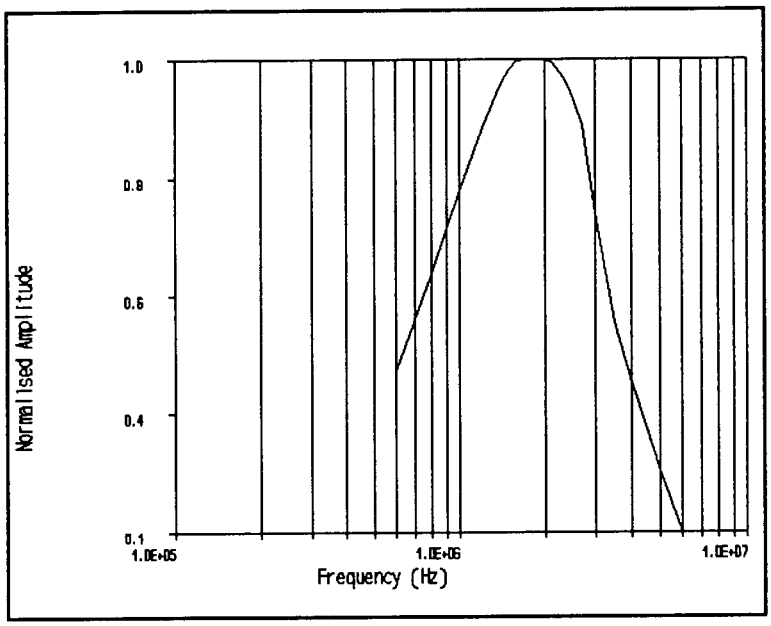


FIGURE F.6: FREQUENCY RESPONSE OF THE RECEIVER

F.3 Choice of suitable component values in a passive envelope detector

The circuit of a passive envelope detector is given in FigF.7. The aim of this section is to obtain component values for R and C so the signal is optimally demodulated. The

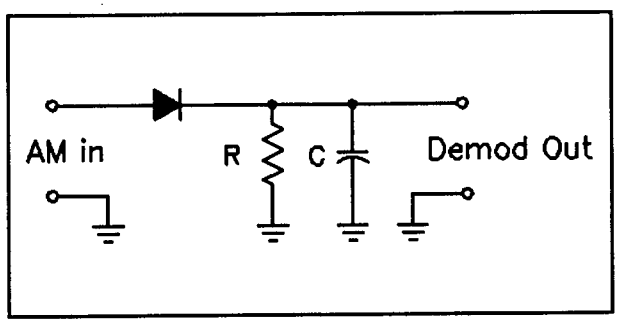


FIGURE F.7: A PASSIVE ENVELOPE DETECTOR

design criteria are :

- $4RC \gg \frac{1}{f_c}$ where f_c is carrier frequency and $4RC$ is the time taken for the envelope detector to reach 99% of its final value.
- Discharge of the capacitor must be faster than rate of change of the modulating signal.
- $Z_{load} \gg R\sqrt{\frac{1}{1+(wRC)^2}}$ or $Z_{load} \gg R$

The second criterion warrants more attention :
The discharge function of RC network is :

$$V = V_o \exp \frac{-t}{RC}$$

and the charge time is almost instantaneous assuming the driving source resistance and on resistance of the diode is negligible. The rate of change of discharge voltage is :

$$\frac{dV}{dt} = \frac{-V_o}{RC} \exp \frac{-t}{RC}$$

Assuming there is sine wave modulation at f_m then

$$V_m = mV_o \sin(2\pi f_m t)$$

where m is the modulation index. The rate of change of the modulation envelope is thus

$$\frac{dV_m}{dt} = 2\pi f_m V_o \cos(2\pi f_m t)$$

and RC must be found such that

$$\frac{V_o}{RC} \exp \frac{-t}{RC} \geq 2\pi f_m V_o \cos(2\pi f_m t)$$

in the time interval $0 \leq t \leq \frac{1}{2f_m}$. The modulation index, m, is 1 for worst case conditions. For RC small the rate of change of modulation simplifies to

$$\begin{aligned} \frac{dV_{exp}}{dt}_{m} in &= \frac{-V_o}{RC} \exp \frac{-1}{2f_m RC} \\ &\approx \frac{-V_o}{RC} \end{aligned}$$

Also

$$\frac{dV_m}{dt}|_{max} = 2\pi f_m V_o$$
$$\rightsquigarrow f_m \leq \frac{1}{2\pi RC}$$

For $f_m = 400kHz$,

$$RC \leq 3.98 \times 10^{-7}$$

Also,

$$4RC \gg \frac{1}{2 \times 10^6}$$
$$\rightsquigarrow 1.25 \times 10^{-}7 \ll RC \ll 3.98 \times 10^{-7}$$

Choosing $R = 10k\Omega$, $C = 39pF$ implies $RC = 3.9 \times 10^{-7}$ and $Z_{load} \gg 10k\Omega$ which is satisfactory.

F.4 Measurement of the BPX 48 silicon photodetector responsivity and active area

The aim of this section is to measure responsivity and active area of one of the photocells in the silicon photodiode.

F.4.1 Measurement of detector responsivity

The apparatus in Figure F.8 is used to measure responsivity and the HeNe laser is placed in front of a $0.8mm$ diameter pinhole so the imaged spot falls completely on the photocell. The Hewlett Packard averaging power meter is used to measure ambient and signal power. One of the photocells in the BPX-48 photodiode is fully illuminated and the corresponding voltage recorded. The current I in the photodiode is

$$I = \frac{V}{68 \times 10^3}$$

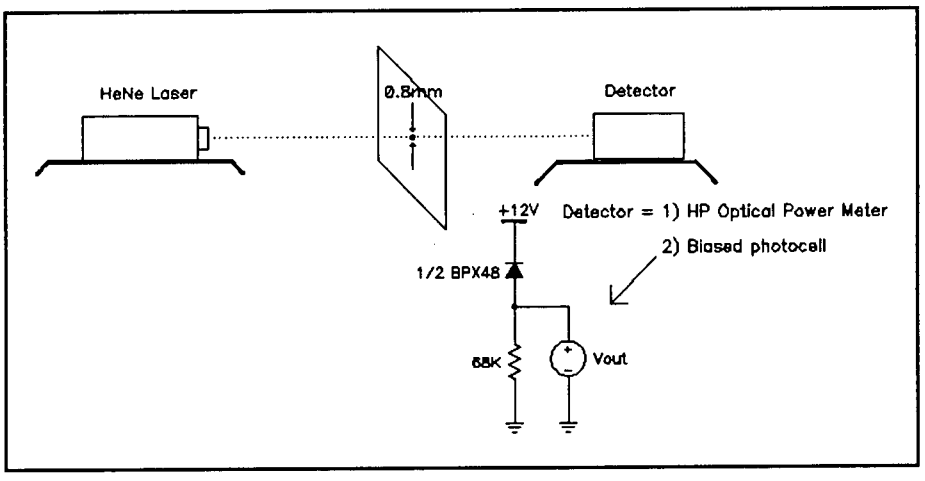


FIGURE F.8: APPARATUS USED TO MEASURE OPTICAL RESPONSIVITY

and responsivity is calculated from

$$R = \frac{I}{P_{meter} - P_{amb}}$$

Note that the signal to noise ratio is significantly high to ignore the noise. The measurements were as follows

$$\begin{aligned} P_{meter} &= 27.3\mu W \\ P_{amb} &= 607nW \\ V &= 1.04V \\ \rightsquigarrow I &= \frac{1.04}{68 \times 10^3} \\ &= 15.3\mu A \\ \Longrightarrow R &= \frac{15.3}{26.7} \\ &= 0.573\frac{A}{W} \end{aligned}$$

This is in agreement with the specifications of the BPX 48 photocell.

F.4.2 Active Detector Area

The apparatus in FigF.9 is used to measure active photocell area. The source is fixed on

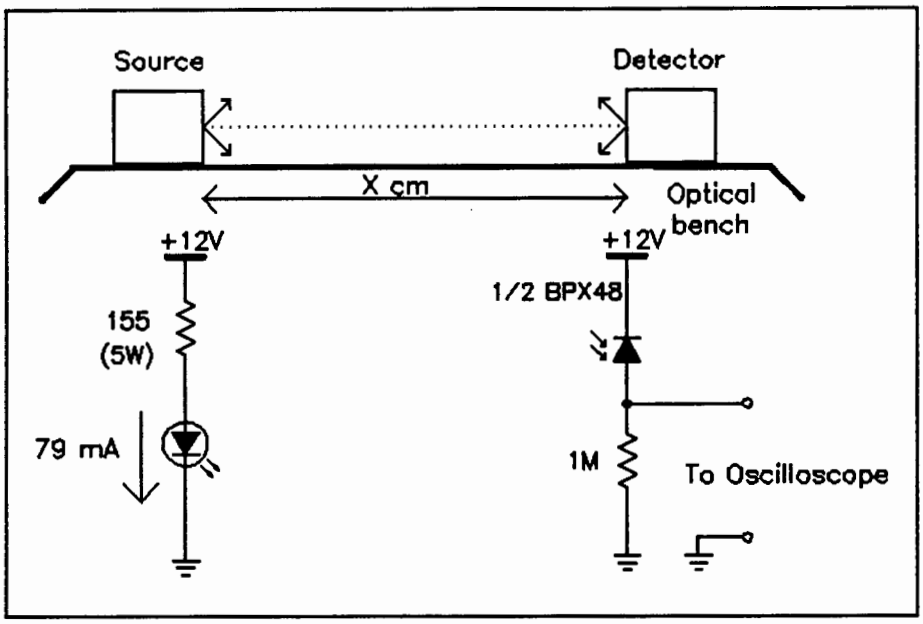


FIGURE F.9: APPARATUS USED TO MEASURE DETECTOR AREA

an optical bench on axis with the photodiode distance Xcm away. The $1M\Omega$ resistance was measured with a digital ohmmeter and voltage V was recorded. The optical power meter is then placed on axis at X and the power is recorded. Detector area is then calculated from :

$$P_{det} = \frac{V_{sig} - V_{amb}}{\rho R}$$

and

$$P_{det} = \frac{IA_d}{r^2}$$

where

$$I = \frac{P_{meter} r^2}{A_{pmeter}}$$

$$\rightsquigarrow A_d = \frac{V_{sig} - V_{amb}}{\rho R} \frac{r^2 A_{pmeter}}{P_{pmeter} r_2^2}$$

But

$$\rho = 0.573$$

$$
\begin{aligned}
V_{sig} &= 0.039 \\
V_{amb} &= 0.039 \\
R &= 1.021 \times 10^6 \\
r &= r_2
\end{aligned}
$$

$$
\begin{aligned}
\rightsquigarrow A_{pmeter} &= \pi \frac{5.1^2}{2} \\
&= 20.4mm^2
\end{aligned}
$$

and

$$
\begin{aligned}
P_{meter} &= P_{sig} - P_{amb} \\
&= 987 \times 10^{-9} - 663 \times 10^{-9} \\
\rightsquigarrow A_d &= 1.51 \times 10^{-6} m^2
\end{aligned}
$$

The specifications give the area as

$$0.7 \times 2.2 = 1.54 \times 10^{-6} m^2$$

and the measured value is within the tolerance.

Appendix G

Measurement of noise and noise equivalent power in the receiver

This appendix presents

- Calculation of noise of the stage following the transimpedance amplifiers.
- Measurement of noise equivalent power.

G.1 Calculation of inverting amplifier input noise

The input noise of the lowpass LRC filter and inverting amplifier in FigG.1 is calculated to show that it insignificantly contributes to the overall noise in the system. The noise voltage is calculated in four phases :

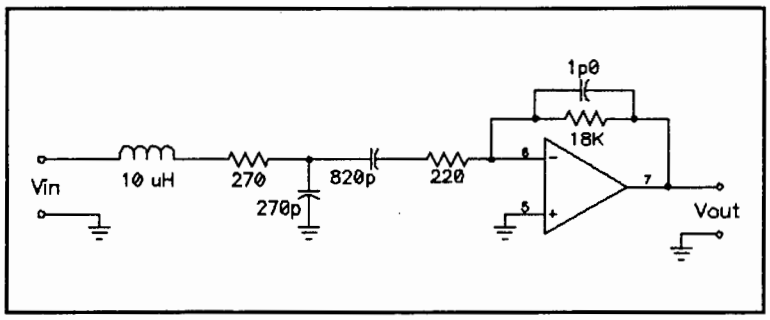


FIGURE G.1: LRC FILTER AND INVERTING AMPLIFIER STAGE IN THE RECEIVER

G.1.1 Inverting amplifier input noise voltage

Current flowing into the 220Ω resistor is the sum of Johnson noise current of the $18k\Omega$ resistor and amplifier noise current.

$$i = \sqrt{i_{j18K}^2 + i_n^2}$$

where

$$\begin{aligned} i_{j18K}^2 &= \frac{4KTB}{18000} \\ &= 1.6 * 10^{-18} A^2 \end{aligned}$$

where $B = 2MHz$

$$\begin{aligned} i_n^2 &= (10^{-12} * \sqrt{2 * 10^6})^2 \\ &= 2 * 10^{-18} \end{aligned}$$

Thus

$$i = 1.9 * 10^{-9} A$$

The input noise voltage is

$$e_{nin} = \sqrt{e_n^2 + (iR_{220})^2}$$

where e_n is amplifier noise voltage

$$\begin{aligned} e_n &= 10 * 10^{-9} * \sqrt{2 * 10^6} \\ &= 14 * 10^{-6} \end{aligned}$$

Thus

$$\begin{aligned} \rightsquigarrow e_{nin} &= \sqrt{(14 * 10^{-6})^2 + 1.74 * 10^{-15}} \\ &= \sqrt{2 * 10^{-10}} \\ &= 14\mu V \end{aligned}$$

G.1.2 Second order filter

The only noise in the LRC filter is the Johnson noise in the resistor

$$\begin{aligned} e_{jnse} &= \sqrt{4KTBR_{270}} \\ &= 3.6\mu V \end{aligned}$$

G.1.3 Total input noise

Total uncorrelated input noise is the vector sum of the noise voltages :

$$\begin{aligned} e_n &= \sqrt{14^2 + 3.6^2}\mu V \\ &= 14.46\mu V \end{aligned}$$

G.1.4 Noise due to the transimpedance and amplifier stage

The total uncorrelated input noise of one photocell amplifier stage is the vector sum of the transimpedance noise ($46\mu V$) and amplifier noise voltages.

$$\begin{aligned} e_{nse} &= \sqrt{e_n^2 + e_{ntrans}^2} \\ &= \sqrt{14^2 + 46^2} \\ &= 48\mu V \end{aligned}$$

The LRC amplifier adds $2\mu V$ of noise which is insignificant.

G.2 Measurement of noise equivalent power

The aim of this section is to measure noise equivalent power of the receiver. The first requirement is to calibrate a suitable source so that its optical power versus distance characteristic is known. The fraction of subcarrier power to total optical output power must also be measured before the NEP can be calculated.

G.2.1 Source Calibration

Source calibration entails measuring the optical power vs distance characteristic for the source. This is done with the apparatus in FigG.2. The optical power meter is fixed and the source is moved away from it at $5mm$ intervals. Average optical power is recorded

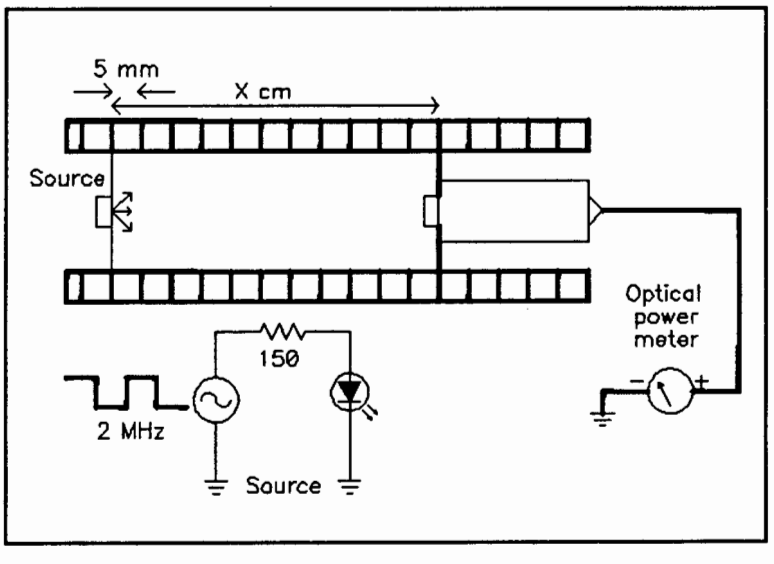


FIGURE G.2: APPARATUS USED TO CALIBRATE TRANSMITTER IN NEP MEASUREMENT

for each interval. The source is an infra-red diode ($\lambda = 950nm$) driven with a $2MHz$ square wave. Peak current through the LED is $150mA$ and it is a pseudo Lambertian emitter. A linear relationship of log power versus log distance is expected since $P = \frac{P_o A}{r^2}$ and $log(P) = log(P_o A) - 2log(r)$.

The power vs distance profile is shown in FigG.3. The linear relationship is clearly

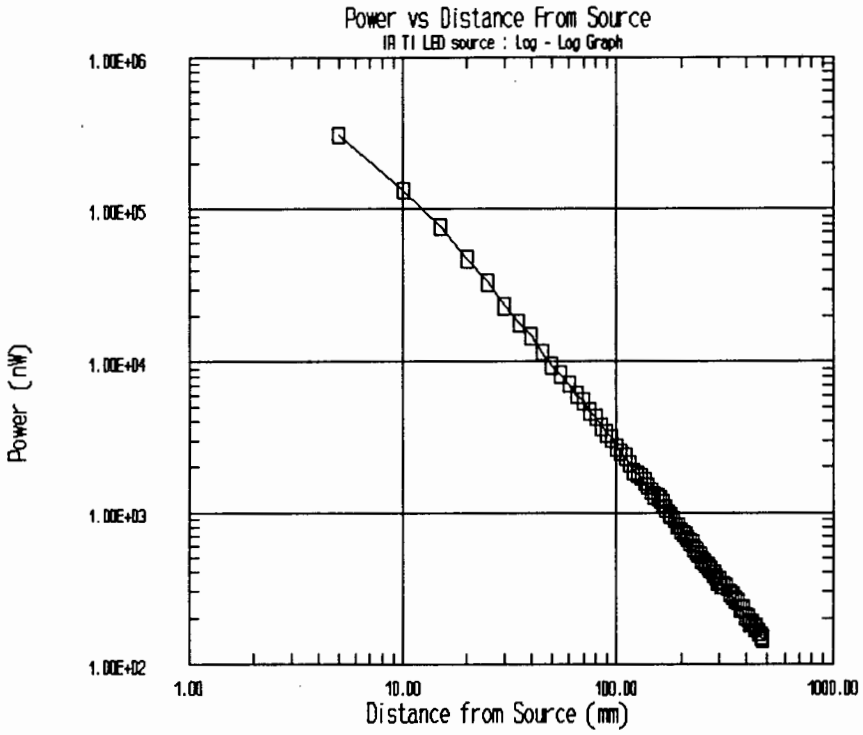


FIGURE G.3: RELATIONSHIP BETWEEN POWER AND DISTANCE

observed.

G.2.2 Noise equivalent power measurement

The power versus distance curve above does not take into account that the area of each of the split photodiodes is smaller than that of the power meter. In addition, the percentage of optical power at the subcarrier frequency is unknown and must be measured. It is therefore necessary to measure the effective power seen by the receiver at at least one point on the curve.

A theoretical estimate of the conversion factor can be calculated by considering the percentage decrease in detector area and the fraction of power in the fundamental of the driving source (modulation depth and harmonic content). For the arrangement above this was $\frac{1}{14.6}$ and $\frac{1}{4.5}$ respectively which means the values in the graph are of the order of

66 times larger than expected.

The apparatus in Figure G.4 was used to calibrate Figure G.3 at a distance of $6.5cm$

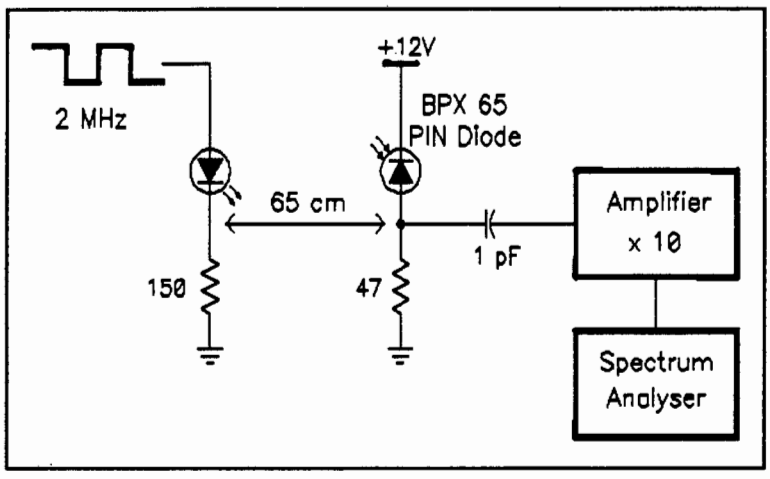


FIGURE G.4: APPARATUS USED TO CALIBRATE THE POWER CURVE

from the source. Amplifier gain is measured and the incident optical power is then calculated using the known responsivity and detector area. This is then related to a BPX 48 photodetector by multiplying by the ratio of detector areas.

The distance of $6.5mm$ was chosen as a compromise between a large signal to noise ratio and the minimum distance for the source to be Lambertian radiating. A large signal to noise ratio means an accurate measurement of the signal power can be made ignoring the noise. On the other hand, the complex near field radiation pattern of the source must not affect the calibration, and the measurement must be taken far enough away so that the source can be treated as Lambertian.

Noise equivalent power of the receiver in Chapter 3 is measured when the signal to noise ratio is unity. The receiver is moved away from the source until this is achieved and the distance recorded. This distance corresponds to a certain source power which, when divided by the square root of bandwidth, corresponds to the NEP. The results of the measurements were as follows :

Amplifier gain

A peak to peak $100mV$ sine wave was amplified to give a voltage of $1.035V$. This corresponds to a measured gain of 10.4.

Power at $65mm$

The voltage measured on the spectrum analyzer is $11\mu V$. Hence the input voltage to the amplifier is

$$\begin{aligned} V_{ia} &= \frac{11}{10.4} \\ &= 1.06\mu V \end{aligned}$$

This corresponds to a signal current of

$$\begin{aligned} i_s &= \frac{1.06}{47} \\ &= 22.6nA \end{aligned}$$

The effective power on the PIN diode is :

$$\begin{aligned} P_{pin} &= \frac{22.6}{0.55}nW \\ &= 41.0nW \end{aligned}$$

Areas of the PIN and BPX 48 photodiodes are :

$$\begin{aligned} A_{pin} &= 1mm \times 1mm \\ &= 1mm^2 \\ A_{BP48} &= 0.7 \times 0.22 \\ &= 1.54mm^2 \end{aligned}$$

Thus the effective power on the split photodetector is $63.14nW$ when it is $6.5cm$ from the source. The powers given in Figure G.3 must be divided by 95 to represent the power seen by the BPX 48 photocell.

NEP measurement

The receiver was $25.5cm$ from the source when the subcarrier signal was of order of the noise. This corresponds to a signal power of $473nW$ in Figure G.3 which when divided by 95 is the effective power seen by the receiver. This is $5.0nW$ and the NEP is thus

$$\begin{aligned} NEP &= \frac{5.0 * 10^{-9}}{\sqrt{2.0 * 10^{6}}} \\ &= 3.5 * 10^{-12} \frac{W}{\sqrt{Hz}} \end{aligned}$$

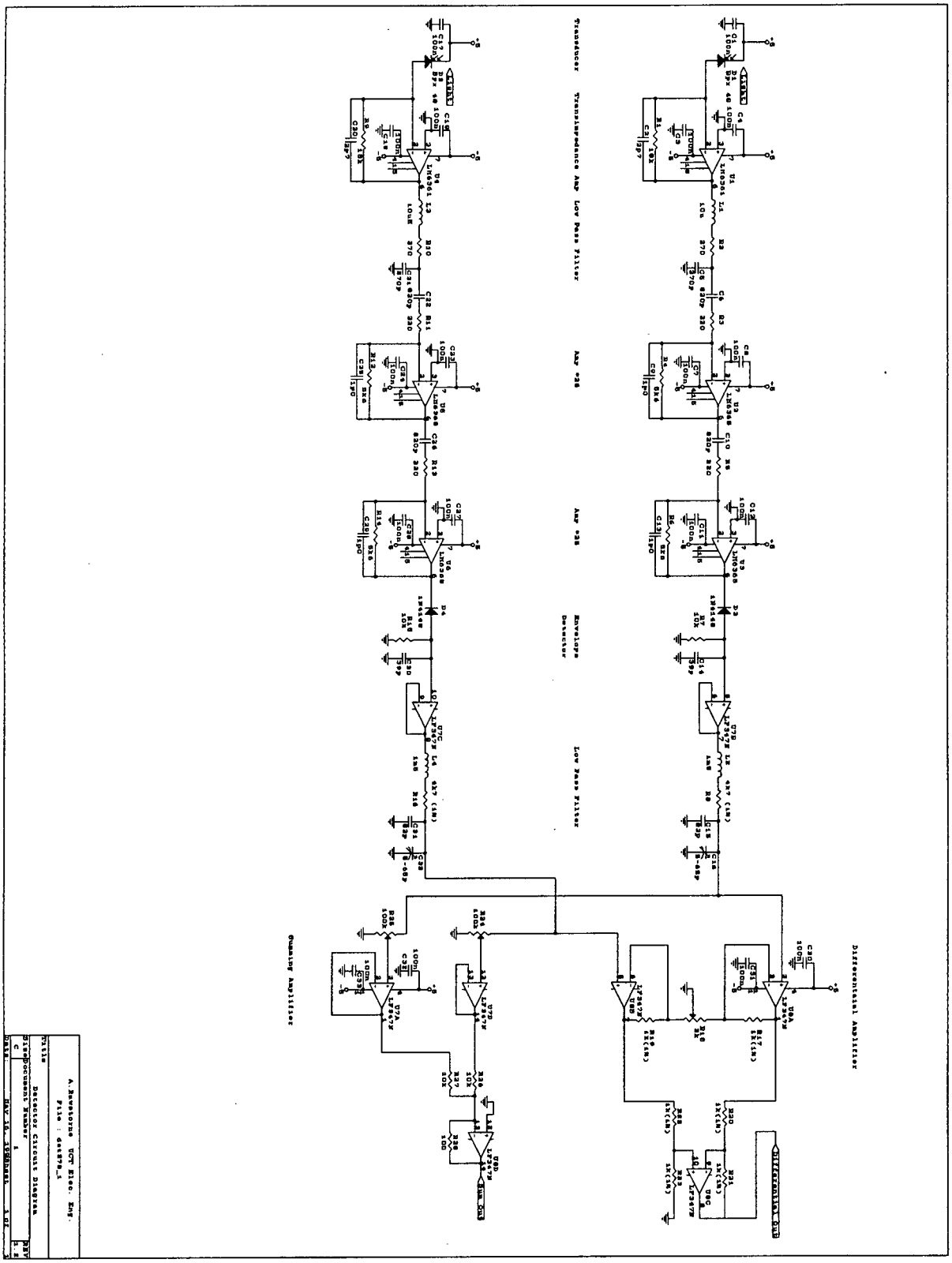
This value is larger than predicted by 20% which is acceptable given measurement inaccuracies and the assumptions made.

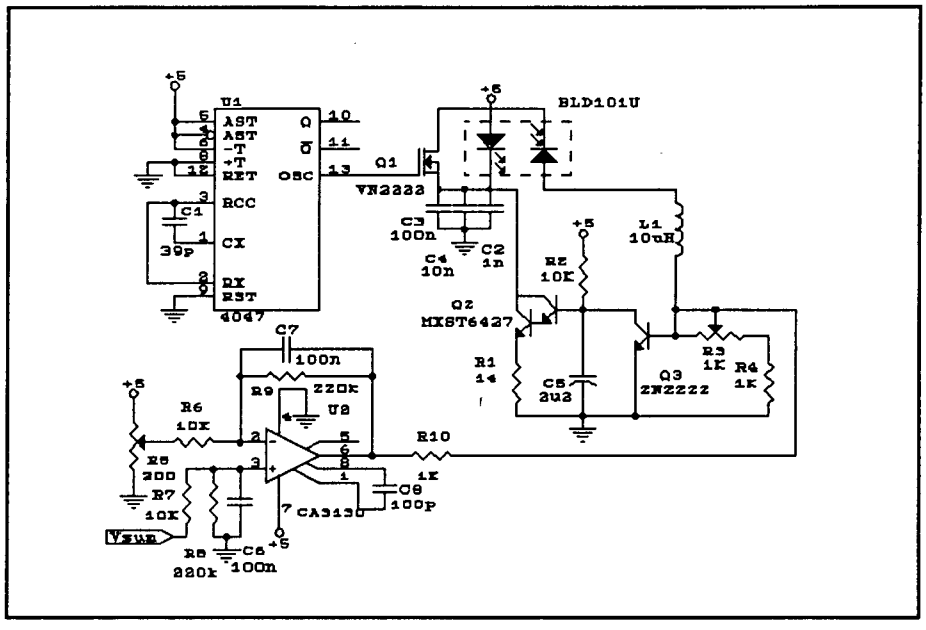
Appendix H

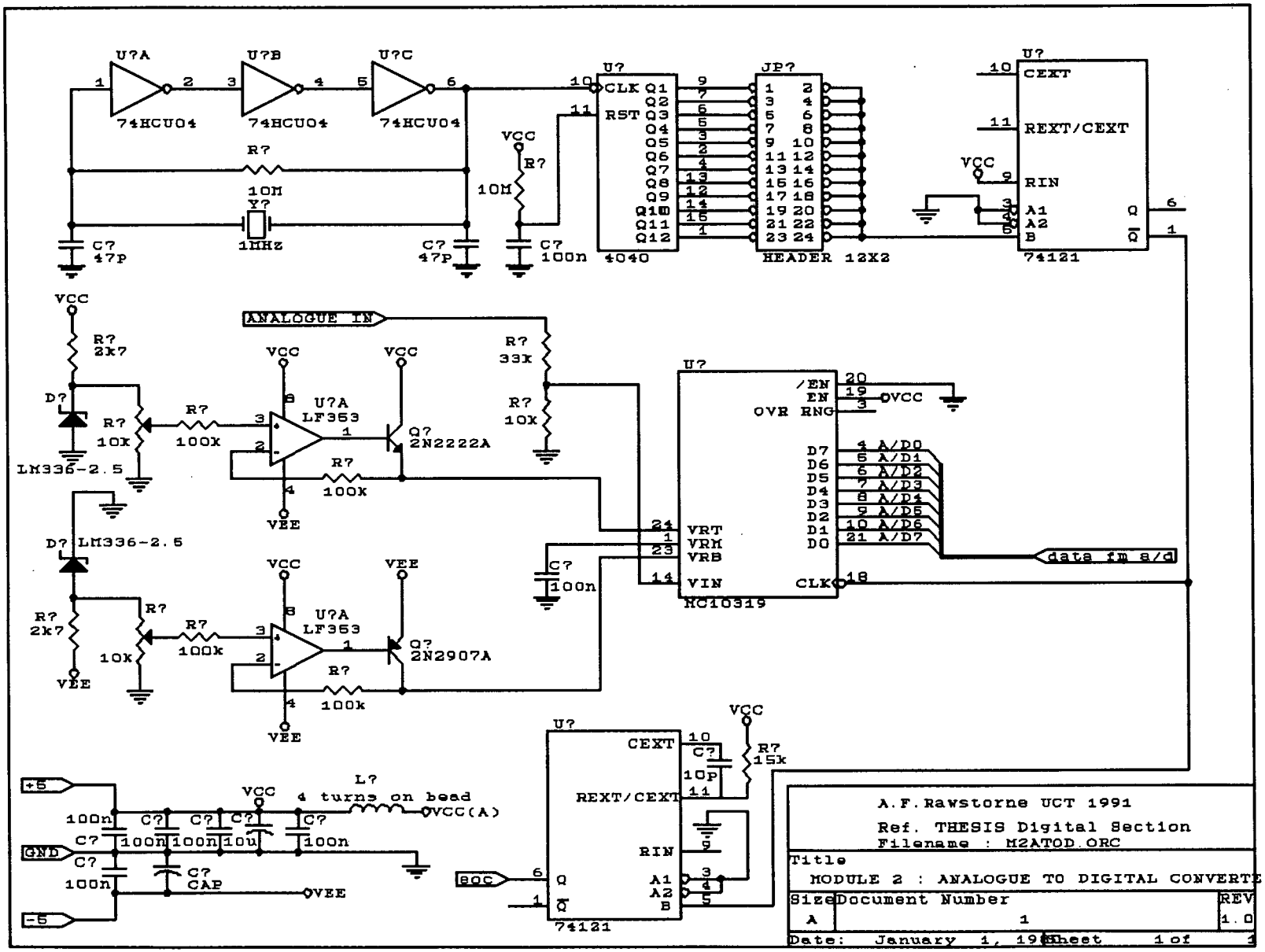
Circuit Diagrams

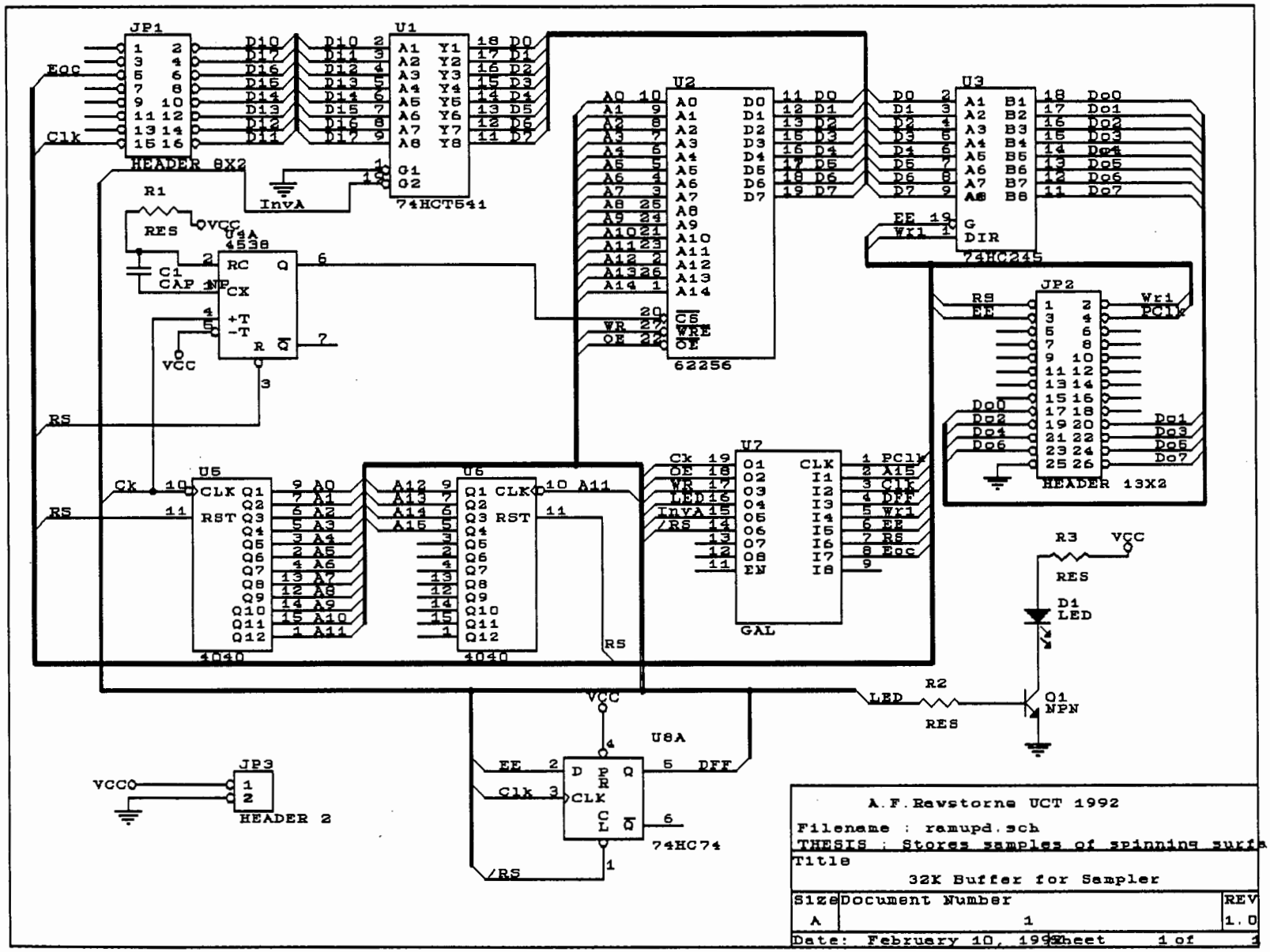
This appendix contains the circuit diagrams for the :

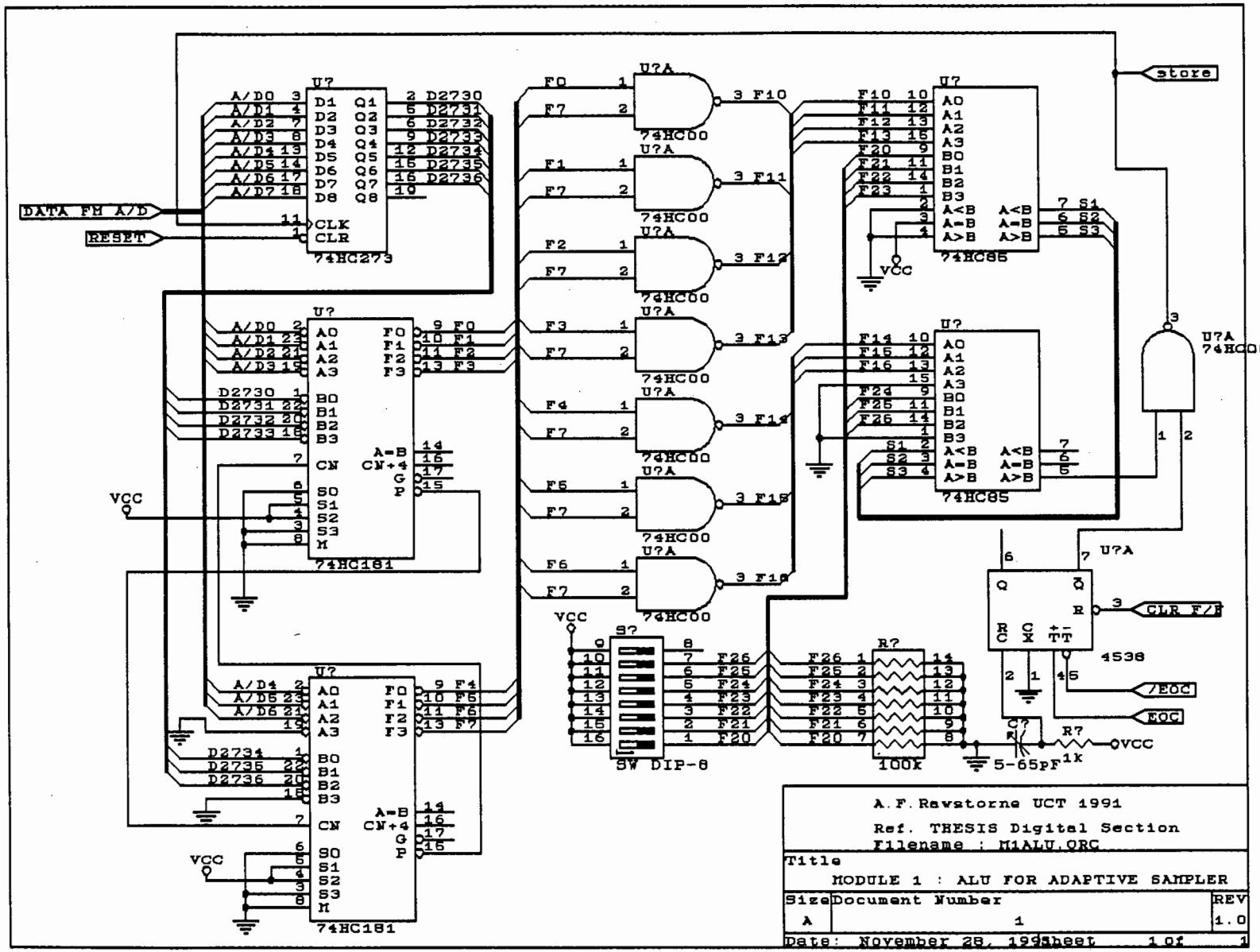
- Receiver electronics
- Transmitter electronics
- Analogue to digital converter module
- Data capture module
- Arithmetic unit for zero order adaptive sampling
- ABEL listing for the GAL in the data capture module




+5
U1
4047
C1
39p
C7
100n
R9
220K
U2
R6
10K
R5
200
R7
10K
R8
220k
C6
100n
CA3130
R10
1K
C8
100p
Vsun
Q1
VN2222
BLD101U
C3
100n
C4
10n
C2
1n
R2
10K
Q2
MXST6427
L1
10uH
R1
1K
C5
2u2
Q3
ZN2222
R3
1K
R4
1K


U?A
U?B
U?C
74HCU04
74HCU04
74HCU04
R?
10M
Y?
1MHz
C?
47P
C?
47P
VCC
R?
10M
C?
100n
U?
CLK
RST
Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8
Q9
Q10
Q11
Q12
4040
JP?
HEADER 12X2
U?
CEXT
REXT/CEXT
RIN
A1
A2
B
Q
74121
VCC
R?
2k7
D?
LM336-2.5
R?
10k
R?
100k
ANALOGUE IN
VCC
VCC
U?A
LF353
Q?
2N2222A
R?
100k
R?
33k
R?
10k
VEE
D?
LM336-2.5
R?
2k7
R?
10k
R?
100k
VCC
VEE
U?A
LF353
Q?
2N2907A
R?
100k
VEE
C?
100n
U?
/EN
EN
OVR RNG
VCC
D7
D6
D5
D4
D3
D2
D1
D0
A/D0
A/D1
A/D2
A/D3
A/D4
A/D5
A/D6
A/D7
data fm a/d
VRT
VRM
VRB
VIN
CLK
MC10319
+5
GND
-5
100n
C?
100n
100n
10u
100n
C?
CAP
VEE
L?
4 turns on bead
VCC(A)
VCC
U?
CEXT
C?
10P
R?
15k
REXT/CEXT
RIN
EOC
A1
A2
B
Q
74121
A.F.Rawstorne UCT 1991
Ref. THESIS Digital Section
Filename : M2ATOD.ORC
Title
MODULE 2 : ANALOGUE TO DIGITAL CONVERTE
Size
A
Document Number
1
REV
1.0
Date: January 1, 19
Sheet 1 of


JP1
HEADER 8X2
U1
74HCT541
InvA
R1
RES
C1
CAP NP
U4A
4538
VCC
RS
U2
62256
U3
74HC245
JP2
HEADER 13X2
U5
4040
U6
4040
U7
GAL
U8A
74HC74
R3
RES
D1
LED
R2
RES
Q1
NPN
LED
JP3
HEADER 2
A.F.Rewstorne UCT 1992
Filename : ramupd.sch
THESIS : Stores samples of spinning surfa
Title
32K Buffer for Sampler
Size
A
Document Number
1
REV
1.0
Date: February 10, 199
Sheet 1 of 1


DATA FM A/D
RESET
store
74HC273
74HC181
74HC00
74HC85
SW DIP-8
100k
5-65pF
1k
4538
VCC
/EOC
EOC
CLR F/F
A.F.Rawstorne UCT 1991
Ref. THESIS Digital Section
Filename : M1ALU.ORC
MODULE 1 : ALU FOR ADAPTIVE SAMPLER
Document Number
1
REV
1.0
November 28, 199
Sheet 1 of 1

```
module rammod

title   'Decoding and Logic for the 32K buffered 500KHz A/D interfaced to
         an IBM PC'

        Rammod device 'P16V8C';

        Pclk,A15,Clk,DFF,WRI,ExtEn,RS,EOC pin 1,2,3,4,5,6,7,8;
        Ck,OE,WR,LED,InvExt,InvRs,InvWri pin 19,18,17,16,15,14,13;

equations
        Ck = (!Clk & !A15 & DFF) # Pclk;
        OE = (WRI # ExtEn);
        WR = (!WRI & !ExtEn) # (!WRI & EOC) # (ExtEn & EOC);
        LED = (!A15 & ExtEn);
        InvExt = !ExtEn;
        InvRs = !RS;
        InvWri = !WRI;
test_vectors
    ([Pclk,A15,Clk,DFF,WRI,ExtEn,RS,EOC]->[Ck,OE,WR,LED,InvExt,InvRs,InvWri])
     [1   ,1  ,1  ,0  ,0  ,0    ,1 ,1  ]->[1 ,0 ,1 ,0  ,1     ,0    ,1     ];

end rammod
```

Appendix I

Pascal Program Listings

This Appendix contains the Turbo Pascal program listings used in this thesis. They are :

- *s.pas : Sampling program*
- binasc.pas : Converts the binary file created by s.pas to the corresponding Datamaster File
- *Find.pas : Finds the point at which data is less than a threshold*
- Average.pas : Outputs the average of successive points in two input data files
- *Adaptive.pas : Zero order adaptive sampling is applied to uniformly sampled input data*

I.1 S.PAS

```
Program A_D_analyzer;
{Reads the RAM in the buffer card and outputs it to the Data.dat bin file}

Uses Crt;
```

```
Var
x  : char;
        i : Word;
        num  : Array[1..$8000] of byte;
data  : File;
        result  : word;

{-------------------------------------------------------------------------}
{---                          Procedure Section                          ---}
{-------------------------------------------------------------------------}
Procedure ResetSampler;

Const
RS = 1;
outp = $200;

Begin
Port[outp] := RS;
End;




{-------------------------------------------------------------------------}
Function ReadByte : Byte;

Const
RS  = 1;
        WRI = 2;
EE  = 4;
        PCL = 8;
        outp = $200;
        inp = $202;
```

```
Begin
Port[outp] := 0;
Readbyte := Port[inp];
Port[outp] := PC1;
End;



{-------------------------------------------------------------------------}
{---                          Main Program                               ---}
{-------------------------------------------------------------------------}
Begin
Assign(input,'');                {Data Master supports Std input }
Reset(input);                    {and Std output only !!         }
Assign(output,'');
Rewrite(output);
ResetSampler;
Writeln('Press Any Key to begin external Sampling');
x := Readkey;
Writeln('Reading ...');
Writeln('Press Any key to Stop');
Port[$200] := 4;                 {initiates sampling on the card}
x:= readkey;                     {Sampler will reset on its own, or}
ResetSampler;                    {by a keypress}
For i := 1 to $8000 do num[i] := ReadByte; {Reads from external RAM}
Assign(data, 'Data.dat');
Rewrite(data,1);
BlockWrite(data,num,SizeOf(num),result);{Write to untyped file Data.dat}
Close(data);
End.
```

I.2 BINASC.PAS

```
Program Binary_To_Ascii;
{Converts the Binary Byte file to Appropriate ASCII file for DataMaster}

Uses Crt, Dos;


Const
Source_option = '-S';
        Destination_option = '-D';
        NoBytes_Option = '-N';
        Frequency_option = '-F';
        Begin_option = '-B';
        Gnuplot_option = '-GNU';
        Help_option = '?';

Var
    Destination_Filename  : String;
no_samples_per_channel  : String;
sample_interval_multiplier  : String;
        no_channels                     : String;
sample_interval_exponent  : String;
data_title  : String;
sample_interval_label  : String;
sample_interval_units  : String;
Y_channel_label  : String;
Y_channel_Units  : String;
Source_Filename  : String;
        st : String;
        i, option,p : Byte;
        code                            : integer;
        value  : Word;
        No_Data_Points : Word;
```

```
        data : text;
        stop                                : Boolean;
        Real_Value                          : Real;
        num : Array[1..$8000] of Byte;
        source  : File;
        j                                   : Word;
        start   : Word;
        num_read : Word;
        Gnu_plot_format : Boolean;


{-------------------------------------------------------------------------}
{---                        Procedure Section                          ---}
{-------------------------------------------------------------------------}
Function Upper(InString : String) : String;
{Converts InSring to Upper Case}

Var
i : Byte;
        st : String;

Begin
st := '';
For i := 1 to Length(InString) do st := st + UpCase(InString[i]);
Upper := st;
End;



{-------------------------------------------------------------------------}
Function FileExists(FileName: string) : Boolean;
{Returns True if file exists; otherwise,it returns False.
Closes the file if it exists. }
var
  f: file;
```

```
begin
  {$I-}
  Assign(f, FileName);
  Reset(f);
  Close(f);
  {$I+}
  FileExists := (IOResult = 0) and
   (FileName <> '');
End;

{------------------------------------------------------------------------}
{---                          Main Program                              ---}
{------------------------------------------------------------------------}
Begin
Assign(Input,'');
Reset(Input);
Assign(Output,'');
Rewrite(Output);
Gnu_plot_format := false;
no_samples_per_channel := '32768';
no_channels := '1';
sample_interval_multiplier := '1';
sample_interval_exponent :='0';
data_title := 'Sampled Data';
sample_interval_label := 'Sample Number';
sample_interval_units := '';
Y_channel_label := 'Voltage';
Y_channel_Units := '0-255';

{The datafile consists of the consecutive entries above preceded by an 'A'
and followed by a blank line. Each sample follows on a new line.}
Destination_Filename := '';
Source_Filename := 'Data.dat';
No_Data_Points := $8000;
stop := false;
```

```
If ParamCount >= 1 then
Begin
        option := 0;
        For i := 1 to Paramcount do
           Begin
           st := Upper(ParamStr(i));
           If st =  Source_Option then option := 1
           Else
             If st = Destination_Option then option := 2
               Else
                If st = NoBytes_Option then option := 3
                Else
                  If st = Frequency_Option then option := 4
                  Else
                   If st = Begin_option then option := 5
                   Else
                    If st = Gnuplot_option then Gnu_plot_format := true
                    Else
                    If st = Help_Option then
                      Else
                     Begin
                if option <> 0 then
                        Case option of
                               1 : Begin
                                   st := ParamStr(i);
                                   p := Pos('.',st);
                                   If p > 0 then Delete(st,p,Length(st)-p+1);
                                   Source_Filename := ParamStr(i)+'.dat';
                                   End;
                              2 : Begin
                                   st := ParamStr(i);
                                   p := Pos('.',st);
                                   If p > 0 then Delete(st,p,Length(st)-p+1);
   Destination_Filename := ParamStr(i)+'.txt';
                                   End;
```

```
                                3 : Begin
    Val(ParamStr(i), value, code);
                                    If (code = 0) and (value <= $8000) then
          No_Data_Points := value;
                                    End;
                                4 : Begin
                                       Str(Real_Value,st);
                                       sample_interval_multiplier :=
          Copy(st,1,Pos('.',st)+1);
                                       sample_interval_exponent :=
          Copy(st,Pos('E',st)+1,Length(st)-
   Pos('E',st));
                                    End;
                                5 : Begin
                                    Val(ParamStr(i), start, code);
                                    If (code <> 0) then start := 1;
                                    End;
                                Else
                                End;
                        End;
          End;
        End;
If not(stop) and FileExists(source_filename) then
Begin
        {Read Datafile}
        Assign(source,source_filename);
        Reset(source,1);
        Blockread(source,num,$8000,num_read);
        Close(source);

        If (start >= num_read) or
((Round(start/2)+Round(No_Data_points/2)) > num_read) then
   start := 1;
        str(No_Data_Points:5,no_samples_per_channel);
      Assign(data,Destination_filename);
```

```
Rewrite(data);
        If not(Gnu_plot_format) then
         Begin
Writeln(data,'A');
Writeln(data,no_samples_per_channel);
        Writeln(data,no_channels);
Writeln(data,sample_interval_multiplier);
Writeln(data,sample_interval_exponent);
Writeln(data,data_title);
Writeln(data,sample_interval_label);
Writeln(data,sample_interval_units);
Writeln(data,Y_channel_label);
Writeln(data,Y_channel_Units);
Writeln(data);
                End;


        For j := start to start+No_Data_points-1 do
         Begin
                str(num[j]:5,st);
Writeln(data,st);
                End;

        Close(data);
End
Else
Begin
Writeln('-s sourcefile          -- default is Data.dat');
Writeln('-d destination file    -- default is screen');
        Writeln('-n No Values converted  -- default is 32768');
        Writeln('-f Sampling frequency   -- default is 500000');
        Writeln(' ? Help Screen ');
        End;
End.
```

I.3 FIND.PAS

```
Program Find;
{This program finds the datapoints at which a threshold -t is exceeded. In
addition, this has a hysteresis of -h}

Const
Source_Option = '-S';
Thresh_Option = '-T';
        NoBytes_Option = '-N';
        Hyst_Option = '-H';
        Help_Option = '?';


Var
source  : File;
        source_filename : String;
        num   : Array[1..$8000] of Byte;
        no    : Word;
        threshold : Byte;
        hysteresis  : Byte;
        i : Word;
        below_thresh : Boolean;
        count   : Word;
        option : Byte;
        st   : String;
        stop : Boolean;
value : Word;
code  : integer;
        No_Data_Points  : Word;
        p : Byte;

{-------------------------------------------------------------------------}
{---                        Procedure Section                          ---}
```

```
{-------------------------------------------------------------------------}
Function Upper(InString : String) : String;
{Converts InSring to Upper Case}

Var
i : Byte;
        st : String;

Begin
st := '';
For i := 1 to Length(InString) do st := st + UpCase(InString[i]);
Upper := st;
End;



{-------------------------------------------------------------------------}
Function FileExists(FileName: string) : Boolean;
{Returns True if file exists; otherwise,it returns False.
Closes the file if it exists. }
var
  f: file;
begin
  {$I-}
  Assign(f, FileName);
  Reset(f);
  Close(f);
  {$I+}
  FileExists := (IOResult = 0) and
   (FileName <> '');
End;



{-------------------------------------------------------------------------}
```

```
{---                          Main Program                                  ---}
{-------------------------------------------------------------------------------}
Begin
Assign(Input,'');
Reset(Input);
Assign(Output,'');
Rewrite(Output);

{Initial settings}
source_filename := 'Data.dat';
threshold := 60;
hysteresis := 60;
below_thresh := false;
count := 0;

{Read Datafile}
Assign(source,source_filename);
Reset(source,1);
Blockread(source,num,$8000,no);
Close(source);
stop := false;

If (ParamCount >= 1) then
Begin
        option := 0;
        For i := 1 to Paramcount do
           Begin
           st := Upper(ParamStr(i));
           If st =  Source_option then option := 1
           Else
             If st = Thresh_Option then option := 2
               Else
                If st = NoBytes_Option then option := 3
                Else
                  If st = Hyst_Option then option := 4
```

```
                Else
                  If st = Help_Option then stop := true
                    Else
                   Begin
                if option <> 0 then
                      Case option of
                            1 : Begin
                                st := ParamStr(i);
                                p := Pos('.',st);
                                If p > 0 then Delete(st,p,Length(st)-p+1);
                                Source_Filename := ParamStr(i)+'.dat';
                                End;
                            2 : Begin
    Val(ParamStr(i), value, code);
                                If (code = 0) and (value <= $ff) then
        threshold := value;
                                End;
                            3 : Begin
                                Val(ParamStr(i), value, code);
                                If (code = 0) and (value <= $8000) then
                  no := value;
                                End;
                            4 : Begin
                                Val(ParamStr(i), value, code);
                                If (code <> 0) then hysteresis := value;
                                End;
                            Else
                            End;
                    End;
          End;
        End;
If not(stop) and FileExists(source_filename) then
Begin
For i := 1 to no do
Begin
```

```
If num[i] < threshold then
Begin
        If not(below_thresh) then
        Begin
Write(i:5,'  ');
               inc(count);
                If (count <> 0) and (count mod 10 = 0) then Writeln;
                End;
below_thresh := true;
        End
Else
If num[i] > (hysteresis + threshold) then below_thresh := false;
End;
Writeln;
Writeln;
Writeln('Total number of detections : ',count);
End
Else
Begin
Writeln('-s sourcefile          -- default is Data.dat');
Writeln('-t threshold 0 - 255    -- default is 60');
        Writeln('-n No Values converted  -- default is 32768');
        Writeln('-h Hystersis            -- default is 60');
        Writeln(' ? Help Screen ');
        End;
End.
```

I.4 AVERAGE.PAS

```
Program Average;

Var
s1,s2  : Text;
```

```
        d : Text;
        s1_file  : String;
        s2_file  : String;
        d_file : String;
        st1,st2  : String;
        i : Word;
        valu1,valu2  : Word;
        valu3 : Real;
        code1,code2 : Word;
        error : Boolean;

Begin
s1_file := 'd1.txt';
s2_file := 'd2.txt';
d_file := 'dd.txt';

Assign(s1,s1_file);
Reset(s1);
Assign(s2,s2_file);
Reset(s2);
Assign(d,d_file);
Rewrite(d);

For i := 1 to 11 do
Begin
Readln(s1,st1);Readln(s2,st2);Writeln(d,st1);
        End;
error := False;

{Read and Average data}
While not(Eof(s1)) and not(Eof(s2)) and not(error) do
Begin
Readln(s1,st1);Readln(s2,st2);
        Val(st1,valu1,code1);Val(st2,valu2,code1);
        If (code1 = 0) and (code2 = 0) then
```

```
        Begin
                valu3 := valu1 + valu2;
                valu2 := Round(valu3/2);
                Str(valu2,st1);
                Writeln(d,st1);
                End
Else
        error := True;
       End;
Close(d);
Close(s2);
Close(s1);
End.
```

I.5 ADAPTIVE.PAS

```
Program Zero_ord_Adaptive;
{Performs zero order adaptive sampling with the source file and stores the
result in the destination file}

Const
Source_option  = '-S';
        Destination_option  = '-D';
        Threshold_option = '-T';
        Points_option = '-P';
        Help_option = '?';

Var
option  : Byte;
        st, st1 : String;
        s_file : String;
        d_file  : String;
        threshold,i,p  : Byte;
```

```
        stop : Boolean;
        value : Byte;
        Code : Integer;
        total_num : Word;
        comp_num : Word;
        error  : Boolean;
        datapt : Byte;
uplimit,lolimit : Word;
        s,d  : Text;
        distribution : Boolean;

{-------------------------------------------------------------------------}
{---                          Procedure Section                          ---}
{-------------------------------------------------------------------------}
Function Upper(InString : String) : String;
{Converts InSring to Upper Case}

Var
i : Byte;
        st : String;

Begin
st := '';
For i := 1 to Length(InString) do st := st + UpCase(InString[i]);
Upper := st;
End;



{-------------------------------------------------------------------------}
Function FileExists(FileName: string) : Boolean;
{Returns True if file exists; otherwise,it returns False.
Closes the file if it exists. }
var
  f: file;
```

```
begin
  {$I-}
  Assign(f, FileName);
  Reset(f);
  Close(f);
  {$I+}
  FileExists := (IOResult = 0) and
   (FileName <> '');
End;


{------------------------------------------------------------------------}
{---                          Main Program                              ---}
{------------------------------------------------------------------------}
Begin
{Default Options}
s_file := 'd.txt';
d_file := 'dd.txt';
threshold := 10;
total_num := 0;
comp_num := 0;
stop := false;
distribution := false;

If ParamCount >= 1 then
Begin
        option := 0;
        For i := 1 to Paramcount do
           Begin
           st := Upper(ParamStr(i));
           If st =  Source_Option then option := 1
           Else
             If st = Destination_Option then option := 2
             Else
               If st = Threshold_option then option := 3
```

```
                Else
                  If st = Help_Option then stop := true
                  Else
                    If st = Points_Option then distribution := true
                       Else
                      Begin
                  if option <> 0 then
                          Case option of
                                  1 : Begin
                                      st := ParamStr(i);
                                      p := Pos('.',st);
                                      If p > 0 then Delete(st,p,Length(st)-p+1);
                                      S_File := ParamStr(i)+'.txt';
                                      End;
                                 2 : Begin
                                      st := ParamStr(i);
                                      p := Pos('.',st);
                                      If p > 0 then Delete(st,p,Length(st)-p+1);
       D_File := ParamStr(i)+'.txt';
                                      End;
                                  3 : Begin
    Val(ParamStr(i), value, code);
                                      If (code = 0) and (value <= $ff) then
              threshold := value;
                                      End;
                                  Else
                                  End;
                        End;
           End;
        End;
If not(stop) then
Begin
        Assign(s,s_file);
        Reset(s);
        Assign(d,d_file);
```

```
        Rewrite(d);
        error := False;

        {First Data point}
        Readln(s,st);
        Val(st,datapt,code);
        If code <> 0 then error := True
        Else
         Begin
                inc(total_num);
                inc(comp_num);

                {Write to file}
                str(total_num:5,st1);
If not(distribution) then Writeln(d,st1+'  '+ st)
Else
                        Writeln(d,st1+'    '+'100');


{Create thresholds}
uplimit := datapt + threshold;
                If uplimit > $ff then uplimit := $ff;
                If threshold > datapt then lolimit := 0
                        Else
                        lolimit := datapt - threshold;
                End;

        While not(Eof(s)) and not(Error) do
         Begin
                Readln(s,st);
                Val(st,datapt,code);
                If code <> 0 then error := True
                Else
                If not(Eof(s)) and (st <> '')then
                 Begin
```

```
                        inc(total_num);
                        If (datapt > uplimit) or (datapt < lolimit) then
                         Begin
                                {Write to file}
                                inc(comp_num);
                                str(total_num:5,st1);
                                If not(distribution) then Writeln(d,st1+'  '+
  st)
Else
                                 Writeln(d,st1+'    '+'100');

                                {Update thresholds}
                                uplimit := datapt + threshold;
                                If uplimit > $ff then uplimit := $ff;
                                If threshold > datapt then lolimit := 0
                                 Else
                                        lolimit := datapt - threshold;
                                End;
                        End

                End;
        Close(d);
        Close(s);
        Writeln('Threshold : ', threshold);
        Writeln('Initial Sample Count : ', total_num);
        Writeln('Compressed Sample Count : ',comp_num);
        If error then Writeln('Process terminated : Error in input File');
        End
Else
Begin
        Writeln('Zero-order Adaptive Sampling Program');
Writeln('-s sourcefile          -- default is D.txt');
Writeln('-d destination file    -- default is dd.txt');
        Writeln('-t Threshold    -- default is 10');
        Writeln('-p plot distribution  -- default is false');
```

```
Writeln(' ? Help Screen ');
        End;
End.
```