



OBJECT IDENTIFICATION

USING

LOW-FREQUENCY PASSIVE TRANSPONDERS

IN

IMPULSIVE NOISE ENVIRONMENTS

A thesis submitted to the Faculty of Engineering of the
University of Adelaide for the degree of

DOCTOR OF PHILOSOPHY

BY

Ashim Kumar ROY B.E.(Hons.), M.Tech.

Department of Electrical and Electronic Engineering
University of Adelaide

August 1982

TO MY PARENTS

PREFACE

This thesis contains no material which has been accepted for the award of any degree or diploma in any University or educational institute. To the best of the author's knowledge and belief, this thesis contains no material previously published or written by another person, except when due reference is made in the text.

Adelaide, 1982

Ashim Roy

ACKNOWLEDGEMENTS

The author is deeply indebted to Dr. P. H. Cole for his continued guidance and encouragement throughout the course of this work. His practical approach to research, and his experience in the field of object identification have been invaluable both to the author's research activities and to his professional career.

There are many people who have directly or indirectly contributed to this work, and to list them all would not be practicable. Mention must, however, be made of Dr. B. R. Davis and Dr. K. Eshraghian, with whom many fruitful discussions have been held.

The academic, technical and secretarial staff of the Department of Electrical Engineering are gratefully acknowledged for their constant support. Thanks are due to Mr. N. R. Blockley for his most capable assistance during the research program. Thanks are also due to Miss Etta Clark for the skilful and efficient typing, and to Mr. J. Zollo for proof-reading and many helpful suggestions.

Last, and most certainly not least, the author is grateful to his wife, Alpana, for her support and patience, without which this work would never have been completed.

PUBLICATIONS

Co-authors: P. H. Cole and K. Eshraghian, "Theory and Operation of the Passive Subharmonic Transponder", IREE International Convention Proceedings, pp 51-54, 1979.

Co-authors: P. H. Cole and K. Eshraghian, "Measurement and Characteristics of Environmental Noise Affecting Low-Frequency Transponder Operations", IREE International Convention Proceedings, pp 456-459, 1979.

Co-inventors: P. H. Cole and K. Eshraghian, "Efficient Object Identification System", European Patent Spec. 7901035.6, 1980.

Co-author: P. H. Cole, "Error Control in a Vehicle Identification System", IREE International Convention Proceedings, pp 381-383, 1981.

"Monolithic IC Design Considerations for Automotive Applications", IREE International Convention Proceedings, pp 414-416, 1981.

Co-authors: K. Eshraghian and P. H. Cole, "Electromagnetic Coupling in Passive Subharmonic Transponders", Journal of Electrical and Electronics Engineering, Australia, Vol. 2, No. 1, pp 28-35, 1982.

SUMMARY

This thesis is primarily concerned with theoretical development and experimental study of some important aspects in the design and operation of a low-frequency object identification system. Object identification has been studied by several workers in the past few years. An overview of these identification schemes is presented. The requirements desired in an object identification scheme mainly depend upon the application. An object identification system employing near-field inductive coupling fulfils most of these requirements and therefore may be adapted to many identification applications. The near-field identification scheme is therefore selected for an in depth study.

The low-frequency object identification is first approached as a communication engineering problem. Electromagnetic object identification is based on coupling of energy over an electromagnetic path between the object and interrogation

equipment. A transponder carrying object when interrogated by a CW signal returns a code modulated reply signal identifying the object. In another section a two-port passive transponder system employing near-field coupling is investigated. This investigation leads to the design of a more efficient one-port system where a single coupling element is used for simultaneous extraction of power from the interrogation source at one frequency and subsequent retransmission of this energy as reply at another frequency.

The requirements of passive operation and communication over electromagnetic propagation path result in very weak reply signals. Environmental noise therefore becomes a major consideration in the design of such systems. The third and major aspect of the research program involves experimental study of the noise in a variety of indoor and outdoor environments. Statistical measures to describe the noise are proposed and a technique to evaluate the system error performance from these statistical measures is presented.

Attention is focussed on practical error control strategies to combat the environmental noise in another section of the work. Some encoding and decoding techniques are reviewed and

a strategy suitable for the noise environment of one application is implemented.

Yet another aspect of the work involves investigation of the technologies suitable for the manufacture of such transponders. A comparison of some low power IC technologies is also presented.

Finally, conclusions are reached from the research program and recommendations for further work are made.

CONTENTS

	Page
PREFACE	i
ACKNOWLEDGEMENTS	ii
PUBLICATIONS	iii
SUMMARY	iv

CHAPTER I

INTRODUCTION

1.1	Remote Object Identification	1.1
1.2	A Communications Engineering Problem	1.6
1.3	Review of Previous Work	1.7
1.3.1	Microwave Identification Systems	1.7
1.3.2	Surface Acoustic Wave Identification Systems	1.13
1.3.3	Low-frequency Identification Systems	1.14
1.4	Objects of Thesis	1.17

CHAPTER II

LOW-FREQUENCY PASSIVE TRANSPONDERS

2.1	Introduction	2.1
2.2	Two-port Low-frequency Passive Transponder	2.2
2.2.1	Passive Subharmonic Transponder	2.4
2.3	Operational Requirements and Constraints	2.7
2.3.1	Optimization of Two-port Transponder	2.9

		Page
2.4	Experimental Results - Two-port System	2.19
2.5	The One-port System	2.24
2.5.1	Limitations of Two-port Transponders	2.25
2.5.2	The One-port PST	2.32
2.5.3	Transponder Coil Considerations	2.34
2.5.4	Control Circuit Considerations	2.39
2.6	Conclusions	2.48

CHAPTER III

ANTENNA STRUCTURES

3.1	Introduction	3.1
3.2	Power and Propagation Constraints	3.3
3.2.1	Transmitter Antenna Considerations	3.9
3.2.2	Receptor and Transponder Antenna Considerations	3.19
3.2.3	Receiver Antenna Considerations	3.22
3.3	Electromagnetic Coupling	3.23
3.4	Coupling Volume Approach	3.24
3.4.1	The Coupling Volume	3.34
3.4.2	The Dispersal Volume	3.39
3.4.3	The Power Coupling Ratio	3.43
3.5	Antenna Considerations for One-port Systems	3.57
3.6	Transceiver Antenna Design	3.59
3.6.1	Isolation Characteristics of Transceiver Antenna	3.63
3.6.2	Filters for the Diplexer	3.66
3.6.3	Diplexer Design Considerations for Two-port Transponder Systems	3.68
3.7	Conclusions	3.71

CHAPTER IV

STUDY OF ENVIRONMENTAL NOISE

4.1	Introduction	4.1
4.2	Literature Survey	4.2
4.2.1	Atmospheric Noise	4.3
4.2.2	Man-made Noise	4.11
4.3	Statistical Noise Measures	4.25
4.3.1	Allan Variance Analysis	4.29
4.3.2	Amplitude Probability Distribution	4.31
4.3.3	Noise Amplitude Distribution	4.31
4.3.4	Pulse Width Distribution	4.32
4.3.5	Time Probability Distribution	4.32
4.4	Environmental Noise Measurements	4.33
4.4.1	Preliminary Noise Measurements	4.35
4.4.2	Noise Data Acquisition Instrumentation	4.37
4.4.3	Data Acquisition System Calibration	4.42
4.4.4	Noise Data Measurements	4.42
4.5	Noise Analysis	4.45
4.6	Error Analysis	4.66
4.6.1	Error Prediction Analysis	4.75
4.6.2	Error Prediction by Means of Simulation	4.79
4.7	Conclusions	4.85

CHAPTER V

ERROR CONTROL

5.1	Introduction	5.1
5.1.1	Transponder Communication System	5.3
5.1.2	Real Data Channel Models	5.8
5.2	Types of Code	5.14
5.3	Review of Linear Algebra	5.18
5.4	Linear Codes	5.39
5.4.1	Definitions	5.39
5.4.2	Cyclic Codes	5.43
5.4.2.1	Error Detection and Correction for Cyclic Codes	5.46
5.4.3	Bose-Chaudhuri-Hocquenghem (BCH) Codes	5.47
5.4.3.1	Construction of BCH Codes	5.49
5.4.3.2	Encoding of BCH Codes	5.54
5.4.3.3	Encoder Design	5.61
5.4.3.4	Decoding of BCH Codes	5.64
5.4.3.5	Decoder Building Blocks	5.75
5.4.4	Disadvantages of BCH Codes	5.90
5.5	Optimization of Encoder and Decoder Parameters	5.97

CHAPTER VI

SYSTEM DESIGN CONSIDERATIONS

6.1	Introduction	6.1
6.2	Practical System Considerations	6.2
6.3	Design of the Transmitter	6.4

		Page
6.4	Selection of Modulation Method	6.14
6.4.1	Binary Amplitude Shift Keying	6.17
6.4.2	Binary Frequency Shift Keying	6.24
6.4.3	Binary Phase Shift Keying	6.30
6.4.4	Comparison of Modulation Methods	6.34
6.4.4.1	Noise Performance	6.35
6.4.4.2	Spectral Occupancy	6.41
6.4.4.3	Equipment Design Tolerance	6.42
6.4.4.4	Effects of Transmission Medium	6.43
6.4.5	Modulation System for PST Applications	6.44
6.5	Signal Synchronization	6.47
6.5.1	Carrier Synchronization	6.48
6.5.2	Bit Synchronization	6.51
6.5.3	Word Synchronization	6.55
6.6	Receiver Design Considerations	6.57
6.6.1	Analogue Processing of the Reply Signal	6.57
6.6.2	Digital Signal Processing	6.61
6.6.3	Receiver Performance Evaluation	6.65
6.7	Conclusions	6.72

CHAPTER VII

TECHNOLOGIES FOR TRANSPONDER REALIZATION

7.1	Introduction	7.1
7.2	Choice of Technology	7.2
7.2.1	NMOS Technology	7.5
7.2.2	CMOS Technology	7.12
7.2.3	I ² L Technology	7.18

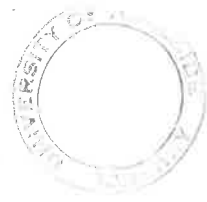
		Page
7.3	Comparison of the IC Technologies	7.20
7.4	Integration of the Two-port PST Label	7.25
7.5	Future Considerations	7.31

CHAPTER VIII

CONCLUSIONS	8.1
-------------	-----

APPENDIX A

RF INTERFERENCE REGULATIONS	A.1
-----------------------------	-----



CHAPTER I

INTRODUCTION

1.1 Remote Object Identification

The increasing use of automation in various fields of human activity has led to widespread interest in many forms of active and passive electromagnetic transponders for remote object and personnel identification. The objectives in the use of such transponders include the elimination of human error in data recording, economy of operation, and the creation of new system organizations which are made possible by the ability of automated operations to cope with speed and complexity.

Transponders for object identification applications may be broadly classified according to whether they are active or passive devices, suitable for interrogation either from near to or far from interrogation apparatus, free or otherwise

of registration and orientation constraints between themselves and the interrogation apparatus, and according to the capacity of their coding.

This study is concerned with object identification by means of passive digitally encoded transponders of registration free type, suitable for interrogation over short distances but capable of returning long codes. The requirement of freedom of registration and severe environmental conditions encountered in some of the identification applications makes the choice of technology an important consideration. Several object identification techniques have been proposed by various workers in the past few years. An overview of these identification techniques is presented later in this chapter.

A number of practical requirements must be met for an automatic object identification system to work satisfactorily. The degree to which a particular requirement affects the selection of a system in a given application depends upon such operational parameters as indoor or outdoor usage, interrogation range, moving or stationary object, object speed, number of objects to be identified and cost considerations. The more specific requirements which emerge from the study of the

systems proposed by various authors include the following:

- i) The interrogation system should not cause interference to the other users of the electromagnetic spectrum or cause health hazard to anyone in the vicinity of the interrogation region.
- ii) The transponders and the interrogation units should be inexpensive to allow for their widespread usage in a variety of applications.
- iii) The transponder system should utilize the less congested portions of the electromagnetic spectrum.
- iv) The reply signal from the transponder should not be impaired by environments such as rain, snow, clothing, wood or plastics so that the identification system is suitable for a large number of applications.
- v) The identification system should be maintenance free as far as possible. This requirement leads to the need for the transponders to be passive.
- vi) The transponder and the interrogator should have convenient physical dimensions to be cost effective.
- vii) The interrogator should be able to accurately "read" the identity of either a moving or a stationary transponder in a well defined inter-

rogation zone.

viii) Some applications require unique identification of a large number of objects. The transponder should, therefore, have a large coding capacity.

ix) The interrogation apparatus should easily blend with the environment.

x) The transponders should be free from any orientation constraint.

xi) The transponders should not have any effect on the object it is attached to. This is to ensure that in automatic personnel identification situations, for instance, persons carrying artificial pacemakers are not adversely affected.

In consideration of these requirements it appears that the ultrasonic and optical systems are best suited for indoor applications with short interrogation range. Health hazards associated with radio-active or high power microwave systems render themselves unsuitable. Further consideration of these requirements and practical constraints leaves only low power microwave and low-frequency inductive technologies as viable options for general purpose object identification applications. Some of the identification systems employing these techniques are described later in this chapter. The potential applications of such object identification systems include the following:

- a. Automatic Vehicle Identification
 - i) Automatic tolling system
 - ii) Rental car identification
 - iii) Electronic license plate
 - iv) Restricted parking areas
- b. Automatic Material Identification
 - i) Containerized shipping
 - ii) Automated freight handling
 - iii) Mailbag routing
 - iv) Aircraft baggage sorting
 - v) Library book sorting
 - vi) Supermarket checkouts
- c. Personnel Identification
 - i) Military and industrial security
 - ii) Employee identification
 - iii) Electronic keys

Other applications of remote identification systems could include patient monitoring at hospitals and cattle monitoring at farms. Adaptability of an object identification system to a large number of applications with minimum modifications is a useful measure to compare different identification systems.

1.2 A Communications Engineering Problem

Although it is true in principle that when seeking to design a remote identification system, all system parameters are available for selection by the designer, in practice this viewpoint is inconveniently broad in forming an initial understanding of how transponder systems may be designed and how various parameters may be optimised in view of several conflicting constraints. The approach used in the analysis presented in the chapters to follow is to select particular system operating principles, assign fixed values to some of the system parameters and then analyse variations to the rest of the parameters.

The requirements of passive operation and interrogation over an electromagnetic propagation path are sufficient to ensure that the reply signal is very weak in comparison to the interrogation signal. The requirement of freedom of registration and orientation, on the other hand, ensures that the electromagnetic field created by the interrogation signal is very strong, not only in the vicinity of the transponder but also in the region of the receiver antenna. The separation of the reply signal from the strongly coupled interrogation signal is therefore of vital importance.

Techniques of providing this separation in a cost effective manner are described in chapter three.

Passivity of the transponders and need to operate these systems in severe electromagnetic environments in some applications requires an understanding of the characteristics of the environmental noise, an essential consideration of the design of such systems. Attention is focussed on the problem of environmental noise as part of this research program and techniques to predict the degradation in the system performance in the presence of noise are discussed in chapter four, while the error control strategies to counter the degradation are described in chapter five

1.3 Review of Previous Work

Remote object identification systems have been studied by several researchers in the past few years. A brief review of these systems is presented below.

1.3.1 Microwave identification systems

This class of identification systems perform the energy and information exchange at micro-

wave frequencies and the separation between the interrogation and reply signals is provided in the frequency domain. One microwave identification system using the second harmonic reflections is proposed by Klensch et al (1973). The system, as shown in Figure 1.1, in its basic form includes an interrogator operating at X-band to illuminate a passing transponder which radiates back a coded signal at the second harmonic of the X-band input. The transponder performs the frequency doubling action by virtue of a Schottky-barrier silicon diode connected across the terminals of the tag antenna via a matching section. The transponder antenna is in the form of a flat, printed-circuit array of dipoles and is tuned to the geometric mean of the X-band interrogation frequency and Ku-band reply frequency. The reply signal is on-off keyed and thus provides the identity code. The keying is performed by the application and removal of reverse bias on the doubler diode. The reply signal is envelope detected and suitably displayed. Similar identification systems have been described by Shefer et al (1973) and Sterzer (1974).

Another system for identifying moving objects in which a passive label having a number of dipoles in accordance with a code is attached to the object to be identified is described by

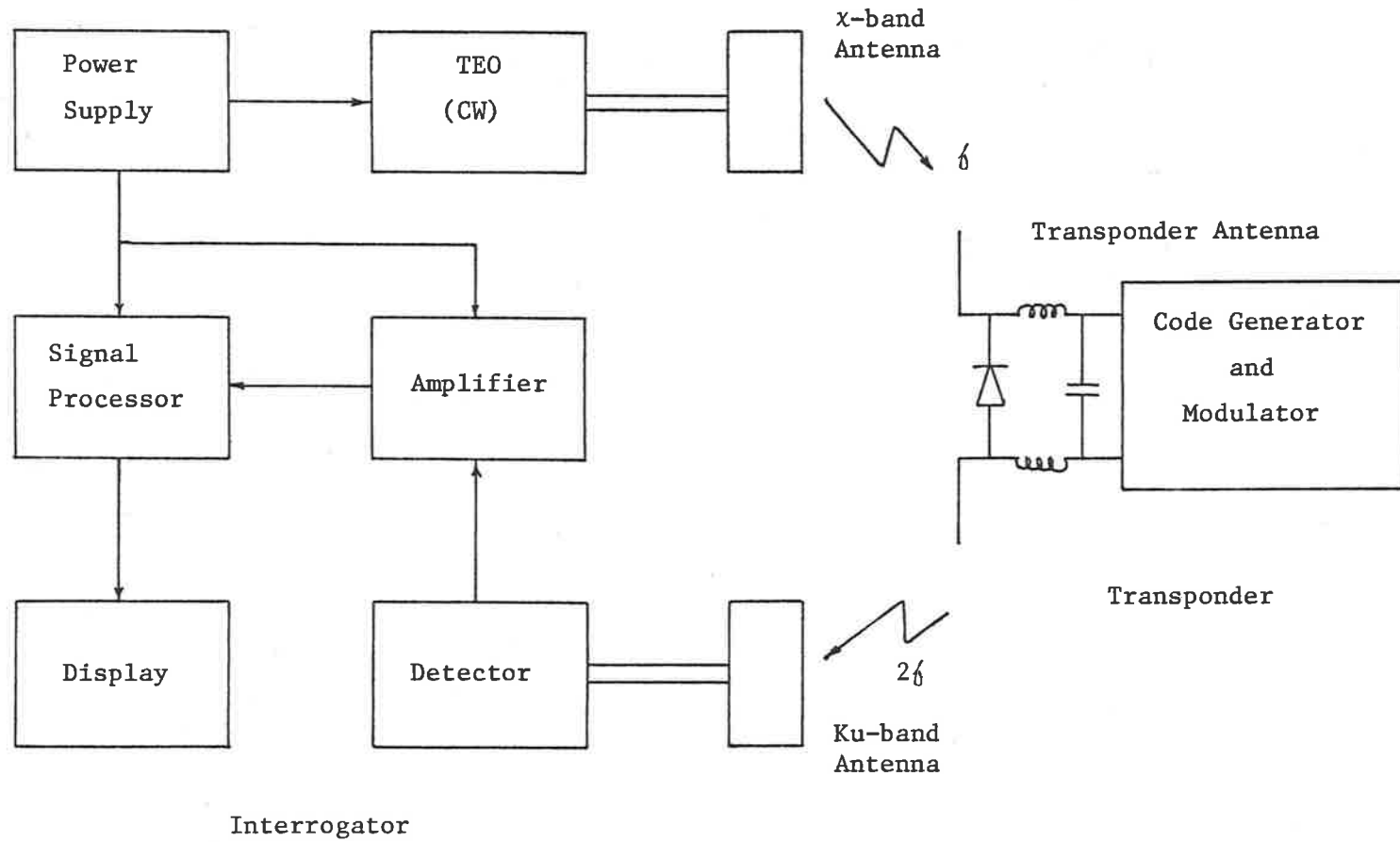


Figure 1.1 A Basic Microwave Identification System

Constant (1974). The system, as shown in Figure 1.2, uses a synthetic aperture radar consisting of a Gunn diode mounted within a resonant cavity and connected to a horn to illuminate the label. The Gunn diode operates as a homodyne so that signals from the label are also detected by the diode as a number of overlapping chirp signals. This signal is then appropriately processed to produce a pulse train corresponding to the dipole structure on the label. The passive dipole resonators for this identification system can take the form of a dipole wire or a slot, the latter punched into a flat surface. The Doppler effect which results in the chirp signal is used

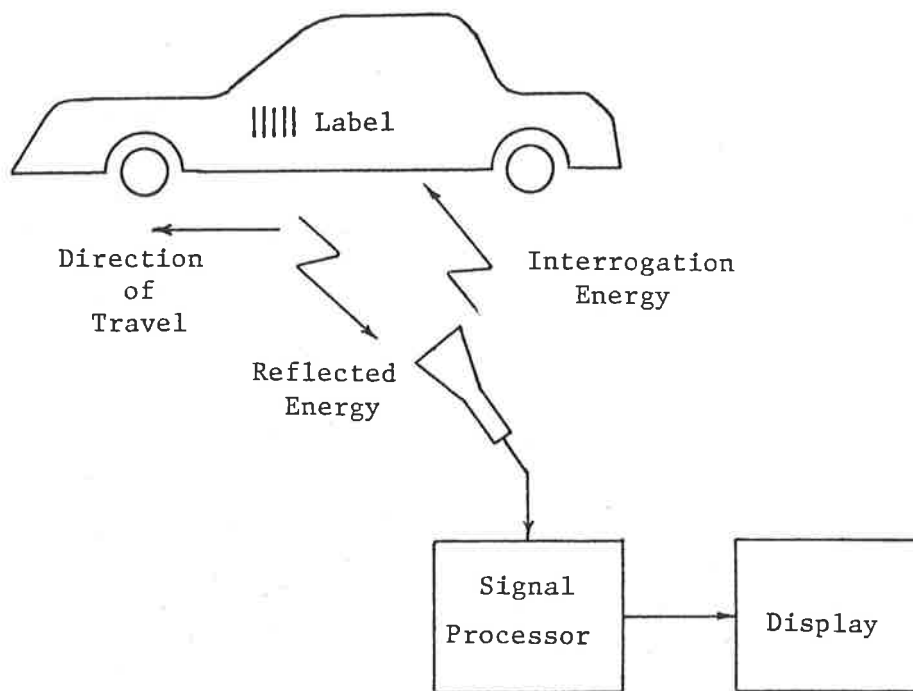


Figure 1.2 Microwave Identification System Using Doppler Effect

as the basis of coding where the frequency difference between the interrogation energy and the energy reflected from individual dipoles in a moving label is continuously processed to identify the label.

Another object identification system, shown in Figure 1.3, which uses modulated backscatter from an r.f. beam-powered transponder, is described by Koelle et al (1975). A continuous wave radio frequency beam is used to illuminate the transponder antenna. Some of the received r.f. is rectified to operate the transponder circuitry, which includes a voltage regulator, a 20 KHz sub-carrier oscillator, and a code generator. The re-radiated r.f. is modulated by varying the load on the rectifier. The load is switched at the sub-carrier frequency and the degree of loading is switched between two levels by the code generator. This results in scattered r.f. being amplitude modulated at 20 KHz with a secondary modulation for the code. A variation of this identification scheme is suggested by McEwen (1975) and uses a battery powered transponder to reduce the power of the incidental radio frequency beam and the complexity of the receiver circuitry.

One of the disadvantages associated with the

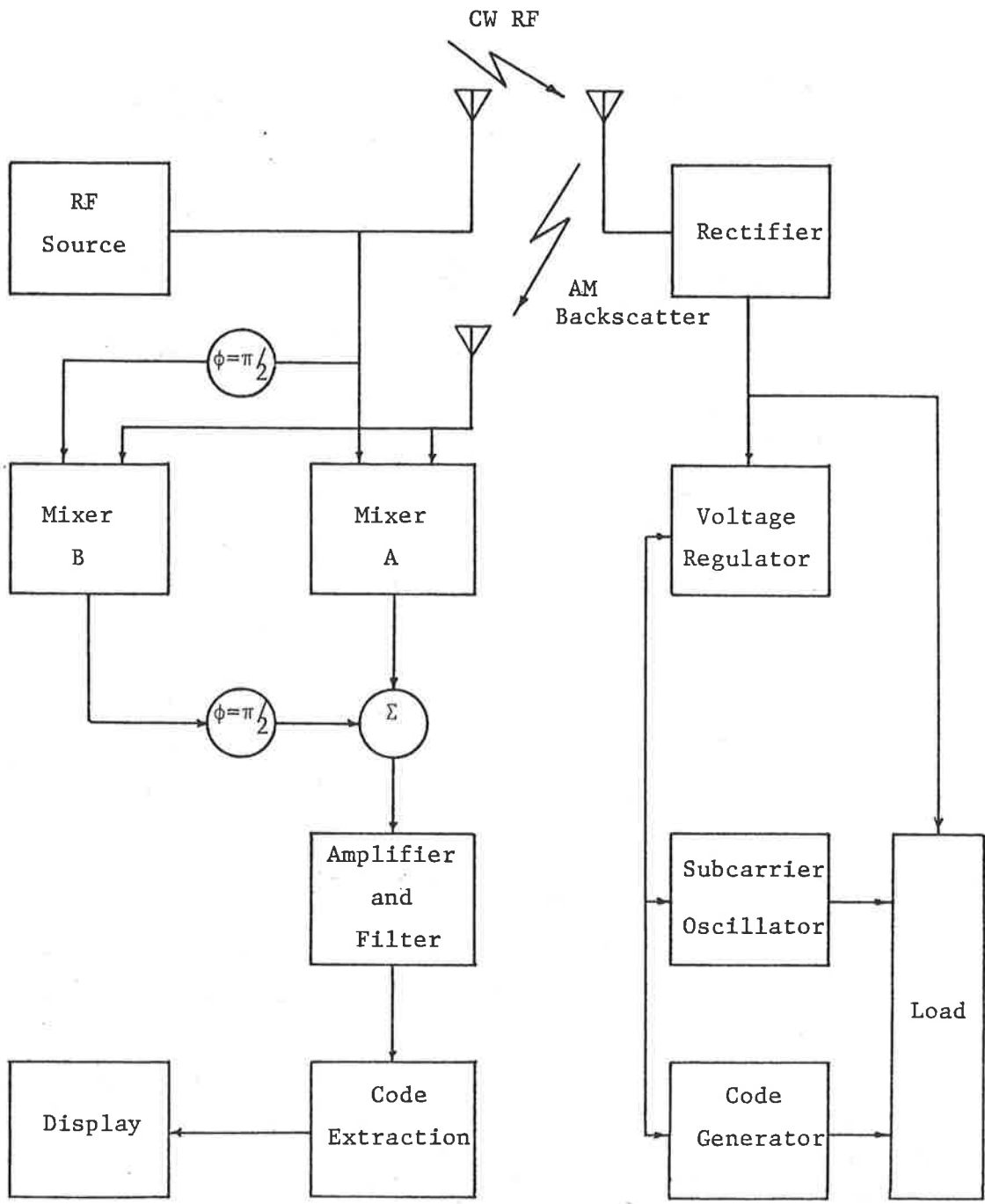


Figure 1.3 Object Identification Using RF Backscatter

microwave systems arises from difficulties in sufficiently confining the interrogation energy. Other disadvantages of the microwave systems include the stringent limits imposed by the statutory bodies on unattended sources of microwave energy and, in some cases, background clutter from extraneous objects.

1.3.2 Surface acoustic wave identification systems

Another class of object identification system which uses surface acoustic wave (SAW) delay lines has been described by Davies et al (1975). The separation between the strong interrogation signal and weak reply signal in this class of identification systems is provided in time domain. The identification system, as shown in Figure 1.4, involves a receiving antenna feeding a delay line having n separate coded taps. The combined output of these taps is either connected to a reply antenna in the case of a two-port transponder or returned to the original antenna in the case of a one-port transponder system. The passive transponder is interrogated at a convenient radio frequency. The time delay to the first tap of the delay line is much longer than the duration of the interrogation pulse. The taps are equally spaced at a time delay corresponding to the in-

terrogation pulse duration. A single interrogating pulse therefore results in a train of reply pulses being returned from the transponder, each associated with the corresponding output of the delay line taps. This train of pulses is used to represent the identity of the transponder by encoding the taps of the delay line in terms of either presence or absence of the taps. Further sophistication in coding may be provided by the variations in the tap geometry. The transponder system described by Davies et al uses a 21 finger delay line operating at 100 MHz.

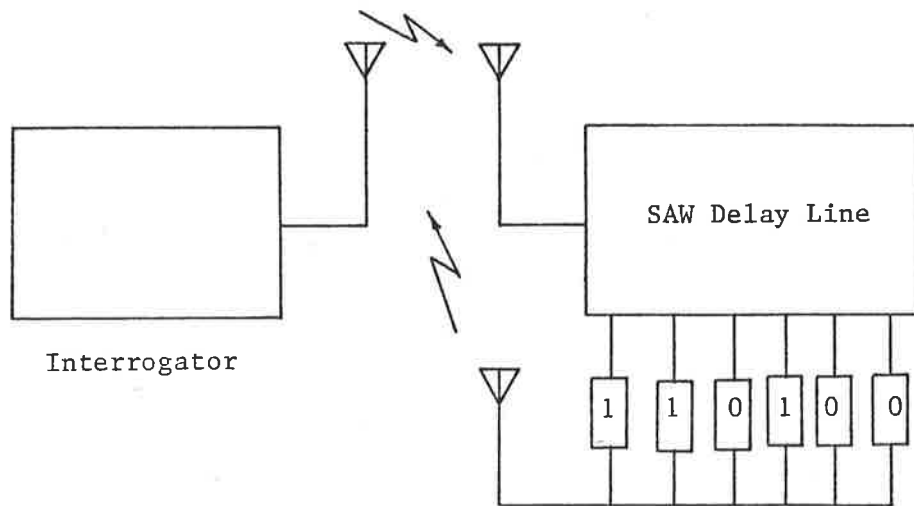


Figure 1.4 SAW Object Identification System

A variation of the above system which operates at 915 MHz with a bandwidth of 50 MHz is described in British Patent number 1298381 by Unisearch Limited (1972).

Difficulties are encountered, in SAW object identification systems, with the delay line lengths when long codes are required to be generated. Furthermore, because the reply signals from this type of transponder are weak, complex signal processing is necessary to extract the reply signal from noise. The SAW system also has the disadvantage that, at most suitable frequencies for its operation, unconfined radiation of high power signals is not permitted.

1.3.3 Low-frequency identification systems

Low-frequency transponders for object identification employing inductive coupling between the transponder and interrogator have been described by Vinding (1967), Ryley (1970), Leaver (1975), Hutton et al (1976) and Cole et al (1979). Low-frequency identification system in its basic form works on the principle that when an object carrying a transponder comes in the vicinity of an interrogator a weak coupling between the transponder

and interrogator is established. The transponder, as shown in Figure 1.5, utilizes a part of the interrogation energy to energize the electronic circuitry on board to generate a code modulated reply signal. The separation between the interrogation and reply signals is achieved spatially and in frequency domain. A detailed description of this system, which forms the basis of the research program, is presented in the next chapter.

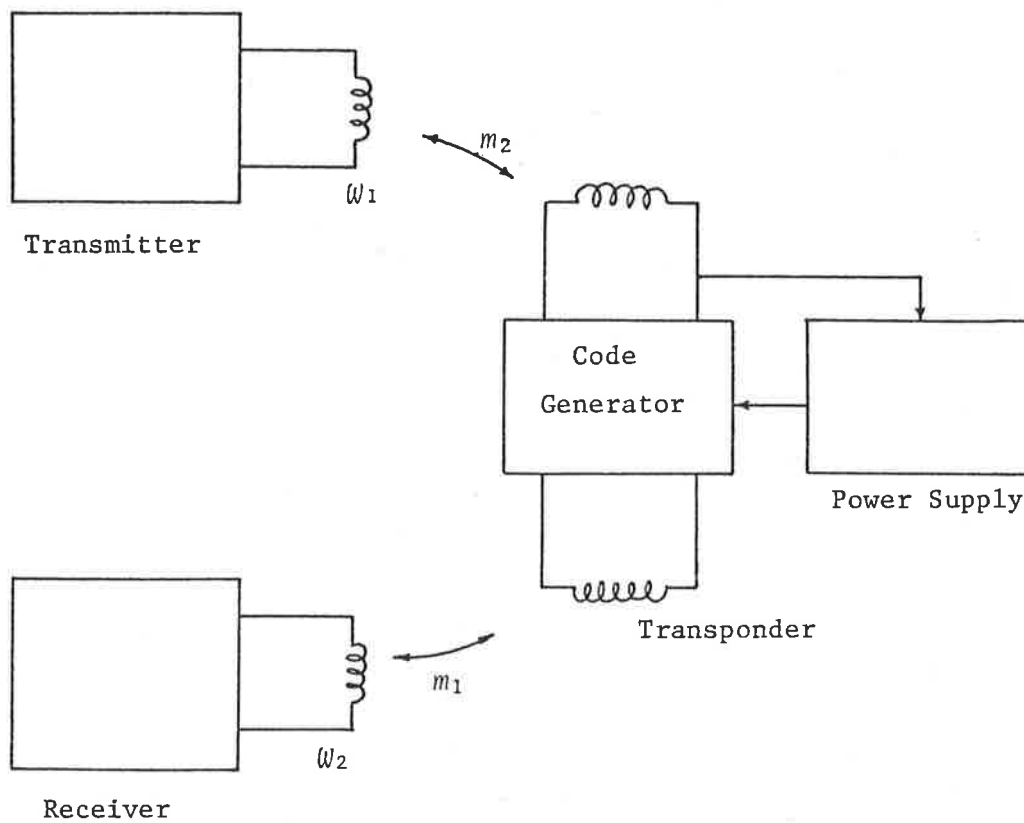


Figure 1.5 Low-frequency Inductive Identification System

1.4 Object of Thesis

The work presented in this thesis is the result of efforts to develop a remote object identification system which may be used in a wide variety of applications. One object of the work presented here is to derive a fundamental understanding of various subsystems leading to an optimal design of the overall system. Another object of the thesis is to study the environmental noise characteristics affecting the operation of such identification systems and develop techniques to predict the system performance in presence of noise. A further object is to develop encoding and decoding strategies to combat the noise and improve the error performance of the system. Another aim of this research is to investigate the suitability of various semiconductor technologies such as CMOS, I²L, NMOS etc. for the integration of the transponder. Finally, it is the aim of this work to substantiate the theoretical development of various subsystems of the object identification system by extensive experimental evidence.

CHAPTER II

LOW-FREQUENCY PASSIVE TRANSPONDERS

2.1 Introduction

Among various techniques of remote object identification, the one employing transponders operating in the low frequency range of the radio-frequency spectrum has been chosen as the focus of interest of this study. Very few of the basic surveillance systems described in the previous chapter have been brought to practical realization. The reasons for this include the cost of manufacture of complex circuits, unacceptability of systems requiring bulky antennae in some applications, difficulties in licensing equipment operating at high frequencies, and health hazards involved in the transmission and reception of large amounts of power, especially at microwave frequencies. Most of the systems operating in the microwave band have difficulties in sufficiently confining the interrogation energy and also suffer in signal quality due to background clutter. Some

of the other systems, by nature, are incapable of accommodating long codes, which is an essential requirement in most of the present-day identification applications. Most of these problems are overcome in the low-frequency transponder system which, therefore, is capable of being used in a variety of applications.

2.2 Two-Port Low-Frequency Passive Transponder

The class of transponder systems which is free of most of the objections mentioned in the previous section is the low-frequency passive transponders. The transponders in this class of system make use of low-frequency communication substantially by magnetic fields. The rules and regulations set by the statutory bodies responsible for the management of the radio spectrum are less restrictive in the frequency band of interest (Jackson, 1980; F.C.C., 1959). The passive transponder is a device which does not have a built-in source of energy, but utilises the interrogation energy to operate the transponder circuitry. The energy to operate, in one type of passive transponder, is provided at a frequency different from that used for interrogation. This approach not only adds more design variables but also introduces problems associated

with uncoupling the power signal, the interrogation signal and the reply signal, and thereby increases the overall cost. An alternative approach is to provide the energy to operate the transponder circuitry at the interrogation frequency. This scheme is the one employed in the research programme.

In general, the reply signal, which is usually modulated with an information signal, may be at any frequency, independent of the interrogation frequency, within the low-frequency band. However, from the system design point of view, it is much more convenient to select the centre frequency of the reply signal to be either an harmonic or subharmonic of the transmitter frequency. The principal problem associated with the harmonic operation is that most transmitters generate some number of harmonics beside their resonant frequency. These harmonics, although filtered out to a large extent by the transmitter antenna resonant circuit, can still be comparable with very low level reply signals. Another problem encountered when the reply is an harmonic, is that of spurious replies generated by the objects in the vicinity of interrogation equipment. The subharmonic transponder does not suffer from these difficulties and, therefore, performs better than the harmonic transponder in most instances.

The electromagnetic coupling required for the interrogation to take place in a two-port object identification system can be provided by either magnetic or electric fields. The electric fields produced by charged bodies are usually monopole fields. The noise picked up by the receiver from monopole fields created by statically charged bodies in the interrogation range sometimes can be large in relation to the reply signal. Furthermore, the electric field produced by the transmitter can easily be shielded by nearby conducting objects from reaching the transponder. Electric field coupling at low frequencies produced by mutual capacitance cannot be provided by coupling elements of practically convenient sizes. Magnetic coupling provides an alternative which is free from most of these problems.

2.2.1 Passive Subharmonic Transponder

The basic structure of the two-port passive subharmonic transponder (PST) identification system is shown in Figure 2.1. The principle of operation of the system is as follows. When an object carrying a transponder appears in the field of the transmitter, a part of the interrogation energy is inductively coupled to the receptor antenna at one frequency. Although the diagram illustrates the normal situation of coupling by magnetic fields, electric field

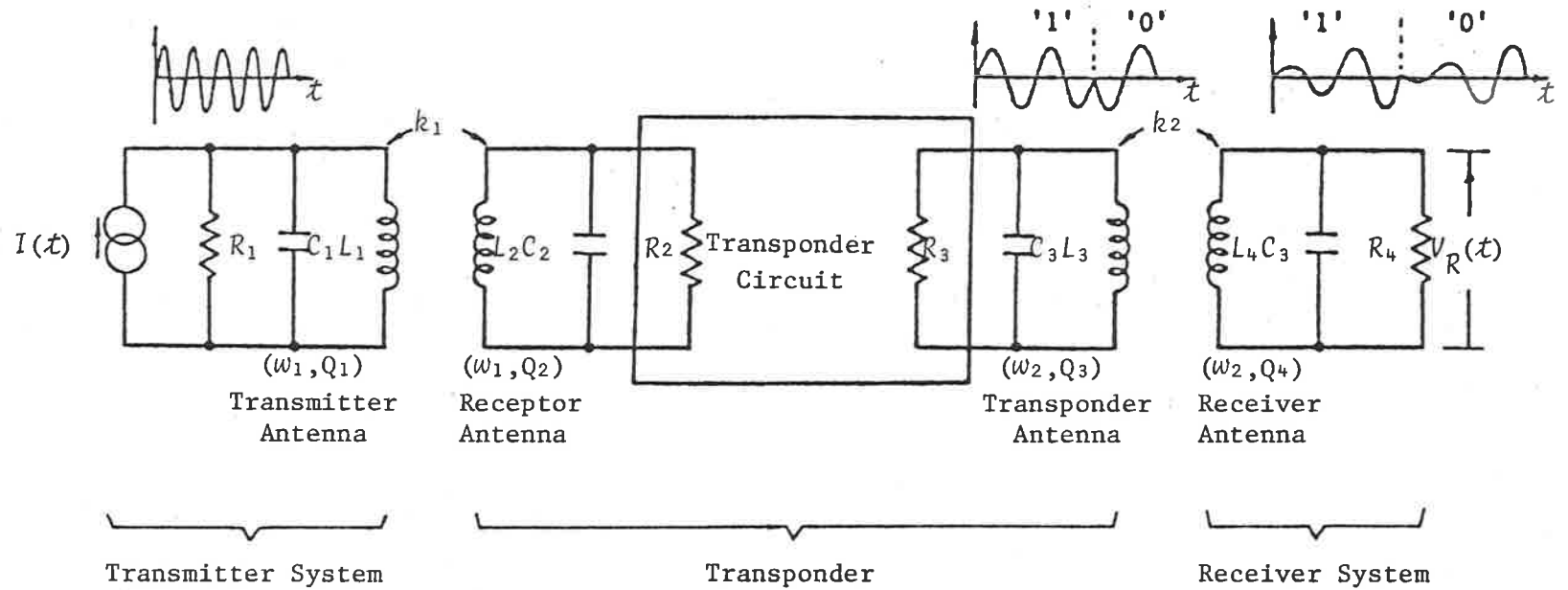


Figure 2.1 Block Schematic of Two-port Transponder System

coupling is also possible in some circumstances. A part of this coupled energy is rectified and filtered to provide power for the code generating circuitry. The other part of the interrogation signal is used to generate a binary code. The code thus generated then modulates a subharmonic of the interrogation frequency, which is also derived from the interrogation signal. The code modulated subharmonic is fed to the transponder antenna. The reply is magnetically coupled to the receiver antenna. This signal is then processed to determine the code and, hence, the object. The four resonant circuits shown in Figure 2.1 are not antennae in the strict sense, as the field created by them is not a radiation field but the energy exchange is by means of near field coupling. It is, however, convenient to use the terms antenna, coupling element, and resonant circuit interchangeably within the text.

Although in theory it is possible to construct a transponder system as described above without regard to any other consideration, in a practical application one must consider the large number of constraints and requirements of the application. The object of this study is to implement a general purpose object identification system which may be adapted to various applications. One

approach to such design is to consider all the design variables and then perform the necessary optimization to reach a design solution. An alternative approach is proposed in the section to follow. The theories developed in the various parts of the thesis are illustrated by the example of a vehicle identification system design.

2.3 Operational Requirements and Constraints

The requirements and constraints involved in the design and operation of a practical object identification system using low-frequency transponders must be considered carefully to achieve optimum system performance. When seeking an optimum design of such a system, the broad viewpoint of optimizing all the design variables available for selection is very inconvenient not only from the point of view of initial understanding but also from the point of view of mathematical tractability. The approach used in the analysis to follow is to initially assume a set of design variables to be fixed while the optimization is performed on the remaining design parameters. The system parameters, such as the interrogation frequency, reply message length, and modulation parameters, are considered fixed while the optimization is performed

on parameters such as the antenna shape and size, sensing distance, transmitter power, transponder properties and receiver design. This approach is not meant to imply that a question such as *what is the optimum frequency* should not be considered, rather, it suggests that this very important question can best be posed when the results of other optimizations are properly understood.

In order to optimize the performance of a two-port transponder system, one must carefully consider the operational characteristics of such a communication system. In the case of the two-port system shown in Figure 2.1, to maximize the power transfer from the transmitter to receiver, for example, design of antennae which provide the electromagnetic coupling, matching of the transponder and receptor antennae to their source or load impedance, and design of very low loss transponder circuitry are some of the major considerations.

Many other system parameters need to be optimized for correct operation of the system. One of the tasks is to get a sufficient amount of power delivered to the transponder circuitry for its proper functioning. It is also necessary to provide the power at an optimum voltage for the transponder to operate efficiently. The system para-

meters affecting the efficiency of the transponder must be investigated. An important factor for consideration is the optimum transmitter power. The sensing distance is another important parameter in object identification applications. The electromagnetic losses in the transmitter and reply paths depend principally on sensing distance. The design of the antennae must be optimized with respect to the sensing distance required. An optimum performance of the transponder also requires a minimum coupling between the interrogation and reply signals. This minimization must be performed at all levels of system design where any possibility of interference from the interrogation signal affecting the reply signal exists. Environmental noise also plays a major role in the passive object identification systems. The optimum information transfer rate is governed by the acceptable level of errors in the reply signal. The optimization of the parameters to achieve the desired performance is carried out in view of the constraints imposed in a practical system.

2.3.1 Optimization of Two-Port Transponder

The optimization of various parts of the identification system can be performed conveniently if the system is partitioned into subsystems such that

closely related parameters are grouped together. This approach allows local optimization to be performed within each subsystem without the need for complex mathematical derivations. Initially, the two-port transponder system has been divided into two subsystems, one containing all the antennae required for the communication to take place, and the other containing the transponder circuitry. The considerations for efficient functioning of the transponder and related optimizations are considered in this chapter, while the optimization of antenna structures is considered in Chapter Three. The interfaces of the two subsystems also require careful consideration for efficient operation of the overall system. The interfaces have been considered as part of the transponder and, therefore, included in this chapter.

The functional requirements of a two-port passive transponder for object identification application include a power supply, a code generator, and a modulator. A functional block diagram of such a transponder is shown in Figure 2.2. The receptor antenna characteristics are optimized such that the electromagnetic coupling between the transmitter and the receptor is maximized. These are discussed in the chapter on antennae. The resistor R_2 in Figure 2.1 represents the sum of the powers dissi-

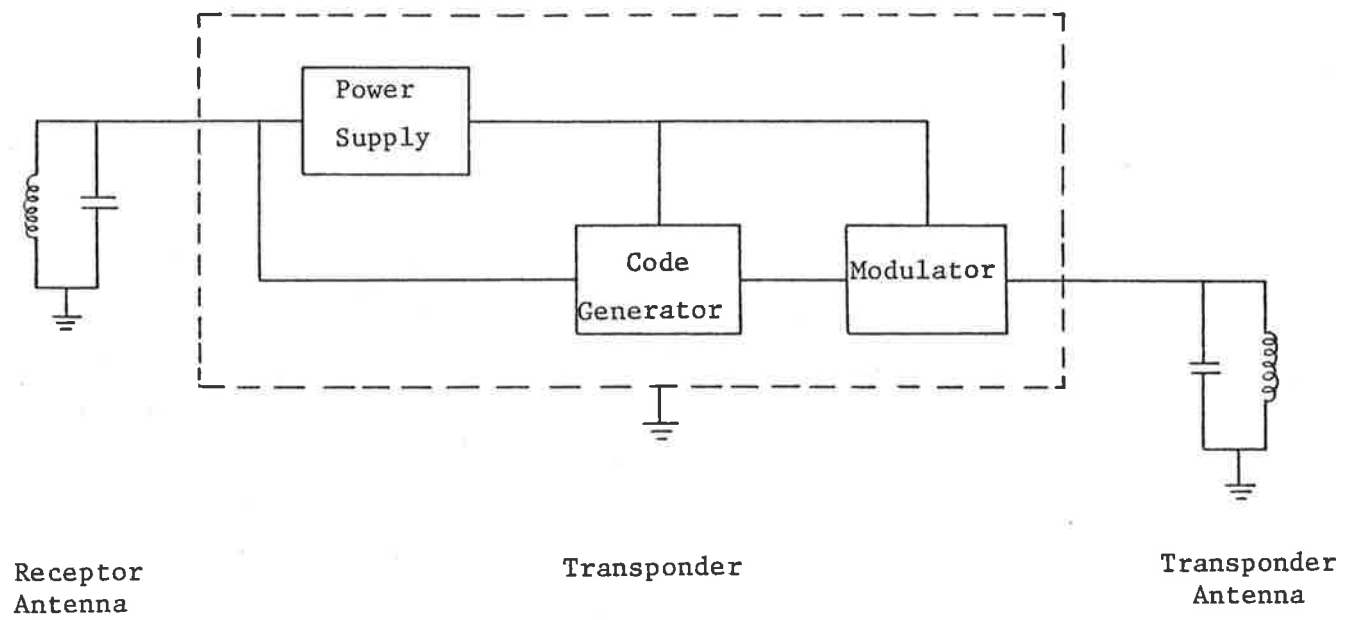


Figure 2.2 Functional Block Diagram of a Two-port Transponder

pated in receptor coil and transponder electronics, while the resistor R_3 represents the power dissipated to establish the desired stored energy in the transponder antenna. Again, some of this power is dissipated in the transponder coil, and the rest in the electronic circuitry.

The principal objective of the optimization of the transponder parameters is to design a two-port system with its antennae such that the power transfer from the receptor antenna to the transponder antenna is maximized. Under its normal operating conditions, the two-port transponder acts as a power converter which receives power from the source at one frequency, ω_1 , and outputs it at a subharmonic of the input frequency, ω_2 , as illustrated in Figure 2.2. Practical transponder circuits, however, do not operate as linear power converters, particularly at the low power levels, but rather exhibit a threshold behaviour: that is, the circuit ceases to function below a certain threshold voltage. This threshold voltage depends upon the technology employed for the transponder circuitry.

Considerations of proper power matching in its general sense between the transponder circuit and its input and output ports is complicated by several factors. One problem stems from the fact

that at low power levels the general theorems of linear network theory are inapplicable due to the non-linear behaviour of the transponder circuitry. Another important characteristic of the low-frequency transponders, that the coupling between the near field antennae occurs principally through their stored energies rather than any dissipated or radiated power, implies that the question of optimum antenna design strategy is, therefore, one of maximizing the stored energies rather than matching of power between a source and a radiation resistance.* The situation is further complicated by the fact that practical transponders operate as neither extremely lossy nor lossless converters of energy and the complex behaviour of practical transponders makes generalization extremely difficult.

An analysis of the two-port transponder based on the assumption that the energy conversion process is sufficiently inefficient for the input and output ports to be substantially isolated leads to the simple conclusion that a power match between the intrinsic losses of the receptor antenna and the transponder loading should be achieved. This conclusion, however, is devoid of any practical

* This view is valid for transmitter and receiver antennae. Transponder loading must be considered for the receptor and transponder antennae.

interest because the assumption of low efficiency of energy conversion within the transponder can, in fact, be violated in practical transponders, and it is highly desirable that it be so.

In view of the considerations outlined above, an appropriate design procedure for practical transponders is outlined here. The analysis is based on the assumption that the receptor antenna is excited by a fixed level of magnetic field created at a given sensing position by the transmitter antenna.

A reasonable description of conditions at the interface between the transponder output port and the transponder coil, under these circumstances, is provided by the equivalent circuit shown in Figure 2.3. Such an equivalent circuit is appropriate because the output circuit of the transponder may be considered as interconnection of logic gates of suitable type. The technologies suitable for integration of passive transponders is investigated in a later chapter. The prototype transponder is designed using CMOS MSI devices. The output of a CMOS gate may be represented by a squarewave source and a source resistance R_s . The value of the source resistance depends upon the output voltage. Typical dc output resistance variations with output voltage, and the supply voltage for a CMOS inverter is shown

in Figure 2.4 (Redfern, 1975).

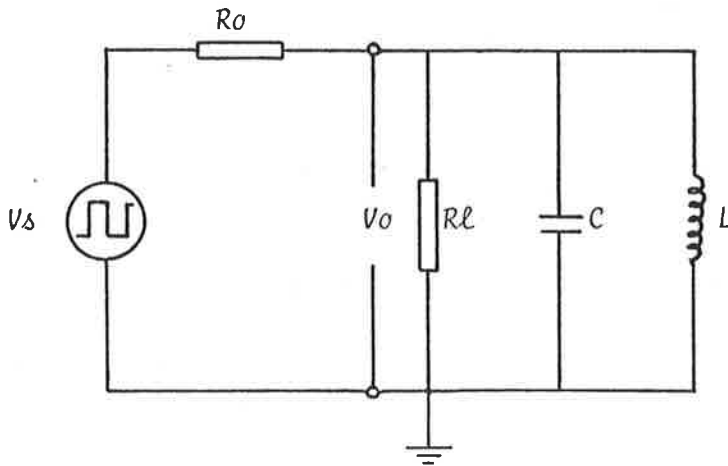


Figure 2.3 Transponder Output Equivalent Circuit

In order to derive the best performance from the transponder output circuit and transponder antenna, the stored energy in the inductor should be maximized, subject to the constraint of fixed quality factor. It can be easily shown that the optimization is provided by making the antenna intrinsic losses as small as possible, and ensuring that the damping required by the communication bandwidth provided by the effective output impedance of the transponder circuit. The range of supply and output voltages for which the desired output resistance is obtained may be found from the characteristic plots of the type shown in Figure 2.4. In practical transponder circuits it is not possible to keep the output resistance constant, in view of the changing

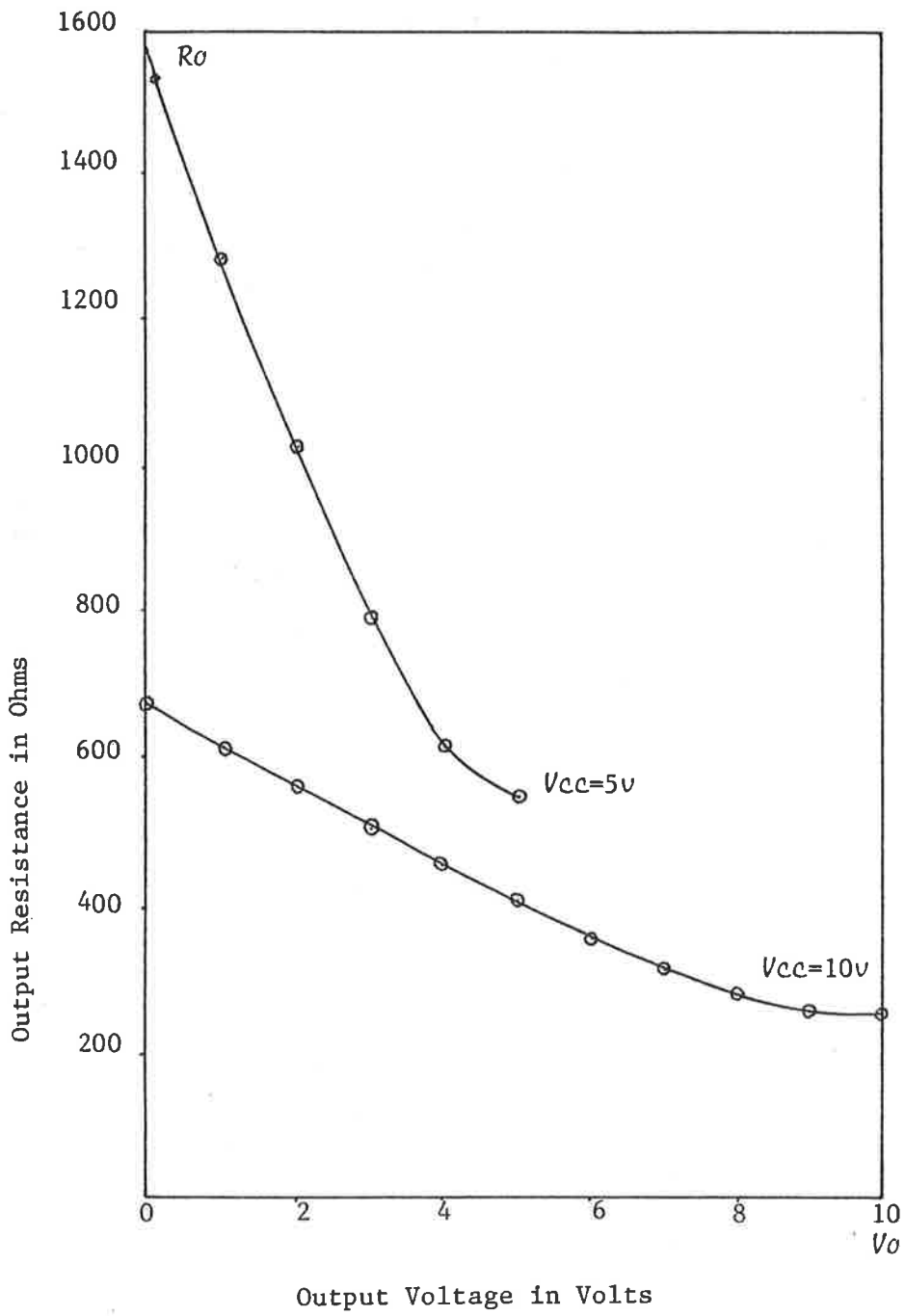


Figure 2.4 Typical CMOS Gate Output Characteristics

supply and output voltages, without introducing further losses in the transponder.

The optimization of the remaining parameters of the transponder circuit should now be considered. The forward voltage drop in the rectifier, the control circuit loading, and the losses in the transponder output switch are the major contributors to the losses in the frequency conversion performed by the transponder. Some degree of matching is required to maximize the power transfer from the receptor to the transponder antenna because of the losses in the frequency conversion.

An analytical procedure to determine the optimum antenna and transponder parameters can be very cumbersome, due to the non-linear behaviour of the transponder and the interaction between the two tag antennae. In this situation, the optimum operating parameters for the two-port transponder are obtained, and a better understanding of the behaviour of the transponder is achieved, by experimental means. The procedure requires determination of transponder output resistance variations with excitation level. The optimum transponder antenna design, as discussed earlier, requires an inductance with high intrinsic quality factor and the loaded quality factor of the resonant circuit as demanded by the communication

bandwidth is provided by the output resistance of the transponder circuitry. The nature of the transponder output circuit is such that a parallel resonant circuit is suitable for the transponder coil. The mid-range value of the output resistance is used as the initial approximation for the determination of the optimum transponder antenna parameters using an iterative technique. This value of the output resistance and known quality factor are used to calculate the self-inductance of the transponder coil. The efficiency of the power conversion by the transponder is plotted against excitation level for each iteration. The excitation level which results in maximum efficiency is used to calculate the output resistance to be used for the next iteration. The procedure is repeated until convergence is achieved. The receptor coil is designed to provide maximum coupling between the transmitter and the receptor. The impedance at which the input coupling is provided is matched to the losses in the transponder input circuitry at the excitation level previously determined. It should be noted, at this point, that the above procedure produces values for impedance levels of the antennae, and excitation level for the transponder for an optimum performance of the two-port system, but in the case of moving object identification application, where the transponder experiences a varying level of excitation as

it travels through the non-uniform field created by the transmitter, it is not possible to ensure that these parameters provide the most efficient power conversion from the transmitter frequency to the receiver frequency as the transponder experiences varying levels of excitation. Although this difficulty is inherent in all moving object identification schemes, this does not imply that the calculation of the optimum excitation level is of no practical benefit where the variations in the input excitation cannot be controlled without introduction of further losses. The importance of the experimental determination of the optimum excitation level is two-fold. First of all, it provides an opportunity for better understanding of the behaviour of the two-port transponder and, secondly, it allows the calculation of optimum antenna parameters.

2.4 Experimental Results - Two-Port System

A two-port passive subharmonic transponder was constructed using CMOS circuitry in accordance with the block schematic of Figure 2.5. The transponder was operated at an interrogation frequency of 100 MHz, and returned a continuously recycled reply code of length 64 bits in the form of PSK modulated carrier frequency of 50 kHz. Eight car-

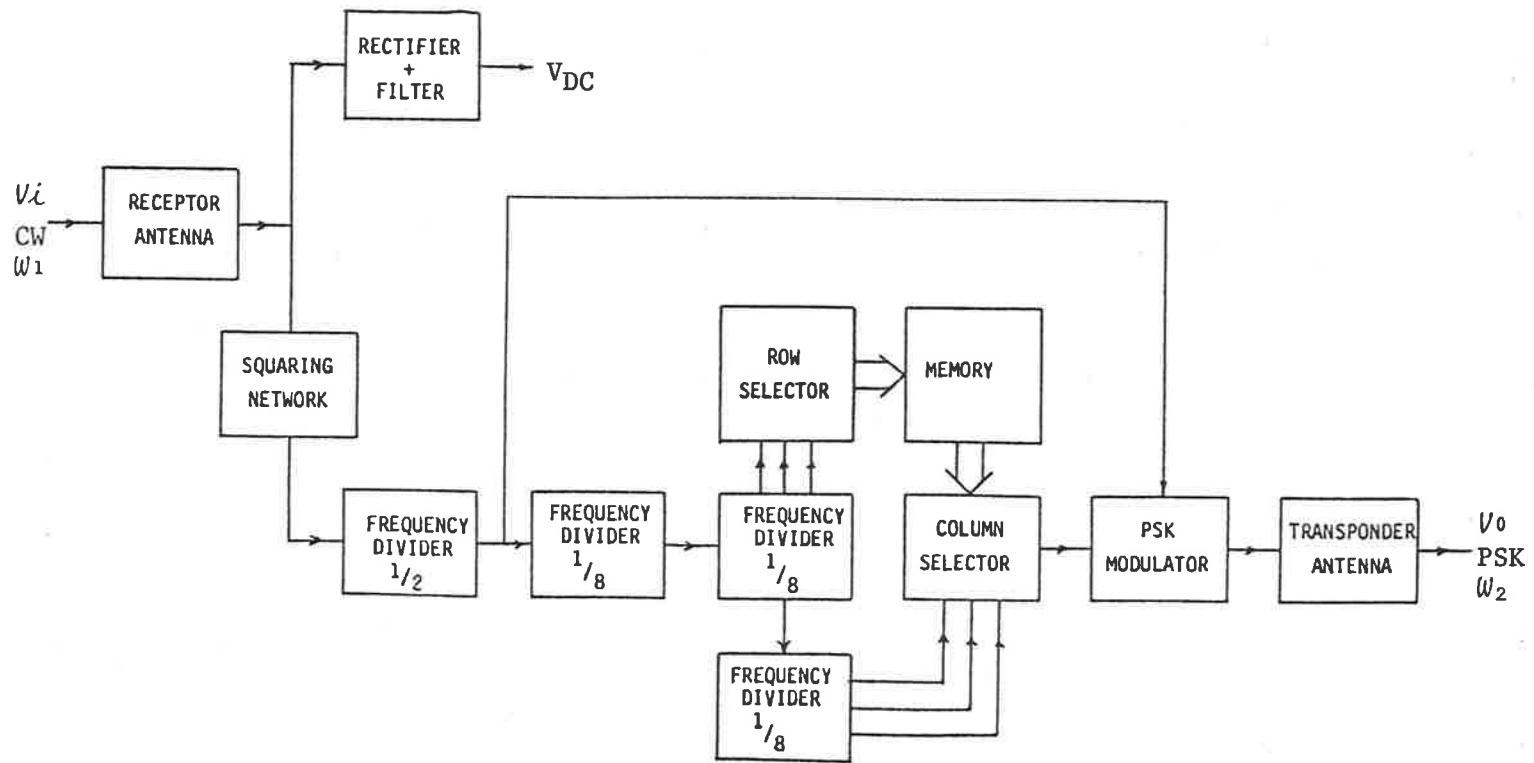


Figure 2.5 Two-port PST Block Diagram

rier cycles constituted one bit of the reply code. The system was designed to operate with air cored antennae.

One of the initial experimental studies involved the investigation of the voltage dependence the output resistance of the transponder. For this purpose the transponder is excited by a sinusoidal source at the interrogation frequency. The output of the transponder is terminated in a resistive load. For a given level of excitation the output voltage of the transponder is plotted against the load resistance, as shown in Figure 2.6. The output resistance of the transponder is found from the value of the load resistor for which the output voltage is half of its no-load value. A plot of the variations of the output resistance with the dc output of the transponder internal power supply is shown in Figure 2.7.

The efficiency, η , of the two-port passive subharmonic transponder is defined as the ratio of power output measured at the transponder coil at the subharmonic frequency to the power input to the receptor coil at the transmitter frequency. For the measurement of efficiency the modulation circuit is disabled to allow for output power measurement at the reply carrier frequency. The efficiency of the two-port PST is plotted against the transponder in-

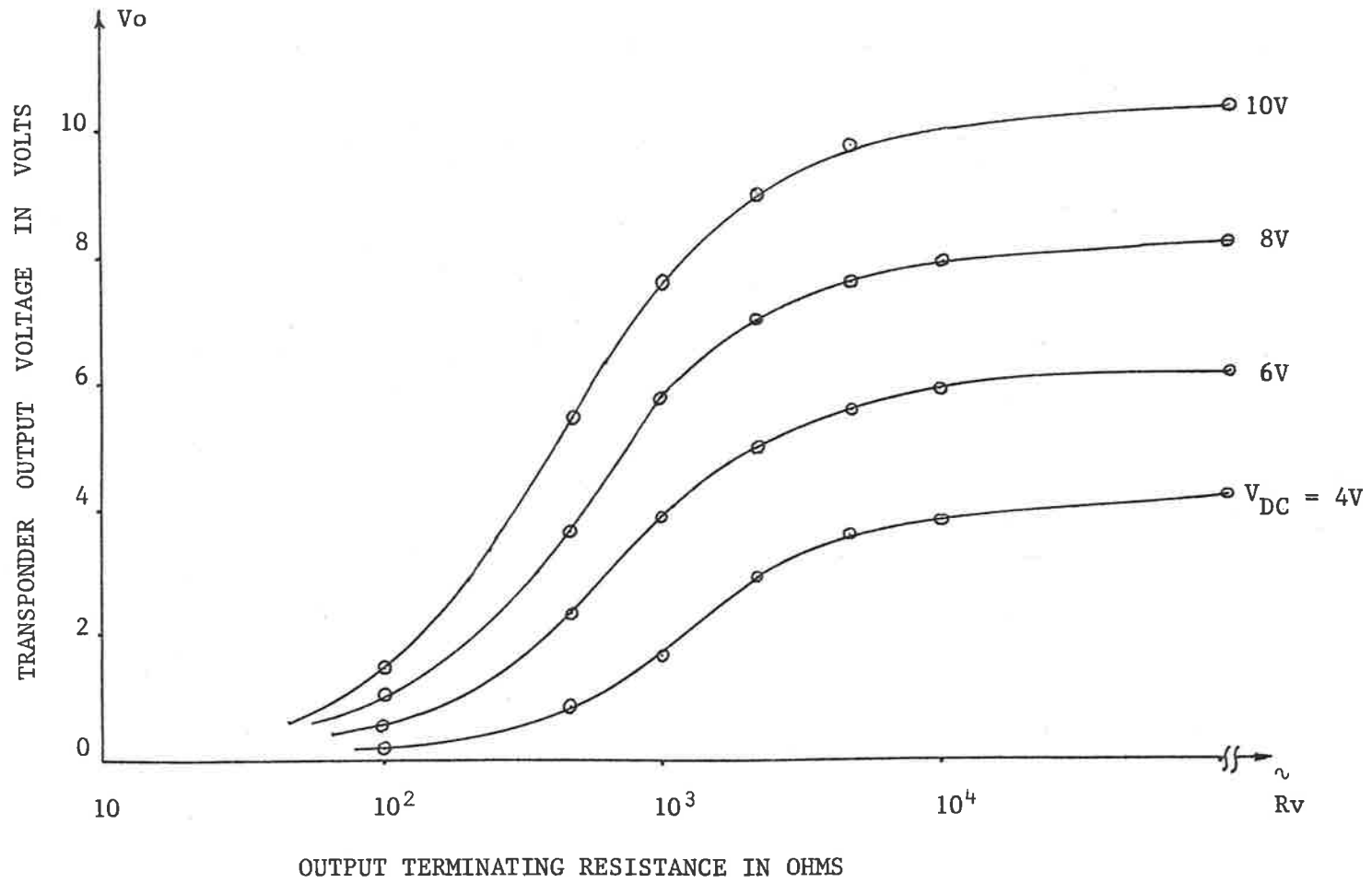


Figure 2.6 Measurement of Output Impedance of Transponder (Two-port)

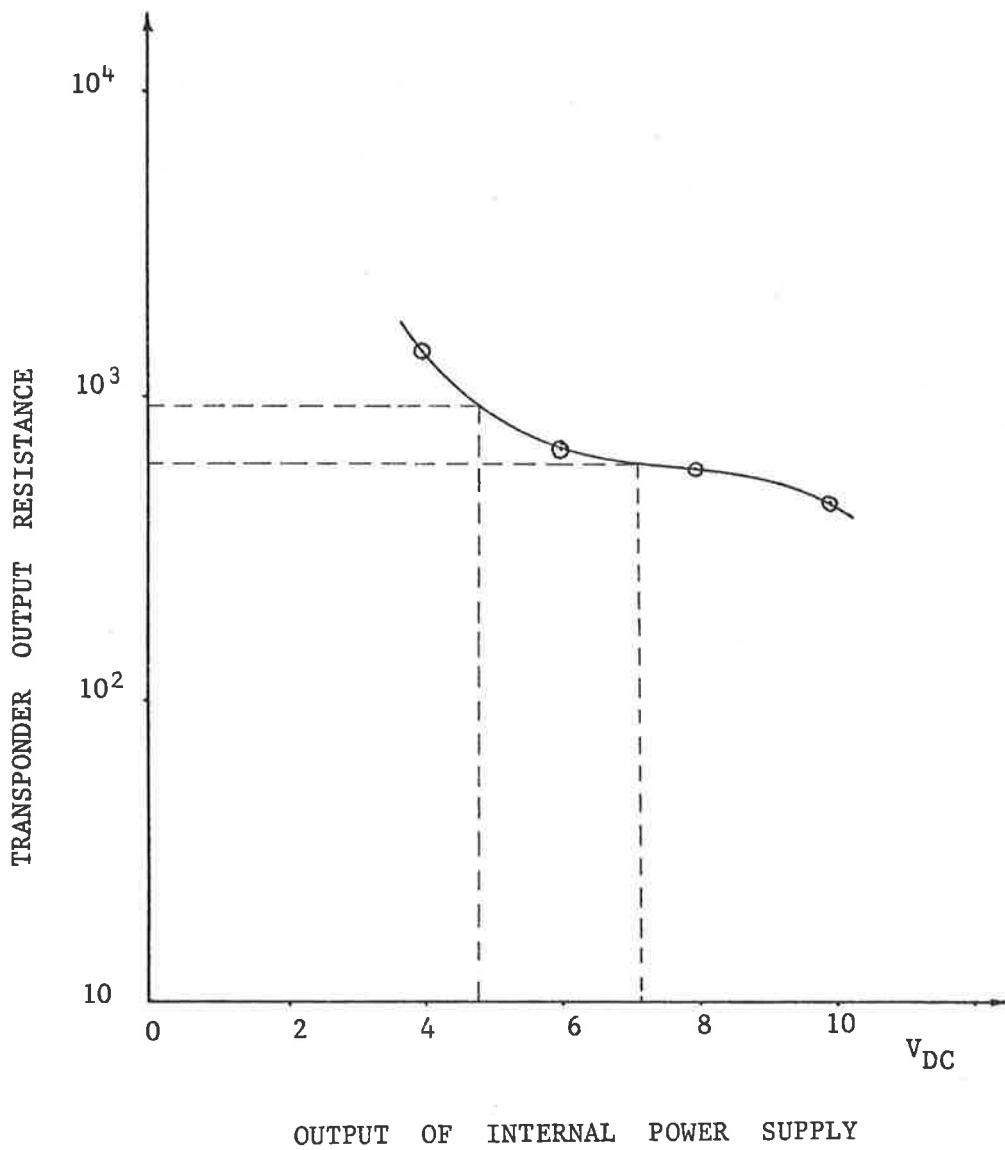


Figure 2.7 Transponder Output Resistance Variations with Excitation

ternal supply voltage in Figure 2.8. The two curves shown in this figure correspond to the initial and the final designs of the transponder coil. It should be pointed out that the apparently low efficiency is not due to the losses in the power conversion in the transponder, but is mainly the result of conversion of the squarewave output of the transponder into sinusoidal waveform, which appears across the transponder resonant circuit.

The prototype two-port transponder is then operated by energizing via the electromagnetic link. The system parameters for this system are given in Table 2.1. The resulting waveforms for two information sequences are shown in Figures 2.9 and 2.10, respectively.

2.5 The One-Port System

One aspect of the research on low-frequency object identification is to study ways of improving the overall efficiency of the identification system in a cost effective manner. A careful study of the limitations of the two-port passive subharmonic transponder reveals the areas where possibility of improvement exists.

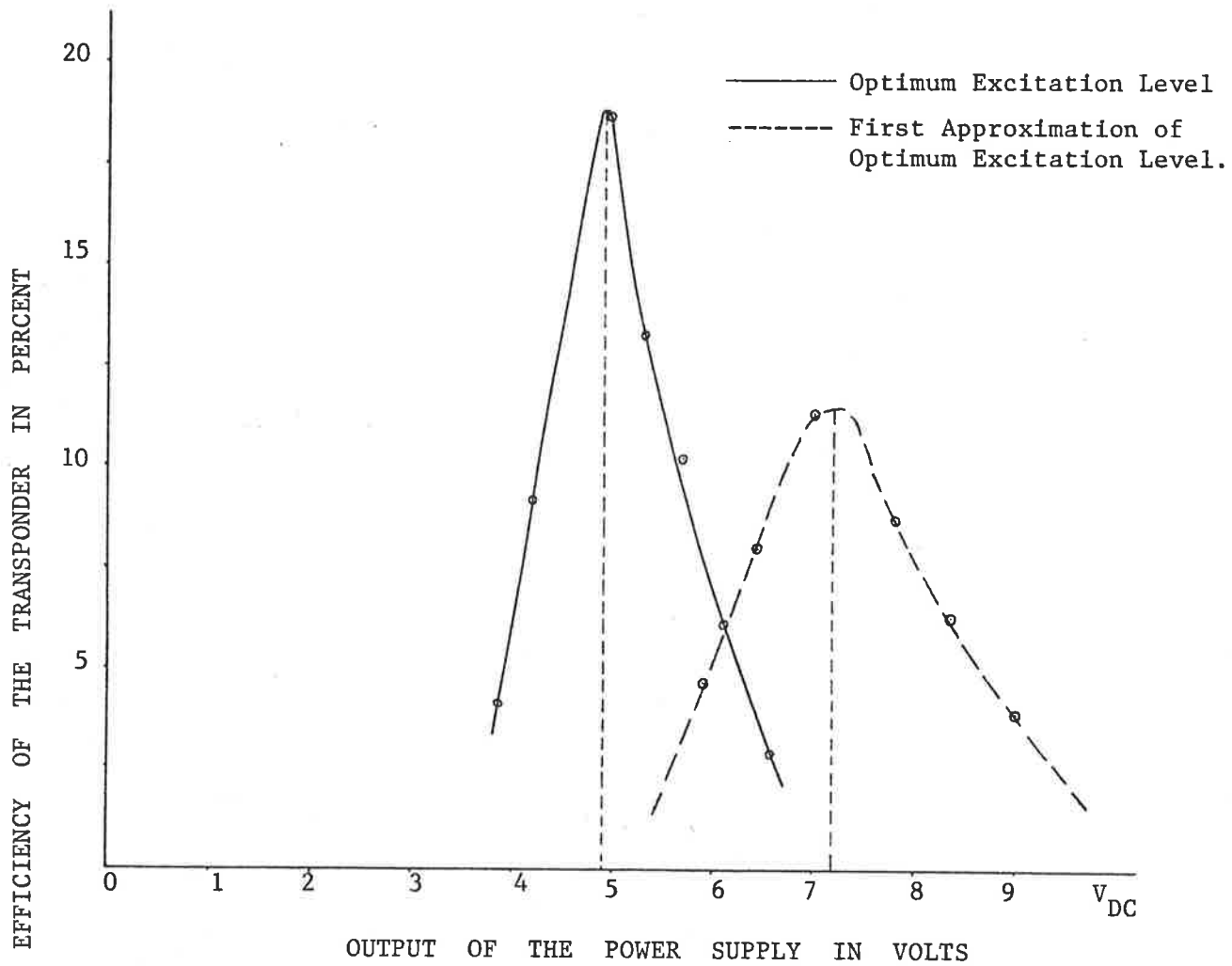


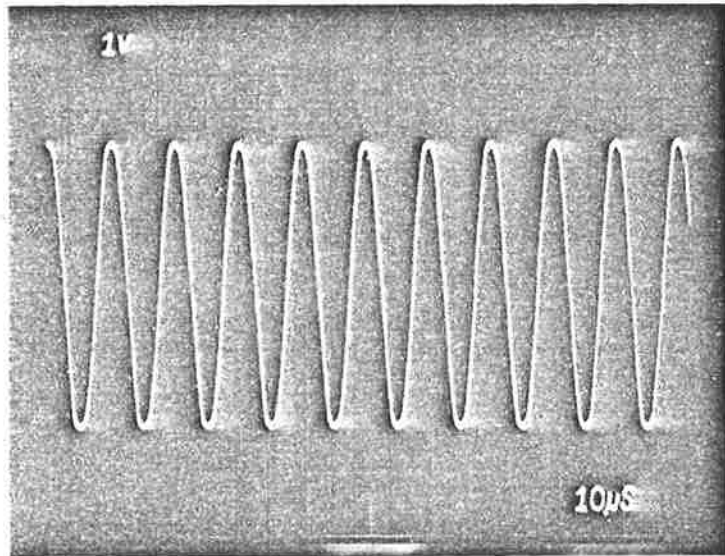
Figure 2.8 Optimization of Transponder Excitation Level

2.5.1 Limitations of Two-Port Transponders

Several factors which limit the performance of two-port passive subharmonic transponders become evident from the experimental study of the two-port transponders, and related analysis. One limitation, as discussed in the chapter on antennae, is that the device performance is sensitively dependent upon the total volume which can be allocated to its antennae,

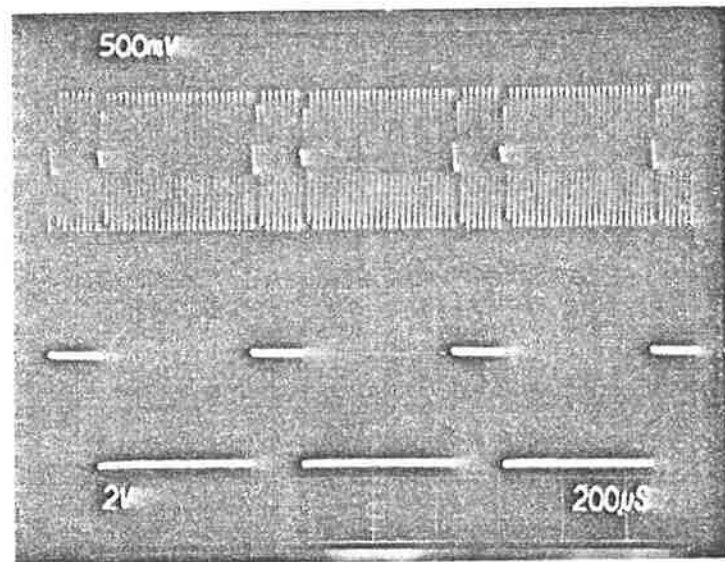
Table 2.1 Two-port Transponder Parameters

PARAMETER	SYMBOL	VALUES	UNITS
Transmitter Antenna Self-inductance	L_1	3.53	μH
Receptor Antenna Self-inductance	L_2	1.42	mH
Transponder Antenna Self-inductance	L_3	0.358	mH
Receiver Antenna Self-inductance	L_4	3.57	μH
Transmitter Frequency (CW)	f_1	100	kHz
Receiver Band Centre Frequency (PSK)	f_2	50	kHz
Transmitter Antenna Quality Factor	Q_1	34	1
Receptor Antenna Quality Factor	Q_2	37	1
Transponder Antenna Quality Factor	Q_3	8	1
Receiver Antenna Quality Factor	Q_4	8	1
Transmitter Antenna Dynamic Resist. at f_1	R_1	75.4	Ω
Receptor Antenna Dynamic Resist. at f_1	R_2	33.01	$k\Omega$
Transponder Antenna Dynamic Resist. at f_2	R_3	900	Ω
Receiver Antenna Dynamic Resist. at f_2	R_4	8.97	Ω
Power Dissipated in the Transmitter Coil	P_1	0	dBW
Power Received by the Receptor Coil	P_2	-22.15	dBW
Power Dissipated in the Transponder Coil	P_3	-29.42	dBW
Power Received by the Receiver Coil	P_4	-62.37	dBW
Transponder Sensing Distance	S_T	0.75	met.
Transmitter Link Coefficient of Coupling	k_1	6.96×10^{-3}	1
Receiver Link Coefficient of Coupling	k_2	7.87×10^{-3}	1
Transponder Conversion Efficiency	η	18.76	%



Receptor
Waveform

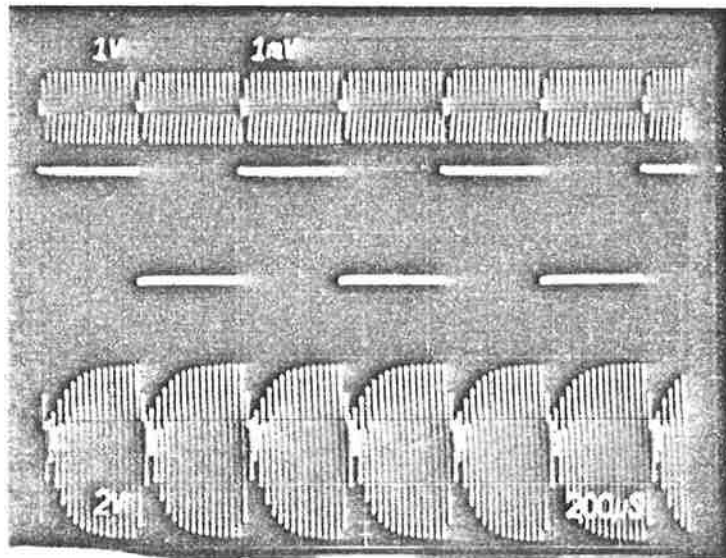
Figure 2.9a Voltage across the Receptor Antenna



Transponder
Waveform

Cyclic
Code

Figure 2.9b Modulating and Modulated Signals

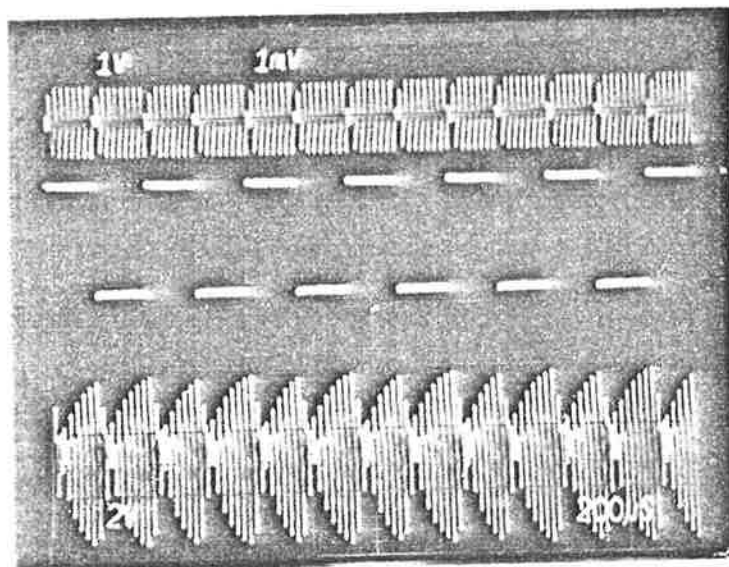


Transponder
Waveform

Cyclic
Information
Sequence

Received
Waveform

Figure 2.10a Information, Transponder and Receiver Waveforms

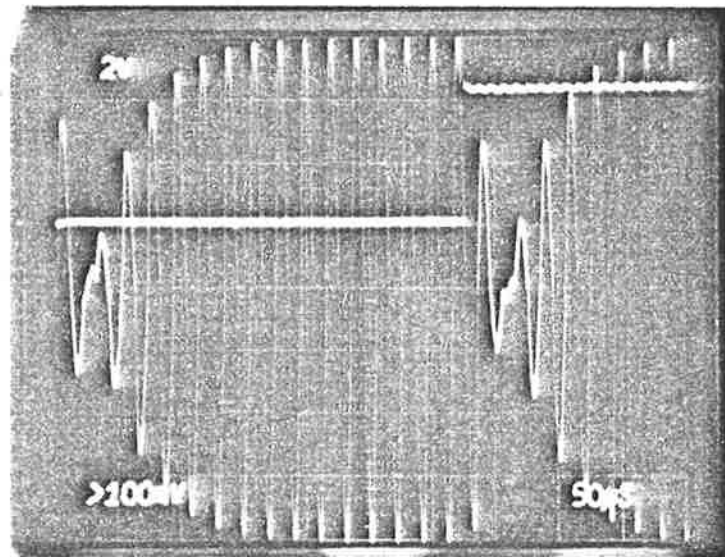


Transponder
Waveform

Cyclic
Information
Sequence

Received
Waveform

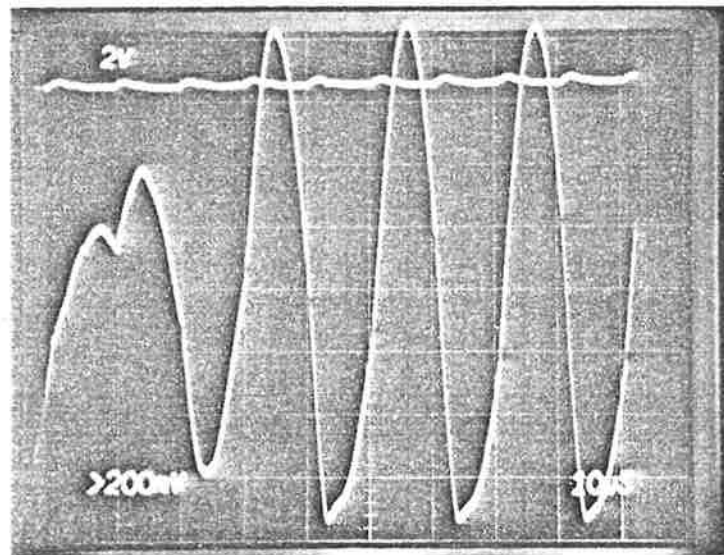
Figure 2.10b Information, Transponder and Receiver Waveforms



Information
Waveform

Transponder
Waveform

Figure 2.10c Expanded View near the Point of Switching



Information
Waveform

Transponder
Waveform

Figure 2.10d Expanded View near the Point of Switching

and in applications where the transponder volume must be kept small the competition between the receptor and transponder antennae for the available space is undesirable. A second factor is that the conversion of the received energy first to dc and then reconversion of this energy to the subharmonic frequency is made inefficient by the forward voltage drop in available rectifier diodes, particularly in applications where low levels of excitation voltage prevail. A third, and very significant, limitation is that in the preceding design the quality factor of the transponder coil is limited by the consideration of communication bandwidth to a value much less than the values that can be achieved in resonant circuits of the allowed size at the working frequency. This limitation on the quality factor is seen to be especially undesirable in view of the fact that the strength of coupling between near-field antennae is established by the stored energy rather than the dissipated power within them, and operation of the transponder coil at a higher quality factor would enhance the transfer of power across the reply link. Another factor which contributes to the overall inefficiency of the system is the modulation. In the binary PSK modulation, when a change in the information bit from a "1" to a "0" occurs, the oscillations across the transponder coil are first allowed to decay to zero and then build up again in

the opposite phase. This process is essentially a lossy one.

Some of the limitations of the basic two-port system, as discussed above, may be overcome by improved design techniques. The inefficiency in the energy conversion due to the forward voltage drop across the rectifier diode may be overcome by some form of active rectification. The losses introduced by the PSK modulation in its basic form may be eliminated by providing a means of phase reversal of the oscillations without first dissipating the energy in the tuned circuit. This objective may be easily accomplished, while still preserving the essential nature of the resonant circuit and providing large amplitude oscillations of very little power, by inserting a polarity reversal switch between the transponder inductor and its resonating capacitor, the switch being designed such that it presents only a low loss to the tuned circuit, and being operated at the zeroes of inductor current or capacitor voltage.

Although all these ideas may be pursued in the context of the two-port transponder, their essential nature is preserved in the broad concept of the one-port transponder, which is discussed, briefly, in the next section.

2.5.2 The One-Port PST

The general form of the one-port passive sub-harmonic transponder employing inductive coupling is shown in Figure 2.11 (Cole et al, 1979). Here a single inductor L_2 is energized by a transmitter, and coupled to a receiver by mutual inductance M_1 and M_2 , respectively. The transmitter and receiver operate at frequencies ω_1 and ω_2 , respectively, with $\omega_2 = \frac{\omega_1}{2}$. The transponder inductor is resonated, in

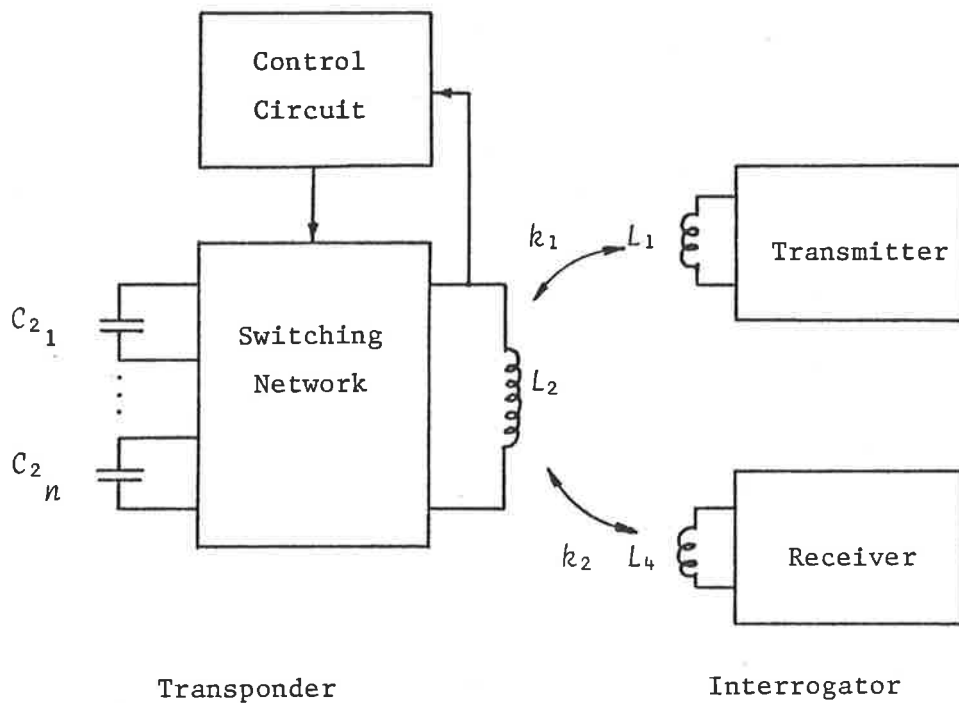


Figure 2.11 General Form of One-Port Transponder

the general sense, by an array of capacitors $C_1 . . . C_n$, which is connected to the inductor by a low loss switching network. The network is operated by a control circuit such that the energy exchange between the inductor and the capacitor array takes place in a cyclic fashion where the fundamental period is some multiple of the period of CW transmitter signal, and the current in the inductor consists of a series of harmonically related frequencies along with ω_1 and ω_2 . These two frequencies are responsible for transfer of energy from the transmitter to the general resonant system in the transponder, and transfer of a reply code from the transponder to the receiver. In most applications where the transponder is required to provide a reply code of more than one bit, some further variation of the switching sequence is required to provide the modulation (Eshraghian et al, 1981). In all of these operations the switching network is designed to be low loss and controlled in such a way that excessively dissipative transients do not occur. Reply code modulation can thus be provided while the energy required to maintain the oscillation is kept to the minimum which can be achieved consistent with the quality factors of the components used.

In its simplest form of realization, the one-port transponder contains a single inductor, capa-

citor and single pole switch with an appropriate control circuit. The parallel connected version of one-port transponder is shown in Figure 2.12. The waveforms of voltages at the transmitter, transponder and receiver coils are also shown in the figure. The transmitter and transponder coils are tuned with high quality factor, while the receiver quality factor is limited by the information transfer rate. The transponder is not limited by the communication bandwidth constraint because of the way in which energy exchange between the energy storage elements of the transponder takes place. The waveforms are shown for an unmodulated reply signal. PSK modulation of the reply carrier may be achieved by displacing the switch control voltage by one half period with a resultant interchange of the on and off periods of the switch.

2.5.2 Transponder Coil Considerations

The understanding of the mechanism which allows the use of a reasonably high Q transponder coil is of importance. There are two important features of this mechanism. The first is that the ratios of stored energies at the various harmonics of the inductor current are determined by the waveform imposed on the transponder resonant system by the control circuit, rather than the details of the loss

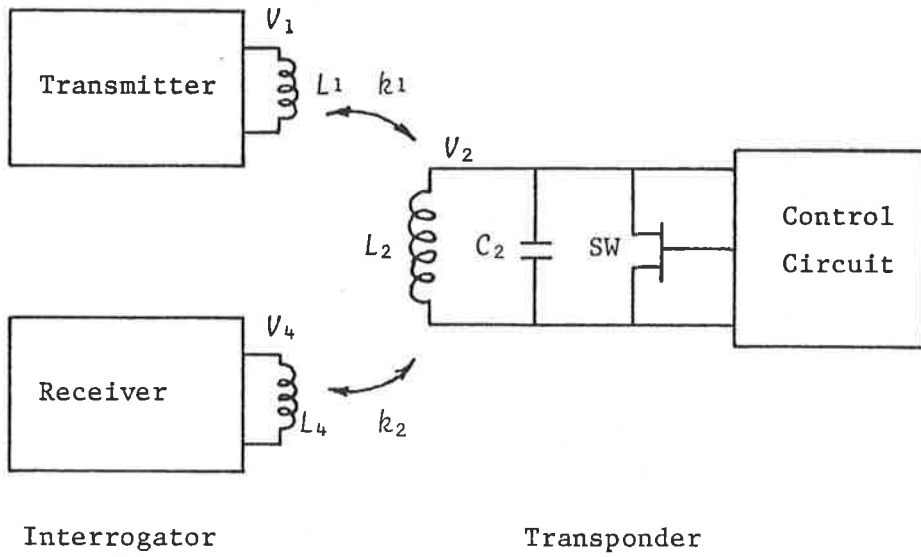


Figure 2.12a Basic One-Port Transponder

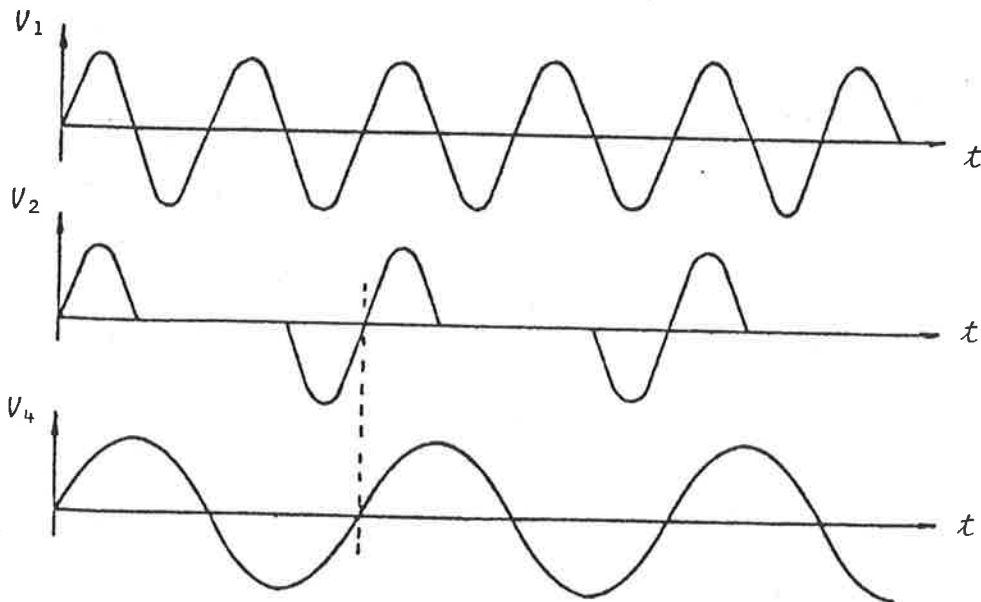


Figure 2.12b One-Port Transponder Waveforms

mechanisms which occur at the various harmonic frequencies in the individual circuit elements. The second feature is that the energy needed to sustain the oscillation at any particular level is related to the amplitude of that oscillation, and an overall effective quality factor which is defined by the total stored energy in the resonant system, and total losses in one complete cycle of oscillation of the waveform imposed by the control circuit. The fundamental quality factor of the resonant circuit is thus defined,

$$Q_F = 2\pi \frac{\left| \begin{array}{c} \text{Energy stored in the} \\ \text{transponder resonant circuit} \end{array} \right|}{\left| \begin{array}{c} \text{Energy dissipated in one} \\ \text{fundamental period} \end{array} \right|} \quad (2.1)$$

or
$$Q_F = \pi L_2 I_m^2 / P_2 T \quad (2.2)$$

where

L_2 is self-inductance of the transponder coil

I_m is peak value of the inductor current

P_2 is mean power delivered to the transponder

T is $2\pi/\omega_2$ fundamental period of oscillation

For the transfer of energy over the electromagnetic coupling links, in passive transponder systems, one is not concerned with the total stored energy in the transponder, but only that associated with the harmonics of the inductor current at ω_1 and ω_2 , respectively. In view of the relation

$Q = \omega LI^2 / P$ between the quality factor Q , rms current I in the inductance L and mean rate of power dissipation P for a simple tuned circuit, the power transfer between the transponder and the interrogation antennae may be easily visualized by defining two effective quality factors for the transponder coil at the transmitting and receiving frequencies

$$Q_2' = \omega_1 L_2 I_2^2 / P_2 \quad (2.3)$$

$$Q_3' = \omega_2 L_2 I_3^2 / P_3 \quad (2.4)$$

The importance of these quality factors is that they allow the transfer of power across the two electromagnetic reply links to be calculated by means of simple relationships given below

$$\frac{P_2}{P_1} = k_1^2 Q_1 Q_2' \quad (2.5)$$

$$\frac{P_4}{P_3} = k_2^2 Q_3' Q_4 \quad (2.6)$$

- where P_1 is the power dissipated in the transmitter coil
 P_2 is the power dissipated in the receptor coil
 P_3 is the power dissipated in the transponder coil
 P_4 is the power dissipated in the receiver coil
 Q_1 is the quality factor of the transmitter antenna
 Q_2 is the quality factor of the receiver antenna
 k_1 is the coefficient of coupling between the transmitter and the transponder
 k_2 is the coefficient of coupling between the transponder and the receiver

In the practical realization of one-port transponders, it is not difficult to make the control power consumed within the transponder negligible in relation to the power which sustains the oscillation and, hence, contributes to the transfer of energy. Therefore, to a good approximation, P_2 and P_3 may be considered equal

$$P_3 / P_2 = 1 \quad (2.7)$$

Equations (2.5), (2.6) and (2.7) are now combined to give the overall power transfer ratio

$$P_4 / P_1 = k_1^2 k_2^2 Q_1 Q_2' Q_3' Q_4 \quad (2.8)$$

In view of the fixed relationships between the harmonics of the oscillation waveform, and the fact that Q_F , Q_2' and Q_3' are all defined in relation to the same mean power dissipated in the resonant circuit, it is possible to express Q_2' and Q_3' in terms of Q_F

$$I_2 / I_m = K_2 \quad (2.9)$$

$$I_3 / I_m = K_3 \quad (2.10)$$

Thus

$$Q_2' / Q_F = 4K_2^2 \quad (2.11)$$

$$Q_3' / Q_F = 2K_3^2 \quad (2.12)$$

The overall power transfer ratio between the transponder and the receiver then becomes

$$P_4 / P_1 = k_1^2 k_2^2 X_p Q_1 Q_F^2 Q_4 \quad (2.13)$$

where

$$X_p = 8K_2^2 K_3^2 = 8 / 9\pi^2 \quad (2.14)$$

The power transfer relation and the losses in the control circuit for one-port transponder has been considered in greater detail by Eshraghian (1980). It must be pointed out that in the power transfer relationship given by the Equation (2.12) the designer is at liberty to choose both Q_1 and Q_F as high as practicable, and only Q_4 is limited by modulation bandwidth considerations.

2.5.4 Control Circuit Considerations

The control circuit for the one-port transponder is required to perform a number of important functions in order to operate the transponder efficiently. One function is that of synchronization of the switch operations with the transponder excitations such that the switch is closed accurately at the zeroes of the resonant circuit voltage so as to avoid undue dissipation of the stored energy.

The opening of the switch to permit resumption of the oscillation is usually, but not necessarily, one period of the transmitter frequency later. The control circuit must also perform its operations coherently for both phases of the modulation waveform so that the phase of the transmitter frequency component of the inductor current is not altered by the occurrence of bit changes in the reply code, and a continual flow of energy from the transmitter to the transponder is assured. One method by which these functions can be realized is by means of a phase-lock-loop, as shown in Figure 2.13 (Cole et al, 1980).

It is also possible to achieve the desired results by more straightforward logic circuits which monitor the inductor voltage and the derivative of the switch current during the off and on periods of the switch respectively. In this case it is desirable to incorporate circuit functions which compensate for the delays between the recognition of the zeroes of the waveforms monitored and the consequent opening or closing of the switch.

Independent of the control circuit design, the amplitude of the transponder oscillation builds up over a period of $T_D = Q_F/\omega_2$, which is normally quite large compared with the period of reply code modula-

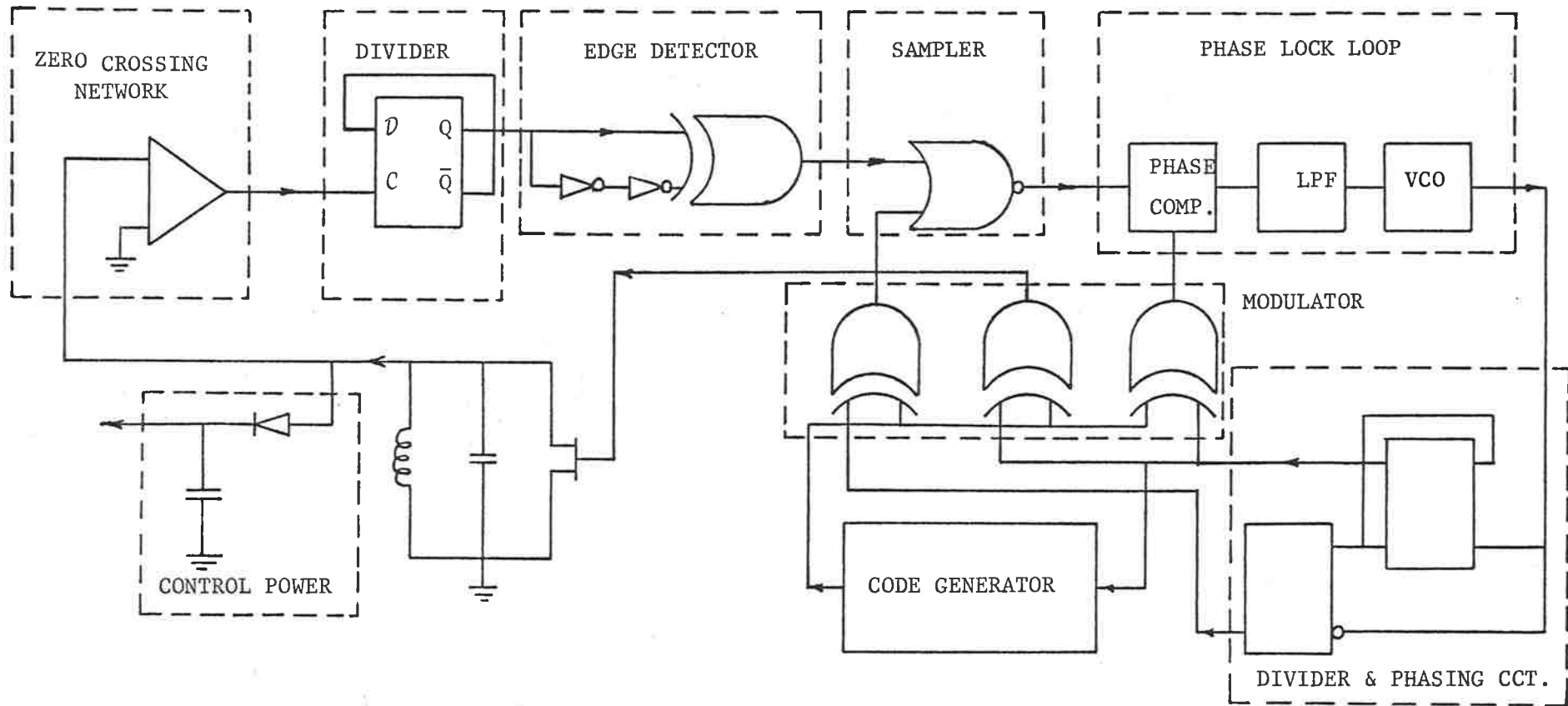


Figure 2.13 One-port Transponder Realization

tion. It is, therefore, desirable to delay, when the transponder is first energized, the commencement of the modulation until the transponder stored energy has reached a reasonable fraction of its maximum level.

A comparison of signal-to-noise ratios for the one-port and two-port passive subharmonic transponder is shown in Figure 2.14 (Eshraghian, 1980). It may be noticed that one-port and two-port with equal size transponder antennae exhibit similar performances. However, if the transponder antenna of the one-port system is increased such that its dimensions are the same as the overall dimensions of the receptor and transponder antennae for the two-port system, then an improvement of approximately 10 dB in the system performance is achieved.

The reasons for not achieving a much better performance from the one-port transponder may be explained by considering the circuit shown in Figure 2.15a. The transistors S_1 and S_2 are assumed to be ideal switches to simplify the analysis. Initially, the signal is stored either as a current through the inductor or a charge on the capacitor. A transition of the circuit from one state to another is illustrated in Figure 2.15b. Waveforms for the voltage across the capacitor and the current through the

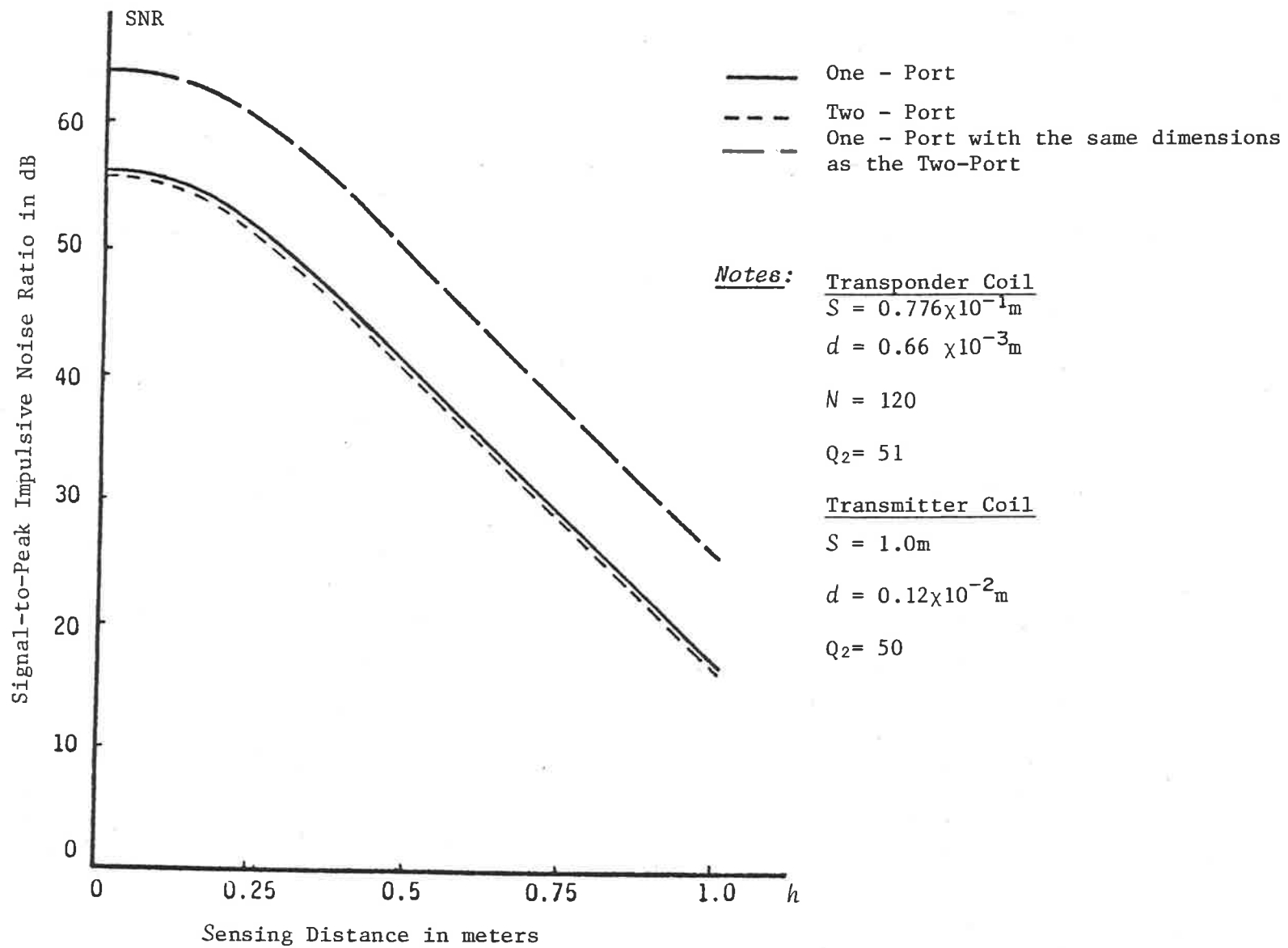
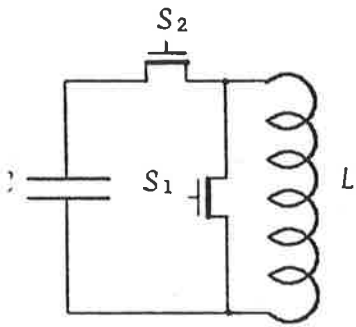


Figure 2.14 Comparison of the Performance of One-port and Two-port Transponders

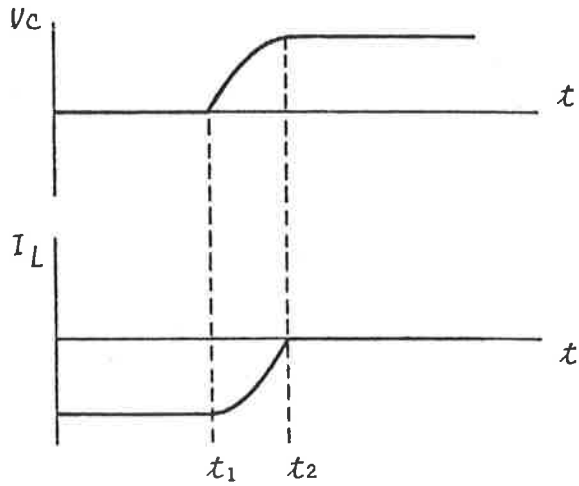
inductor are shown. Initially, the signal is stored as an inductor current, while the capacitor voltage is zero. At the instant t_1 , the series switch S_2 is turned on, while the parallel switch S_1 is turned off. At time t_2 , when the voltage on the capacitor reaches its maximum value, the series switch S_2 is turned off, and the charge representing the signal is stored on the capacitor. The current in the inductor is now zero. This quiescent state may be maintained on the capacitor for an infinite period in the case of a lossless capacitor.

In order to restore the circuit to its original state, at the instant t_3 , the switch S_2 is turned on, thereby initiating a current flow from the capacitor into the inductor, as shown in Figure 2.16a. The current will increase, reach a peak, and then oscillate back until time t_4 , when it reaches its maximum negative value once again. At this instant, the parallel switch S_1 is turned on, and the series switch is turned off. This completes one cycle of the circuit. If the transistors S_1 and S_2 were perfect switches, this circuit would be dissipationless, and one would be able to run an indefinite number of switching events without losing the signal energy. However, in order for the logic to function, the gate voltages which switch the transistors on and off must come from a signal such as the capacitor voltage V_C .

The details of the capacitor voltage and the inductor current in the proximity of time t_4 are shown in Figure 2.16b. The inductor current is nearing its maximum negative value, and the capacitor voltage is approaching zero. When the capacitor voltage reaches some small value $-V_0$, the switch S_2 begins to open. Thus, the voltage, instead of following a straight trajectory as it would if the transistor were not activated, will follow an exponential curve, eventually reaching zero voltage after a very long interval. This is due to the fact that no transistor is able to change from a completely on state to a completely off state without having a finite voltage applied to its control gate. In this instance, it is assumed that the transistor has zero resistance for gate voltages greater than V_0 , and an infinite resistance for gate voltages less than or equal to zero. The transistor, during the switching period, does not act as either a perfect short circuit or a perfect open circuit, instead it has a finite voltage across it, together with a finite current flowing through it. The energy dissipated in an elementary switching event described above is proportional to the current multiplied by the voltage, integrated over the switching transient.

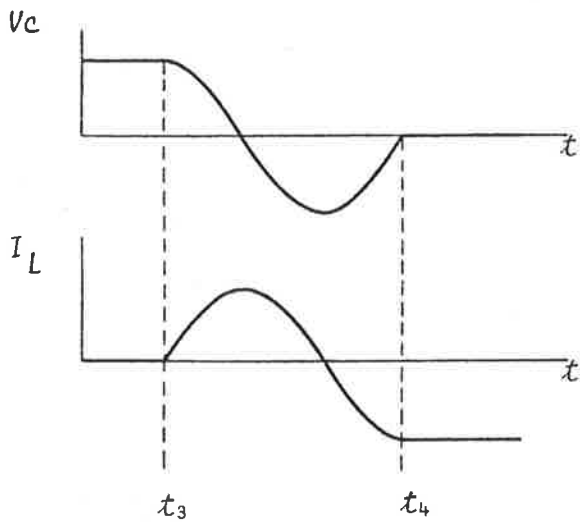


a. LC Network

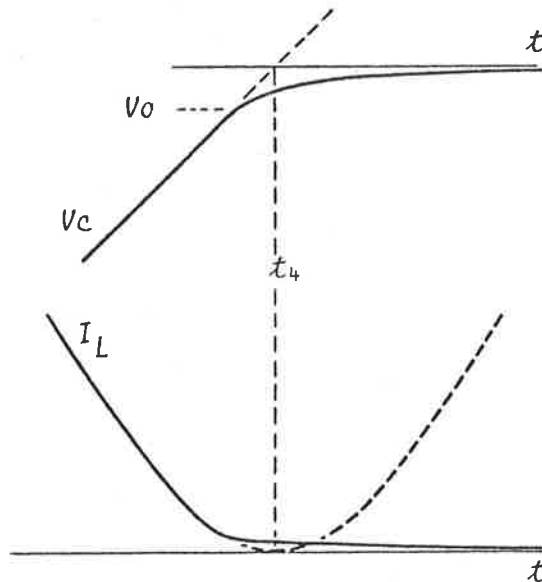


b. Waveforms when switching from one state to another

Figure 2.15 LC Network and Related Waveforms



a. Switching back to original state



b. Expanded View near switching point

Figure 2.16 Switching State of the LC Network

$$E_{SW} = \int v \cdot I dt \quad (2.15)$$

or

$$E_{SW} = \int v \cdot c \left(\frac{dv}{dt} \right) \cdot dt$$

or

$$E_{SW} = \frac{1}{2} c V_0^2 \quad (2.16)$$

where

E_{SW} is the energy dissipated in the switching transient

The total switching energy is, therefore, equal to that stored on the capacitor with an applied voltage V_0 . This is the maximum amount of energy required by any switch that cannot sense an infinitesimal potential difference.

It may be concluded from the above illustration that, in any switching network, a certain amount of energy, E_0 , is required to change the state of a switching device in the network. Furthermore, no other circuit constructed using that switching device can dissipate less energy than E_0 per switching event.

In the case of the one-port transponder, the losses occur in the control circuit, and during the switching event, because of the finite on and off resistances of the switch. However, more efficient use of the available area by the transponder antenna results in a

significant improvement in performance over the two-port version.

2.6 Conclusions

Some aspects in the design of low-frequency passive transponders employing near-field coupling have been considered in this chapter. An approach aimed at optimizing the various system parameters, and which allows separation of different parameters into suitable groups so that local optimization on these parameters may be performed, is suggested. An understanding of the limitations of the two-port transponders by experimental means has led to the concept of the more efficient one-port system. The antennae which provide the electromagnetic coupling and, therefore, form an essential part of the communication system, are considered in greater detail in the next chapter. However, because of difficulty in completely separating the antenna considerations from the transponder considerations arising from interaction of the antenna parameters with the transponder parameters, the optimization of some of the antenna parameters has already been considered in the above discussion. Other major considerations of the electromagnetic coupling are considered in the next chapter.

CHAPTER III

ANTENNA STRUCTURES

3.1 Introduction

Antennae are an integral part of any communication system, and the low frequency object identification systems are no exception. The communication in the low-frequency band is principally by means of near-field coupling rather than far-field. The antennae used, therefore, may be described by their quality factor and coefficient of coupling in a given system rather than their radiation resistance. The electromagnetic coupling required for the communication may be provided either by magnetic or electric fields. The electric field coupling systems utilize the mutual capacitance between the coupling elements, whereas the magnetic field coupling systems utilize the mutual inductance between the coupling elements.

The advantages offered by magnetic field coupling outweigh the ones offered by electric field

coupling in most applications, and hence its selection as a means of coupling for the object identification system.

The four antennae associated with a two-port transponder system are the transmitter, receptor, transponder and the receiver antennae. To clearly identify the relative position of the antennae, and because these terms will be regularly used in the text from here on, a typical two-port transponder system with the different antennae clearly distinguished is shown in Figure 3.1 below.

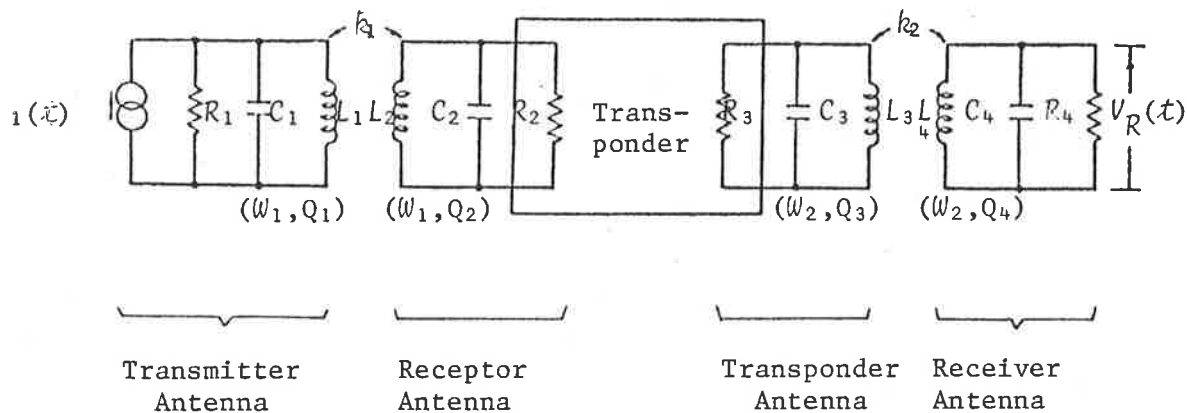


Figure 3.1 Antenna Terminology for Two-port Transponder System

3.2 Power and Propagation Constraints

The essential characteristics of the electromagnetic coupling link between the transmitter and the receptor antennae in a two-port passive transponder system is shown in Figure 3.2. The coupling between the antennae is by magnetic field alone. The resonant circuits are tuned to a frequency ω_1 radians per second with quality factors Q_1 and Q_2 , respectively, given by the following relationships.

$$\omega_1 = \frac{1}{\sqrt{L_1 C_1}} = \frac{1}{\sqrt{L_2 C_2}} \quad (3.1)$$

$$k_1 = \frac{M_{12}}{\sqrt{L_1 L_2}}$$

$$Q_1 = \frac{R_1}{\omega_1 L_1} \quad (3.2)$$

$$Q_2 = \frac{R_2}{\omega_1 L_2}$$

where

ω_1 is the frequency of resonance of the transmitter and receptor tuned circuits

k_1 is the coefficient of coupling between the transmitter and receptor coils

Q_1 is the quality factor of the transmitter resonant circuit

Q_2 is the quality factor of the receptor resonant circuit

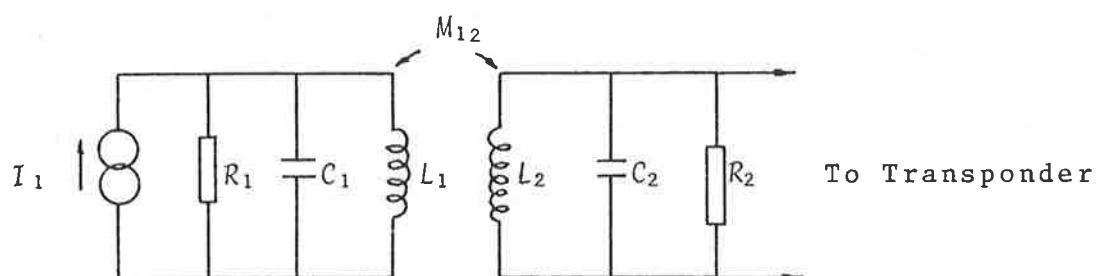


Figure 3.2 Electromagnetic Coupling between Transmitter and Receptor

Although the resonant circuits shown in the diagram are parallel tuned, in principle, either series or parallel tuning may be employed. The selection of series or parallel tuning in any given application depends upon practical considera-

tions such as deriving sufficient voltage for the operation of transponder circuitry. The damping resistor R_1 represents the overall effect of the losses in the capacitor C_1 , inductor L_1 , and those due to the environment. The damping resistor R_2 represents the losses in the capacitor C_2 , inductor L_2 , environmental losses, and the loading due to transponder input circuitry.

The reply link between the transponder and the receiver antennae is shown in Figure 3.3. The resonant circuits in the reply link are tuned to a frequency ω_2 radians per second with quality factors Q_3 and Q_4 , respectively, given by the following relationships.

$$\omega_2 = \frac{1}{\sqrt{L_3 C_3}} = \frac{1}{\sqrt{L_4 C_4}} \quad (3.3)$$

$$k_2 = \frac{M_{34}}{\sqrt{L_3 L_4}}$$

$$Q_3 = \frac{R_3}{\omega_2 L_3} \quad (3.4)$$

$$Q_4 = \frac{R_4}{\omega_2 L_4}$$

where ω_2 is the frequency of resonance of the transponder and the receiver tuned circuits

k_2 is the coefficient of coupling between the transponder and receiver coils

Q_3 is the quality factor of the transponder resonant circuit

Q_4 is the quality factor of the receiver resonant circuit

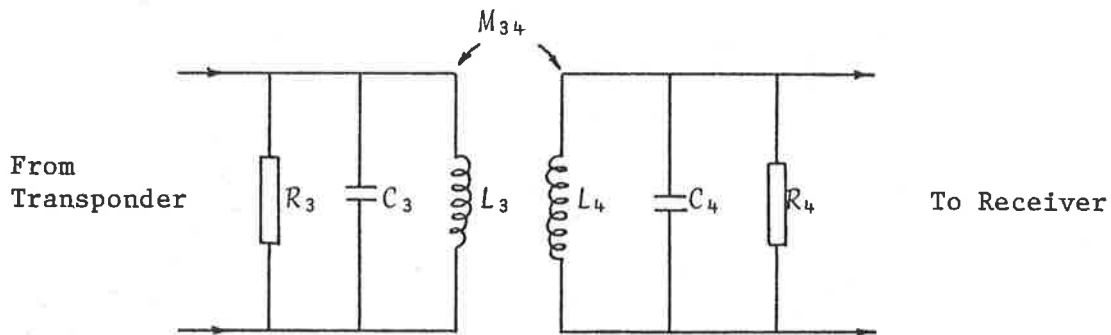


Figure 3.3 Electromagnetic Coupling between Transponder and Receiver

The damping resistors R_3 and R_4 are used to provide sufficient bandwidth for the desired information transfer rate. The coefficient of coupling is invariably small in the situations described above. The power transfer relationship under these circumstances can easily be shown to be:

$$P_2 / P_1 = k_1^2 Q_1 Q_2 \quad (3.5)$$

$$P_4 / P_3 = k_2^2 Q_3 Q_4$$

where P_1 is the power dissipated in the transmitter damping resistor

P_2 is the power delivered to the receptor damping resistor

P_3 is the power dissipated in the transponder damping resistor

P_4 is the power delivered to the receiver damping resistor

These relationships show the reciprocity in the roles of two antennae in a pair. If η is the fraction of the energy received by the receptor coil which is available to excite the transponder coil, the overall power transfer ratio is given by the following relationship.

$$P_4 / P_1 = \eta k_1^2 k_2^2 Q_1 Q_2 Q_3 Q_4 \quad (3.6)$$

In seeking to maximize the received power, the quality factors Q_1 and Q_2 of the transmitter and receptor coils can be made as high as conveniently practicable, the relevant constraints being the intrinsic losses in inductors of convenient size, and the undesirability of employing extremely high Q factors in coils which are subject to detuning by changing environments. The quality factors Q_3 and Q_4 of the transponder and receiver antennae are limited by the same factors as just mentioned, as well as by constraints imposed by the communication bandwidth requirements to accommodate the bit rate present in the reply code. In practice, it is not difficult to design a transponder control circuit which operates on very low power levels so that the majority of the energy received by the receptor antenna is available for exciting the transponder antenna.

The Equation (3.6) illustrates the importance of high quality factor and coefficient of coupling. The overall power transfer ratio may be calculated by evaluating the coefficient of coupling and the design values of quality factors.

3.2.1 Transmitter Antenna Considerations

The requirements of the transmitter antenna include high quality factor, appropriate impedance level for safe operation, and size to provide sufficient interrogation range. The constraints in achieving high quality factor Q of the transmitter antenna arise mainly from the detuning effects due to the antenna surroundings. Several experiments were conducted to investigate these effects. Although these measurements are more relevant to one particular application, namely, vehicle identification, similar effects of different magnitude may be noticed in other applications. Model A transmitter antenna, which was used in the prototype identification system, is shown in Figure 3.4. The variations in inductance and quality factor of two models of the transmitter antenna, when placed in the vicinity of a conducting plane, are plotted in Figure 3.5.

Practical values of quality factor Q_1 of the transmitter antenna under these constraints seem to be of the order 30. If the proportions of the transmitting antenna are assumed to be fixed, then the problem of maximizing the energy density at the point of sensing* is one of maximizing the transmitter power.

*The requirement of small transponders in most object identi-

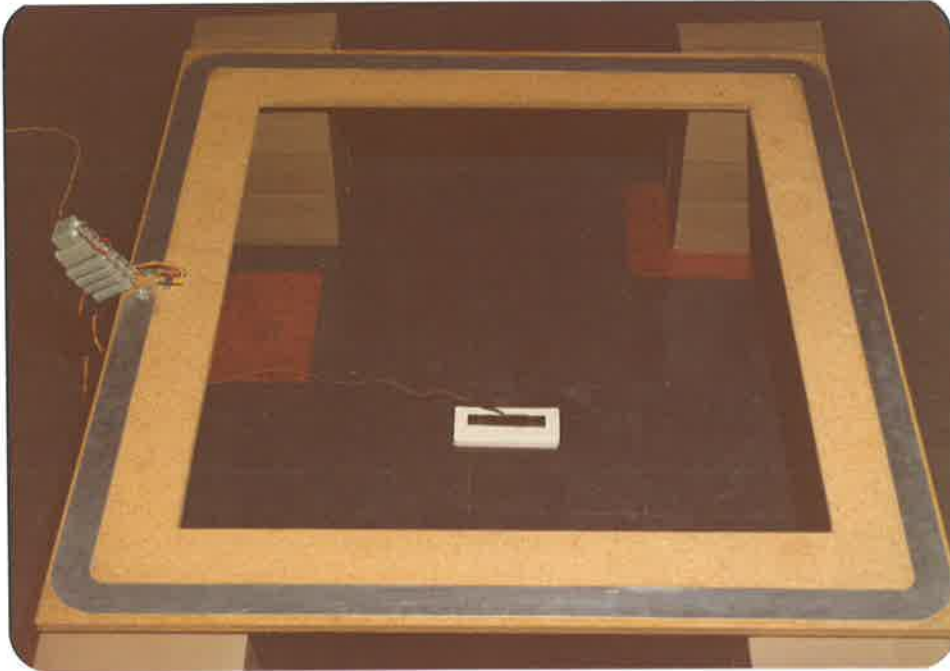
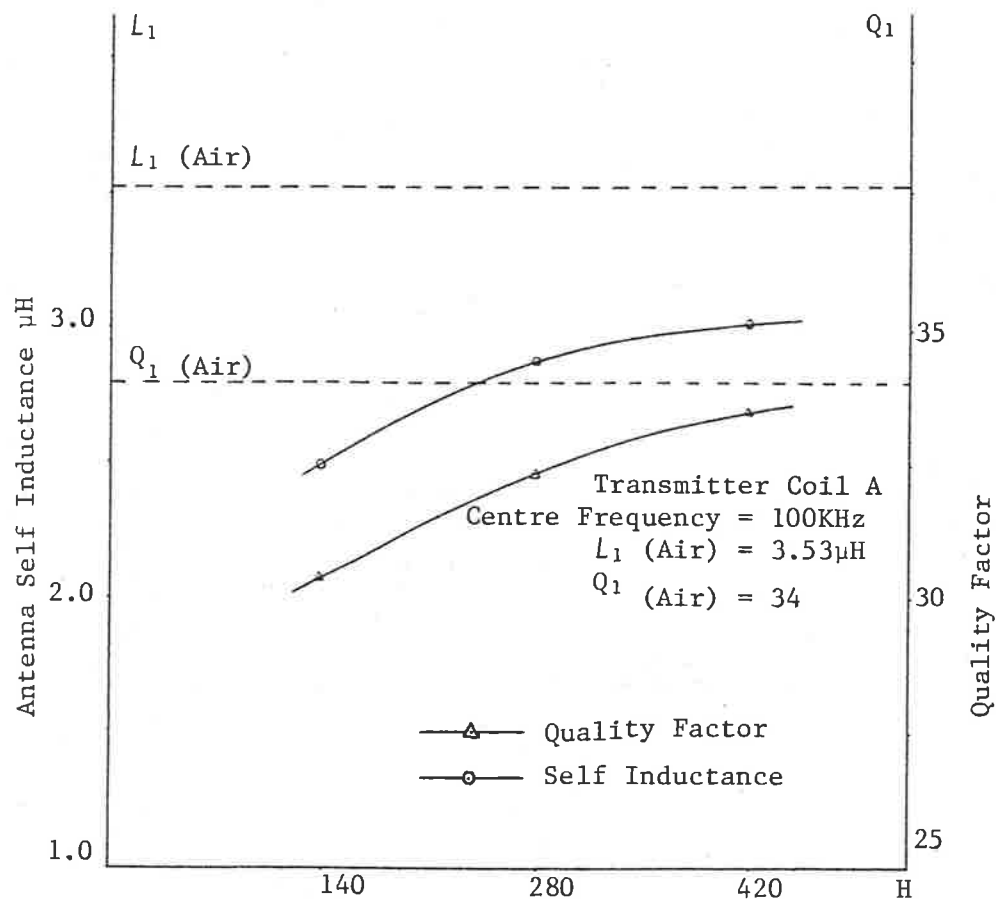


Figure 3.4 Model A Transmitter Antenna for
the Prototype Identification
System

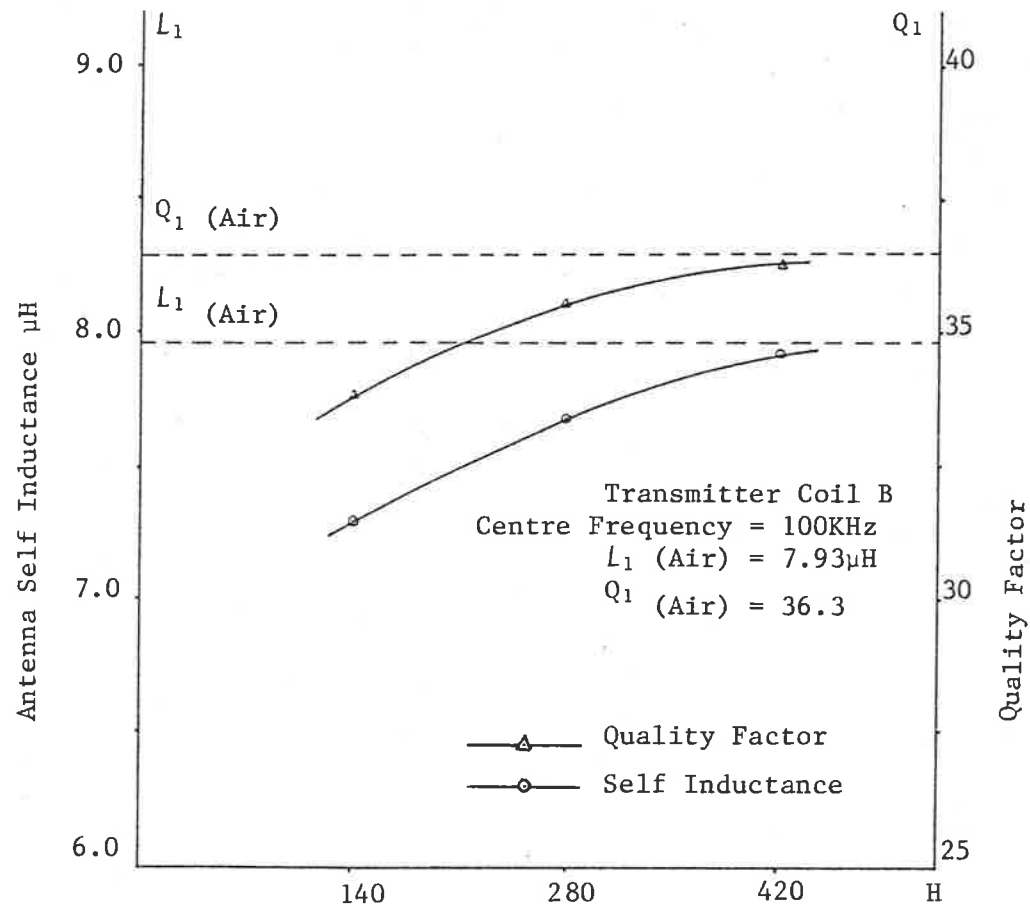
The antenna dimensions were:

External 1200 × 1300mm

Internal 1000 × 1200mm



Antenna Separation from Conducting Plane mm
 Figure 3.5a Variations in Self-Inductance and Quality Factor of Transmitter Antenna Model A



Antenna Separation from Conducting Plane mm
 Figure 3.5b Variations in Self-Inductance and Quality Factor of Transmitter Antenna B

In practice, the allowable values of transmitter power are limited by several factors. These include the power level which can be conveniently provided, biological hazards which may arise from the creation of intense magnetic fields, statutory limitations on radiated or near fields produced by antennae, and considerations of electromagnetic compatibility where other electronic instruments are installed within the field of interrogation apparatus.

The most easily quantifiable among the above constraints are the statutory limitations imposed by licensing authorities. An important constraint is that imposed by the United States Federal Communications Commission rules and regulations (FCC, 1959; Head, 1982). One such regulation requires that the field created at a standard distance of 305m (1000ft) from the source by the low-frequency apparatus operating without a license should be less

fication applications results in small receptor and transponder antennae. A large interrogation range, on the other hand, requires large transmitter or receiver antennae. Therefore, the field created by the transmitter is uniform over the receptor antenna area. Hence, for all practical purposes, the transponder may be considered a point in space.

than the maximum value E_M given by the equation below.

$$E_M = \frac{2.4}{f} \text{ V/m} \quad (3.7)$$

where f is the frequency of transmitter in Hz.

It is sometimes convenient from the point of view of analysis to convert this field to an energy density. In this case, if one takes as an objective the selection of antenna size which maximizes the magnetic field energy density at the transponder position, subject to the radiated field constraints, but without regard for the transmitter power required, we find that the transmitter antenna dimensions become vanishingly small. Such a small transmitter antenna cannot be used in practice, because the power required to provide sufficient interrogation energy density at the sensing position then tends to increase without bound. As a result of these considerations, we conclude that an optimum situation is reached when the transmitter power is made as large as economic constraints allow, and then the transmitter antenna is made small enough to meet the radiation constraint expressed above.

The comparison of electric field produced by the transmitter at the standard distance with the

maximum allowable field strength at the same point, although suitable for assessment of the radiation characteristics of the proposed system, does not lend itself readily to the purpose of the various optimizations involved in the design of a transmitter antenna suitable for the passive transponder application. For such applications, a direct comparison of magnetic field energy density at the point of sensing with the electric field energy density at the standard distance is more fruitful, in that it usually reduces to mathematically tractable functions of the transmitter antenna dimensions, sensing distance, and antenna shape. These calculations are performed in Appendix A. As is evident from these calculations, it is convenient to express the electrical performance of each antenna in terms of (a) a characteristic length indicating the overall size; (b) number of turns; and (c) series of dimensionless factors whose magnitude depends on the particular shape of the antenna.

The two basic electrical parameters of an antenna which are determined by its size are the flux collecting area A and the self-inductance L , which are expressed by the following relationships.

$$L = \mu_0 N^2 FS \quad (3.8)$$

$$A = NGS^2 \quad (3.9)$$

where μ_0 is the permeability of free space
 N is the number of turns of the coil
 F is the antenna size parameter
 S is the characteristic length of the
antenna
 G is the antenna shape parameter

To calculate the value of shape parameter G for different shapes, we define a square coil to be the reference with $G = 1$. Some of the useful antenna shapes for object identification systems are shown in Figure 3.6. It may be seen from the definition of the area in Equation (3.9) that simplification is achieved if the shape parameter G for a circular antenna is defined to be $\pi/4$. It then becomes evident from Equation (3.9) that the characteristic length for the circular antenna is its diameter. A similar argument, for a rectangular antenna, leads to the definition of shape parameter to be unity, and characteristic length given by

$$S = (\delta_1 \delta_2)^{1/2} \quad (3.10)$$

where δ_1 and δ_2 are the dimensions of the antenna along two orthogonal axes, as shown in Figure 3.6.

The characteristic length, due to its definition, does not provide any information about the

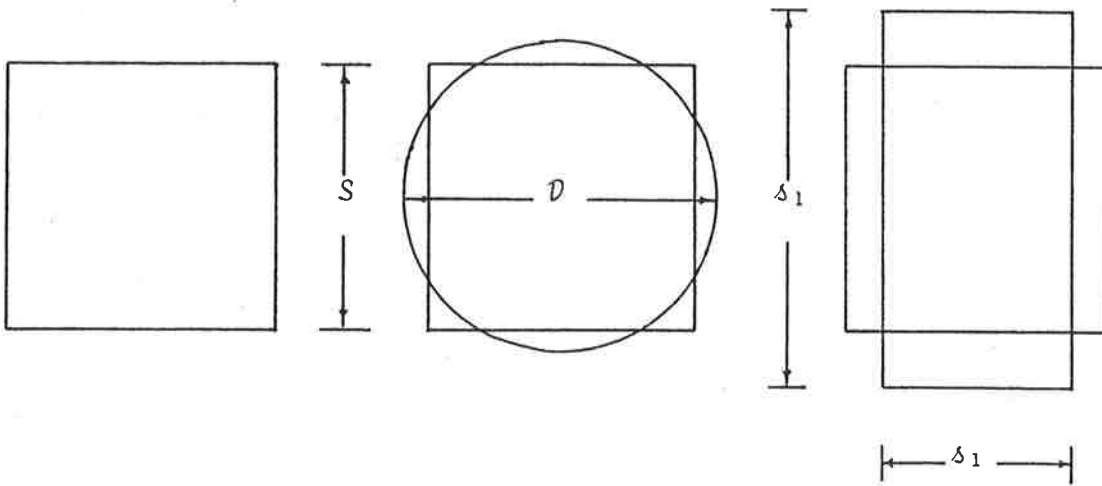


Figure 3.6 Useful Antenna Shapes

aspect ratio of the different antenna shapes. This problem is overcome by defining two additional parameters characterizing various antenna dimensions and shapes. The perimeter p is defined as the total circumference of a planar antenna structure, and asymmetry ratio λ is defined as a function of the aspect ratio of the antenna dimensions as follows:

$$\lambda = \frac{1}{2} \left(\frac{\delta_1}{\delta_2} + \frac{\delta_2}{\delta_1} + 2 \right)^{1/2} \quad (3.11)$$

$$p = 4\lambda GS \quad (3.12)$$

From this definition the asymmetry ratio for the circular and square antenna geometries is calculated to be unity.

The self-inductance and size parameters for some of the popular antenna geometries are shown below.

i) The self-inductance of an antenna formed by a circular coil of diameter \mathcal{D} comprising N turns of wire forming a circular bundle of diameter d , in the skin current situation, is given by the following relationship (Terman, 1947).

$$L = \frac{\mu_0 N^2 \mathcal{D}}{2} \left[\log_e \left(\frac{8\mathcal{D}}{d} \right) - 2 \right] \quad (3.13)$$

A comparison of this relationship with Equation (3.8) yields

$$F = \frac{1}{2} \left[\log_e \left(\frac{8\mathcal{D}}{d} \right) - 2 \right] \quad (3.14)$$

With the value of size factor as given by Equation (3.14), the self-inductance of the coil can be expressed by the simpler relationship described by Equation (3.8).

ii) The self-inductance of a square coil of side S comprising N turns of wire forming a circular bundle of diameter d is given by

$$L = \frac{2\mu_0 N^2 S}{\pi} \left[\log_e \left(\frac{8S}{d} \right) + \left(\frac{d}{2S} \right) - 2.16 \right] \quad (3.15)$$

From this, the size parameter can be found

$$F = \frac{2}{\pi} \left[\log_e \left(\frac{8S}{d} \right) + \left(\frac{d}{2S} \right) - 2.16 \right] \quad (3.16)$$

iii) Another structure of interest is a square of rectangular wire. Self-inductance of this coil of side S and consisting of N turns of wire forming a rectangular bundle of cross-sectional dimensions b and c is given by

$$L = \frac{2\mu_0 N^2 S}{\pi} \left[\log_e \left(\frac{8S}{b+c} \right) + \left(\frac{b+c}{4.5S} \right) - 1.354 \right] \quad (3.17)$$

The size parameter is found to be

$$F = \frac{2}{\pi} \left[\log_e \left(\frac{8S}{b+c} \right) + \left(\frac{b+c}{4.5S} \right) - 1.354 \right] \quad (3.18)$$

The approximations involved in the derivation of the relationships given for self-inductance and the size parameter are valid for the situation when the ratio of antenna characteristic length to wire diameter is very large. A comparison of the expressions of size parameter given by Equations (3.14), (3.16) and (3.18) reveals that they are of similar form.

The basic equations relating the real power P_x , energy stored U , quality factor Q , circulating reactive power P_y , resonant frequency ω_0 , and rms voltage E across the inductance L of a resonant circuit are given by the following relations.

$$P_x = \omega_0 U \quad (3.19)$$

$$Q = P_y / P_x \quad (3.20)$$

$$E = \sqrt{\omega_0 L P_y} \quad (3.21)$$

When such a coil has induced in it a voltage e by near-field coupling, the voltage at resonance across the inductance for a parallel tuned situation is given by

$$E = e \cdot Q \quad (3.22)$$

It should be noted here that the amount of reactive power present as a result of such a series induced voltage is magnified by a factor Q^2 over the value which results when the coil is short circuited instead of resonated.


3.2.2 Receptor and Transponder Antenna Considerations

The principle requirement of the receptor antenna is to couple sufficient amount of interroga-

tion energy to the transponder for subsequent retransmission. It will be shown in a later section that the transfer of power to the receptor in a near-field coupling situation depends mainly upon its physical volume and the quality factor which may be achieved when it is resonated, and is not affected, except for second order effects, by the number of turns with which the coil is wound; this number merely serves to establish the impedance level at which the power is obtained. This does not imply that the impedance level of the receptor coil is not an important design parameter. As the coupling between near-field antennae occurs through their stored energy rather than dissipated or radiated power, the question of optimum impedance level of the tag antennae in the two-port* transponder system is therefore one of maximizing these stored energies rather than matching of power between a source and a radiation impedance.

Another important factor in the design of the receptor antenna, as described in the previous chapter, is the fact that practical transponder circuits do not operate as linear power converters, particularly at low power levels, but rather show a thre-

*The considerations for design of tag antennae in one-port systems will be described separately in a later section.



shold characteristic, below which the operation of the circuit fails. The problem of deriving sufficient voltage from the receptor coil for satisfactory operation of the circuit is another major consideration.

In order to maximize the available voltage, the receptor coil is always parallel tuned, and the voltage available to the transponder circuit is therefore greater than the induced voltage by the quality factor of the receptor resonant circuit. The problem, in practice, is, therefore, that of employing sufficient numbers of turns in the receptor coil to obtain the necessary voltage at the transmitter power level used, the transmitter power itself being selected on the basis of propagation loss in the complete reply path between the transmitter and receiver, transponder conversion loss, and extraneous noise coupled into the receiver from the interrogation environment.

The requirement of the transponder antenna is to optimally couple the reply energy to the receiver under the constraints of size, and quality factor, to accommodate the information transfer rate. The design criteria for the transponder antenna are very similar to the ones described for the receptor antenna. As described in the previous chapter, the quality factor required for the transmission of the

information bearing signal is provided by the output impedance of the transponder electronic circuitry. One of the problems faced in the design of the two-port transponder system is that, in a large number of applications, the need to minimize the transponder dimensions results in competition for space between the receptor and transponder antennae. An optimal solution to this problem may be found by considering the two tag antennae together with the problem of minimizing the mutual coupling between them.

3.2.3 Receiver Antenna Considerations

The considerations governing the design of the receiver antenna include quality factor, noise shielding, and signal-to-noise ratio. In many object identification systems, the reply signal must compete against environmental noise coupled to the receiver by means of the receiver antenna, while the internal noise of the receiver is of no major consequence. The optimum design of the receiver antenna, under these circumstances, results from maximizing the signal-to-noise ratio. The problem of noise pick-up by the receiver antenna from electric fields in the vicinity of the interrogation system may be solved by providing adequate electrical shielding to the magnetic dipole used for the reception of reply signals.

3.3 Electromagnetic Coupling

In low-frequency communication systems of the type considered in this thesis, the information exchange takes place by means of the near-field coupling between the antennae involved in the communication process. The coupling achieved is weak in most applications. The electromagnetic propagation loss, in a situation like this, is determined by the coefficient of coupling between the two antennae and their quality factors, as described by Equation (3.5). The coefficient of coupling between the transmitter and receptor antenna pair is described by Equation (3.1). The coefficient of coupling between the antenna pair in the reply link is described by a similar equation. In order to obtain mathematically tractable results, this study is restricted to simple antenna geometries. The analyses are further simplified by performing the initial calculations on mathematically convenient shapes. The analyses for other shapes are performed by introducing the appropriate size and shape factors in the final results.

Although the Equation (3.6) illustrates the importance of high quality factor and coefficient of coupling, it does not provide an appropriate insight into the influence of antenna geometry in establishing the level of coupling achieved. For this

reason, the concepts of coupling volume and dispersal volume are developed. These concepts enable the separate influences of the two sets of antenna parameters on the coupling mechanism to be highlighted, and appropriate local optimizations to be performed. (Cole and Roy, 1977).

3.4 Coupling Volume Approach

The observation that in low-frequency object identification systems, the difference in sizes of the antennae involved, the loose coupling situation for which the power transfer relationships given by Equation (3.5) hold, and the reciprocity of the power transfer relation, have important consequences in the design of the antennae. One of the consequences is that various physical quantities used in the calculation of power transfer between the antennae may be divided into suitable groups, so that the effects of these quantities on the power transfer relationship can be studied separately.

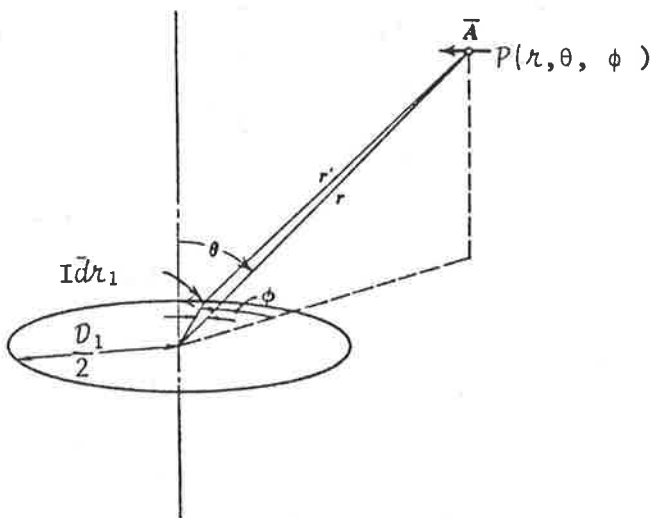
For example, transmitter antenna dimensions, sensing distances, and receiver antenna dimensions may be placed in one group, while transponder and receptor antenna dimensions may be placed in another group. Functionally, this type of grouping may, at

first, appear incompatible. By invoking the reciprocity theorem, as it will be shown later, the grouping is quite logical.

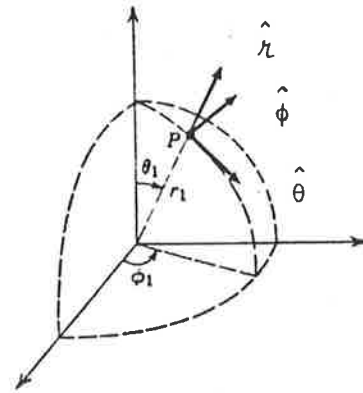
The first step in the process of separating the two sets of dimensions is achieved by recognizing that in a weak coupling situation the basic excitation process for the receptor antenna is the magnetic field created by the transmitter at the position of the transponder. The magnetic field under these circumstances is determined by the transmitter antenna parameters, excitation level and the sensing distance alone, and is not influenced by the small power level established in the receptor coil. Furthermore, in most applications it is valid to assume that the magnetic field is uniform over the comparatively small volume of the identification label.

Let the rms current in the transmitter loop antenna shown in Figure 3.7 be I_1 . The magnetic field H_1 created by this current at the tag position may be derived by first calculating the vector potential.

The vector potential \bar{A}_1 at the sensing point P due to a current I_1 , in the transmitter loop has the form:



a. Circular Loop



b. Co-ordinate system

r = distance of the test location from origin

r' = distance between the current element and test point

Figure 3.7 Magnetic Field at a Sensing Position Due to a Circular Loop

$$\bar{A}_1(\bar{r}) = \frac{\mu_0 I_1}{4\pi} \oint_c \frac{d\bar{r}_1}{r'} \quad (3.23)$$

where

$$r' = \frac{1}{4}(4r^2 + D_1^2 - 4rD_1 \sin\theta \cdot \cos\phi)$$

and

$$d\bar{r}_1 = \frac{1}{2} [(-D_1 \sin\phi d\phi) \hat{r} + (D_1 \cos\phi d\phi) \hat{\phi}]$$

where \hat{r} and $\hat{\phi}$ are the unit vectors in the radial and azimuthal directions at the sensing position. Since r' is an even function of ϕ , the radial component of \bar{A}_1 integrates to zero and we are left with

$$\bar{A}_1(\bar{r}) = \frac{\mu_0 I_1}{4\pi} \hat{\phi} \int_0^{2\pi} \frac{D_1 \cos\phi d\phi}{(4r^2 + D_1^2 - 4rD_1 \sin\theta \cdot \cos\phi)^{1/2}} \quad (3.24)$$

Although it is possible to accurately solve the integral in the above equation by numerical means, for most practical situations some simple assumptions lead to mathematically tractable solutions without much loss of generality. In most object identification applications, the relative position of an identification label with respect to the transmitter antenna is expected to be close to the vertical axis passing through the centre of the transmitter antenna. Under these circumstances, $1/r'$ may be approximated to:

$$\frac{1}{r'} \approx \frac{2}{(4r^2 + D_1^2)^{1/2}} + \frac{4rD_1 \sin\theta \cdot \cos\phi}{(4r^2 + D_1^2)^{3/2}} + \dots$$

Substituting this value into Equation (3.22) we obtain for the vector potential

$$\bar{A}_1(\bar{r}) = \mu_0 I_1 \frac{D_1 r \sin\theta}{(4r^2 + D_1^2)^{3/2}} \hat{\phi} \quad (3.25)$$

The magnetic field H_1 may then be derived by using the following relationship:

$$\int_{\mathbf{s}} \bar{H}(\bar{r}) \cdot d\bar{S} = \frac{1}{\mu_0} \oint_c \bar{A}(\bar{r}) \cdot d\bar{r} \quad (3.26)$$

For the situation when the sensing point is reasonably close to the Z-axis, the above integral is evaluated over a circle centred on the Z-axis.

$$\therefore \bar{H}_1(Z) \approx \frac{I_1 D_1^2}{(4Z^2 + D_1^2)^{3/2}} \hat{Z} \quad (3.27)$$

For the near-field situation, the energy density at the tag position is given by

$$u = \mu_0 H_1^2 \quad (3.28)$$

The energy density given by Equation (3.28) may be evaluated by substituting the value of H_1 from Equation (3.27).

$$u = \mu_0 \frac{I_1^2 D_1^4}{(4Z^2 + D_1^2)^3} \quad (3.29)$$

The real and reactive powers established in the receptor antenna due to this magnetic field are proportional to this energy density and the quality factor of the resonance, but to separate the effects of receptor antenna dimensions and resonance quality factors the coupling volume V_c of the receptor antenna is defined as the ratio

$$V_c = \frac{\left[\begin{array}{l} \text{peak value of stored magnetic energy} \\ \text{in receptor self-inductance when it} \\ \text{is short-circuited} \end{array} \right]}{\left[\begin{array}{l} \text{peak value of the magnetic energy} \\ \text{density produced at the tag position} \\ \text{by the transmitter} \end{array} \right]} \quad (3.30)$$

It may be shown, by simple mathematical manipulations, that for a single turn plane coil, the coupling volume is given by

$$V_c = \frac{\mu_0 A^2}{L} \quad (3.31)$$

where

A is the flux collecting area of the antenna

L is the self-inductance of the antenna

It may be noticed that the coupling volume has units of volume. The power to the receptor loading resistance P_2 , may be calculated by using the fact that the power at resonance is $\omega_1 Q_2$ times the peak value of stored magnetic energy in the receptor coil self-inductance when it is short-circuited. That is,

$$P_2 = \omega_1 Q_2 \left[\begin{array}{l} \text{peak value of stored magnetic} \\ \text{energy in the tag self-inductance} \\ \text{when short-circuited} \end{array} \right] \quad (3.32)$$

where ω_1 is the resonant frequency of the transmitter and receptor antennae

Combining Equations (3.30) and (3.32), we obtain

$$P_2 = \omega_1 Q_2 V_{c_1} \left[\begin{array}{l} \text{peak value of energy density} \\ \text{produced at the transponder} \\ \text{position by the transmitter} \end{array} \right] \quad (3.33)$$

where V_{c_1} is the coupling volume of the receptor coil or

$$P_2 = (\omega_1 Q_2 u) \cdot V_{c_1} \quad (3.34)$$

This relation clearly illustrates the importance of the coupling volume. The power to the receptor loading resistor may be maximized by maximizing the coupling volume V_{c_1} of the receptor coil. In fact, the coupling volume serves as a figure of merit for various antennae, all loaded to give the

same quality factor. Furthermore, coupling volume is a function of the antenna parameters only, and has dimensions of volume. The optimization can therefore be performed on the receptor antenna regardless of the transmitter antenna properties.

The development of a relationship between the energy density at the transponder position and the transmitter power is based on the observation that for a given transmitter geometry, this density is proportional to the total stored energy of the antenna. This fact is utilized in the definition of the dispersal volume V_d for the transmitter antenna.

$$V_d = \frac{\left[\begin{array}{l} \text{peak value of the total stored} \\ \text{magnetic energy in the transmitter} \\ \text{antenna} \end{array} \right]}{\left[\begin{array}{l} \text{peak value of the magnetic energy} \\ \text{density produced at the transponder} \\ \text{position by the transmitter} \end{array} \right]} \quad (3.35)$$

It may be observed that dispersal volume has dimensions of volume. Because of the complex nature of the transmitter antenna field structure, it is not possible to derive a simple expression for the dispersal volume analogous to Equation (3.31). However, a relationship analogous to Equation (3.32) may be derived.

$$P_1 = \frac{\omega_1}{Q_1} \left[\begin{array}{l} \text{peak value of stored magnetic} \\ \text{energy in the transmitter} \\ \text{self-inductance} \end{array} \right] \quad (3.36)$$

$$\text{or } P_1 = \frac{\omega_1}{Q_1} V_{d_1} \left[\begin{array}{l} \text{peak value of magnetic} \\ \text{energy density produced} \\ \text{at the transponder} \\ \text{position by the} \\ \text{transmitter} \end{array} \right] \quad (3.37)$$

$$\text{or } P_1 = \left(\frac{\omega_1 u}{Q_1} \right) \cdot V_{d_1} \quad (3.38)$$

The power transfer relationship between the transmitter and the receptor antennae may be derived by combining Equations (3.34) and (3.38):

$$\frac{P_2}{P_1} = \left(\frac{V_{c_1}}{V_{d_1}} \right) \cdot Q_1 Q_2 \quad (3.39)$$

The power transfer ratio is therefore proportional to the ratio of the receptor antenna coupling volume to the transmitter antenna dispersal volume. This result shows that to improve the coupling a larger coupling volume in the receptor antenna and a smaller dispersal volume in the transmitter antenna is required. The transmitter antenna behaves, in so far as the transponder is concerned, as if the transmitter antenna stored energy were uniformly distributed over a volume equal to the transmitter dispersal volume. A comparison between the coupling and dispersal volumes shows that, while the former depends only upon the receptor antenna characteristics, the dispersal volume depends, not only upon

the transmitter antenna, but also on the position of the transponder within the field of the transmitter antenna. It is this characteristic of the coupling and dispersal volume which enables the local optimization on individual antenna parameters to be performed.

By using the reciprocity theorem, and an analysis similar to the one used in the derivation of Equation (3.39), an analogous result for the transponder-receiver coupling link is derived.

$$\frac{P_4}{P_3} = Q_3 Q_4 \left(\frac{V_{c_2}}{V_{d_2}} \right) \quad (3.40)$$

where

V_{c_2} is the coupling volume of the transponder antenna

V_{d_2} is the dispersal volume of the receiver antenna

A comparison of Equations (3.39) and (3.40) with Equation (3.5) gives

$$k_1^2 = \frac{V_{c_1}}{V_{d_1}} \quad (3.41)$$

$$k_2^2 = \frac{V_{c_2}}{V_{d_2}}$$

To gain further insight into these concepts, expressions for the coupling and dispersal volumes of some simple antenna structures are developed.

3.4.1 The Coupling Volume

The expressions for self-inductance and the flux collecting area in terms of number of turns and shape and size parameters are given by Equations (3.8) and (3.9). Substituting these expressions into Equation (3.31) we get


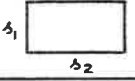

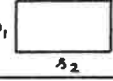




$$V_c = \frac{\mu_0 (NGS^2)^2}{\mu_0 N^2 FS}$$

or

$$V_c = \frac{S^3 G^2}{F} \quad (3.42)$$

The expressions for electrical parameters for some simple antenna geometries are given in Table 3.1. These parameters are then calculated for four geometries. The results of these calculations are given in Table 3.2. It may be seen from Table 3.2 that, for circular, square and rectangular antenna geometries with equal area, the circular shape has the highest coupling volume. In most object identification applications, the receptor antenna is required to be placed in a small space. In these situations, it is possible to employ antennae with

Table 3.1 Electrical Parameters of Various Antenna Geometries

Parameter \ Shape		Circle 	Rectangle 	Square 	Rectangle 	Universal
Wire Bundle Cross-section	d					d
Number of Turns	N	N	N	N	N	N
Shape Parameter	G	$\pi/4$	1	1	1	G
Characteristic Length	S	D	$(a_1 a_2)^{0.5}$	S	$(a_1 a_2)^{0.5}$	S
Assymetry Ratio	λ	1	$0.5 \left(\frac{a_1}{a_2} + \frac{a_2}{a_1} + 2 \right)^{0.5}$	1	$0.5 \left(\frac{a_1}{a_2} + \frac{a_2}{a_1} + 2 \right)^{0.5}$	λ
Perimeter	p	πD	$2(a_1 + a_2)$	4S	$2(a_1 + a_2)$	$4 \lambda G S$
Area	A	$\pi D^2 N/4$	$N a_1 a_2$	$N S^2$	$N a_1 a_2$	$G N S^2$
Size Parameter	F	$0.5 \left[\log_e \left(\frac{8D}{d} \right) - 2 \right]$	$\frac{2\lambda}{\pi} \left[\log_e \left(\frac{8S}{d} \right) - 2.16 \right]$	$\frac{2}{\pi} \left[\log_e \left(\frac{8S}{d} \right) + \frac{d}{2S} - 2.16 \right]$	$\frac{2\lambda}{\pi} \left[\log_e \left(\frac{8S}{b+c} \right) - 2.16 \right]$	$\frac{2\lambda G}{\pi} \left[\log_e \left(\frac{8S}{d} \right) - \theta \right]^*$
Self-Inductance	L	$\frac{\mu_0 N^2 D}{2} \left[\log_e \left(\frac{8D}{d} \right) - 2 \right]$	$\frac{2\lambda \mu_0 N^2 S}{\pi} \left[\log_e \left(\frac{8S}{d} \right) - 2.16 \right]$	$\frac{2\mu_0 N^2 S}{\pi} \left[\log_e \left(\frac{8S}{d} \right) + \frac{d}{2S} - 2.16 \right]$	$\frac{2\lambda \mu_0 N^2 S}{\pi} \left[\log_e \left(\frac{8S}{b+c} \right) - 2.16 \right]$	$\mu_0 F N^2 S$
Coupling Volume	V_c	$\frac{\pi^2 D^3}{8 \left[\log_e \left(\frac{8D}{d} \right) - 2 \right]}$	$\frac{\pi S^3}{2\lambda \left[\log_e \left(\frac{8S}{d} \right) - 2.16 \right]}$	$\frac{\pi S^3}{2 \left[\log_e \left(\frac{8S}{d} \right) + \frac{d}{2S} - 2.16 \right]}$	$\frac{\pi S^3}{2\lambda \left[\log_e \left(\frac{8S}{b+c} \right) - 2.16 \right]}$	$G^2 S^3 / F$
Geometrical Figure of Merit	γ_g	$\frac{\pi^2}{8 \left[\log_e \left(\frac{8D}{d} \right) - 2 \right]}$	$\frac{\pi}{2\lambda \left[\log_e \left(\frac{8S}{d} \right) - 2.16 \right]}$	$\frac{\pi}{2 \left[\log_e \left(\frac{8S}{d} \right) + \frac{d}{2S} - 2.16 \right]}$	$\frac{\pi}{2\lambda \left[\log_e \left(\frac{8S}{b+c} \right) - 2.16 \right]}$	G^2 / F

* θ is a shape dependent factor of the order 2

Table 3.2 Comparative Study of Different Antenna Geometries

PARAMETER	SHAPE	SYMBOL	CIRCLE	CIRCLE	RECTANGLE	SQUARE	UNITS
	→						
WIRE BUNDLE DIAMETER	↓	d	5×10^{-3}	5×10^{-3}	5×10^{-3}	5×10^{-3}	m
NUMBER OF TURNS		N	92	92	92	92	1
SHAPE PARAMETER		G	$\pi/4$	$\pi/4$	1	1	1
CHARACTERISTIC LENGTH		S	87.5×10^{-3}	77.6×10^{-3}	77.6×10^{-3}	77.6×10^{-3}	m
ASYMMETRY RATIO		λ	1	1	1.0185	1	1
PERIMETER		p	0.275	0.244	0.316	0.310	m
AREA		A	0.554	0.435	0.554	0.554	m ²
SIZE PARAMETER		F	1.471	1.411	1.726	1.716	1
SELF-INDUCTANCE		L	1.369×10^{-3}	1.164×10^{-3}	1.424×10^{-3}	1.416×10^{-3}	H
COUPLING VOLUME		V_c	0.281×10^{-3}	0.204×10^{-3}	0.270×10^{-3}	0.272×10^{-3}	m ³
GEOMETRICAL FIGURE OF MERIT		γ_g	0.419	0.437	0.579	0.583	1

different geometries, with equal characteristic length, but not equal area. For this reason, the coupling volume of different shapes with equal characteristic length must be compared. If this criterion is used for the comparison of different shapes, then the figure of merit for the geometry of different antennae, from Equation (3.42), becomes

$$\gamma_g = \frac{G^2}{F} \quad (3.43)$$

where γ_g is the geometrical figure of merit

The comparison of γ_g for the elementary shapes with equal characteristic length is shown in Figure 3.8. It is evident, from the plot, that the circular shape has the lowest figure of merit of all the simple shapes considered. For the antenna sizes of interest for passive transponder applications, an improvement of approximately 33% is achieved by using either a square or a rectangular antenna instead of a circular geometry. In some object identification applications, i.e. vehicle identification application, rectangular geometry utilizes the available space more effectively, although a small loss in the coupling volume as compared with square shapes results. For this reason, the rectangular antenna with the parameters described in Table 3.2 is used as the receptor antenna, called model A

receptor, for the prototype two-port identification system. An identical geometry is used for the transponder. The antennae used in the investigation of optimum shape for vehicle identification application are shown in Figure 3.9.

For the one-port transponder system, where only one antenna is used, the coupling between the transmitter and the receptor may be improved over the planar coils by using ferrite cored solenoids with circular or rectangular cross-section (Eshraghian, 1980). The use of ferrite cored antennae is avoided in two-port transponder systems, due to the requirement of minimum coupling between the two antennae on the transponder. The physical and electrical characteristics of the ferrite cored antenna called the model B receptor antenna is given in Table 3.3. This receptor antenna has been used in the coefficient of coupling experiments discussed later. The theoretical calculation of coefficient of coupling is based on Equation (3.41).

Table 3.3 Ferrite Antenna Characteristics

Type	Ferrite Slab
Antenna Dimensions	121mm × 115mm × 6.5mm
Number of Turns	196
Self-Inductance	6.845 mH
Coupling Volume	$1.23 \times 10^{-4} \text{ m}^3$
Quality Factor	61
Operating Frequency	100 kHz

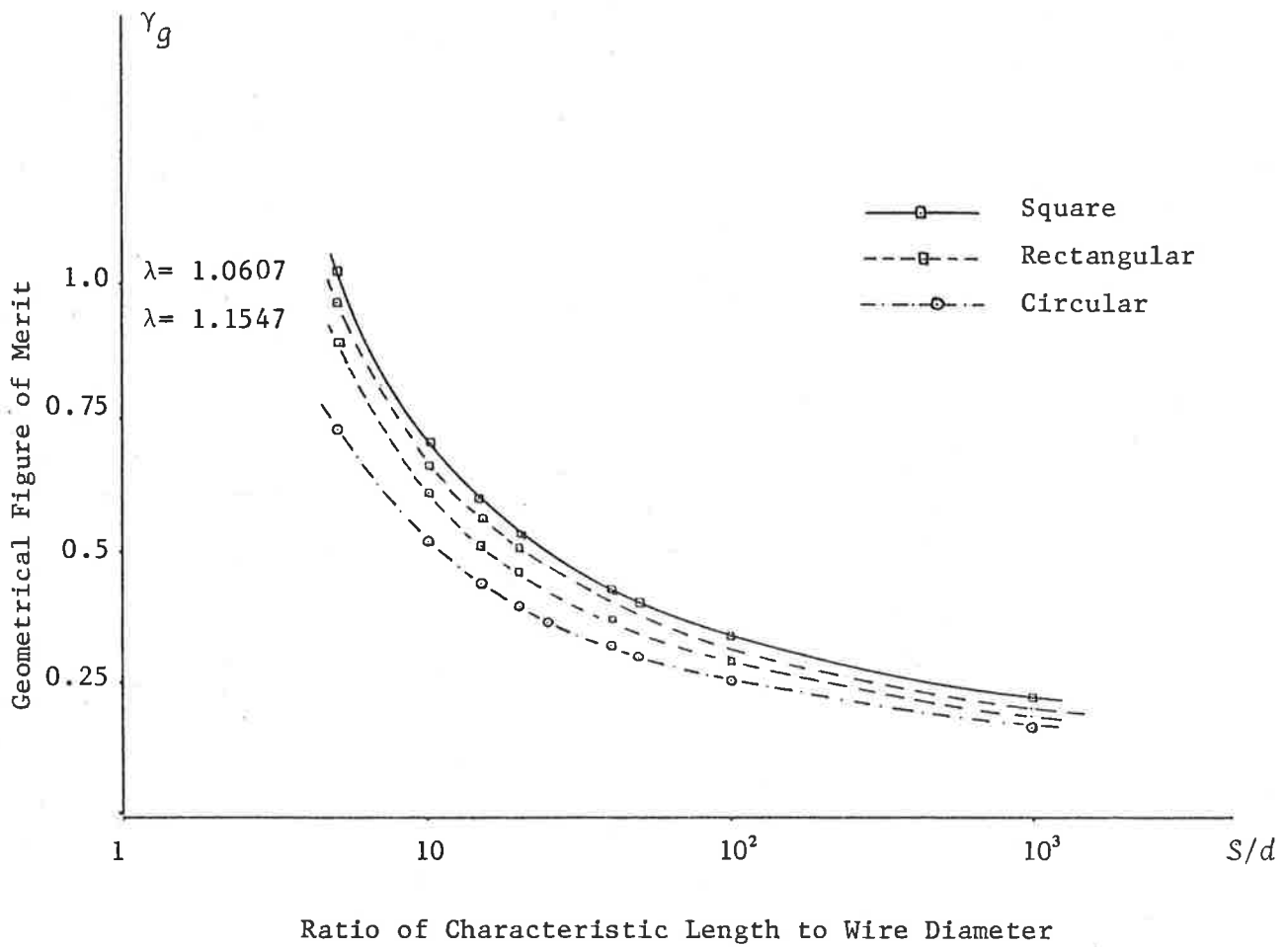


Figure 3.8 Comparison of Geometrical Figure of Merit γ_g for Various Shapes

3.4.2 The Dispersal Volume

To calculate the coefficient of coupling, the dispersal volume of the transmitter antenna must be evaluated. The peak value of the total stored magnetic energy in the transmitter antenna is given by the product of antenna self-inductance and the square of the current, and the peak value of the magnetic energy density produced at the transponder position by the transmitter is given by the product of the square of magnetic field and the permeability of air. The dispersal volume V_d is therefore given by

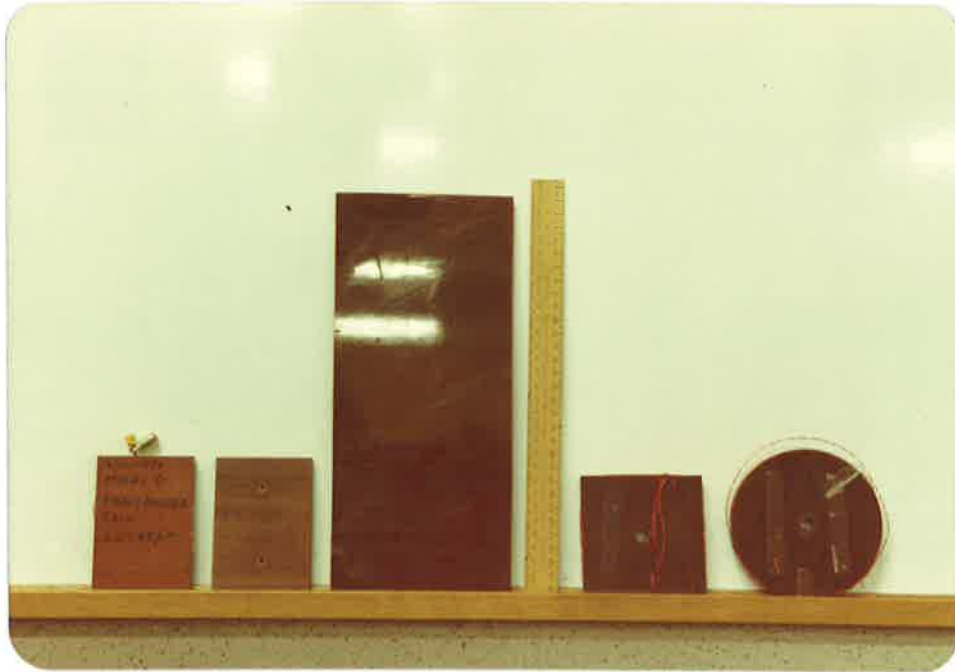


Figure 3.9a Various Antenna Geometries Considered for Vehicle Identification System

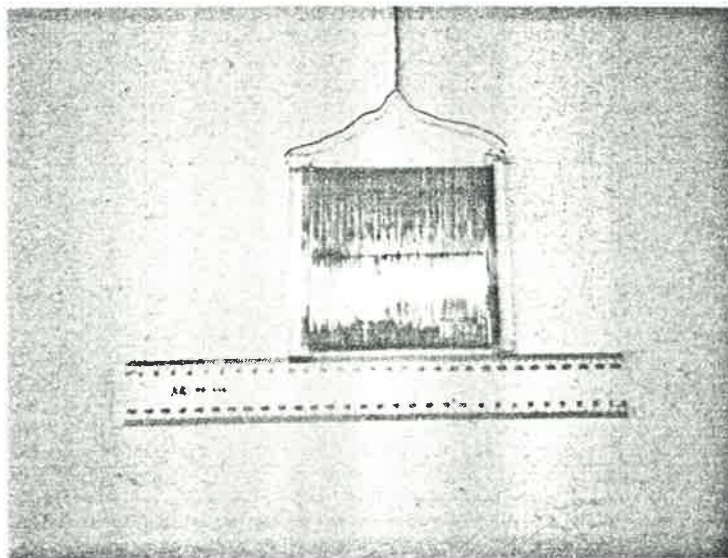


Figure 3.9b Model B Receptor Antenna

$$V_d = \frac{LI^2}{\mu_0 H^2} \quad (3.44)$$

The magnetic field H , created at a sensing position P , distant S_T along the axis from the centre of a single turn circular antenna of diameter D , carrying a current I , is given by Equation (3.27).

$$H = \frac{ID^2}{(D^2 + 4S_T^2)^{3/2}} \quad (3.45)$$

The magnetic field due to a square antenna may be approximated by considering a square antenna of an area equal to that of the circular one, and substituting the characteristic length for diameter. The desired calculations may be performed using this approximation without the need for complex mathematical manipulations. The magnetic field due to a square antenna of characteristic length S , at the sensing point distant S_T along the axis from the centre of the coil is then given by

$$H \approx \sqrt{\frac{\pi}{4}} \cdot \frac{IS^2}{(S^2 + \pi S_T^2)^{3/2}} \quad (3.46)$$

The dispersal volume of the square antenna may then be expressed as

$$V_d = FS^3 \chi \quad (3.47)$$

where

X is a dimensionless factor which depends upon the sensing distance normalized with respect to the characteristic length of the antenna

F is the antenna size parameter as defined in Table 3.1

$$X = \frac{4}{\pi} \left[1 + \pi \left(\frac{S_T}{S} \right)^2 \right]^3 \quad (3.48)$$

In most object identification situations, square or rectangular antennae for the transmitter and receiver are preferred over the circular antenna. The preference is mainly for practical reasons, such as ease of installation and placement of antennae without the need to reshape or re-organize the interrogation environment in a major way. For example, in a personnel identification situation, a rectangular transmitter antenna is most suitable, as it may be conveniently placed around a door.

The structure of the expression for V_d is of interest, because it allows the local optimization to be performed. For a given characteristic length of the transmitter antenna, the minimum value of dispersal volume is found when the distance degradation factor X is $4/\pi$. This is achieved when

$$\left[1 + \pi \left(\frac{S_T}{S_1} \right)^2 \right]^3 = 1 \quad (3.49)$$

It can be readily seen that this is achieved when the sensing position is chosen to be at the centre of the transmitter antenna, i.e. S_T equals zero. The extent to which the dispersal volume increases, thereby resulting in a decrease in coupling between the transmitter and receptor due to increase in sensing distance, is proportional to the distance degradation parameter λ , and is plotted for normalized value of sensing distances in Figure 3.10. It is evident from this plot that the optimum position of the transmitter antenna in relation to the motion of the transponder is when the transmitter antenna encircles the transponder rather than lies parallel to the motion of the transponder.

The electrical parameters of the transmitter, receptor, transponder, and the receiver antenna are given in Table 3.4.

3.4.3 The Power Coupling Ratio

In order to investigate the effect of various environmental conditions on the coupling achieved between an antenna pair, the power coupling ratio κ of an antenna pair in a given propagation link is defined as the received power in watts when the transmitted power is one watt. For the transmitter-

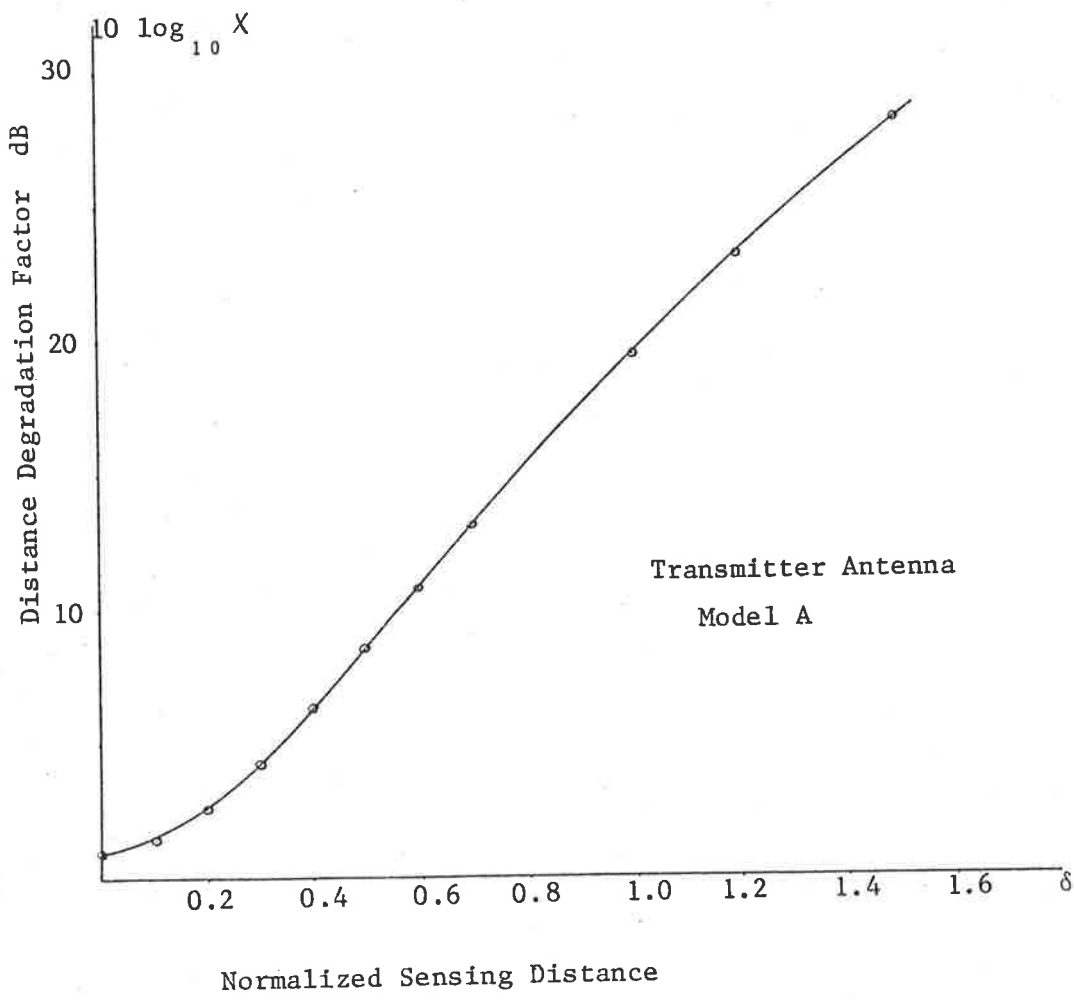


Figure 3.10 Increase in Dispersal Volume with Distance

Table 3.4 Antenna Parameters for a Two-port Transponder System

PARAMETER	ANTENNA	TRANSMITTER MODEL A	RECEPTOR MODEL A	TRANSPONDER MODEL A	RECEIVER MODEL A	UNITS
SHAPE		APPROX. SQUARE	RECTANGLE	RECTANGLE	SQUARE	
NUMBER OF TURNS	N	1	92	46	1	1
SHAPE PARAMETER	G	1	1	1	1	1
CHARACTERISTIC LENGTH	S	1.25	77.6×10^{-3}	77.6×10^{-3}	1.1	m
ASYMMETRY RATIO	λ	1	1.0185	1.0185	1	1
SIZE PARAMETER	F	2.24	1.726	1.726	2.57	1
SELF-INDUCTANCE	L	3.53×10^{-6}	1.42×10^{-3}	0.358×10^{-3}	3.57×10^{-6}	H
GEOMETRICAL FIGURE OF MERIT	γ_g^*	-	0.579	0.579	-	1
COUPLING VOLUME	V_c	-	0.27×10^{-3}	0.27×10^{-3}	-	m ³
DISTANCE DEGRADATION FACTOR	χ^\dagger	$4/\pi$	-	-	$4/\pi$	1
DISPERSAL VOLUME	V_d	5.57	-	-	4.36	m ³
QUALITY FACTOR	Q	34	37	8	8	
OPERATING FREQUENCY	f	100	100	50	50	kHz

* The parameters relevant to the particular antennae only are calculated

† Calculated at the centre of the antenna

receptor link, the power coupling ratio is expressed as

$$\kappa_1 = \left(\frac{V_{c_1}}{V_{c'_1}} \right) Q_1 Q_2 \quad (3.50)$$

Substituting the expressions of coupling and dispersal volume derived earlier, we get

$$\kappa_1 = \frac{G_2^2 S_2^3 / F_2}{X_1 F_1 S_1^3} \cdot Q_1 Q_2$$

where

the subscript 2 refers to the receptor
antenna parameters

the subscript 1 refers to the transmitter
antenna parameters

As discussed in Section 3.2, the most useful receptor antenna geometry for object identification applications is a rectangle for which the shape parameter G_2 is unity. The power coupling ratio then becomes

$$\kappa_1 = \frac{4}{\pi F_1 F_2 \left[1 + \pi \left(\frac{S_T}{S_1} \right)^2 \right]^3} \cdot \left(\frac{S_2}{S_1} \right)^3 \cdot Q_1 Q_2 \quad (3.51)$$

or

$$\kappa_1 = \frac{4}{\pi F_1 F_2 (1 + \pi \delta^2)^3} \left(\frac{S_2}{S_1} \right)^3 \cdot Q_1 Q_2 \quad (3.52)$$

where δ is the sensing distance normalized with respect to the transmitter antenna characteristic length S_t .

A family of plots of power coupling ratio κ_1 is shown in Figure 3.11 for various sizes of receptor and transmitter antennae, the values of size parameter given in Table 3.4 being used for these curves. The power coupling ratio is also plotted for the case when the model B receptor antenna is used in conjunction with model A transmitter antenna. The relative position of the two antennae is shown in Figure 3.12, and the power coupling ratio is plotted in Figure 3.13.

One of the major problems encountered in many object identification applications is the effect of the interrogation environment on the coupling process. For instance, in personnel identification applications, when the interrogation takes place inside a building, the cumulative effect of steel structures and other metal objects in the vicinity of the interrogation apparatus on the coupling mechanism must be carefully considered. Theoretical evaluation of the resultant electromagnetic coupling under these circumstances invariably leads to cumbersome mathematical manipulations. The approximations required to make the calculations simple of-

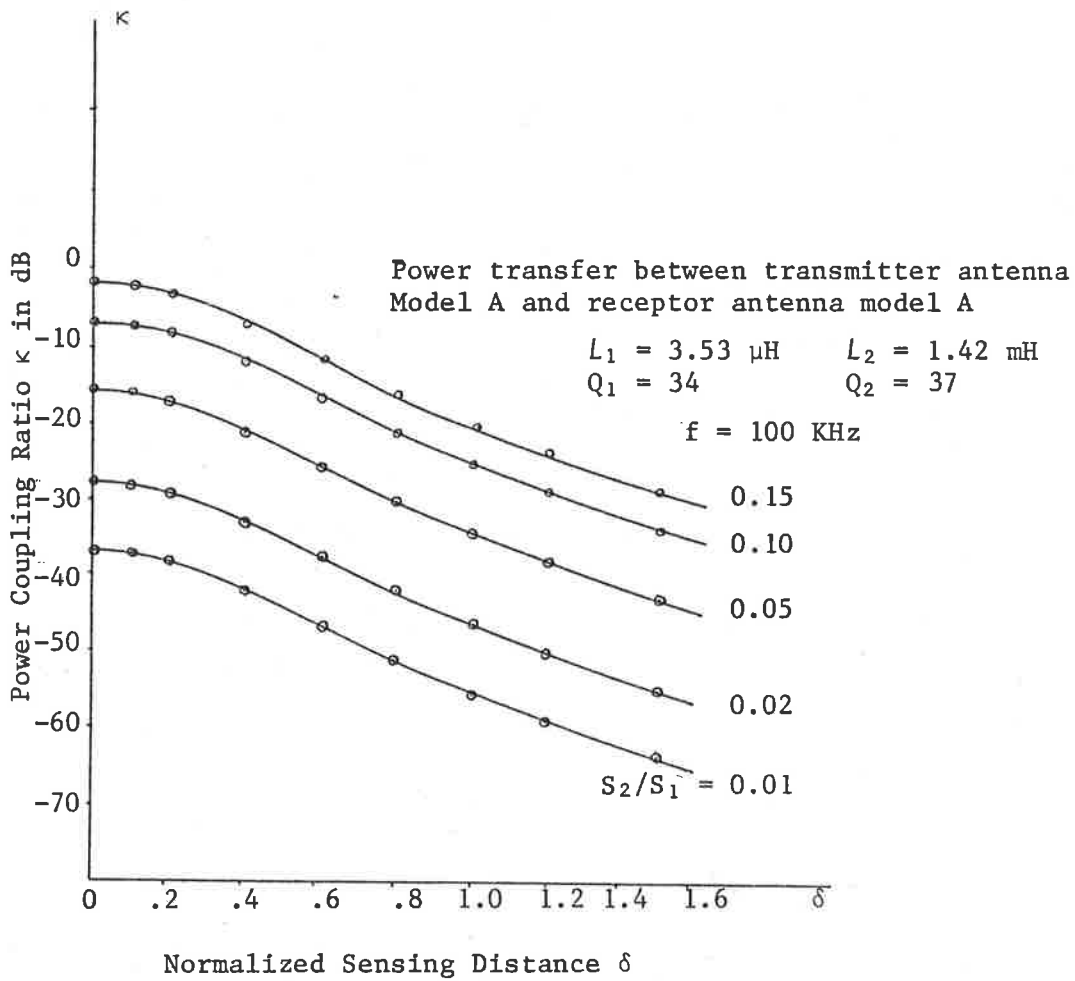


Figure 3.11 Variation of Power Coupling Ratio with Sensing Distance



Figure 3.12 Relative Position of the Two Antennae
for the Measurement of Power Coupling
Ratio

In the above experiment, the model B receptor antenna was placed in a horizontal plane on a wooden table top and the model A transmitter antenna in a horizontal plane various sensing distances above it.

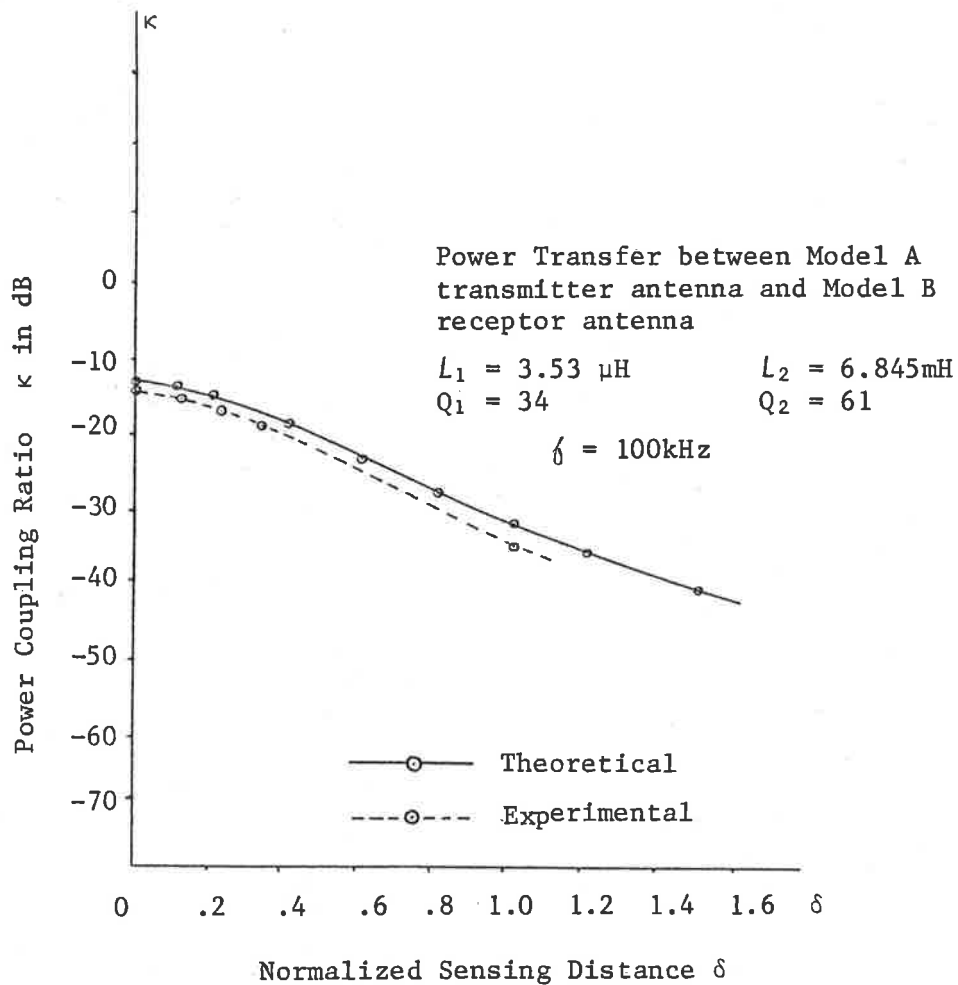


Figure 3.13 Power Coupling Ratio for Model B Receptor

ten make validity of the resultant expressions questionable. Furthermore, a generalized model of the interrogation environment for different applications is almost impossible to construct. The approach used in this study has been to obtain an understanding of some of these effects in a vehicle identification application by experimental means.

The effect of vehicle undercarriage on the electromagnetic coupling between the transmitter and the receptor was experimentally investigated by simulating the undercarriage with a sheet of mild steel of the dimensions of a medium-sized automobile. The variations in the self-inductance and quality factor of the transmitter antenna when used in the proximity of a large conducting surface have already been shown in Figure 3.5. Similar variations are also experienced with the receptor and transponder antennae.

In the vehicle identification application, the most convenient location on the body of the vehicle for the placement of the transponder, from the point of view of optimum coupling, is the vehicle undercarriage. The self-inductance and the quality factor of the model B receptor antenna were measured in the vicinity of the mild steel sheet, the results being plotted in Figure 3.14. Large variations in

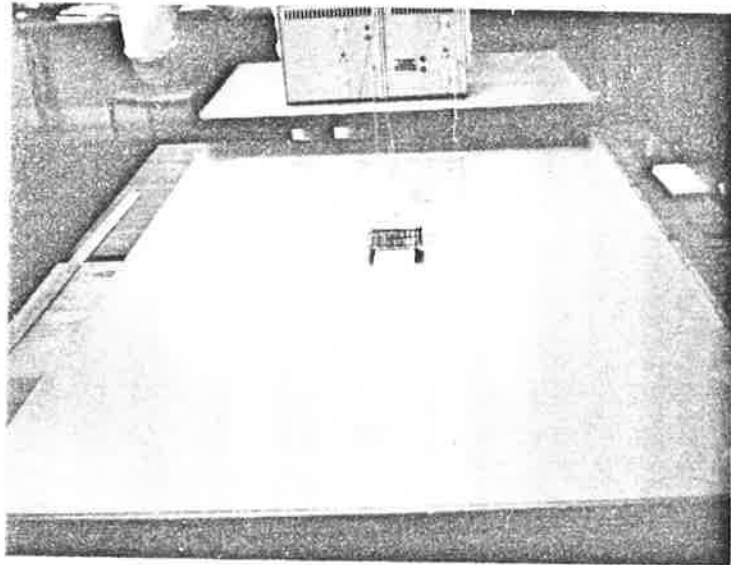


Figure 3.14a Study of Receptor Antenna Characteristics in the Vicinity of a Large Conducting Surface

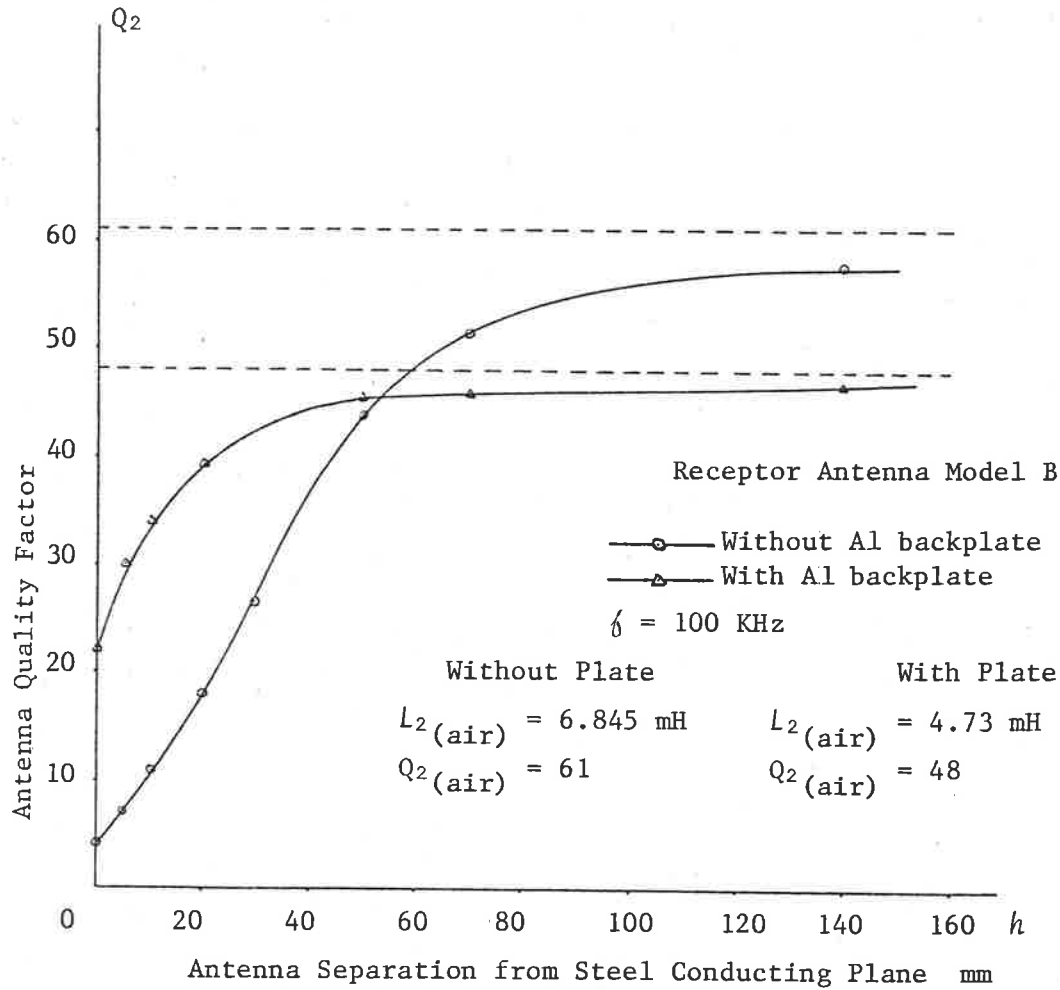


Figure 3.14b Variations in Receptor Antenna Quality Factor Near a Large Conducting Surface

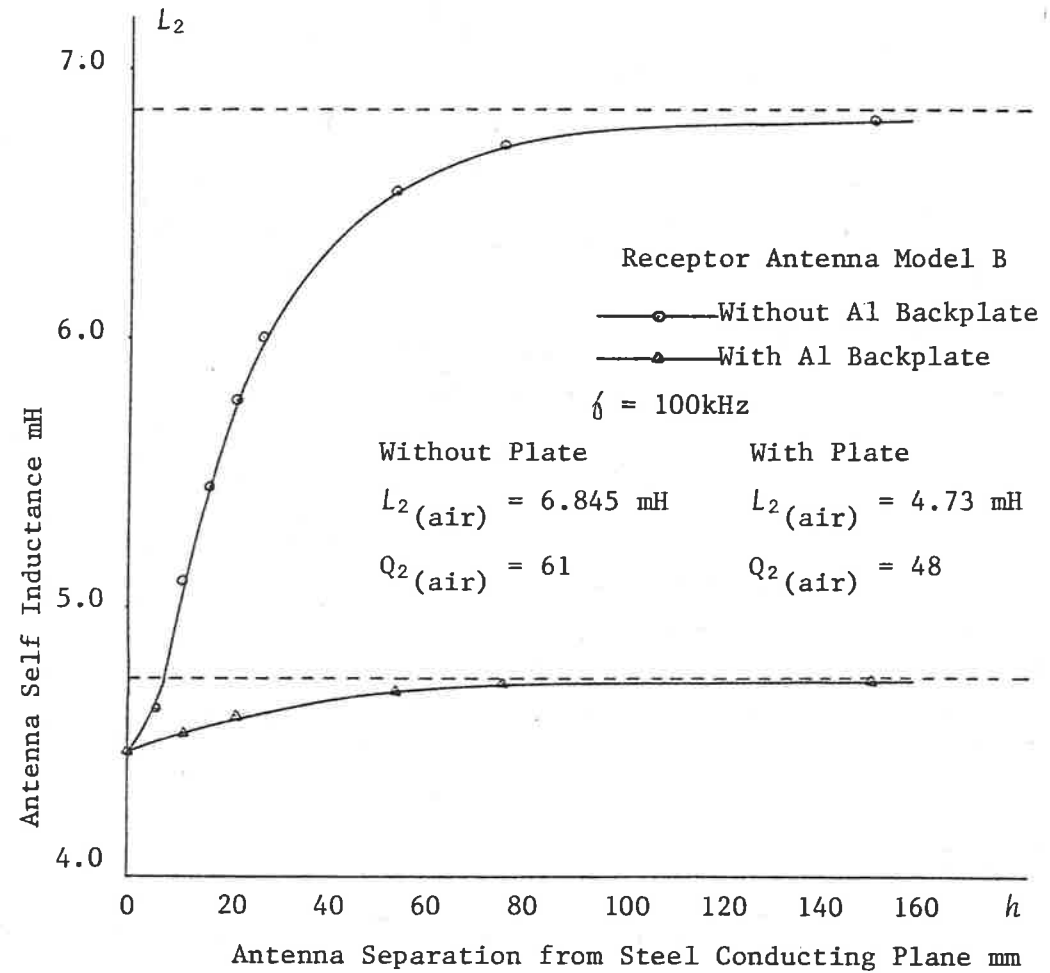


Figure 3.14c Variations in Receptor Antenna Self-Inductance Near a Large Conducting Surface

both self-inductance and quality factor of the antenna are noticed with small variation in distance between the antenna and the conducting plane. These variations would be even more pronounced in practice, due to additional losses introduced by the uneven shape and rusted surface of the undercarriage. The variations in self-inductance and quality factor close to the body of the vehicle may be reduced if the antenna is placed on an aluminium plate of the same size as the antenna. However, for practical materials, the reduction in the variation is achieved at the cost of reduction in the absolute value of the parameters concerned. The effect of using an aluminium backplate is also shown in Figure 3.14.

The effects of the large conducting plane on the power coupling ratio κ between the transmitter and the receptor were also investigated experimentally. The measurement set-up was similar to the one shown in Figure 3.12, except that the conducting plane was placed such that it was at a distance H from the model A transmitter and a distance h from the model B receptor antenna.

The power coupling ratio under these circumstances may be determined by the application of image theory. If the edge effects are neglected, the conducting plane, by introducing an image of the

transmitter antenna, helps in enhancing the field created by the transmitter at the transponder position. In the absence of other effects of the conducting surface, coupling between the transmitter and the receptor is improved. Further, a large reduction in the self-inductance of the receptor antenna as compared with the transmitter also enhances the coupling. The degradation in power coupling ratio between the antenna pair which was in fact observed is principally due to the reduction in quality factors of the antennae. The variations in the power coupling ratio with the normalized sensing distance is plotted in Figure 3.15. A comparison of the power coupling ratio in the presence of a large metal body with the same, in the absence of metal structures in the vicinity, reveals that, when the receptor antenna is placed very close to the metal surface, the reduction in quality factor of the receptor antenna negates all the benefits gained by the image and the reduction in self-inductance. This loss is appreciably reduced when the receptor antenna with an aluminium backplate of the same dimensions as that of the receptor is used. In applications such as vehicle identification, where the receptor antenna must be placed in close proximity to large metal bodies, the use of receptor antennae with metal backplates not only improves the power coupling ratio, but also reduces the variations in

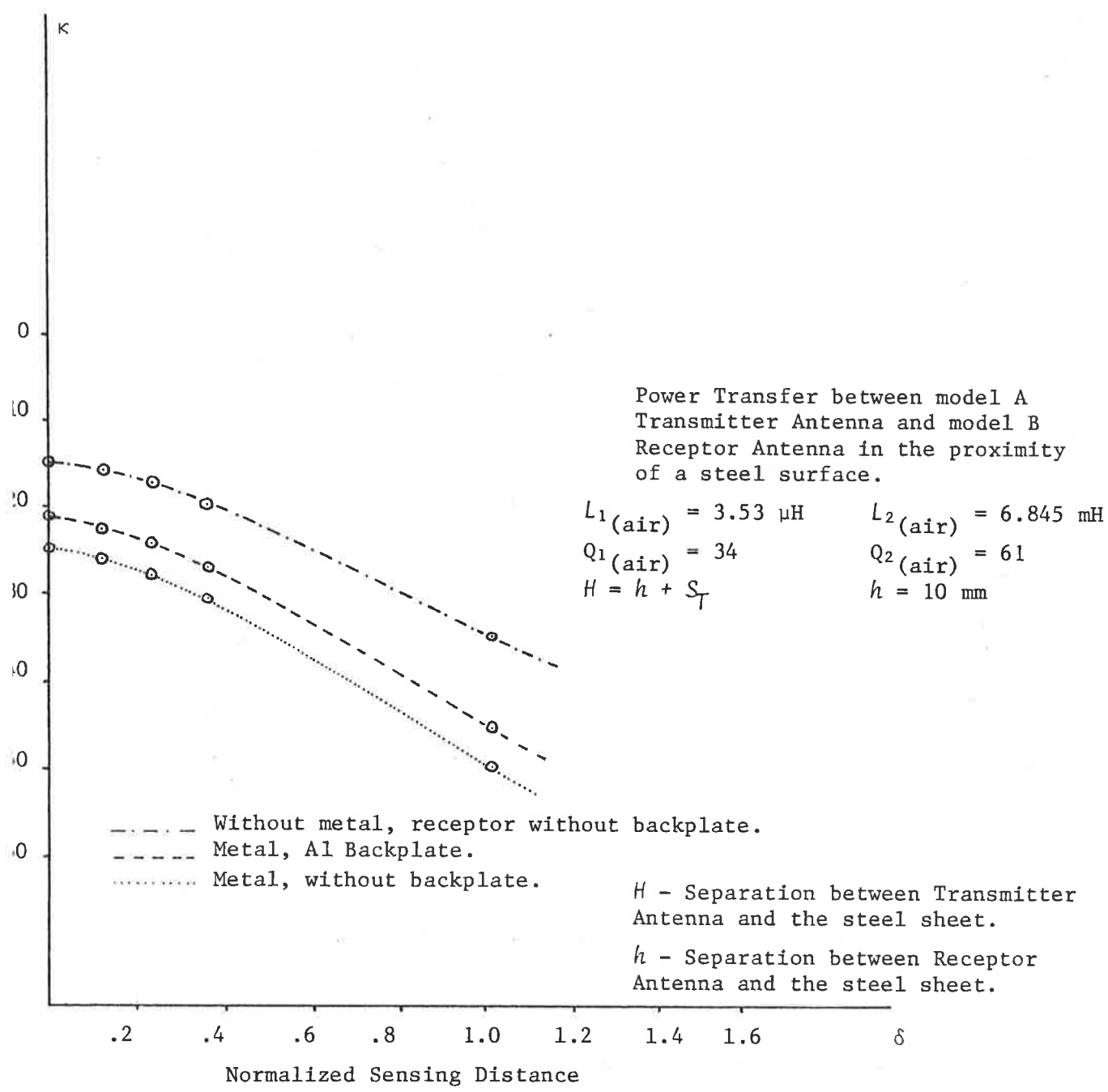


Figure 3.15 Power Transfer in the Vicinity of Steel Sheet

the self-inductance and quality factor of the antenna with variations in the distance between the transponder and the body of the vehicle. This decrease in sensitivity has practical benefits when such an identification system is required to be implemented. Although the use of other metals for the receptor backplate is possible, the suitability of different metals for the backplate is of secondary importance, and has not been investigated in any detail.

3.5 Antenna Considerations for One-Port Systems

The antenna considerations for the one-port transponder system are somewhat different from those for a two-port system. The single antenna on the one-port label, which serves as the receptor as well as the transponder antenna, is referred to as the transponder antenna. As described in the previous chapter, the basic constraint on the quality factor of the antenna imposed by the information bandwidth does not exist in the case of the one-port system. As the interrogation and reply signals are at two different frequencies, the commonly used definition whereby the quality factor of the antenna is simply defined as the ratio of centre frequency to bandwidth is no longer applicable. Furthermore, a

single antenna on the identification label eliminates some of the other problems, such as coupling between the receptor and transponder coils, competition for space between the two antennae, etc.

The quality factor of the transponder antenna for the one-port system has no constraint other than the practical ones. The optimum quality factor under these circumstances is only limited by the antenna environment, as described in the previous section.

Another consideration in the transponder antenna design is the impedance level of the transponder antenna. Because of the very low level of the signals involved, in order to maximize the available voltage, the transponder antenna is usually parallel tuned. Antennae suitable for one-port systems, therefore, should provide the maximum electromagnetic coupling at both the transmit and receive frequencies, with the only other losses arising in the switch.

3.6 Transceiver Antenna Design

The two-port object identification system described in the previous chapter requires four antennae for the exchange of energy and information between the interrogator and the transponder. The techniques to optimize the coupling between the transmitter and receptor antennae, and between the transponder and receiver antennae, have been discussed earlier in the chapter. The optimal operation of the two-port passive transponder system also requires that the interrogation signal and reply signals be uncoupled. This requirement, in a practical system, may be satisfied by minimizing coupling between the interrogation and reply signals, with cost and hardware complexity as the major constraints.

The coupling between the interrogation and reply signals arises because of the coupling between the transmitter and receiver antennae, and between the receptor and transponder antennae. Basically, the coupling between two signals can be minimized by providing the desired separation between the signals in space, frequency or time domain. Separation in time domain is achieved by switching the interrogation signal off during the reply signal period. The surface acoustic wave object identification system (Unisearch, 1972) is an example of a system where

this technique is employed. If the two signals are sufficiently separated in frequency, and the respective antennae are tuned to narrow bands around the respective resonant frequencies, then the signals may be sufficiently isolated from each other. This technique of isolating the interrogation and reply signals is employed in the microwave and low-frequency type identification schemes. An alternative technique of providing the separation is by spatial means. This technique of isolating signals may be implemented in several ways. The antennae responsible for unwanted coupling may be physically separated. The physical separations required, for most antenna geometries, to achieve sufficient isolation between the signals, make this approach practically unattractive. In the case of magnetic coupling, antennae of similar geometries may also be uncoupled by controlling the amount of overlap between their inductive elements to achieve zero mutual inductance. The principle of orthogonality may also be employed to isolate the signals. In this case, the antennae are positioned such that the flux-linkage is zero due to orthogonality. The interrogation signal may be further isolated from the reply signal by means of appropriate filtering. The performance requirements of the system, in a given application, may dictate the use of one or more of the above-mentioned techniques to provide the iso-

lation between the strong interrogation signal and weak reply signal.

Experimental studies of the antennae for the two-port passive transponder system indicate that sufficient isolation within the transponder between the interrogation and reply signals may be achieved either by orthogonality or by minimizing the mutual inductance between the receptor and transponder antennae by the controlled overlap method.

To sufficiently isolate the strong interrogation signal from the weak reply signal in a cost effective manner requires several of the above techniques used in conjunction with appropriate cost reduction schemes. The use of a single antenna for simultaneous transmission and reception is one of the most cost effective schemes and, therefore is the one considered here.

As the number of systems in a given area increases, the distance between the various antenna sites decreases, thus reducing the isolation between the antennae even further. In addition, larger numbers of systems in operation also cause a higher level of interference. This interference may be reduced to an acceptable level by the insertion of appropriate filters in the antenna feed lines.

However, it is often more economical to connect both transmission and reception equipment to a common antenna, known as the transceiver antenna, by means of a diplexer. This arrangement can often be made to provide the required isolation, and to eliminate the need for multiple antennae and associated hardware.

The problems which are generally encountered in transceiver type installation are interference and waveform distortion. The problem of finding a single favourable antenna location for both transmission and reception is less often encountered. When a transceiver antenna is to be used in a communication system, it is first necessary to establish the amount of isolation needed over the frequency band of interest. The amount of isolation required may be calculated from the known characteristics of the transmitter and receiver, and other system requirements. Once the isolation-frequency relationship has been established, the next consideration is the manner in which the isolation is achieved. Some techniques to provide the isolation using passive linear networks are proposed in the next section.

3.6.1 Isolation Characteristics of Transceiver Antenna

The isolation characteristics of a transceiver antenna depend upon the frequency separation between the transmitter and the receiver centre frequency, transmitter output voltage, and level of interfering voltage the receiver can withstand without desensitizing. The isolation required at any frequency can be expressed as a function of the above variables.

Let

f_1 be the transmitter centre frequency

f_2 be the receiver centre frequency

$v_1(f_1)$ be the amplitude of the transmitter output at the resonant frequency f_1

$v_2(f_2)$ be the level of the interfering voltage which the receiver can withstand without desensitizing, at its resonant frequency f_2

$N_1(f)$ be the ratio of the transmitter output at frequency f to the output at its resonant frequency f_1

$N_2(f)$ be the ratio of interfering voltage at frequency f that the receiver can withstand without desensitizing to the reference level at frequency f_2

$N_{12}(f)$ be the required isolation, in dB, between transmitter and receiver, at frequency f

Then it can be easily shown that:

$$N_{12}(\delta) = 20 \log_{10} \{v_1(\delta_1)\} - 20 \log_{10} \{v_2(\delta_2)\} \\ + 20 \log_{10} \{N_1(\delta)\} - 20 \log_{10} \{N_2(\delta)\} \quad (3.53)$$

or

$$N_{12}(\delta) = 20 \log_{10} \frac{v_1(\delta_1) \cdot N_1(\delta)}{v_2(\delta_2) \cdot N_2(\delta)} \quad (3.54)$$

$N_{12}(\delta)$ provides the isolation-frequency relationship required in the design of the diplexer for the communication system. It is possible to provide the isolation given by Equation (3.54) by a number of techniques. The technique to provide the required isolation considered here is by means of linear passive networks. When it is necessary to design a diplexer from standard off-the-shelf transmitter and receiver equipment, the required isolation characteristic, as given by Equation (3.54), can be derived from the equipment specifications supplied by the manufacturer. Usually, it is a good design approach to add some safety margin to allow for factors such as variations from the typical values of the specifications. Furthermore, additional isolation may be required at the points where intermodulation products are a problem. A block schematic of a transceiver antenna operation using a diplexer where passive filters are used to provide the required isolation is shown in Figure 3.16.

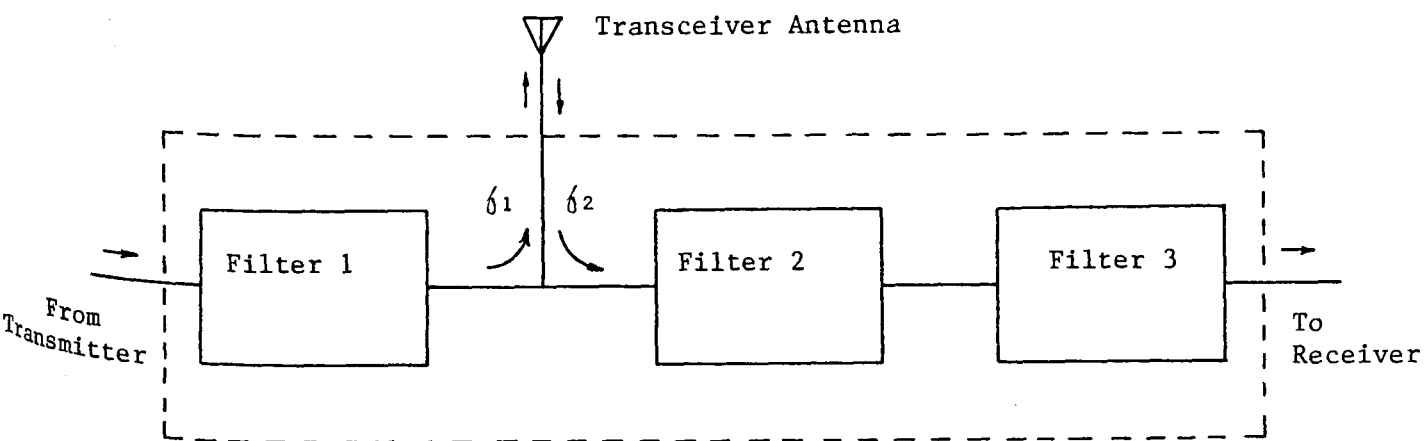


Figure 3.16 A Typical Transceiver Antenna with Diplexer

It is apparent from the block schematic that the system must be designed such that the transmit signal passes from the transmitter to the antenna terminal with very little attenuation, while at the same time being sufficiently isolated from the receiver. Similarly, the receive signal must pass freely from the antenna terminal to the receiver. In order to achieve these objectives, a large variety of passive linear filters are available. The selection of the filter type for the three filters shown in the diplexer arrangement depends upon the relative values of the transmit and receive frequencies, and the separation between the two.

The use of harmonically related frequencies for transmit and receive signals is highly advantageous. The advantages of this approach lie not only in ease of synchronization, but also in a greatly simplified receiver structure. For the reasons described in the previous chapter, the use of a subharmonic frequency as the reply signal carrier is assumed in the discussion to follow.

3.6.2 Filters for the Diplexer

One of the simplest diplexer networks requires a high-pass filter for one signal, and a low-pass filter for another. In this type of arrangement, there are two broad useable frequency bands separated by a relatively narrow crossover region. A good impedance match can usually be obtained by the use of complementary filters. An advantage of this type of system, apart from simplicity in the design, is that within the bands of operation either the transmitter or the receiver frequency can be changed without any change being made in the diplexer. The drawbacks of the technique are that the flexibility of the system is less compared with one using band-pass filters, and narrow crossover regions can only be achieved at the cost of high insertion losses.

Another common approach is to construct a diplexer using band-pass filters. One of the greatest advantages of using band-pass filters is the ease with which a system can be designed. The band-pass filter fundamentally passes one frequency, while attenuating all others. In addition, as the frequencies become further and further removed from the resonant frequency of the filters, they become more and more attenuated. This latter feature is sometimes desirable. However, it is a feature that is paid for with a higher insertion loss than would be obtained with other filter types. One of the important practical considerations in achieving frequency rejection using band-pass filters is the fact that increasingly complex band-pass filter configurations are required, which may be avoided if other filter types are employed.

Band-reject filters provide another very effective means of implementing a diplexer. The band-reject filter attenuates one frequency by a given amount and attenuates all others by a lesser amount. The further removed a frequency is from the resonance, the less it is attenuated. The advantage of the band-reject filter over the band-pass filter is that, in most applications, for a given complexity, it offers a much larger attenuation at the rejection frequency than the band-pass filter. The single

section band-reject filter has the disadvantage, when used in a diplexer, that if the insertion loss of the frequency to be passed is to be kept below a certain limit, then the loaded quality factor of the filter at the rejection frequency must change when the separation between the transmitter and receiver frequencies change.

An alternative to the single section band-reject filter is the twin section band-reject filter. Twin section filters are most suitable in applications where one frequency has to be passed and another rejected. To achieve this objective, it is merely necessary to tune one section for a null at the frequency to be stopped, and to tune the other section so that the frequency to be passed is attenuated as little as possible. Usually, the tuning of one section affects the tuning of the other by a small amount. Therefore, sometimes it is necessary to repeat the tuning procedure on both the sections a few times.

3.6.3 Diplexer Design Considerations for Two-port Transponder Systems

When far-field antennae are employed, the radiation resistance of the antenna acts as the load for one filter, and the source impedance for

the other. In the case of the two-port transponder employing near-field coupling elements, some further consideration must be given to the relationship between the transceiver antenna inductance and the filters in the diplexer. The problem seen here is that of the effect of the reactive antenna impedance on the filters of the diplexer. Furthermore, a filter with reactive source or load impedance is far more difficult to design than one with resistive loads. To overcome these problems, one needs to incorporate the transceiver antenna reactance into the filters for the diplexer such that it provides the electromagnetic coupling in the transmitter and receiver links on the one hand, and provides appropriate transfer functions for the filters on the other hand. This technique has a further advantage that, once the antenna inductance is seen as a part of the filters, the design of filters becomes much simpler, due to the absence of inductive loads.

In passive subharmonic transponder systems, the transmitter signal is very narrow-band, and the receiver signal is relatively broad-band. This asymmetry in the two signals may be utilized very advantageously in the design of a diplexer for such a system. It may be recalled from the discussion in the earlier part of this section that the simplest method of separating a narrow-band signal from an-

other signal is by means of a band-reject filter. These considerations lead to a simple design of a diplexer for the PST system using a band-reject filter for one signal, and a band-pass filter for the other.

Another important factor for consideration is the power matching between the transmitter and the antenna. In the case of the conventional four-antenna two-port object identification system, high resonant resistance of the transmitter antenna must be matched to the low source resistance of the transmitter power amplifier. This may be achieved either by means of split-capacitor impedance transformation or by transformer impedance matching techniques. Another factor which must be considered before implementation of any matching network is that of the impedance of the cabling between the power amplifier and the antenna. Usually, the impedance of the cable for the cable lengths used in practice is small compared with the dynamic resistance of the antenna, but is, nevertheless, significant because of the low source resistance. A matching transformer located at the source overcomes the problem of cable impedance.

The considerations detailed above vary from application to application. It is important to consider all these factors while at the design stage, to avoid problems later.

3.7 Conclusions

The design of various antennae for the object identification system employing low-frequency inductive coupling needs careful consideration of the numerous factors affecting the electromagnetic coupling. The constraints placed on the design of these antennae have been described. Optimization of antenna configurations and parameters under these constraints require a new approach, as described in the coupling volume theory. These techniques have been employed in achieving the local optimization of different antenna parameters, by separating them into suitable groups. Environmental effects on various antenna parameters are experimentally investigated for one application. And, finally, an economic approach to the antenna design is presented, in the form of a single antenna employed for simultaneous transmission and reception using a diplexer. The factors affecting the design of such diplexers for a two-port system are considered. Although all the factors mentioned in the study may not be ap-

plicable in all cases, a clear understanding of these problems and how they effect the system design is very important from the point of view of overall system design.

CHAPTER IV

STUDY OF ENVIRONMENTAL NOISE

4.1 Introduction

Environmental noise is a major source of concern in the design of any communication system. Interest in the radio noise environment of the passive object identification system arises because an optimal selection of transmitted signal power, frequencies of operation, modulation parameters, receiver location, and some of the other design variables, is critically dependent upon the characteristics of the environmental noise encountered in such applications.

The sources of noise most likely to affect a communication system are thermal noise, atmospheric noise, man-made noise, and receiver noise. Thermal noise makes very little contribution in the frequency band of interest. The noise performance of low frequency transponders is mainly affected by environmental noise, such as atmospheric and man-

made noise, with receiver noise playing only a minor part. The passive nature of these transponders, which leads to reply signals being very weak, requires a correct prediction of the predominant sources of noise likely to affect the performance of identification systems. A proper understanding of the characteristics of the environmental noise affecting the transponder operations is therefore essential in achieving an optimum system design.

4.2 Literature Survey

A large number of researchers have studied various aspects of environmental noise (Watt, 1957; Shinde, 1974; Field, 1978; Bolton, 1976; Hsu, 1976; Oranc, 1975; Skomal, 1969). The work includes measurement and characterization of atmospheric, as well as man-made, noise from a number of sources in various frequency bands. A brief summary of the published work will be presented to provide both a framework within which the results of the studies carried out by the author may be expressed, and a background with which the results of these studies may be compared.

4.2.1 Atmospheric Noise

Atmospheric noise has been studied by Watt and Maxwell (1957), Shinde and Gupta (1974), and Field and Lewinstein (1978), in various frequency bands. Watt and Maxwell have studied the statistical characteristics of atmospheric noise in the VLF and LF bands. Shinde and Gupta have developed a statistical model of high frequency impulsive atmospheric noise. Field and Lewinstein, on the other hand, have provided an amplitude probability distribution model of the VLF and ELF atmospheric noise. Among other authors, Ginsberg (1974) and Oh (1969) have contributed very useful results based on their low frequency noise measurements.

The atmospheric radio frequency noise field generally results from the combination of a large number of substantially independent events, and therefore it is readily amenable to statistical treatment. The primary source of atmospheric noise is lightning discharges. Although these discharges are essentially independent events, the variations in the noise field at any given location do not occur entirely at random. This departure from a random time distribution is due to the fact that the lightning discharges usually consist of multiple strokes, and also due to the time correlation introduced by the multiple propagation paths with differ-

ent characteristic delays. The noise mainly arises in the form of large pulses superimposed on a more homogeneous background. The impulsive bursts are caused by local thunderstorm activities, whereas the continuous background arises as a result of the superimposition of a large number of sources distant from the receiving location.

Watt and Maxwell have measured the cumulative distribution of the amplitudes and spacings of pulses in the instantaneous envelope of the atmospheric noise field strength in the very low to low frequency ranges. The block schematic of the equipment used by Watt and Maxwell is shown in Figure 4.1. A shielded loop antenna was used to reduce locally-generated noise. A relatively large loop was employed to keep the contribution from thermal noise in the band of interest low. The cumulative amplitude distribution of the noise envelope was obtained by gating a 10kHz signal for the period of data accumulation, and counting the number of cycles during which the output of the rectifier exceeded a threshold level. The time distributions were obtained by means of a run down unit operating at a specific sensing level, which translated the time spacings between pulses into amplitude variations. Two amplitude discriminators set at different levels were used to record the pulse spacings. Watt and Maxwell

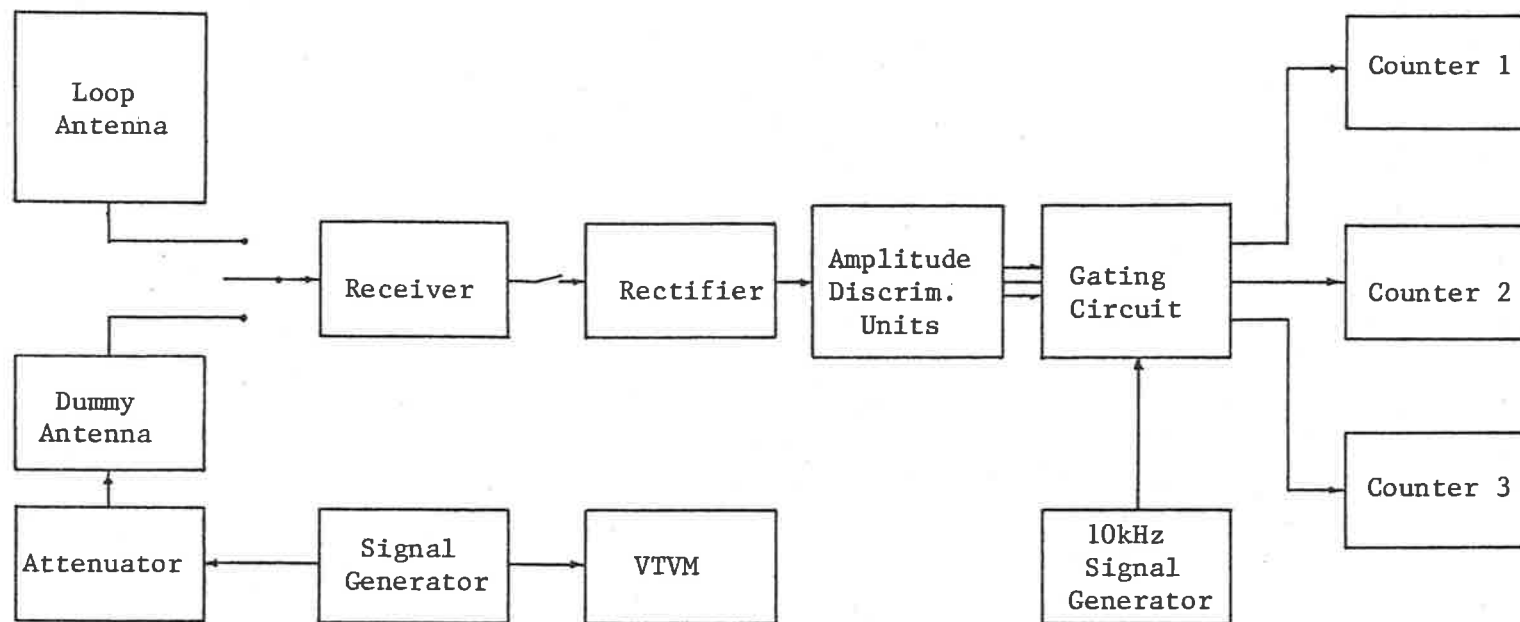


Figure 4.1a Measurement Instrumentation for Noise Envelope Cumulative Amplitude Distributions

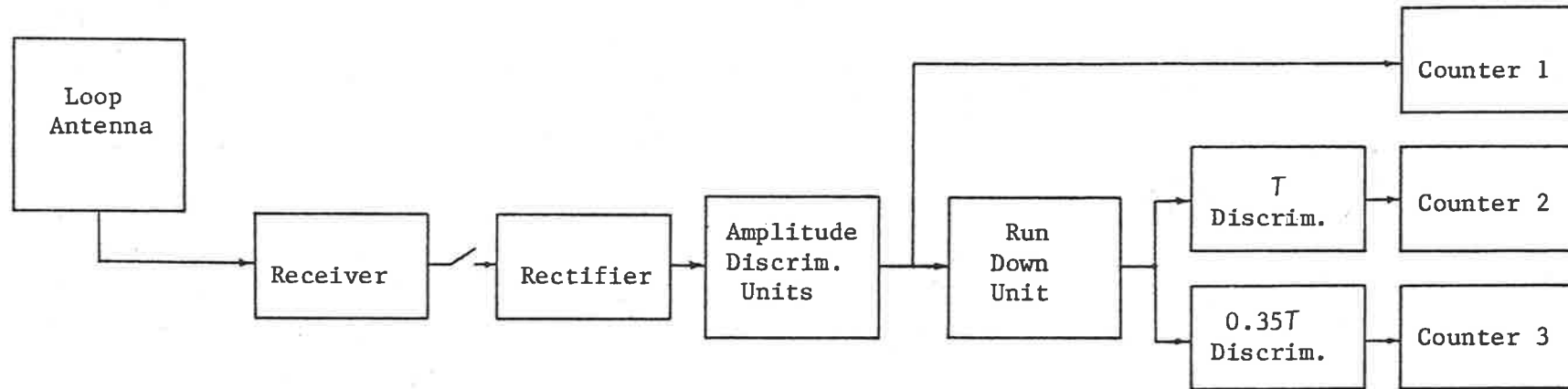


Figure 4.1b Instrumentation for Noise Envelope Pulse Width and Spacing Distribution

have shown that a part of the atmospheric noise behaves in the same way as narrow-band thermal noise. In the low field intensity region of the atmospheric noise distribution, the noise envelope is Rayleigh distributed. At higher amplitude levels, the spacing between the pulses does not appear to be random in temperate and arctic locations, while the pulses in tropical locations appear to be more randomly spaced.

Field and Lewinsein have provided an analytical model for VLF and ELF atmospheric noise amplitude probability distributions (APDs), and have confirmed the validity of their model by direct measurements. They have used a two component representation of the noise, in which $p_1(x)$ and $p_2(y)$ are the probability density functions (PDF) of the background and impulsive components, respectively. Field and Lewinsein have found the background noise component to be a zero-mean Gaussian random process with finite energy σ_0^2 . When such a process is applied to a narrow band filter, the output appears as an amplitude-modulated signal with a double-sided slowly varying envelope. The envelope has a Rayleigh density with average noise power $R_0^2 = 2\sigma_0^2$ and, therefore, the probability density function p_1 is given by

$$p_1(x) = \frac{2x}{R_0^2} \exp\left[\frac{-x^2}{R_0^2}\right] \quad x \geq 0 \quad (4.1)$$

Field and Lewinstein have suggested a power Rayleigh distribution for the non-Gaussian impulsive component, given by

$$p_2(y) = \frac{\alpha y^{\alpha-1}}{R^\alpha} \exp\left[-\left(\frac{y}{R}\right)^\alpha\right] \quad y \geq 0 \quad (4.2)$$

where α is a constant, $0 \leq \alpha \leq 2.0$, and depends on the pulse characteristics of the impulsive component and the system bandwidth. The parameter R corresponds to the average energy in y . Using the above selections for the probability density functions for the two noise components, the probability density function $p(z)$ for the overall noise $z = x + y$ is given by the convolution relation

$$p(z) = \int_{-\infty}^{\infty} p_1(z-y) p_2(y) dy \quad (4.3)$$

$$\text{or } p(z) = 2 \int_0^{\infty} \frac{\alpha y^{\alpha-1}}{R^\alpha} \exp\left[-\left(\frac{y}{R}\right)^\alpha\right] \frac{(z-y)}{R_0^2} \exp\left[-\frac{(z-y)^2}{R_0^2}\right] dy \quad (4.4)$$

and the amplitude probability distribution of the atmospheric noise is then given by

$$\text{APD} = \text{Prob}[Z > Z_T] = \int_{Z_T}^{\infty} p(z) dz \quad (4.5)$$

or

$$\text{APD} = \exp\left[-\left(\frac{Z_T}{R}\right)^\alpha\right] + \int_0^{Z_T} \frac{\alpha y^{\alpha-1}}{R^\alpha} \exp\left[-\left(\frac{y}{R}\right)^\alpha\right] \exp\left[-\frac{(Z_T - y)^2}{R_0^2}\right] dy \quad (4.6)$$

To simplify the calculation of the amplitude probability density, Field and Lewinstein defined a parameter V_D , which is the ratio of rms noise voltage to the average noise-envelope voltage. A good match between the experimental data and the noise model described by Equation (4.6) by appropriate selection of the parameter α has been observed.

Shinde and Gupta have developed a model of high frequency atmospheric noise. The model proposed is based on the observation that the impulsive component of the noise does not deliver energy at a constant rate, and clustering of noise pulses has been observed (Gupta, 1972). The resulting noise is, therefore, non-Gaussian in character, with a large dynamic range. Shinde and Gupta have used the quotient of two independent processes to describe the high frequency atmospheric noise $V(t)$.

$$V(t) = \frac{N(t)}{B(t)} \quad (4.7)$$

where $N(t)$ is a zero-mean Gaussian noise component with probability density function $p_N(x)$ given by

$$p_N(x) = \frac{1}{(2\pi\sigma_1^2)^{1/2}} \exp\left(\frac{-x^2}{2\sigma_1^2}\right) \quad -\infty < x < \infty \quad (4.8)$$

and $B(t)$ is a non-Gaussian random process with strong correlation among its samples. The probability density function used by Shinde and Gupta is

$$p_B(b) = \begin{cases} \left(\frac{2}{\pi}\right)^{1/2} \cdot \frac{b^2}{\sigma(3+1/2n)} \exp\left[\frac{-b^2}{2\sigma^2}\right] & b \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.9)$$

where n is a constant

Then the probability density function $p_Y(y)$ for the overall atmospheric noise is given by

$$p_Y(y) = \int_0^{\infty} x p_B(x) \cdot p_N(yx) dx \quad 0 < y < \infty \quad (4.10)$$

The above integral is simplified when $\sigma_1 = \sigma$, then

$$p_Y(y) = \begin{cases} \frac{2}{\pi \sigma^{(1/2n)} (1+y^2)^2} & 0 < y < \infty \\ 0 & \text{otherwise} \end{cases} \quad (4.11)$$

Shinde and Gupta have found a close agreement between the observed values and the calculated values based on their model. Some of the other important statistical parameters have also been derived from Equation (4.11).

4.2.2 Man-made Noise

As the radio frequency spectrum becomes more and more congested due to the increasing number of users, efficient utilization of the spectrum becomes an important consideration in the design of a communication system. In the low frequency band, where the passive subharmonic transponder system is expected to operate, man-made noise forms the major part of environmental noise. The major sources of man-made radio interference include power transmission lines, automotive ignition systems, gas discharge devices, industrial switching equipment, and radio frequency stabilized arc welders. The waveform characteristics of radio noise emitted by different sources place upon the system designer constraints peculiar to the origin of the noise. In certain applications, correct prediction of the predominant noise source is essential to achieve optimum system operation. In addition to achieving a better system design through knowledge of the relative intensities of the incidental noise sources, one also benefits from the correct identification of the principal sources of interference producing the noise environment.

As a result of the very low reply signal levels, an understanding of the characteristics of environmental noise becomes particularly important in ob-

ject identification applications. In most identification applications, the noise at the receiver input is the cumulative effect of the noise from a large number of sources. The identification of individual noises in such applications is not very critical; rather, the characteristics of overall interference may be utilized in the operation of an optimum passive subharmonic transponder system. Man-made noise has been studied extensively by Bolton (1976), Hsu et al (1976), Middleton (1973), Oranc (1975), Schulz (1974), Skomal (1966, 1967, 1969), and Kanda (1975), mainly in VHF and UHF bands.

Hsu, Maxam et al (1976) have investigated automotive noise due to individual cylinders from the diagnostic viewpoint. The data acquisition system used for their measurements is shown in Figure 4.2. The instrumentation used for the measurement included a radio-frequency receiver, a logarithmic amplifier, a peak detector, single cylinder gating circuitry, a pulse height analyser, an oscilloscope, a teletype, and a dipole antenna used in vertical and horizontal polarisations.

After studying the pulse amplitude density (PAD) of the ignition noise due to the individual cylinders at VHF frequencies, Hsu et al have con-

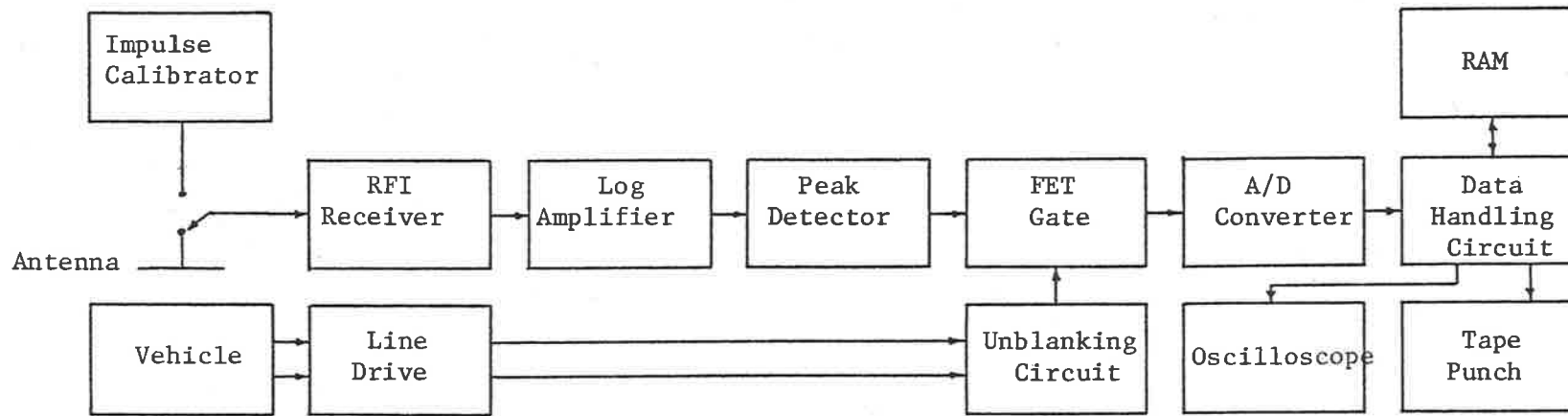


Figure 4.2 Pulse Amplitude Distribution Measurement Instrumentation

cluded that the statistics of the ignition noise due to individual cylinders are nonuniform. Under identical test conditions, the pulse amplitude density of different cylinders has been found to vary by more than 30 dB. The composite pulse amplitude density curves have also been plotted by means of superimposition. The technique developed in this study is quite useful in investigating the ignition noise emanating from individual cylinders.

Hsu et al have also studied the statistical distribution of the amplitude of automotive ignition noise. In particular, they have considered relating the deterioration effects of the ignition noise on various communication systems to a set of statistical parameters. Based on their experimental results, Hsu et al have concluded that noise parameters such as peak level, quasi-peak level, average level, and rms level are inadequate for the evaluation of a communication system's performance. They have also characterised the amplitude probability distribution of automotive ignition noise in the high frequency band by means of best fitting of data on various probability distribution graph papers, in particular, those for the log-normal distribution, the exponential distribution, the Rayleigh distribution, and the Weibull distribution.

This graphical method of investigating the probability distribution of a random process requires, first of all, plotting a set of data with known distribution and, then, transforming the axes to obtain a straight line plot. After the probability paper for a given distribution is ready, the actual noise data is plotted. The study of Hsu et al indicates that, for their measurements, the amplitude probability distribution of ignition noise is given by the Weibull distribution. Further, it is suggested that, possibly, other distributions also exist which provide good fit for the ignition noise. Another conclusion derived from the study is that the ignition noise from a single vehicle is highly reproducible.

Shepherd has studied the amplitude probability distribution of automobile ignition noise-envelope in the HF band. The measurements were conducted on vehicles under stationary and freeway conditions. Measurements were conducted at two engine speeds in the stationary mode. Shepherd's instrumentation, as shown in Figure 4.3, included several phase-stable receivers with coherent quadrature detectors, whose outputs were digitized, and the resulting digitized envelope was stored on a magtape for off-line waveform analysis. The amplitude probability distribution for most vehicles was found to follow Rayleigh

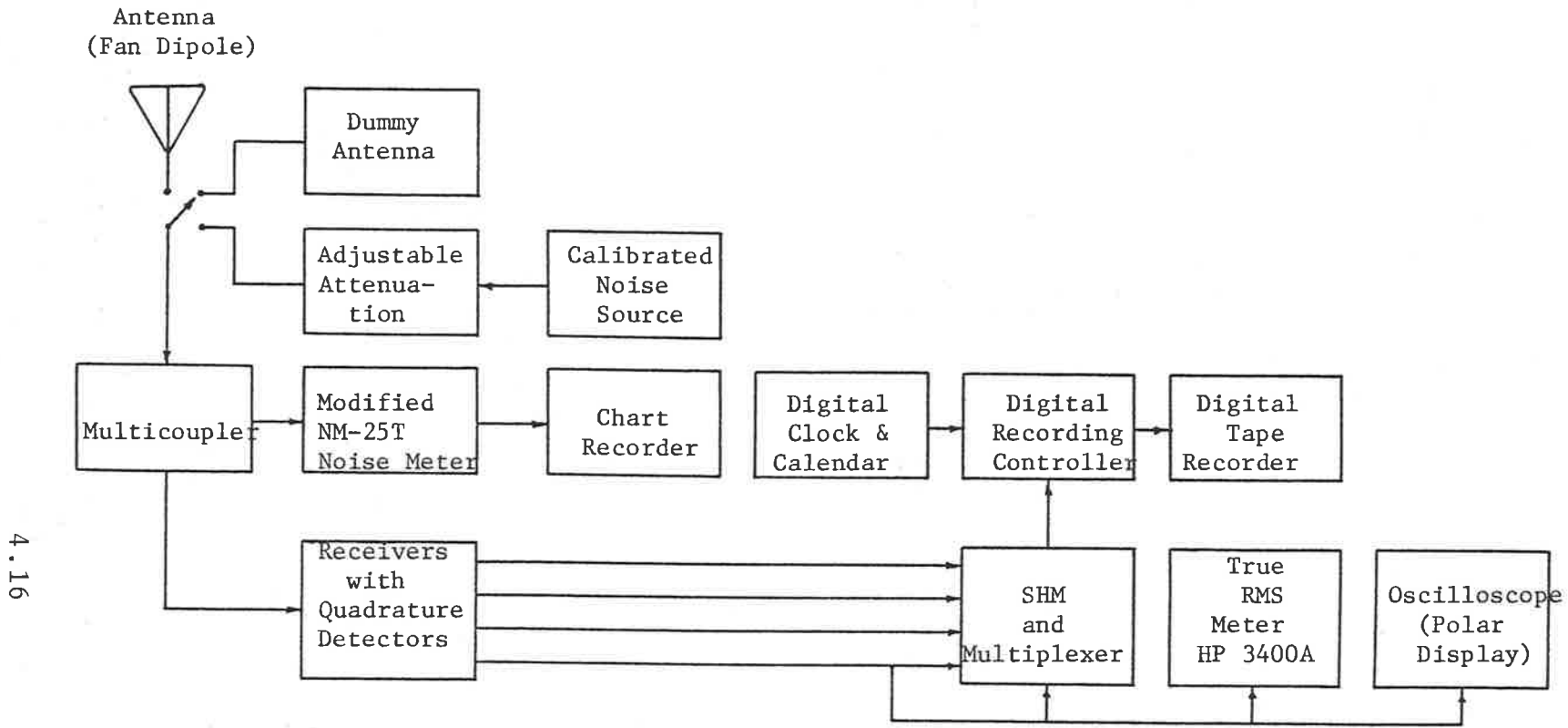


Figure 4.3 Shepherd's Instrumentation for Automobile Ignition Noise Measurement

distribution, except for a small percentage of high amplitude pulses. Shepherd's study shows that the APD's mainly consist of two distinct straight line segments. The breakpoint location depends upon the engine speed and the traffic intensity. The APD for stationary vehicles varies noticeably from one vehicle to another. In the freeway situation, however, there is no appreciable change in the APD with changes in the traffic density. Further, the average noise power radiated from a single stationary vehicle varies with frequency, and increases monotonically with engine speed. For freeway traffic, the average power increases with traffic intensity, and with the proximity of the receiver to the freeway.

Skomal has made a comparative study of man-made radio frequency noise from various sources, such as power transmission lines, automotive traffic and rf stabilized arc welders. In this study, radio interference data of many investigators in the frequency bands ranging from ELF through to UHF were assembled and converted to a common system of units, and plotted collectively to assess the relative interference levels produced by different sources. The major findings of Skomal's study were that in industrial environments, power transmission lines and automotive ignition systems were the major

sources of radio noise. In the 750kHz to 25MHz frequency bands, rf stabilized arc welders produced greater radio interference than either power transmission lines at comparable distances from the receiver or automotive ignition due to heavy traffic. In the HF band, lower voltage power transmission lines produced more noise for distances up to 30 metres than automotive ignition noise due to heavy traffic at the same distances. In the upper VHF and UHF bands, the automotive traffic interference exceeded the noise levels generated by power lines under dry weather conditions. However, under conditions of fog, mist, high relative humidity, or rain, power-line noise became comparable to that of dense traffic. A non-linear dispersion law to describe the frequency variations of the total man-made noise proposed by Skomal showed that the composite noise field had a decreasing negative slope with increasing frequency. In another study of incidental noise-envelopes Skomal established the limiting statistical distribution of the man-made noise by applying the central limit theorem. It was observed that the distribution of the incidental noise-envelope converges to a Rayleigh distribution either with increasing frequency for a given measuring distance, or with increasing distance for a given frequency.

Schulz and Southwick have studied the ignition noise from V-8 automotive engines. Their measurements of the amplitude probability distribution correspond well with the results of Hsu et al. The amplitude probability distribution was determined by real-time calculations of long term measurements of ignition noise. The instrumentation, as shown in Figure 4.4, consisted of an antenna connected to a step attenuator in series with a converter. The output of the converter was connected to an amplitude dwell timer via a spectrum analyser. The amplitude dwell timer, along with the counter, was used to measure, in real time, the percentage of time the noise-envelope exceeded a specific level. Schulz and Southwick have found that measurements were highly repeatable, shapes of amplitude probability distribution curves were essentially independent of the receiver centre frequency, but depended upon the receiver bandwidth and the number of vehicles, while the noise-envelope amplitude was found to be a function of the receiver centre frequency and the number of vehicles, but independent of the receiver bandwidth.

Oranc has studied the automobile ignition noise in the VHF and UHF bands. The aim of the study was to calculate the effect of ignition noise on a mobile data communication link. A new parameter, the

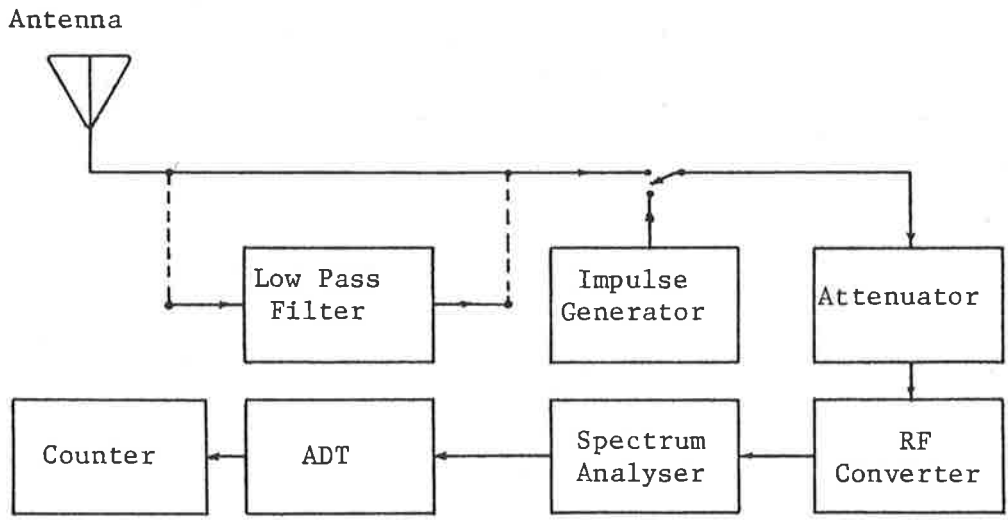


Figure 4.4a APD Measurement Instrumentation of Schulz and Southwick

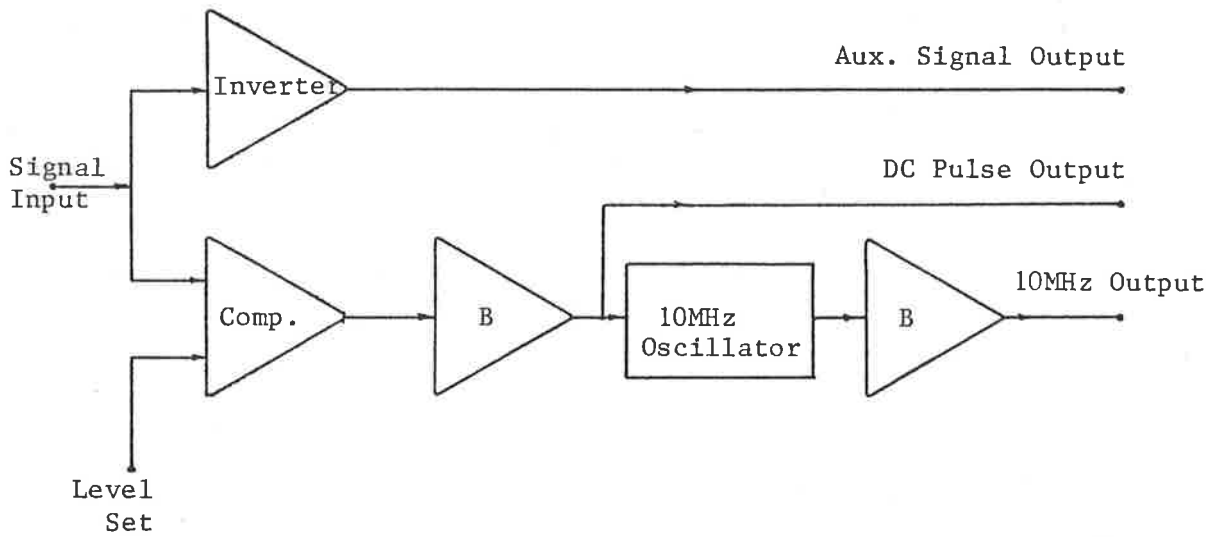


Figure 4.4b Amplitude Dwell Timer (ADT)

noise amplitude distribution, was defined for this purpose. An analytical expression was developed for the noise amplitude distribution under idealized conditions. Oranc also described the method for calculating other useful statistical measures, such as the average repetition frequency and the peak amplitude distribution of the noise pulses. As shown in Figure 4.5, Oranc used a very simple technique for the measurement of the noise amplitude distribution of automotive ignition noise. The system consisted of a conventional receiving system in which noise pulses were passed through an envelope detector and filter arrangement. The noise-envelopes were grouped according to their amplitude levels, and the number of pulses in each group was counted. Although the noise amplitude distribution does not provide information about the distribution of noise pulses in time, it may be used to calculate the average bit error rate.

Bolton has studied the man-made impulsive noise in the low frequency band. The instrumentation used in this study is shown in Figure 4.6. The dummy antenna and the signal generator arrangement were provided for the purpose of system calibration. Two classes of noise were defined in his study: Class A type, where the emission spectrum of the noise is narrower than the bandwidth of the receiving system;

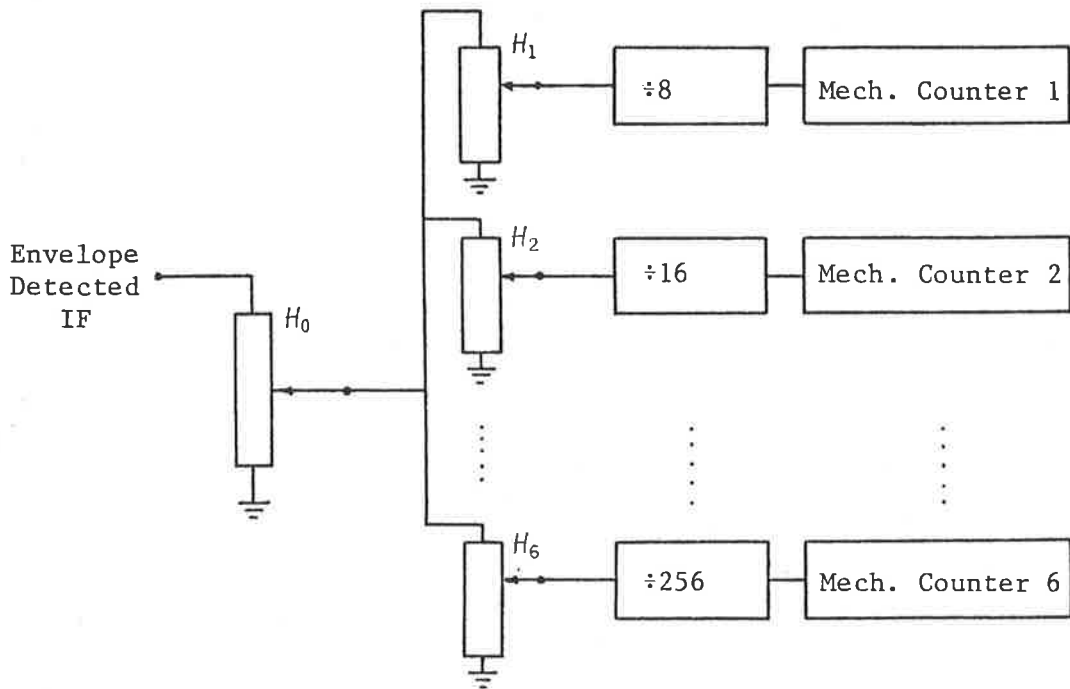
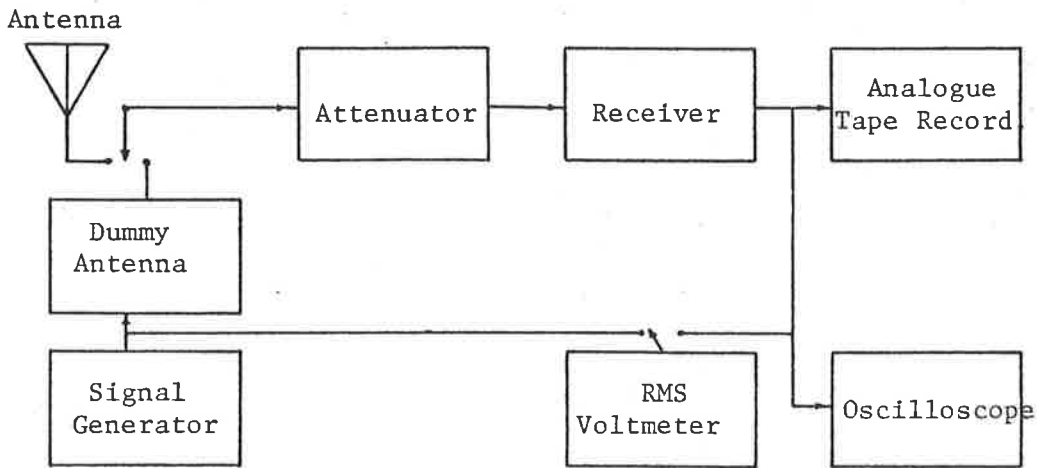
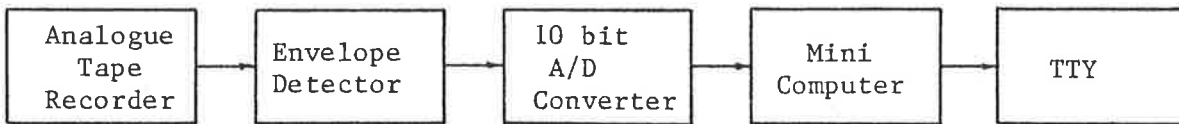


Figure 4.5 NAD Measurement Instrumentation of Oranc



a. Data Recording

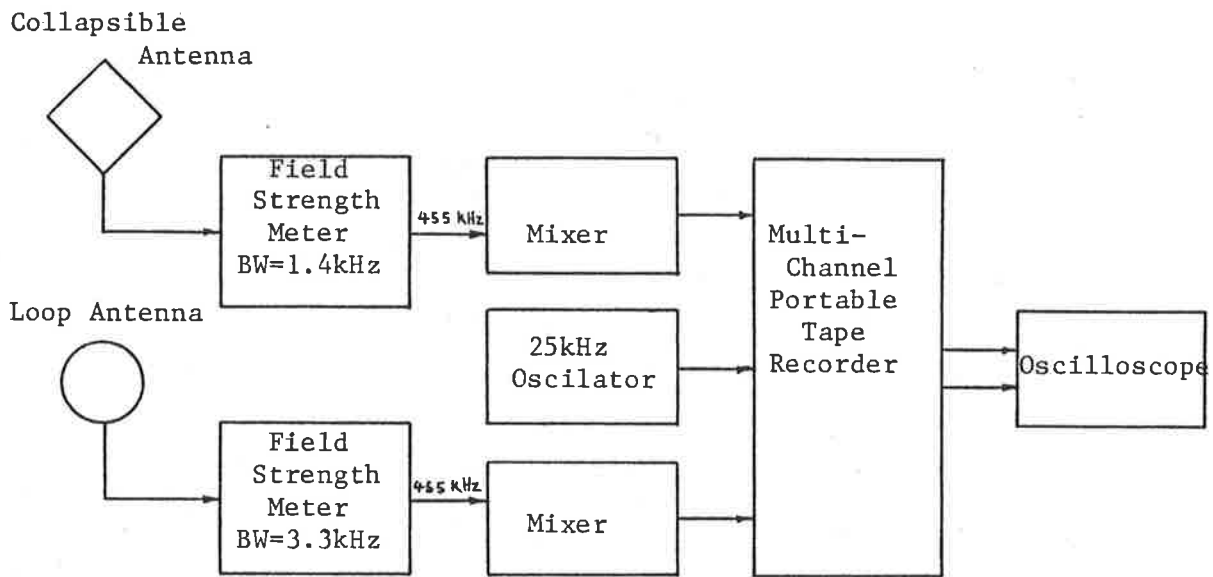


b. Data Analysis

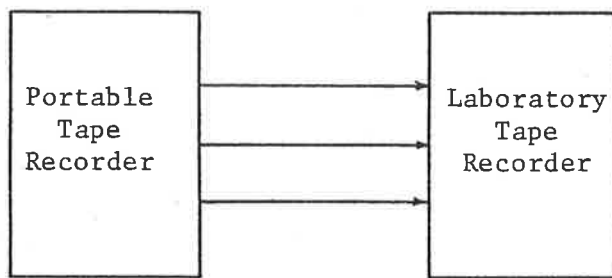
Figure 4.6 Noise Recording and Data Analysis System of Bolton

and Class B type, where the emission spectrum of the impulsive noise is greater than the passband of the receiver. While automotive ignition noise, atmospheric noise, and noise from industrial equipment were found to be of Class B type, the power-line noise was found to be of Class A type. Bolton made a comparative study of the amplitude distribution of man-made impulsive noise at business, industrial and residential sites. His studies revealed that, during severe thunderstorm activity, the rms amplitude of atmospheric noise in the low frequency band was greater than any other man-made noise by more than 20dB. In the absence of thunderstorm activity, man-made noise at a business site, which was mainly due to automotive traffic, was found to be the greatest. Bolton concluded that, due to the very large number of man-made noise-generating sources, it is very difficult to provide a general model of man-made noise. This problem can be overcome by direct measurement of noise data in the area where a communication system is expected to operate.

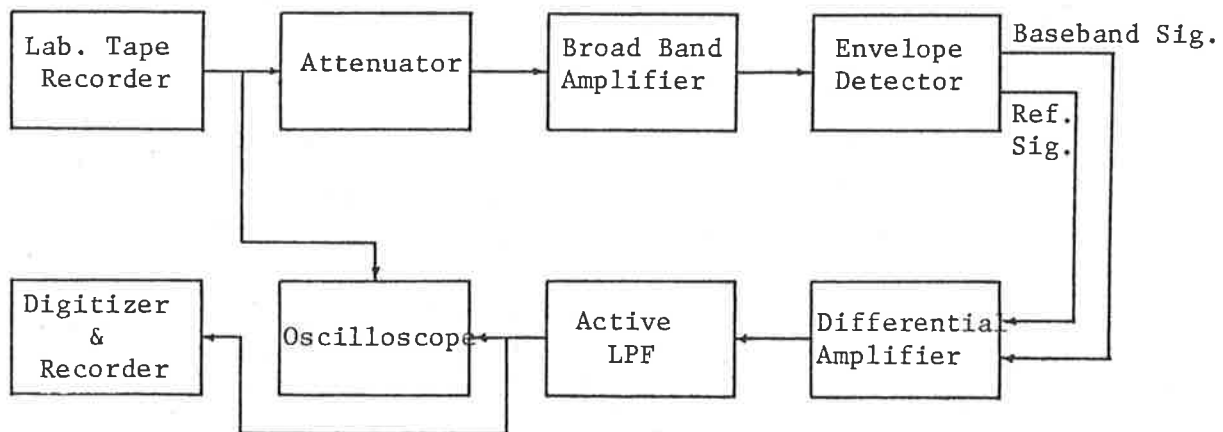
Kanda has studied the time and amplitude statistics of electromagnetic noise in mines, and determined a number of statistical parameters. The measurement and data processing instrumentation of Kanda is illustrated in Figure 4.7. The time and amplitude statistics were obtained from a large



a. Recording System



b. Transcribing System



c. Data Processing System

Figure 4.7 Recording, Transcribing and Processing System of Kanda

sample of raw analogue data in operational mines through computer software techniques. The magnetic field strength was calculated for the frequencies from 10kHz to 32MHz. Kanda used Allan variance analysis to determine the length of time necessary to provide statistical validity.

4.3 Statistical Noise Measures

Environmental noise has been studied by various workers, with different objectives. Signature analysis (Myers, 1963) has been used to correlate noise to its source. Studies have been conducted to develop mathematical models of the random process (Yue et al, 1978; Shinde and Gupta, 1974). While some authors have made a comparative study of the noise from various sources (Skomal, 1967), others have studied the characteristics of the composite noise. Effects of environmental noise on the performance of a communication system have also been considered. Most of the literature to date has been devoted to the study of noise in VHF and UHF bands mainly, because of the great deal of usage of these frequency ranges in various communication applications.

The principal objectives of the study conducted by the author were to measure and characterize environmental noise affecting the operation of low frequency object identification systems, to develop measurement and analysis techniques to derive statistical measures from the raw data, to use these statistical measures in the evaluation of system performance, and to formulate strategies to improve performance of such systems when operated in realistic environments.

Degradation in the performance of a digital communication system of the type used in low frequency object identification applications is mainly due to impulsive noise which has a high peak-to-rms ratio, with thermal and receiver noise playing a minor part. The sources of impulsive noise in the radio frequency spectrum are:

- i) Atmospheric noise
- ii) Man-made noise.

Atmospheric noise in the low frequency band in the non-tropical regions, where thunderstorm activities are very infrequent, is low. This hypothesis was later confirmed by the author via measurements, and found to be consistent with the studies of Bolton (1976). The preliminary environmental noise measurements performed by the author

indicated that impulsive noise from man-made sources was a major source of concern, and efforts were directed towards the measurement and characterization of environmental noise from such sources.

The principal contributors of the man-made noise are industrial switching equipment, such as switching regulators, commutators, rf stabilized arc welders, inverters, etc., power lines, and automobile ignition. The object identification system may be required to operate in the proximity of any of the above-mentioned sources. Measurements were conducted in the vicinity of some of the above sources. The statistical characteristics of the impulsive noise generated by the sources mentioned above mainly depend upon the source and measurement environment.

Wide band impulsive noise is also generated in power lines, due to the corona discharge between the conductors. Power lines also help in propagating noise from other sources. The power line noise has a strong near-field component, and also has some far-field components.

Automotive ignition noise is the result of a high discharge current in the spark plug, which lasts for a few nanoseconds for the most commonly

used spark gap sizes. This process results in a very wide spectrum of radio interference, extending from the low frequency band up to the ultra high frequency band. The noise-envelope at the receiver is possibly affected by the body of the automobile.

Impulsive noise from different sources has special characteristics, due to different generating mechanisms. It is, therefore, very difficult to define a general model of impulsive noise. One technique for the evaluation of system performance in the presence of impulsive noise is to measure the statistical parameters necessary for the prediction. Because some of the statistical parameters are dependent upon the type of communication link, impulsive noise measurements must be carried out on a link similar to the one used for the application. The following data concerning the environmental noise statistics should preferably be known, to evaluate the performance degradation in a communication system:

- i) Amplitude distribution
- ii) Time distribution
- iii) Pulse width distribution
- iv) Impulsiveness of noise
- v) RMS and average amplitude of the noise-envelope.

Several statistical measures may be used to derive the above information. The following measures provide a useful set of statistics of the electromagnetic interference upon which communication system design decisions can be based:

- i) Allan Variance Analysis (AVA)
- ii) Amplitude Probability Distribution (APD)
- iii) Noise Amplitude Distribution (NAD)
- iv) Pulse Width Distribution (PWD)
- v) Time Probability Distribution (TPD).

4.3.1 Allan Variance Analysis (AVA)

It is essential, in a statistical study of a random process, to determine how much data must be collected. Therefore, in the measurement of a statistical phenomenon, the minimum length of time over which the phenomenon is observed should be calculated. Allan Variance Analysis can be used to accomplish this determination (Allan, 1966).

A record of the random process under consideration, $n(t)$ is divided into a number of equal time records of length ζ , and the average value of the data $n(t)$ during each record is calculated. Let n_k be the time average of the process during the k th record. Then

$$n_K = \frac{1}{\zeta} \int_{t_K}^{t_{K+1}} n(t) dt \quad (4.12)$$

where $t_{K+1} - t_K = \zeta$

The sample average and sample variance of successive samples for a sample size of two is then defined as:

$$\bar{n}_K(2, \zeta) = \frac{1}{2}(n_K + n_{K+1}) \quad (4.13)$$

$$\begin{aligned} \sigma_{n_K}^2(2, \zeta) &= \sum_{i=K}^{K+1} (n_i - \bar{n}_K)^2 \\ &= (n_K - \bar{n}_K)^2 + (n_{K+1} - \bar{n}_K)^2 \\ &= \left(\frac{n_K - n_{K+1}}{2} \right)^2 + \left(\frac{n_{K+1} - n_K}{2} \right)^2 \\ &= \frac{1}{2}(n_{K+1} - n_K)^2 \end{aligned} \quad (4.14)$$

The Allan Variance, $\sigma_n^2(\zeta)$ is then defined as the infinite time average of all the sample variances for one selection of sample length ζ .

$$\sigma_n^2(\zeta) = \langle \sigma_{n_K}^2(2, \zeta) \rangle \quad (4.15)$$

In practice, a finite set of sample variance may be used to calculate the Allan Variance. Therefore,

$$\sigma_n^2(\zeta) = \frac{1}{N} \sum_{K=1}^N \sigma_{n_K}^2(2, \zeta) \quad (4.16)$$

The above calculations are repeated for several values of the record length ζ . Allan Variance is then plotted against ζ . For a given maximum allow-

able variation in $n(t)$, the requirement of the minimum average can then be determined from the plot. A graph of Allan Variance that does not decrease with record length ζ indicates that, to achieve the requirement of the minimum average, the record length must be increased.

4.3.2 Amplitude Probability Distribution (APD)

The Amplitude Probability Distribution is the most common statistical measure used in the study of random processes. The APD is defined as the percentage of time the noise-envelope exceeds various amplitude levels. The percentages are calculated with respect to the complete record length.

4.3.3 Noise Amplitude Distribution (NAD)

The Noise Amplitude Distribution is a measure of the cumulative distribution of peak amplitude of the noise pulses. The NAD is defined as the average number of noise pulses per unit time exceeding various amplitude levels, and is plotted as the percentage of noise pulses per unit time that exceed the specified amplitude levels.

4.3.4 Pulse Width Distribution (PWD)

The Pulse Width Distributions show the probability distribution of the duration of the noise-envelopes. A suitable definition of the pulse width may be used. For example, the pulse width may be defined as the time duration from the instant when the instantaneous amplitude of a noise-envelope exceeds 10% of its peak amplitude to the instant when the instantaneous amplitude of the noise-envelope goes below 10% of the peak amplitude. The PWD is plotted as the percentage of pulse widths exceeding various values. The PWD is a function of the peak envelope amplitude.

4.3.5 Time Probability Distribution (TPD)

The Time Probability Distribution is a measure of the cumulative distribution of the time spacing between successive pulses. These distributions are also function of the peak noise amplitude. The TPD plots are normally given as a family of curves, with each curve representing the percentage of pulse spacings exceeding various values for the pulses of a given amplitude.

A complete characterization of impulsive noise requires the measurement of the statistical parameters mentioned above, along with the measurement

of some other parameters, such as rms value. The performance degradation of a communication system can, however, be evaluated with only a few of the above statistical measures. The main objective of the environmental noise study is the prediction of the system error performance in the presence of environmental noise. For this purpose, the noise amplitude distribution, and the time probability distribution appear to be the most suitable measures. Although it is possible to derive the information necessary to predict the error performance of a communication system from a combination of other statistical measures, the NAD and TPD are advantageous, from the point of view of measurements.

4.4 Environmental Noise Measurements

As is evident from the different measurement techniques employed by various authors described in Section 4.2, the measurement of a set of statistical parameters of a random process may be performed by a variety of techniques. Some of the important considerations in the design of suitable measurement techniques are:

- i) the desired statistical parameters
- ii) simplicity of instrumentation
- iii) on or off line data processing

- iv) data storage constraints
- v) data analysis software.

Obviously, the most important considerations in the design of measurement systems and their implementation is the set of statistical parameters required to be measured. However, in the measurement of an unknown, such as environmental noise, the decision on which parameter is more important than the other may be quite difficult. In particular, when the objective of the study is to develop a model of the random process, the relative importance of each statistical measure becomes an important consideration, and the simplicity and accuracy of the model depends entirely on the inclusion of appropriate statistical parameters in the model. One approach to circumvent this problem is by means of a preliminary study of most of the statistical parameters, followed by a detailed study of a smaller subset. An alternative approach is to keep a complete record of the noise data, followed by analysis, to provide the relevant parameters. This approach has the advantage of simpler hardware, as compared with the situation when the hardware is employed to measure the statistical parameters directly. If the analysis is to be performed at a later date, then the raw data must be stored. The size of storage medium then becomes an important

consideration in the design of data acquisition systems. The size of storage medium does not impose any constraints on the system design if the data is processed on-line. Portability of the data acquisition system is another important consideration. All of these considerations are highlighted in the description given below of the noise data acquisition and analysis system developed by the author.

The statistical parameters of environmental noise may be measured by

- i) short or long term analogue recording of the noise-envelope on photographic film or magnetic tape
- ii) short or long term recording of the digitized noise-envelope on magnetic tape or disk with on- or off-line processing
- iii) real-time processing of the received noise for the direct accumulation of the noise statistics desired.

4.4.1 Preliminary Noise Measurements

The study of environmental noise was performed in several stages. The preliminary study included a short term photographic recording of the noise transient, and measurement of the rms noise level at a number of recording sites. The noise recording instrumentation is shown in Figure 4.8.

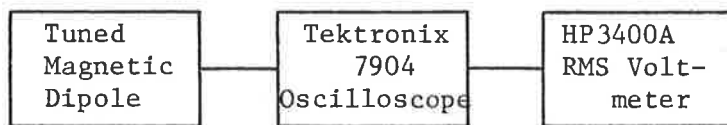


Figure 4.8 Preliminary Study of Environmental Noise

The antenna used was an electrically balanced magnetic dipole of the type proposed for the object identification system. The characteristics of the antenna are described in Chapter 3. The impulsive character of the man-made noise became evident from the display of the noise on the oscilloscope. The impulsive noise resulted in a decaying transient across the antenna tuned circuit, as seen on the oscilloscope. The oscilloscope amplifier was used to amplify the noise before rms noise level was measured. Photographic recording of the man-made noise was performed in a variety of environments, including an automobile parking lot with and without automotive engines running in the vicinity of the

recording set-up, and a range of indoor laboratory situations. One such recording is shown in Figure 4.9. The results of the preliminary study are presented in Table 4.1.

The observations made in this phase of the study directed the course of the next phase. The major finding of the first phase was that the man-made noise in the low-frequency band is impulsive in character, with a large peak-to-rms ratio. The impulsive character of the noise is mainly responsible for errors in a digital communication system.

The second phase of the environmental noise study described below was directed towards obtaining a better understanding of the impulsive behaviour of man-made noise, in order to develop an error-control strategy which utilizes its characteristics.

4.4.2 Noise Data Acquisition Instrumentation

The technique of digital recording of the noise-envelope satisfied the objective of achieving simplicity of data acquisition hardware, while at the same time providing extensive data on which to develop a model for investigation of the characteristics of the man-made impulsive noise. The block schematic of the noise data acquisition system is

Table 4.1 Results of Preliminary Noise Study

ANTENNA ENVIRONMENT	IMPULSE NOISE POWER* dBW	RMS NOISE POWER† dBW
Laboratory C Ambient Noise	-95.6	-115.7
Machines Laboratory	-88.9	-98.0
Universal Logic Operating System	-68.9	-88.9
Operating Calculator Near the Antenna	-67.3	-70.1

* Impulse Noise Power = $V_{pp}^2/8R$
 † RMS Noise Power = V_{rms}^2/R

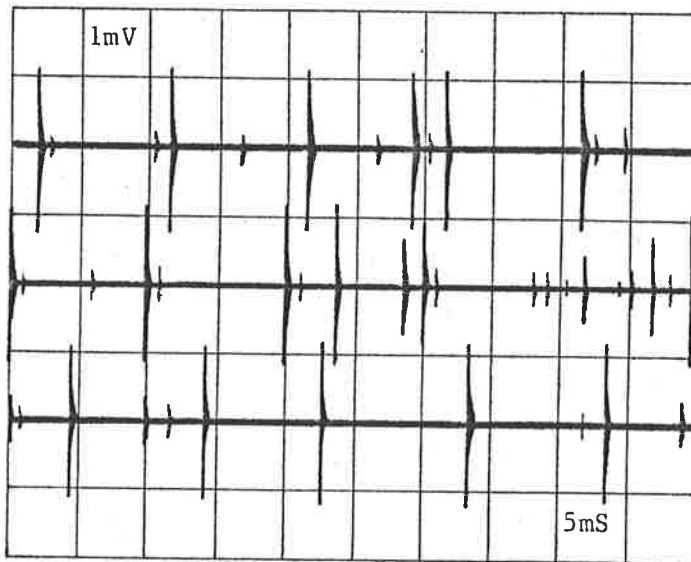


Figure 4.9 A Typical Automotive Noise Transient

shown in Figure 4.10. The antenna used for the measurements was an electrically balanced magnetic dipole of the type proposed for the object identification system. The antenna was tuned to a centre frequency of 50kHz with a quality factor of 8. The output of the antenna was fed to a band-reject filter. The rejection band of the filter was such that noise from outside the band of interest was eliminated. The circuit diagram of the band reject filter is shown in Figure 4.11. Although the parallel tuned circuit of the receiver antenna attenuates most of the signals outside its pass band, the band reject filter is used to attenuate the side bands even further. The output of the band reject filter

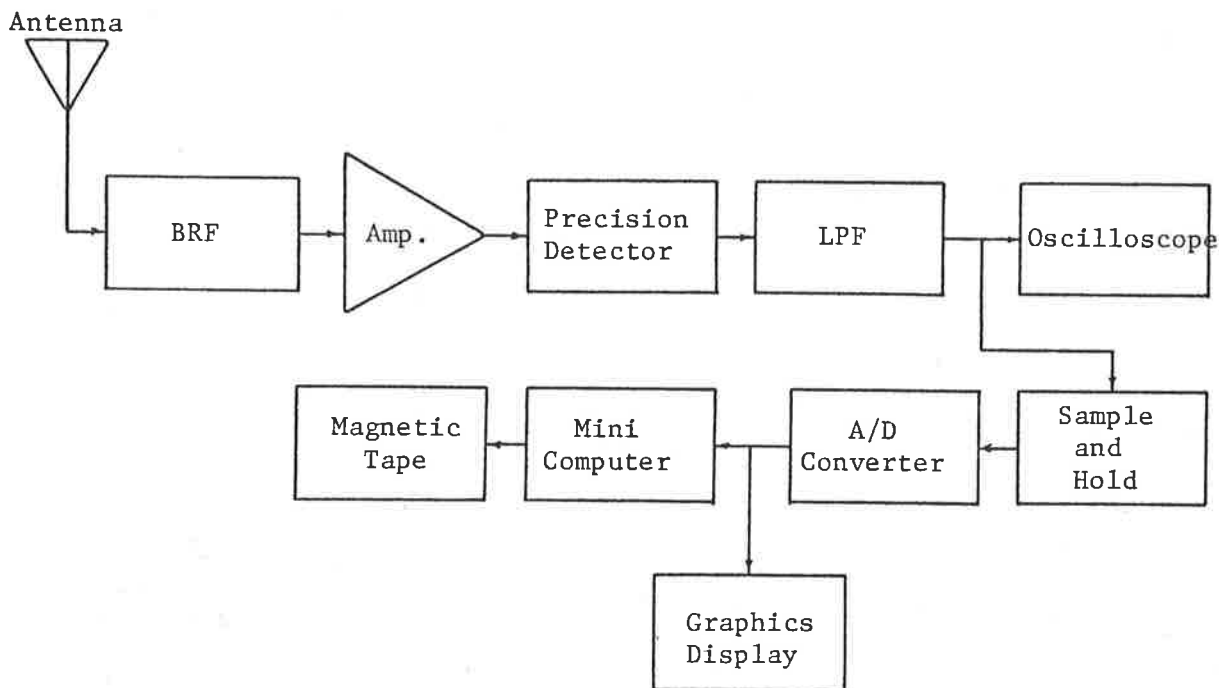


Figure 4.10a Noise Data Acquisition System Schematic

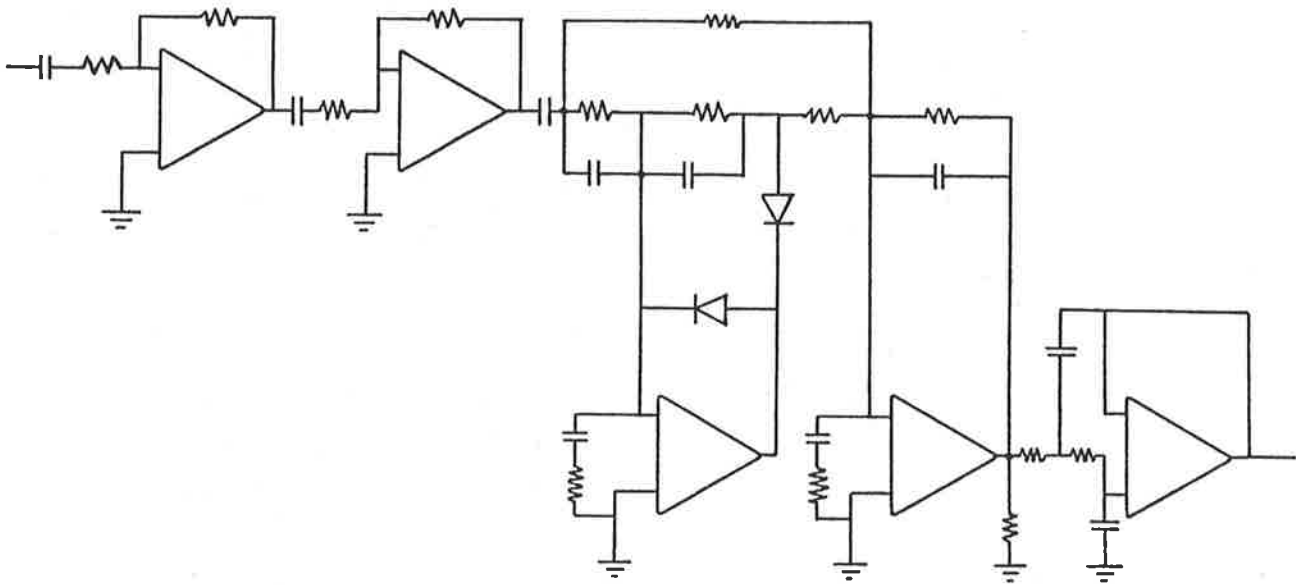
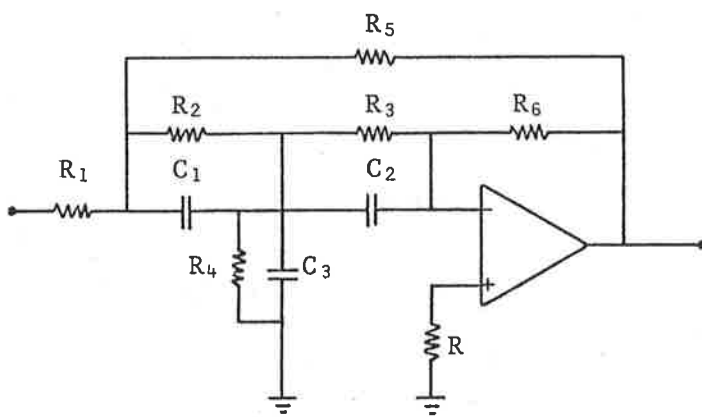


Figure 4.10b Amplifier and Envelope Detector for Noise Data Acquisition System



$$R_2 = R_3 = 2R_4$$

$$C_1 = C_2 = C_3 / 2$$

$$f_0 = (2\pi R_2 C_2)^{-1}$$

Figure 4.11 Band Reject Filter for Noise Data Acquisition System

is sufficiently amplified, in a low noise preamplifier, that the noise signal lies in the linear range of the envelope detector. The circuit configurations of the preamplifier and the envelope detector are shown in Figure 4.10b.

The noise-envelope is then further amplified. The amplified noise-envelope is then fed to a 50Ω coaxial cable to bring it to the analogue-to-digital converter input of the minicomputer. The analogue-to-digital converter board on the minicomputer consists of a sample-and-hold circuit followed by an A-to-D converter. To ensure that the noise-envelope could be reconstructed from the sampled values, the sampling frequency of the A-to-D converter was selected such that it was higher than the Nyquist frequency for the noise-envelope. The 12-bit-output of the A-to-D converter was then stored as two 8-bit words on a magnetic tape. The noise envelope waveform was monitored before and after being digitized, on an oscilloscope and a graphics display respectively. The noise data available on the magnetic tape was then ready for analysis.

The analysis was performed by converting the data to a suitable format for the main-frame computer. The analysis software, discussed in greater detail later in this chapter, used this data for the computation of various noise statistics.

4.4.3 Data Acquisition System Calibration

The measurement of environmental noise at various locations at reasonable distances from the minicomputer was made possible by means of a battery operated detector-amplifier circuit followed by a long coaxial cable. The noise data acquisition system was calibrated before and after noise data recording at each site. It was necessary to ensure that the drift in the amplifier gain was negligible during the period of data accumulation. System calibration was performed by exciting the antenna by a current pulse train of known amplitude. Each pulse excited the antenna into an oscillatory transient, the envelope detector output of such transient being an exponentially decaying curve. Amplitude of this envelope was measured, and the sampled values were stored, and later used for system calibration.

4.4.4 Noise Data Measurements

In view of the possible applications of low-frequency passive transponder systems, environmental noise was measured in a wide variety of indoor and outdoor situations. Among all the sources of man-made impulsive noise, automotive ignition noise was considered more carefully. This was because of the fact that in low frequency identification applications, where the interrogation range is small,

i.e. a few metres, automotive ignition noise affects not only the vehicle identification application, but also some of the other applications, where the presence of vehicles in the vicinity of the interrogation apparatus cannot be avoided.

Outdoor environments from which noise data was gathered include roadway surfaces with underground cables, and a parking lot with and without engines running near the recording site. Roadway traffic conditions were simulated by running several stationary vehicles at speeds corresponding to idle and cruise. The indoor environments studied included a minicomputer laboratory, a machines laboratory, the vicinity of an operating electronic calculator, and that of other household appliances. These measurement conditions are listed in Table 4.2.

Statistical parameters evaluated from the noise data which provide useful information in the calculation of system error performance are described in Section 4.5.

Table 4.2 Noise Measurement Environments

Indoor:

Universal Logic Operating System

Minicomputer Laboratory

Machines Laboratory

Vicinity of Operating Electronic Calculator

Switching Power Supply

Outdoor:

Buried Power Cables

Single Vehicle at Different Speeds

Multiple Vehicles at Idling Speed

4.5 Noise Analysis

The noise data accumulated at various test sites was stored on a magnetic tape. The data was initially checked on the NOVA2 minicomputer for missing end-of-record markers or calibration test results. The 12-bit data words were converted to 64-bit words for use on the CYBER173 computer. First of all, Allan Variance Analysis was performed on the raw data to find the amount of data required for statistically meaningful results. The software for the data analysis was then used to calculate the Noise Amplitude Distribution and the Time Probability Distribution for various amplitude levels. The flow-charts for the calculations of AVA, NAD, and TPD are given in Figures 4.12, 4.13 and 4.14.

Although the sampling frequency was higher than the Nyquist frequency, it was still not possible to sample all the envelope peaks. This problem may be overcome, either by increasing the sampling frequency, or by sampling a larger record length. Both of these solutions result in a large amount of raw data, which increases the memory requirement as well as computing time for analysis. Statistically accurate results for higher amplitude levels were obtained from the original data by extrapolating the position and amplitude of the envelope peaks from the known circuit parameters of the

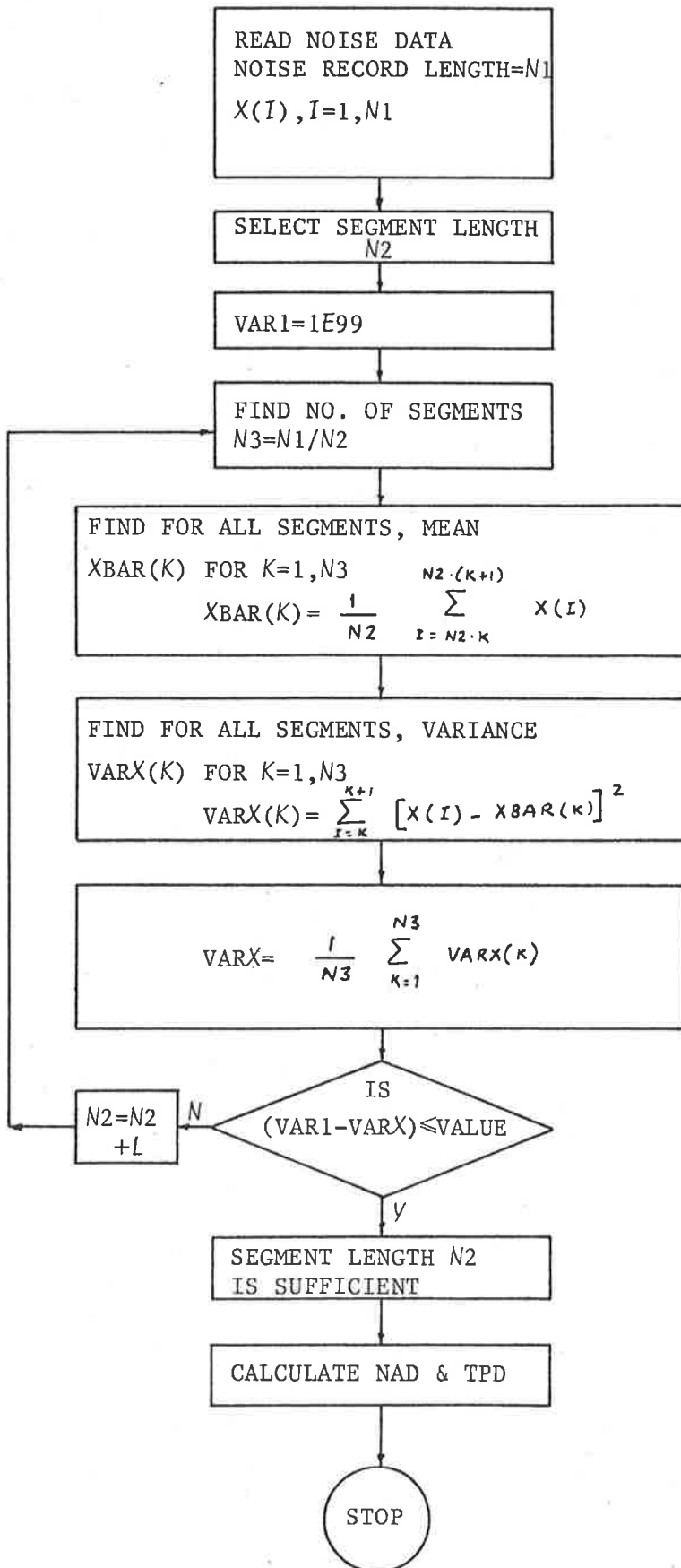


Figure 4.12 AVA Flow Chart

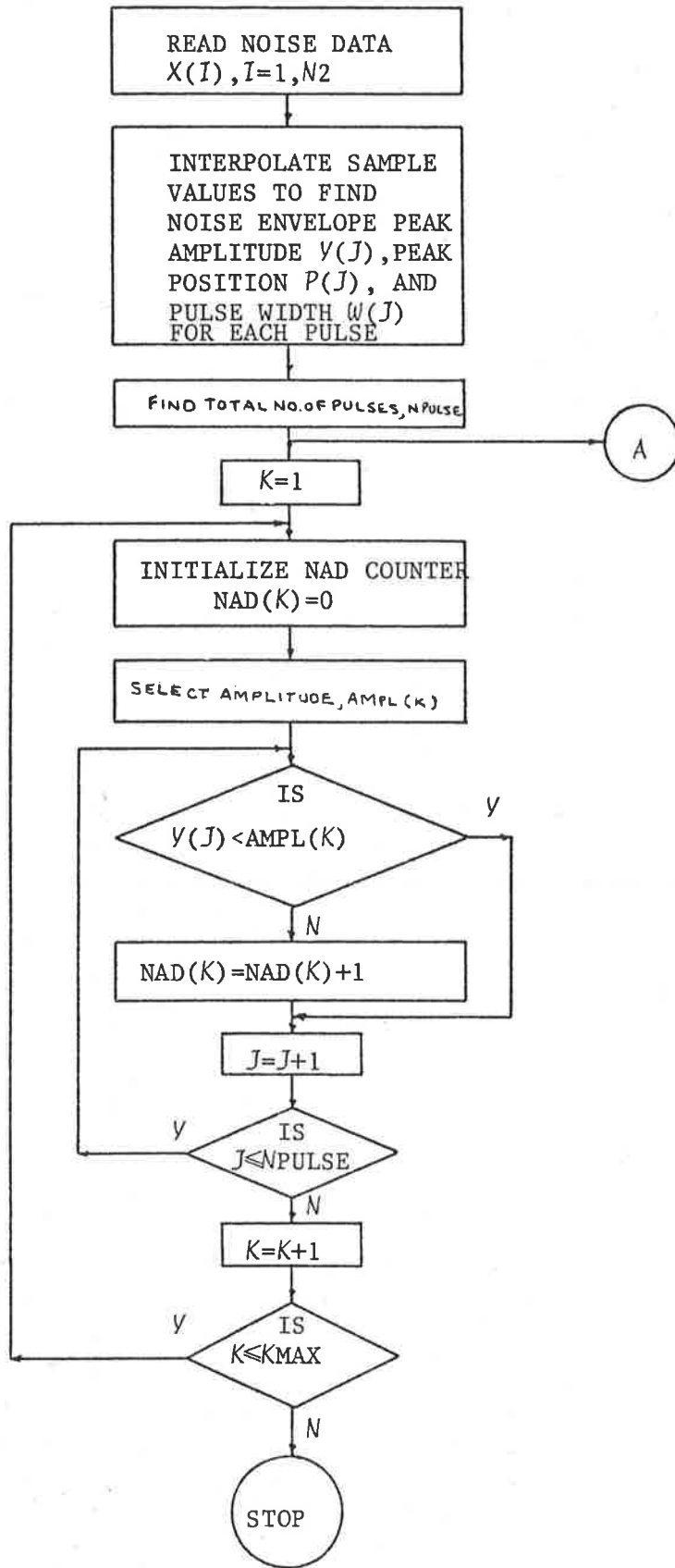


Figure 4.13 NAD Flow Chart

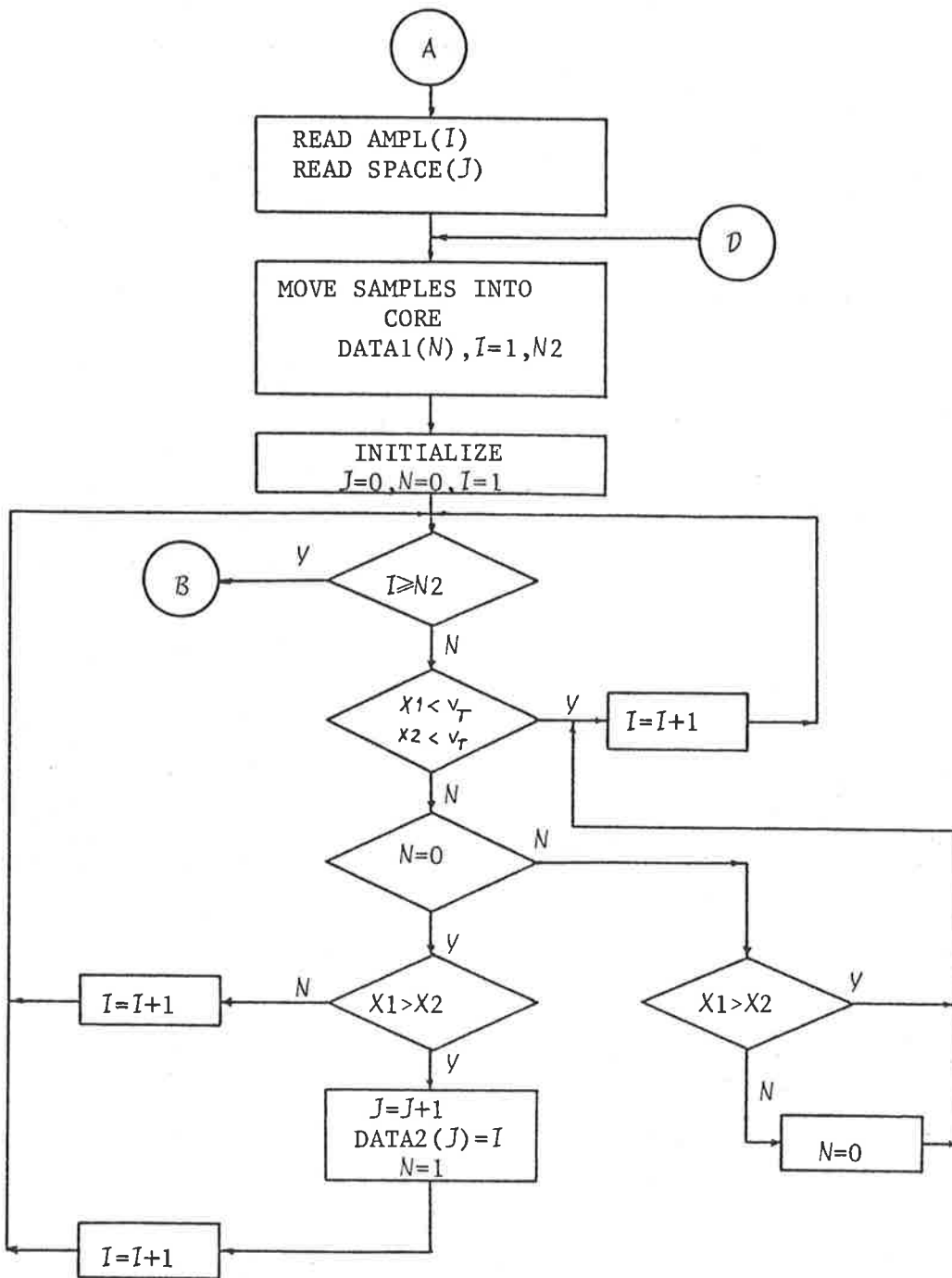
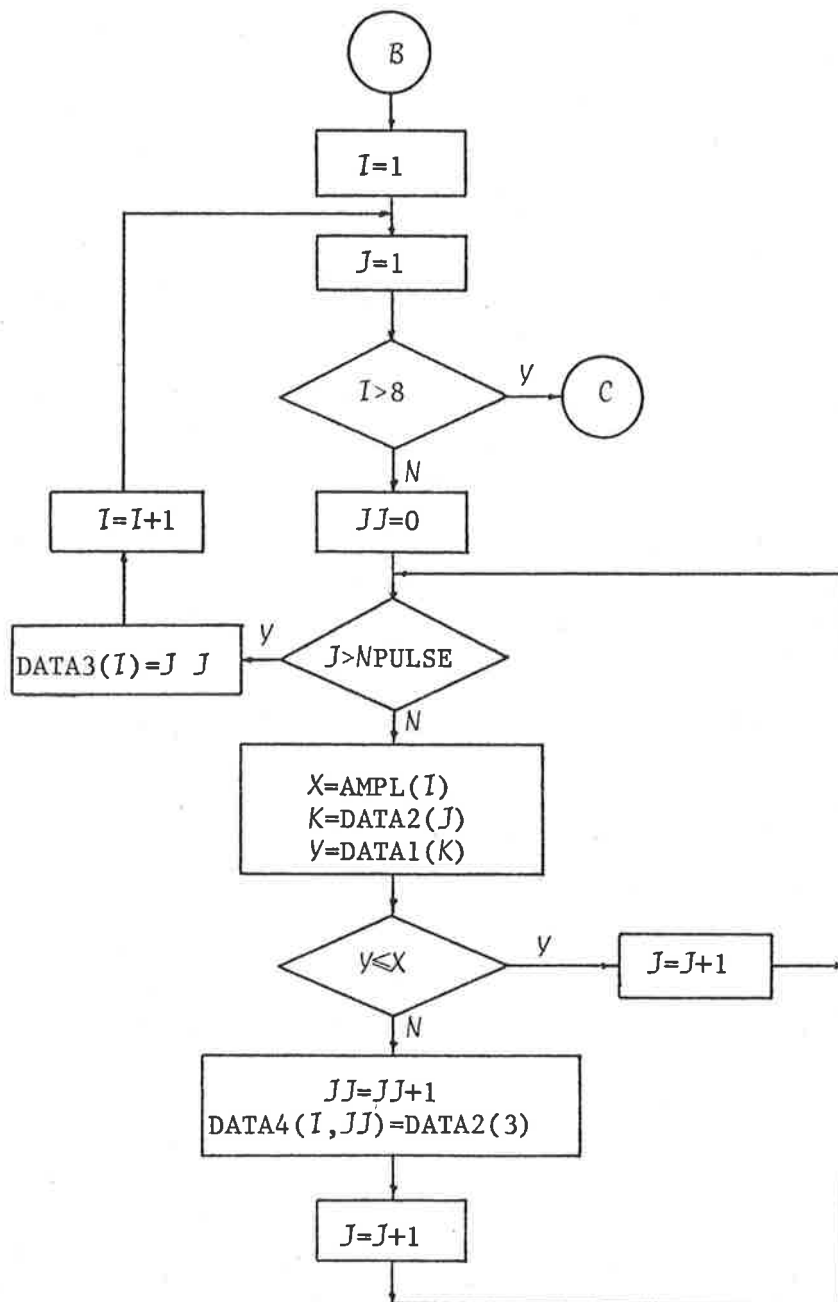
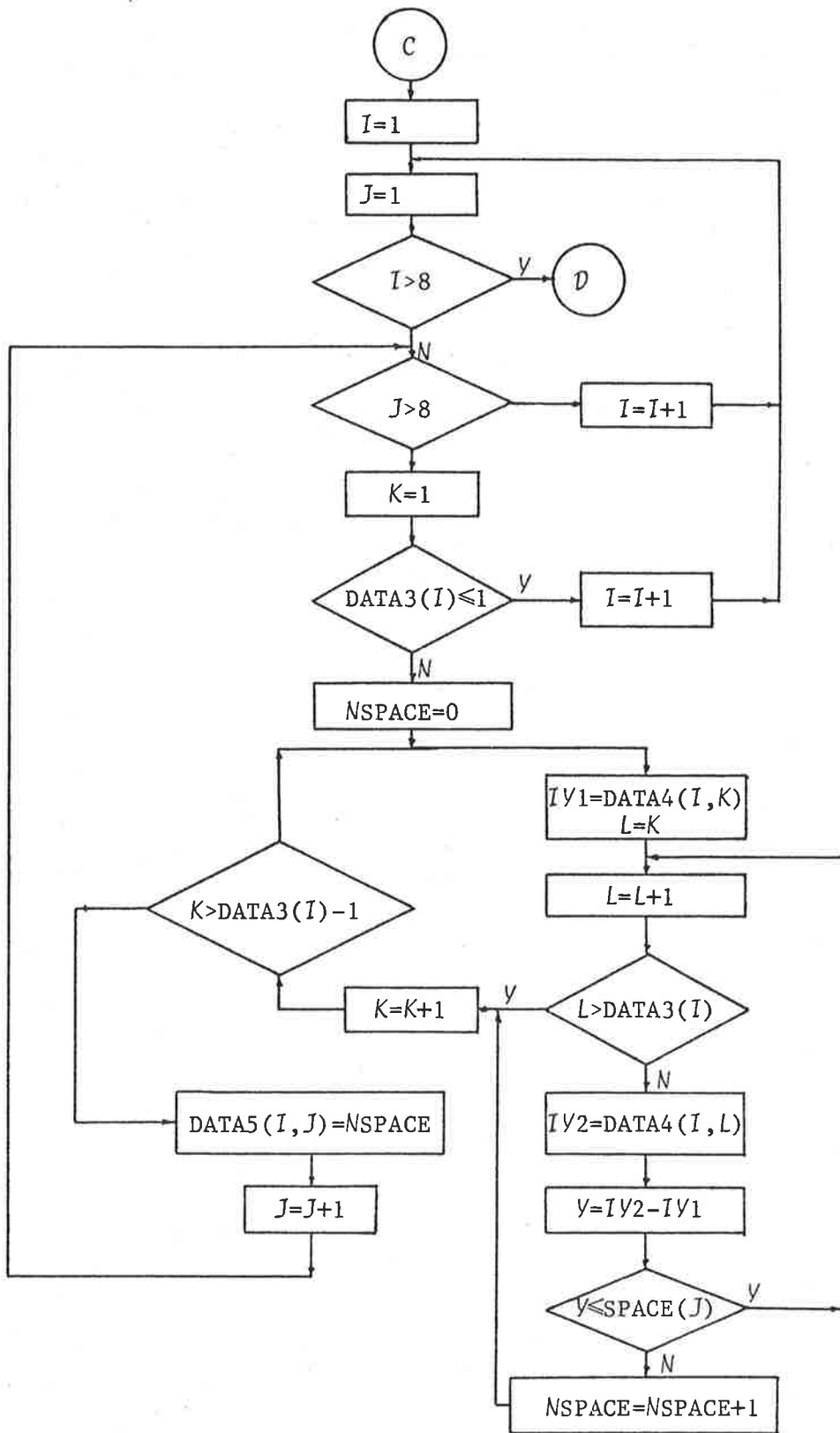


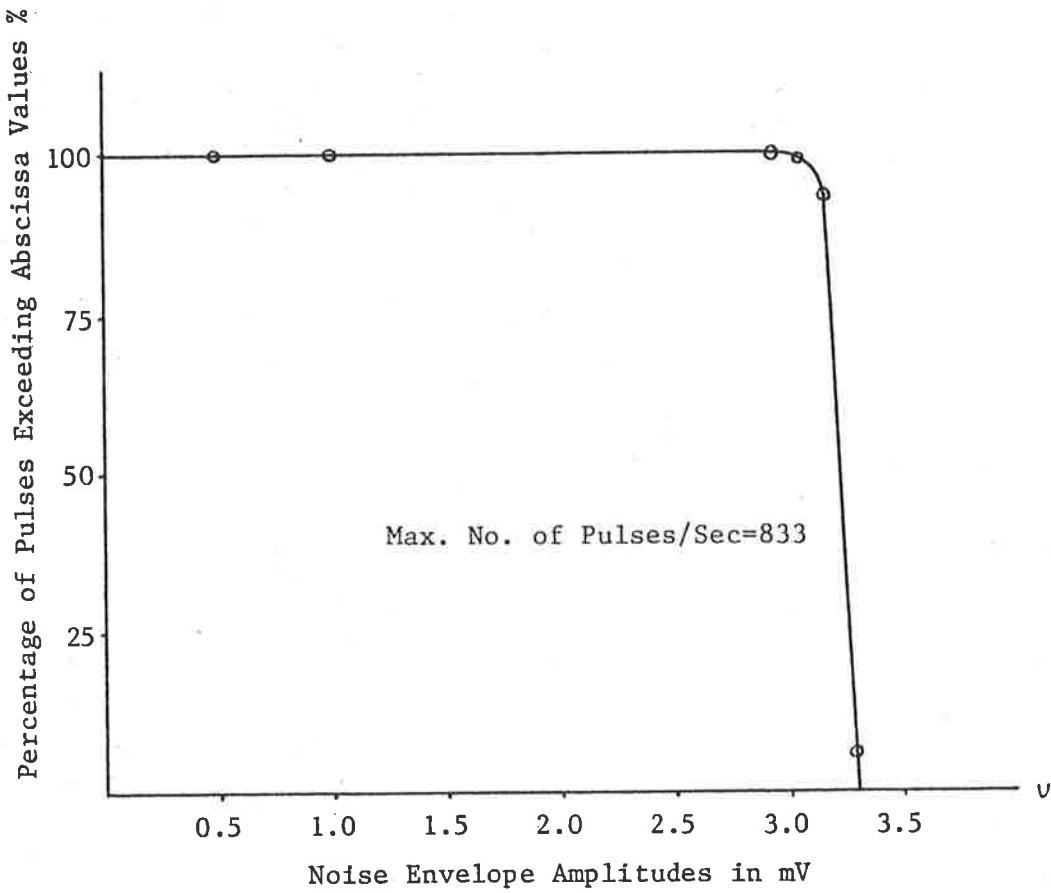
Figure 4.14 TPD Flow Chart



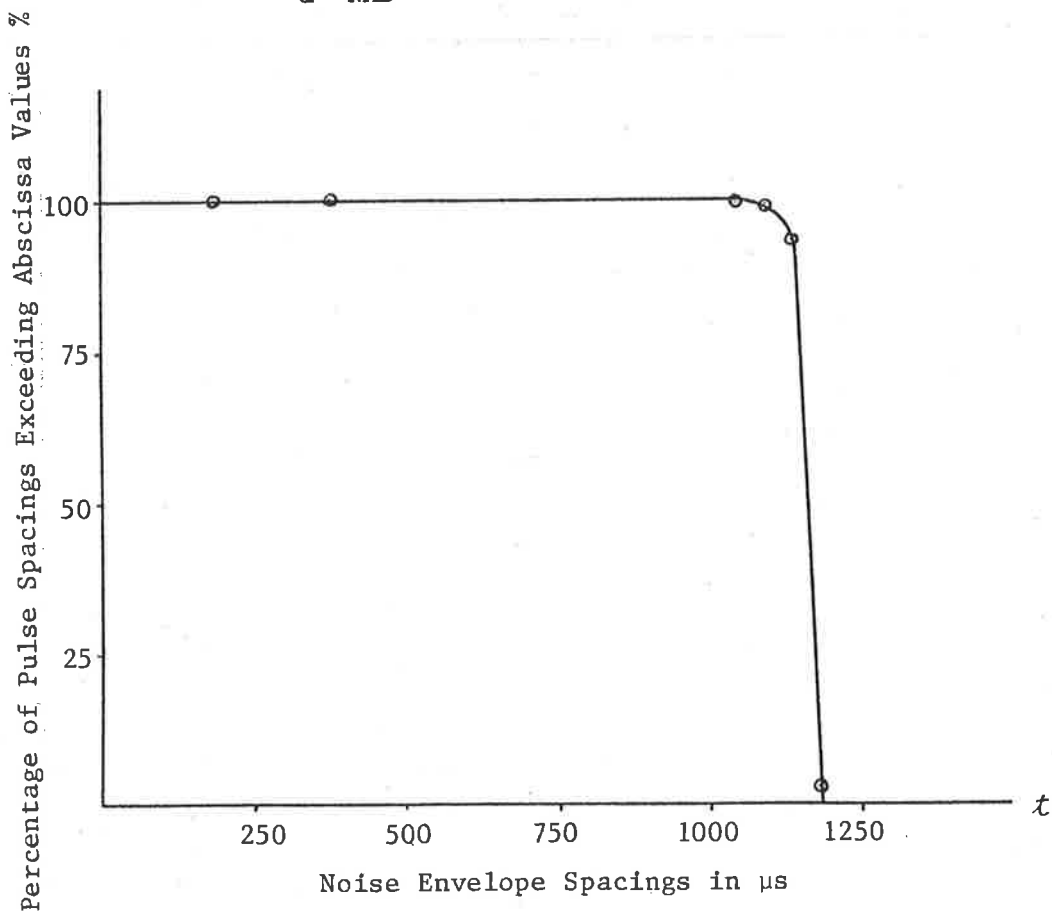


antenna. The NAD and the TPD plots for a periodic pulse train are shown in Figure 4.15. The known characteristics of the NAD and TPD curves for the periodic data were used as a debugging tool for the noise analysis software. The NAD plot is flat up to a certain pulse amplitude, and then falls off very rapidly. The pulses in the pulse train have a uniform amplitude and pulse spacing. For test amplitudes lower than the pulse amplitude, all the pulses pass the test, and for test amplitudes higher than the pulse amplitude, none of the pulses pass the test. Similarly for the TPD: an inverted "L" shaped curve is obtained for all the amplitude levels.

The noise analysis software developed using the periodic data is then used for the calculation of NAD and TPD for environmental noise data. The noise measurement set-up for vehicle ignition noise measurement is shown in Figure 4.16. The ignition noise was measured in two locations, for a six-cylinder and a V-8 vehicle, at speeds corresponding to idle and cruise. The noise from underground power-line cables was also picked up at one of the two locations. The NAD and TPD for automotive ignition noise under the various conditions listed in Table 4.2 are plotted in Figure 4.17.



a NAD



b. TPD

Figure 4.15 NAD and TPD Plots for a Periodic Signal



Figure 4.16a Single Vehicle Noise Measurement Set-up



Figure 4.16b Multiple Vehicles Noise Measurement Set-up

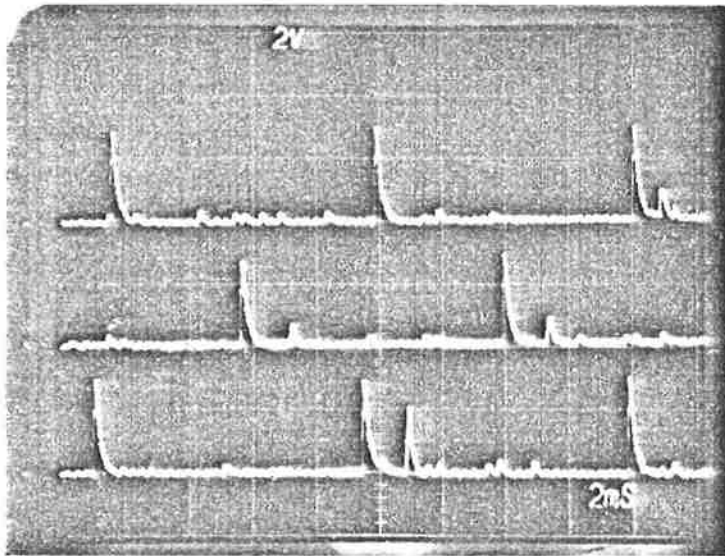


Figure 4.16c 6-Cylinder Vehicle Ignition Noise at Idling Speed

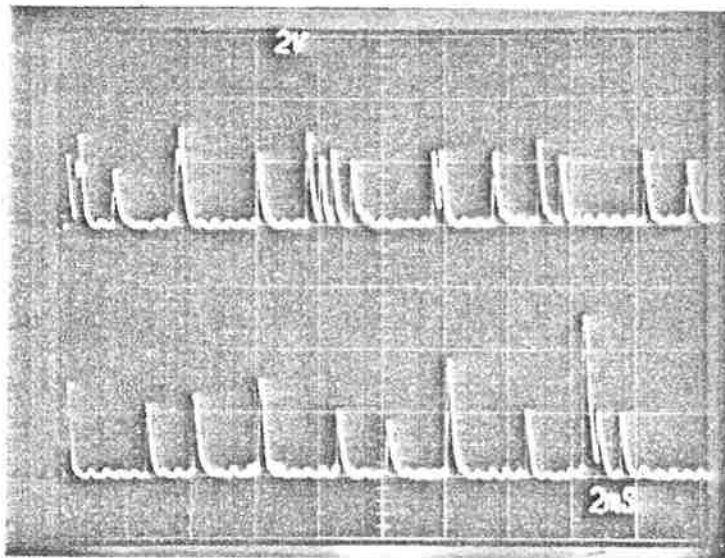


Figure 4.16d V-8 Vehicle Ignition Noise at Cruising Speed

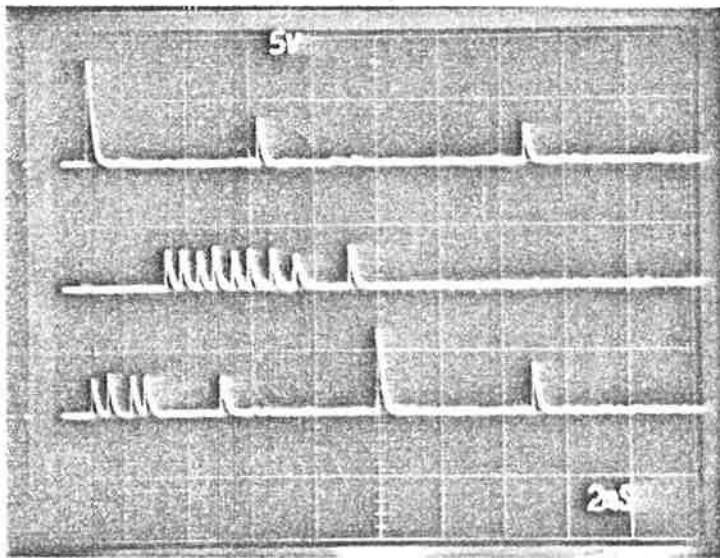
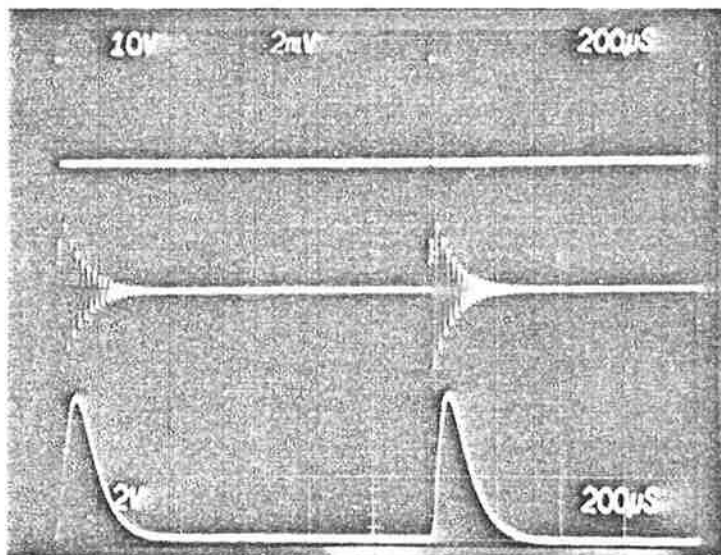


Figure 4.16e ULOS Computer Noise Record



Impulse exciting
wave form.

Antenna response.

Envelope detector
output.

Figure 4.16f Periodic Signal for Noise Data
Acquisition System Calibration

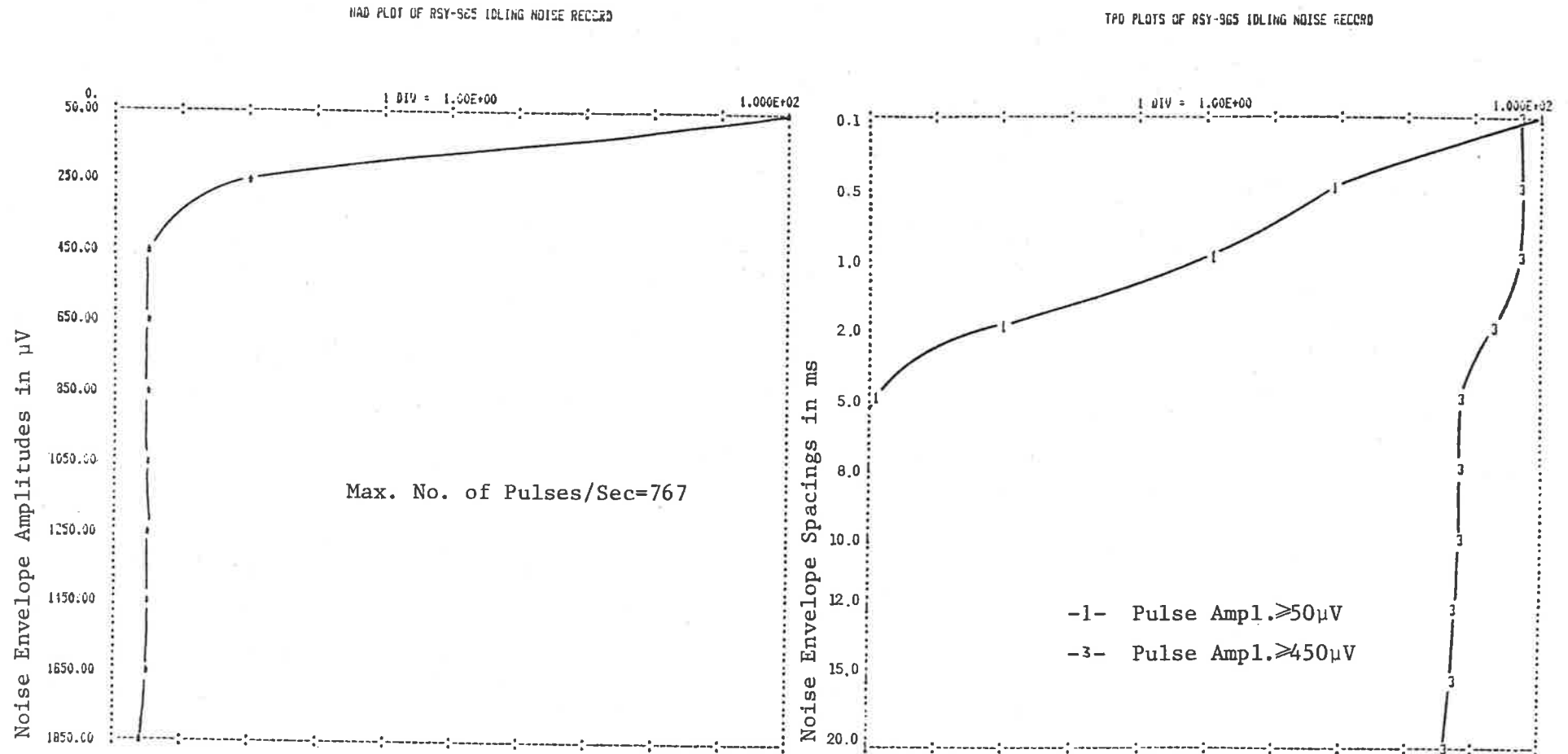
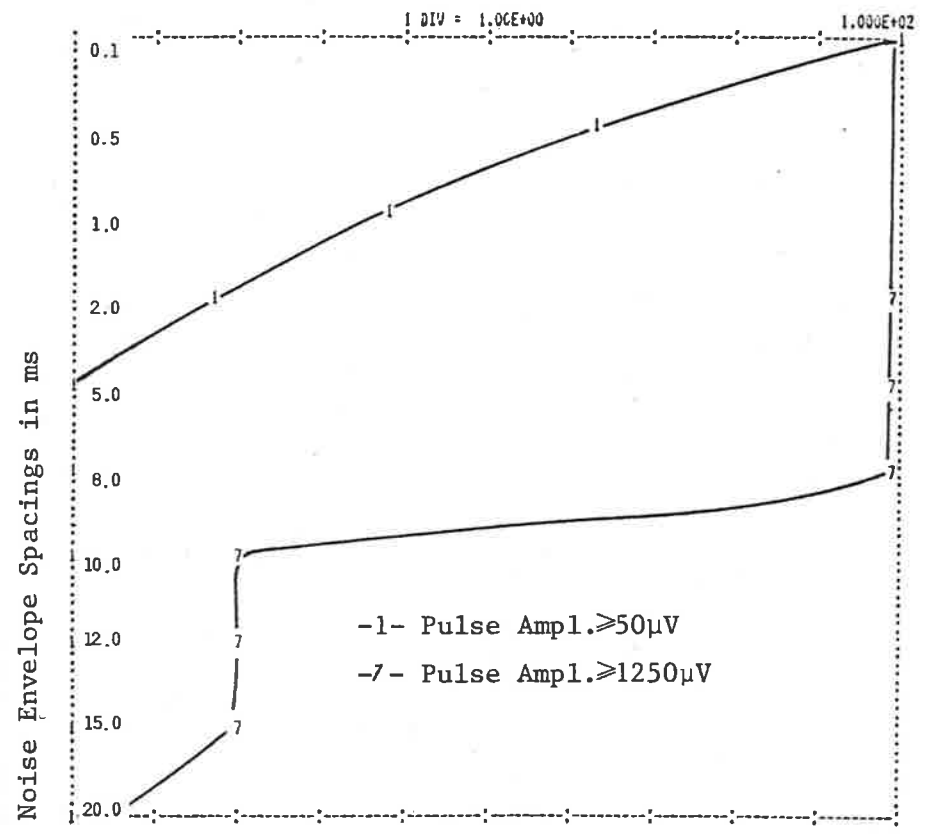
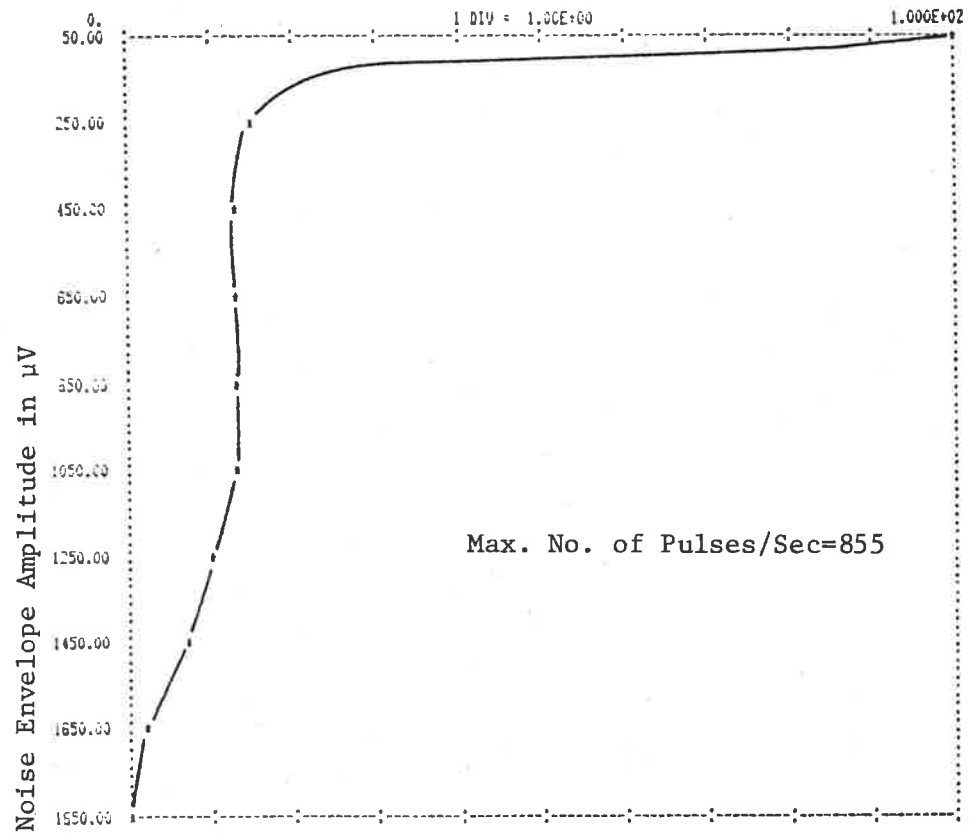


Figure 4.17a Statistical Study of 6-Cylinder Vehicle Noise at Idling Speed

NAD PLOT OF R57-965 REVVED NOISE RECORD

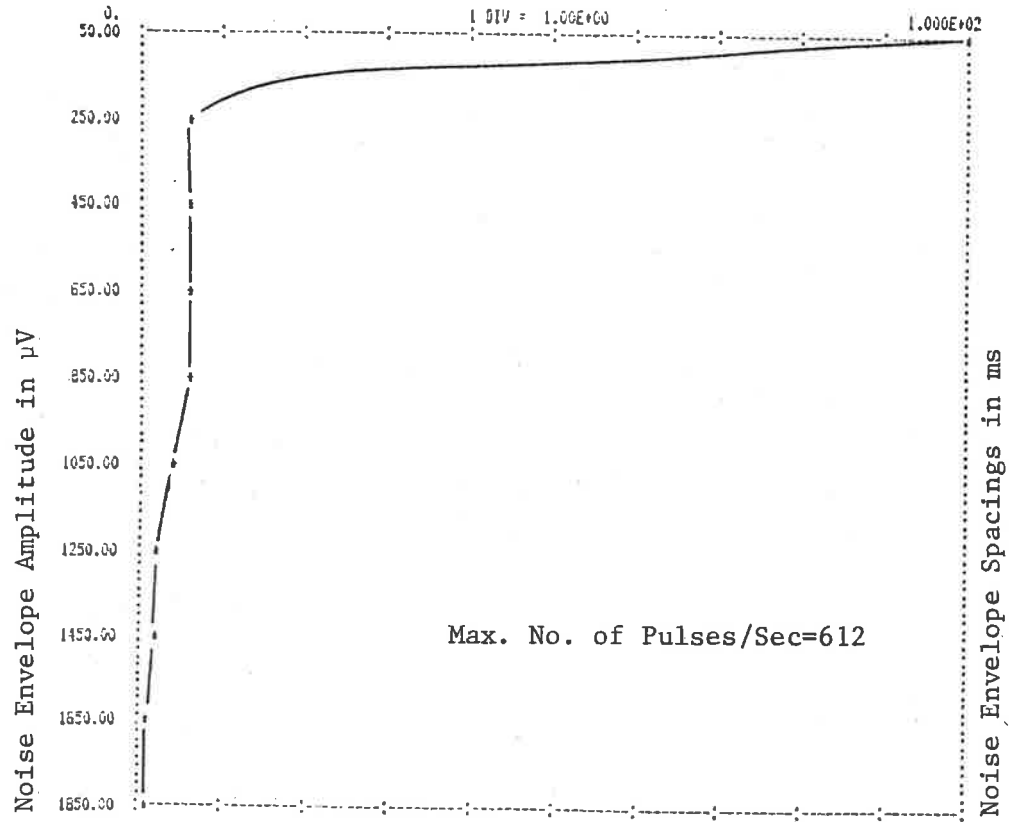
TPD PLOTS OF R57-965 REVVED NOISE RECORD



4.57

Figure 4.17b Statistical Study of 6-Cylinder Vehicle at Cruising Speed

NAD PLOT OF SNN-147 IDLING NOISE RECORD



TPD PLOTS OF SNN-147 IDLING NOISE RECORD

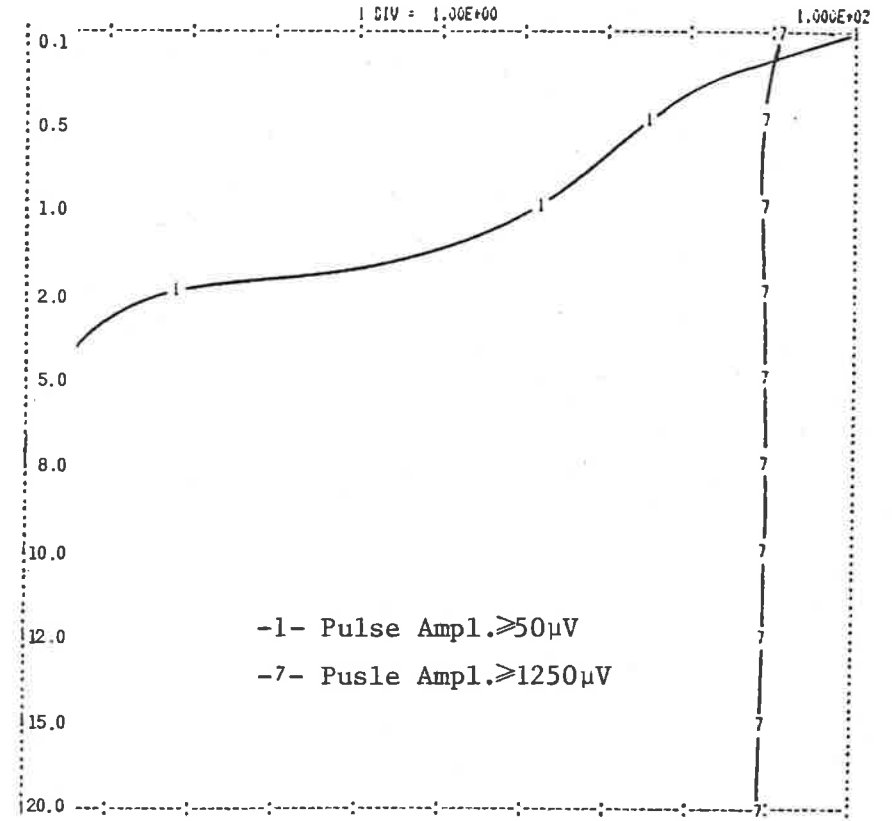


Figure 4.17c Statistical Study of V-8 Vehicle Noise at Idling Speed

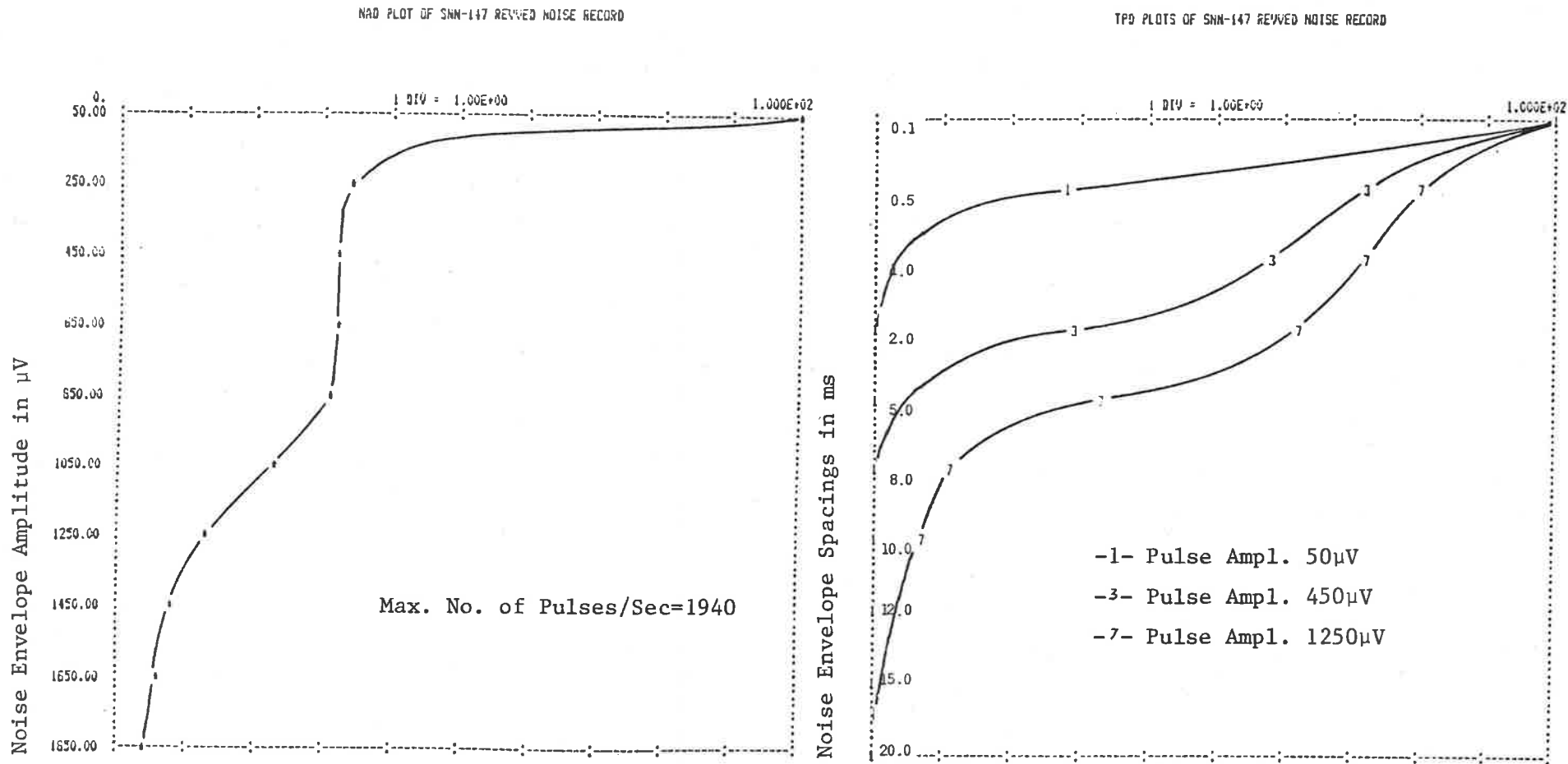
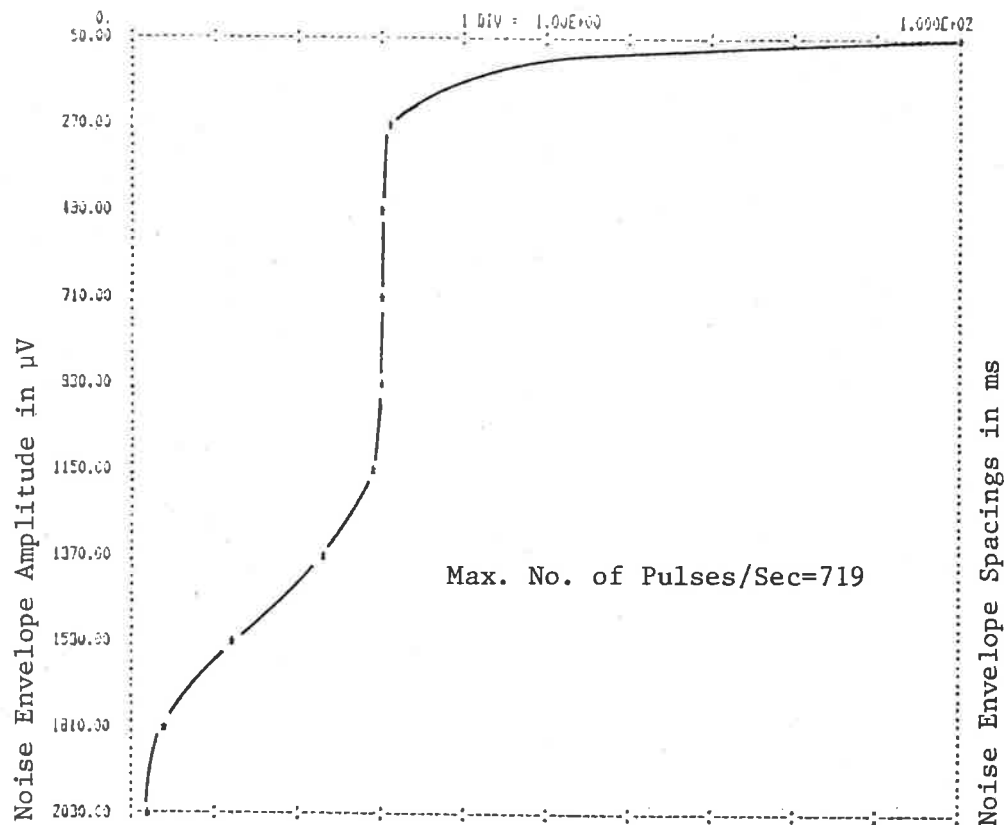
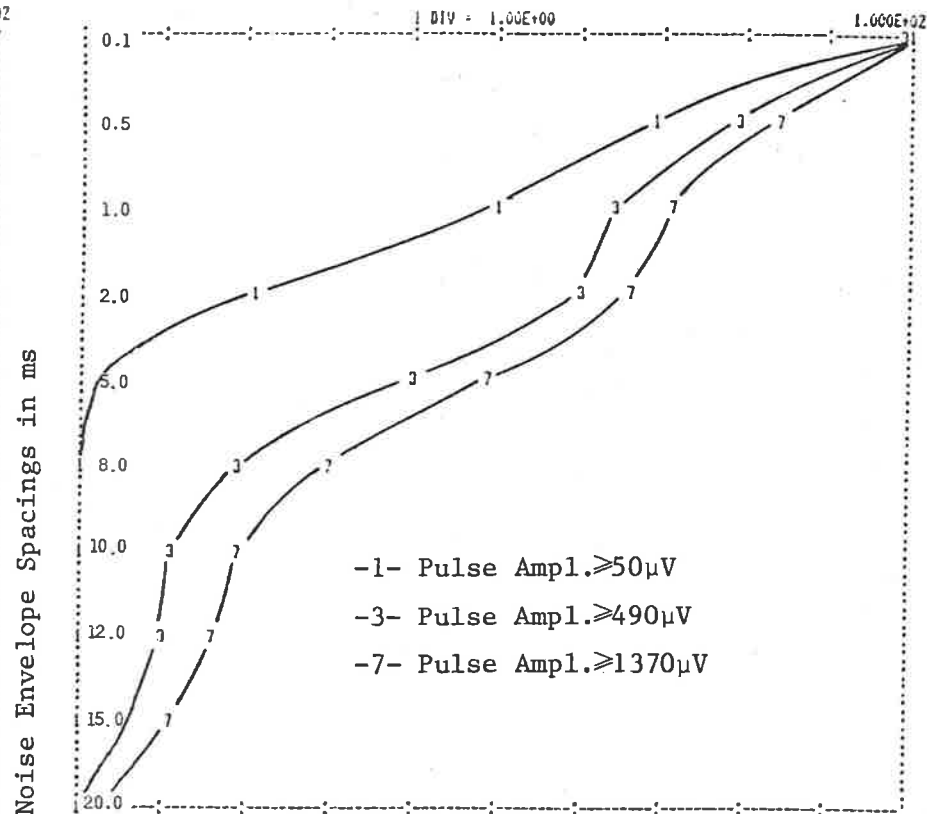


Figure 4.17d Statistical Study of V-8 Vehicle Noise at Cruising Speed

NAD PLOT OF MULTIPLE VEHICLE IDLING NOISE RECORD



TPD PLOTS OF MULTIPLE VEHICLE IDLING NOISE RECORD



4.60

Figure 4.17e Statistical Study of Multiple Vehicles Noise at Idling Speed

MAD PLOT OF ULOS COMPUTER NOISE RECORD

TPD PLOTS OF ULOS COMPUTER NOISE RECORD

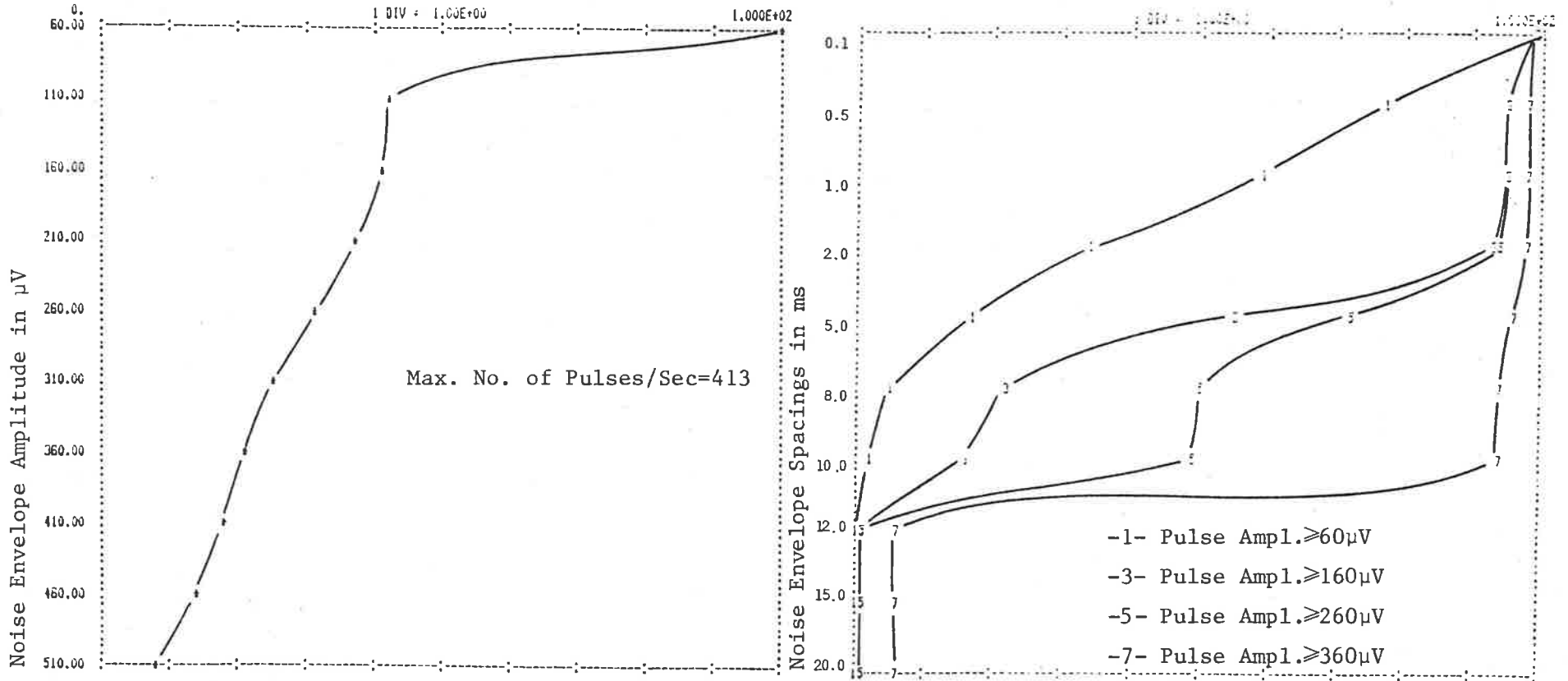


Figure 4.17f Statistical Study of ULOS Computer Noise Record

The antenna was placed under the centre vehicle, as shown in Figure 4.16, to simulate the roadway traffic conditions.

One common feature of most of the NAD curves is that these curves have two distinct slopes. This characteristic of the NAD indicates that the noise emanates from two different sources. This observation is not surprising, in view of the fact that no special care was taken to inhibit the noise from the sources other than the one for which the measurement is intended. However, in a near-field coupling situation such as this, the contribution from sources away from the measurement site is expected to be negligible. This assumption holds for all the noise sources. The location of underground power cables is such that, at the recording site, within easy reach from the minicomputer, it is impossible to completely inhibit the power-line noise from reaching the measurement instrumentation. A careful selection of the noise recording sites reduced the contribution from power-lines, but did not eliminate it completely. The shapes of the NAD plots suggest that the noise consists of a large number of low amplitude randomly occurring pulses superimposed with a small number of large amplitude pulses. The plateauing of TPD at higher amplitudes indicates that the large amplitude pulses have a

contribution from a periodic source. The absence of plateauing for the low amplitude TPD implies that low level pulses are random. A small negative slope for the high amplitude TPD indicates that the higher amplitude pulses have a contribution from a random and a periodic source. While the high amplitude random pulses emanate from the source under study, the periodic pulses emanate from nearby powerlines. The periodicity of these pulses is related to the switching period of a full-wave rectifier.

Apart from the NAD and TPD of the noise, some of the other parameters of the random process were also evaluated. These terms are defined below.

The receiver threshold level, V_T , is defined as the level below which the signal output of the receiver cannot be distinguished from the receiver noise output when all extraneous sources of noise are absent. All noise samples below the threshold V_T are discarded.

The peak noise amplitude, V_p , is defined as the average of all the noise-envelope peaks encountered during the observation period.

$$V_p = \frac{1}{N_p} \sum_{n=1}^{N_p} V_p(n) \quad (4.17)$$

where

N_p is the total number of noise pulses encountered exceeding the level V_T during the observation period
 $V_p(n)$ is the peak amplitude of the n th noise-envelope

The rms noise amplitude, V_r , is defined as the rms value of all the noise samples

$$V_r = \left[\frac{1}{N} \sum_{n=1}^N V_n^2 \right]^{1/2} \quad (4.18)$$

where

N is the total number of noise samples during the observation period above the level V_T

V_n is the amplitude of the n th noise sample

The average noise amplitude, V_a , is defined as the average value of all the noise samples

$$V_a = \frac{1}{N} \sum_{n=1}^N V_n \quad (4.19)$$

For stationary random processes, Allan Variance Analysis ensures that the time average of the analogue noise signal is the same as the ensemble average of the noise samples over the observation

period. Some authors (Bolton, 1976) have used the ratio of rms noise level V_r , and average noise level V_a , as a measure of the impulsiveness of the noise

$$V_d = 20 \log_{10} \left(\frac{V_r}{V_a} \right) \quad (4.20)$$

An alternative measure of the impulsiveness of a random process is defined here as the ratio of peak noise level to rms noise level

$$V_e = 20 \log_{10} \left(\frac{V_p}{V_r} \right) \quad (4.21)$$

The peak, rms and average noise levels, and the respective figures for impulsiveness as defined in Equations (4.20) and (4.21) for the man-made noise environments given in Table 4.2 are listed below in Table 4.3.

Table 4.3 Comparison of Noise from Various Sources

NOISE PARAMETER	V_a	V_r	V_p	V_d	V_e	N_{max}^*
NOISE SOURCE	μV	μV	μV	dB	dB	sec ⁻¹
6-Cylinder Vehicle at Idling Speed	45.30	185.34	255.73	12.24	2.80	767
6-Cylinder Vehicle at Cruising Speed	68.99	204.94	279.95	9.46	2.71	855
V-8 Vehicle at Idling Speed	23.45	120.83	162.66	14.24	2.58	612
V-8 Vehicle at Cruising Speed	212.09	448.46	484.36	6.50	0.67	1940
Multiple Vehicles at Idling Speed	82.87	291.72	538.26	10.93	5.32	719
ULOS: Computer Lab. - Interactive Studio	50.86	75.27	195.21	3.41	8.28	413

* N_{max} is the maximum number of pulses occurring per second.

4.6 Error Analysis

The error performance of a communication system depends on the characteristics of the environmental noise and the parameters of the communication system. The parameters which affect the error performance of the communication system include transmitter power, type of communication link, modulation parameters, receiver bandwidth, antenna sizes, and a number of less critical parameters. The effect of noise on the communication system is determined in terms of the degradation in system performance for a fixed set of system parameters with the noise being the only variable. The prediction of bit error rate of a communication system in the presence of impulsive noise requires an understanding of the mechanism responsible for causing errors. Attention is focussed, in this section, on developing an understanding of the noise mechanisms and analytical formulae for the prediction of error rates, while error control strategies to minimize the bit error rate suitable for the passive subharmonic transponder system are discussed in Chapter 5.

Consider the receiver antenna model shown in Figure 4.18. The transfer function $H(s)$ of this antenna is given by Equation (4.22).

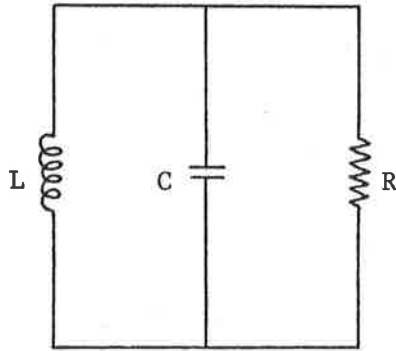


Figure 4.18 Receiver Antenna Model

which expresses the ratio of voltage delivered to the load to a current injected in parallel with the circuit.

$$H(s) = \frac{Ks}{s^2 + 2bs + \omega_2^2} \quad (4.22)$$

where

K is a constant constant

ω_2 is the frequency of resonance

$2b$ is the rf bandwidth

The impulse response of the receiver antenna is therefore written as

$$h(t) = K \cdot \exp(-bt) \left[\cos \omega t - \frac{b}{\omega} \sin \omega t \right] \quad (4.23)$$

where

$$\omega^2 = \omega_2^2 - b^2$$

If the antenna has reasonably high quality factor, then

$$\begin{aligned} \omega &\doteq \omega_2 \\ \text{and } \frac{b}{\omega} &\ll 1 \end{aligned} \quad (4.24)$$

The impulse response, therefore, takes the form

$$h(t) = K \cdot \exp(-bt) \cos \omega t \quad (4.25)$$

Now, let us consider an impulse applied to the antenna input at time α with respect to an arbitrary reference frame. The output of the receiver antenna due to this impulse is

$$v(t) = h(t-\alpha) = K \cdot \exp[-b(t-\alpha)] \cdot \cos[\omega(t-\alpha)] \quad (4.26)$$

If synchronous detection is employed, then this signal may be expressed in terms of its in-phase and quadrature components:

$$v(t) = x(t) \cos \omega_c t - y(t) \sin \omega_c t \quad (4.27)$$

and $x(t)$ is the output of the synchronous detector due to the impulse appearing at the receiver antenna input. The in-phase and quadrature components in Equation (4.27) may be expressed as

$$x(t) = K \cdot \exp[-b(t-\alpha)] \cdot \cos[\omega(t-\alpha) - \omega_c t] \quad (4.28a)$$

$$y(t) = K \cdot \exp[-b(t-\alpha)] \cdot \sin[\omega(t-\alpha) - \omega_c t] \quad (4.28b)$$

In most cases it is reasonable to assume $\omega \approx \omega_c$;
then,

$$x(t) = K \cdot \exp[-b(t-\alpha)] \cdot \cos \omega t \quad (4.29a)$$

$$y(t) = K \cdot \exp[-b(t-\alpha)] \cdot \sin \omega t \quad (4.29b)$$

The output envelope is then given by

$$r(t) = K \cdot \exp[-b(t-\alpha)] \quad (4.30)$$

Let the impulses be independent, i.e. occur at random times and have an amplitude A with probability density function $p(A)$. Let the impulses occur at a rate λ per second. The probability of occurrence of N impulses in T seconds is then given by the Poisson distribution

$$P(N) = \frac{(\lambda T)^N \exp(-\lambda T)}{N!} \quad (4.31)$$

Consider a small time interval $(\alpha, \alpha + d\alpha)$. Probability of presence $P(1)$ and absence $P(0)$ of an impulse in this interval is then given by

$$P(1) = \lambda d\alpha \cdot \exp(-\lambda d\alpha) \quad (4.32a)$$

$$P(0) = 1 - \lambda d\alpha \cdot \exp(-\lambda d\alpha) \quad (4.32b)$$

The probabilities expressed by Equation (4.32) above may be simplified by using the series expansion of the exponential term for small values of $d\alpha$

$$\exp(-x) = 1 - x + \frac{x^2}{2!} + \dots + (-1)^n \frac{x^n}{n!} + \dots \quad (4.33)$$

$$P(1) \approx \lambda d\alpha \quad (4.34a)$$

$$\text{and } P(0) \approx 1 - \lambda d\alpha \quad (4.34b)$$

The contribution of an impulse during the incremental interval $d\alpha$ to $x(t)$ is

$$\begin{aligned} \Delta x(t) &= KA \cdot \exp[-b(t-\alpha)] \cos \omega\alpha && \text{;if impulse present} \\ &= 0 && \text{;if impulse absent} \end{aligned} \quad (4.35)$$

The characteristic function $\Phi_x(\mu)$ of a random variable x with probability density function $p(x)$ is:

$$\Phi_x(\mu) = \int_{-\infty}^{\infty} p(x) \exp(j\mu x) dx \quad (4.36)$$

Using the above definition, the characteristic function $\Phi_{\Delta x}(\mu)$ may be written as

$$\Phi_{\Delta x}(\mu) = \lambda d\alpha \cdot \Phi_A [K\mu \cdot \exp\{-b(t-\alpha)\} \cos \omega\alpha] + (1 - \lambda d\alpha) \quad (4.37)$$

where

$$\Phi_A(\mu) = \int_{-\infty}^{\infty} p(A) \exp(j\mu A) dA \quad (4.38)$$

Let

$$v = K\mu \cdot \exp\{-b(t-\alpha)\} \cos \omega\alpha \quad (4.39)$$

Combining Equations (4.37) and (4.39) we get

$$\Phi_{\Delta x}(\mu) = 1 + \lambda d\alpha [\Phi_A(v) - 1] \quad (4.40)$$

For small $d\alpha$, using the series expansion of exponential function given in Equation (4.33) we get

$$\Phi_{\Delta x}(\mu) = \exp[\lambda d\alpha \{\Phi_A(v) - 1\}] \quad (4.41)$$

The total contribution to $x(t)$ of all the impulses is calculated by summing over the incremental intervals $d\alpha$. Because of the earlier assumption that impulses occur at random times, the impulses in these intervals will be independent. The characteristic function for the total contribution $\Phi_x(\mu)$ is therefore the product of the characteristic functions $\Phi_{\Delta x}(\mu)$ for all values of $d\alpha$.

$$\text{Let } \Psi(\mu) = \log_e [\Phi(\mu)] \quad (4.42)$$

Then

$$\Psi_{\Delta x}(\mu) = \lambda d\alpha [\Phi_A(v) - 1] \quad (4.43)$$

$$\therefore \Psi_x(\mu) = \sum_{\text{all } d\alpha} \lambda d\alpha [\Phi_A(v) - 1] \quad (4.44)$$

$$= \int_{\alpha=-\infty}^t \lambda \left[\Phi_A [K\mu \cdot \exp\{-b(t-\alpha)\} \cos \omega\alpha] - 1 \right] d\alpha \quad (4.45)$$

Substituting $t - \alpha = \tau$

$$\Psi_X(\mu) = \int_{\tau=0}^{\infty} \lambda [\Phi_A \{K\mu \cdot \exp(-b\tau) \cos W(t-\tau)\} - 1] d\tau$$

Now suppose that A is a Gaussian random variable with variance σ^2 , then the characteristic function $\Phi_A(\mu)$ is

$$\Phi_A(\mu) = \exp\left[-\frac{\mu^2 \sigma^2}{2}\right] \quad (4.47)$$

$$\therefore \Psi_X(\mu) = \int_0^{\infty} \lambda \left[\exp\left\{-\frac{\mu^2 \sigma^2}{2} K^2 \exp(-2b\tau) \cos^2 W(t-\tau)\right\} - 1 \right] d\tau \quad (4.48)$$

Using the series expansion of exponential and neglecting the higher order terms, we get

$$\Psi_X(\mu) = \int_0^{\infty} -\frac{\lambda K^2 \mu^2 \sigma^2}{2} \exp(-2b\tau) \cos^2 W(t-\tau) d\tau \quad (4.49)$$

Since $W \gg b$, the above equation may be further approximated to

$$\Psi_X(\mu) = \int_0^{\infty} -\frac{\lambda K^2 \mu^2 \sigma^2}{4} \exp(-2b\tau) d\tau \quad (4.50)$$

or

$$\Psi_x(\mu) = -\frac{\lambda K^2 \mu^2 \sigma^2}{8b} \quad (4.51)$$

For the definition of $\Psi(\mu)$ given in Equation (4.42) and Equation (4.51), the characteristic function $\Phi_x(\mu)$ is calculated to be

$$\Phi_x(\mu) = \exp\left[-\frac{\lambda K^2 \mu^2 \sigma^2}{8b}\right] \quad (4.52)$$

The above equation implies that x is a Gaussian random variable with variance σ_x^2 given by

$$\sigma_x^2 = \frac{\sigma^2 K^2 \lambda}{4b} \quad (4.53)$$

The characteristics of Noise Amplitude Distribution (NAD) of the impulsive noise-envelope may be determined from the above analysis. Equation (4.53) implies that the NAD for the impulsive noise is expected to have a Gaussian curve. Inability of the receiver circuit to resolve the noise pulses of very low amplitude from the receiver noise itself results in a deviation from the Gaussian characteristic. At high amplitude levels, deviation from Gaussian nature occurs as a result of the merging of several pulses due to lack of resolution. If noise from other sources also appears at the receiver input, the NAD is expected to shift right. The way in which a

Gaussian NAD curve would be expected to be modified by these practical factors is shown in Figure 4.19.

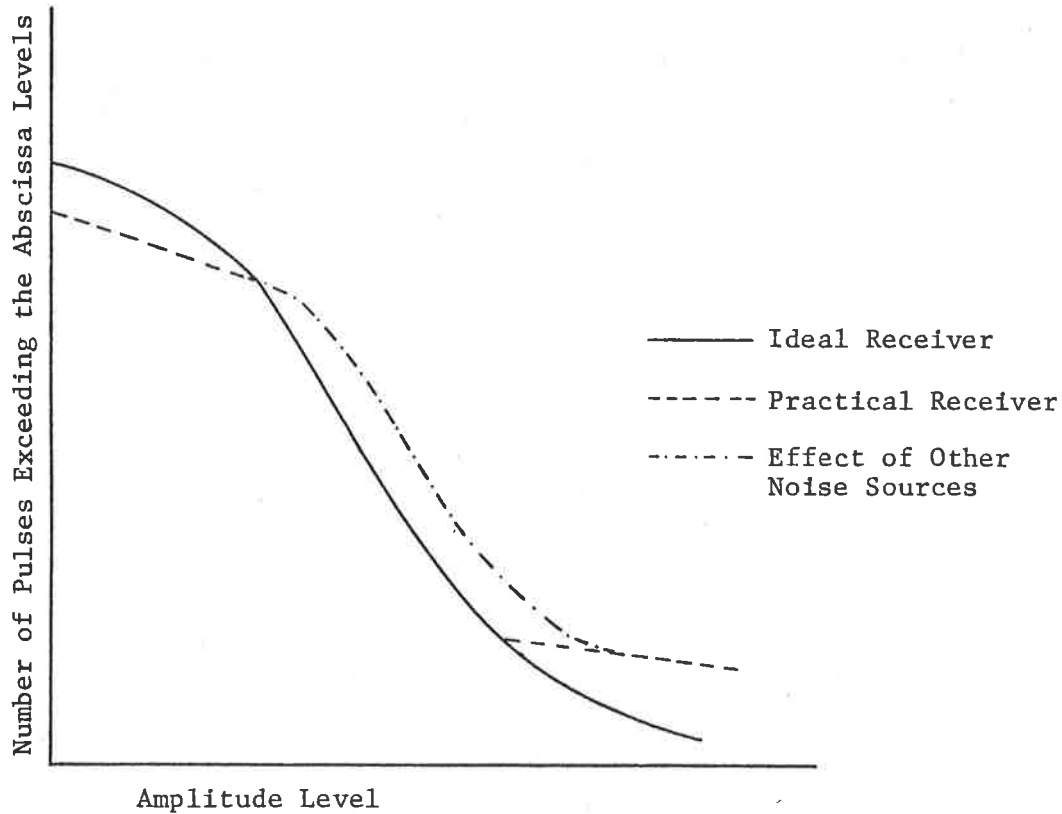


Figure 4.19 Expected NAD for Impulsive Noise

It may be shown that the characteristics of Time Probability Distribution (TPD), when the probability of occurrence of noise pulses has a Poisson distribution, have an exponential distribution as shown in Figure 4.20 below. In a practical receiver, deviation from ideal behaviour occurs due to the inability of the receiver to resolve pulses which are

close together. As the TPD is plotted, a family of curves for pulses of various amplitude level effects similar to the ones shown for NAD are also expected for very low and high amplitude TPDs.

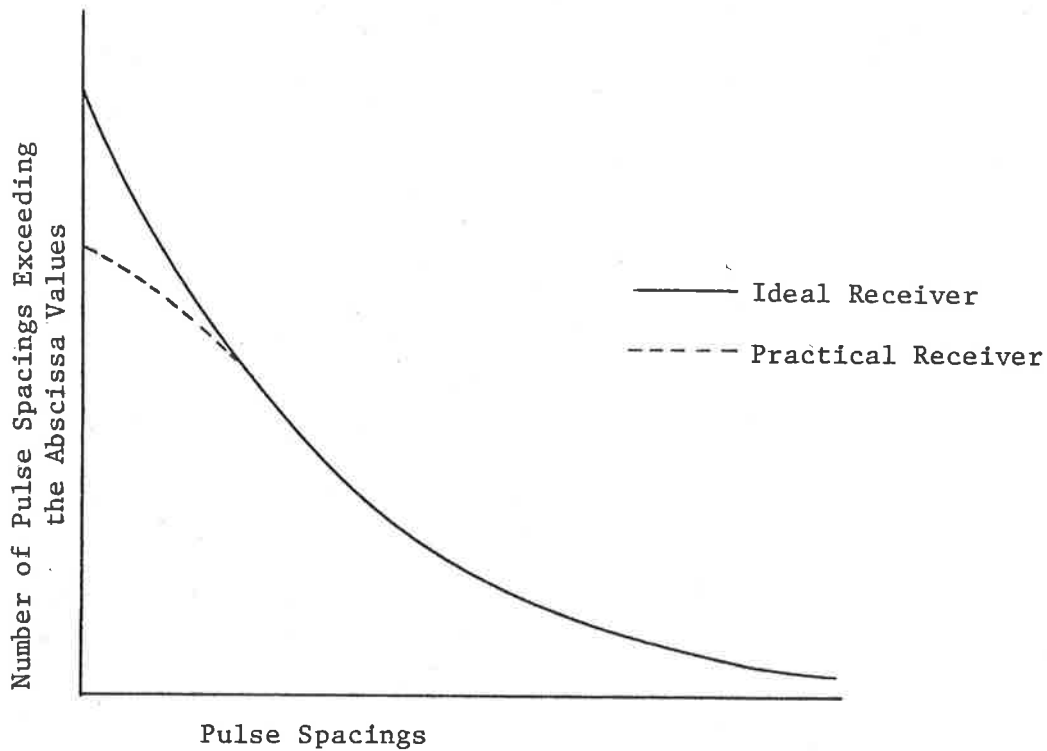
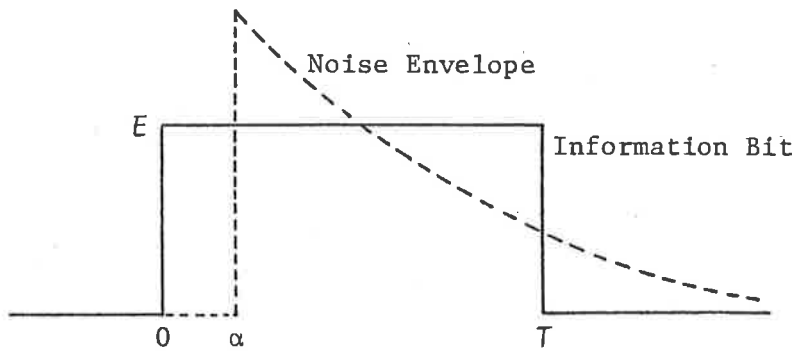


Figure 4.20 Expected TPD for Impulsive Noise in Mid Amplitude Range

4.6.1 Error Prediction Analysis

Consider an impulse occurring at time α . The output of the detector due to this impulse is given by

$$V_n(t) = KA \cdot \exp[-b(t-\alpha)] \cos \omega \alpha \quad (4.54)$$



(see Equation 4.54)

Figure 4.21 Information Bit and Noise Envelope

The information bit due to the reply signal and the noise-envelope at the output of the detector are shown in Figure 4.21. The sum of these signals is fed to an integrate and dump filter. The noise contribution at the output of integrate and dump filter due to the single impulse occurring at time α is Δv_0

$$\Delta v_0 = \frac{KA \cdot \cos \omega \alpha}{b} [\exp(b\alpha) - \exp\{-b(T-\alpha)\}] \quad \alpha \leq 0 \quad (4.55a)$$

$$= \frac{KA \cdot \cos \omega \alpha}{b} [1 - \exp\{-b(T-\alpha)\}] \quad 0 \leq \alpha < T \quad (4.55b)$$

Total noise contribution at the output of the filter is the superposition of all Δv_0 .

$$v_0 = Ag(\alpha) \quad ; \text{pulse present} \quad (4.56a)$$

$$= 0 \quad ; \text{pulse absent} \quad (4.56b)$$

where $g(\alpha)$ is a function of α which results from the superposition of the contribution of individual impulses

By using a procedure similar to the one used for Equation (4.45) we obtain

$$\Psi_{v_0}(\mu) = \int_{\alpha=-\infty}^T \lambda [\Phi_A \{\mu g(\alpha)\} - 1] d\alpha \quad (4.57)$$

By suitable change of variables

$$\Psi_{v_0}(\mu) = \int_{\tau=0}^{\infty} \lambda [\Phi_A \{\mu m(\tau)\} - 1] d\tau \quad (4.58)$$

where substitutions $\alpha = T - \tau$

and $m(\tau) = g(T - \tau)$ are used

The steps involved in the calculation of the probability of error require, firstly, the evaluation of the value of $\Psi_{v_0}(\mu)$ by performing the integral in Equation (4.58). The characteristic function $\Phi_{v_0}(\mu)$ is then evaluated by using

$$\Phi_{v_0}(\mu) = \exp[\Psi_{v_0}(\mu)] \quad (4.59)$$

The probability density function is then calculated

$$p_{v_0}(v_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{v_0}(\mu) \exp(-j\mu v_0) d\mu \quad (4.60)$$

The average probability of error is then found by integrating the probability density functions over an appropriate range

$$\bar{p}_e = \int_{ET}^{\infty} p_{v_0}(v_0) dv_0 \quad (4.61)$$

The drawback of this approach lies in the complexity of mathematical manipulations required to determine the probability of error. The analysis may be simplified by making further assumptions. The accuracy of such analysis suffers because assumptions capable of sufficiently simplifying the analysis are often not strictly valid in most applications. It is for these reasons that another

approach to the error performance prediction has been employed for the passive subharmonic transponder applications. This approach is described in the following section.

4.6.2 Error Prediction by Means of Simulation

An alternative technique to predict the error performance of a communication system under a given environmental noise condition is to subject the noise samples to the detection process of the receiver system, and to compare the resultant output with a presumed signal level. In applying this technique, a long record of noise-envelope samples was first obtained by suitable hardware, and then used as the input to a simulated receiver processor. The noise voltage at the output of the simulated integrate and dump filter was compared with the expected reply signal level, to determine the frequency of errors. For this study, the same noise-envelope as was recorded for the statistical studies was used. A block schematic of the data acquisition system and subsequent signal processing performed for the prediction of error performance by means of direct comparison is shown in Figure 4.22. The signal processing for the environmental noise was mainly performed in hardware followed by some data manipulations in software, while the pro-

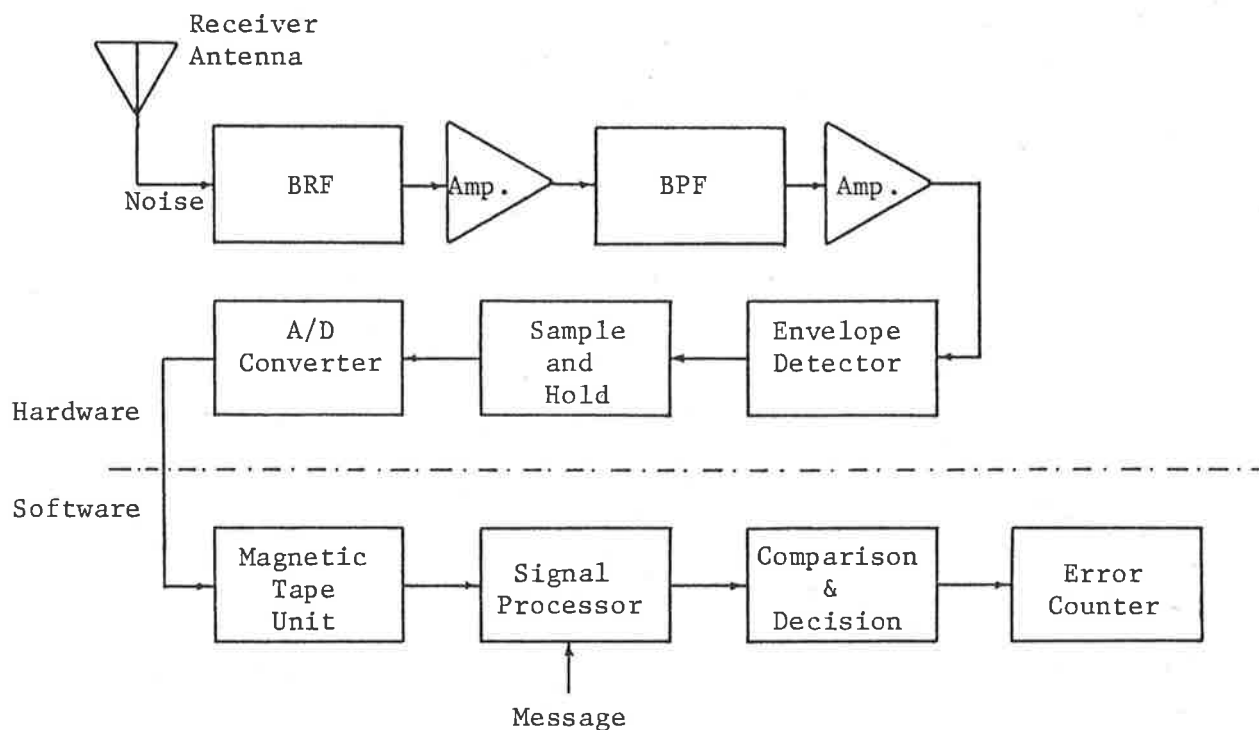


Figure 4.22 Signal Processing for Error Prediction

cessing of the reply signal was entirely performed by the simulated receiver. In the collection of the noise record, the sampling of the noise-envelope was performed at a rate of 10K samples per second, which for the receiver antenna bandwidth of 3125Hz is a rate sufficiently high to enable the practical reconstruction of the original envelope. The 12-bit sample values were then stored on a magnetic tape as two consecutive 8-bit words for later processing. In the subsequent computer processing, a three point

quadratic interpolation of the sample values was performed to determine the peak amplitude and effective duration of the envelope of each noise pulse. This interpolation was also useful in determining the noise contribution at the output of the bit integrator. The measured noise-envelope and an assumed information signal were each integrated over a one bit period, and the signal and the noise outputs of the bit integrator at each bit sampling position were compared. This process was repeated for all possible values of the phase of each noise transient relative to the information signal. For each comparison, errors were recorded when the noise contribution at the output of the bit integration process exceeded the signal contribution. The results of error prediction by this technique for several noise environments are plotted in Figure 4.23.

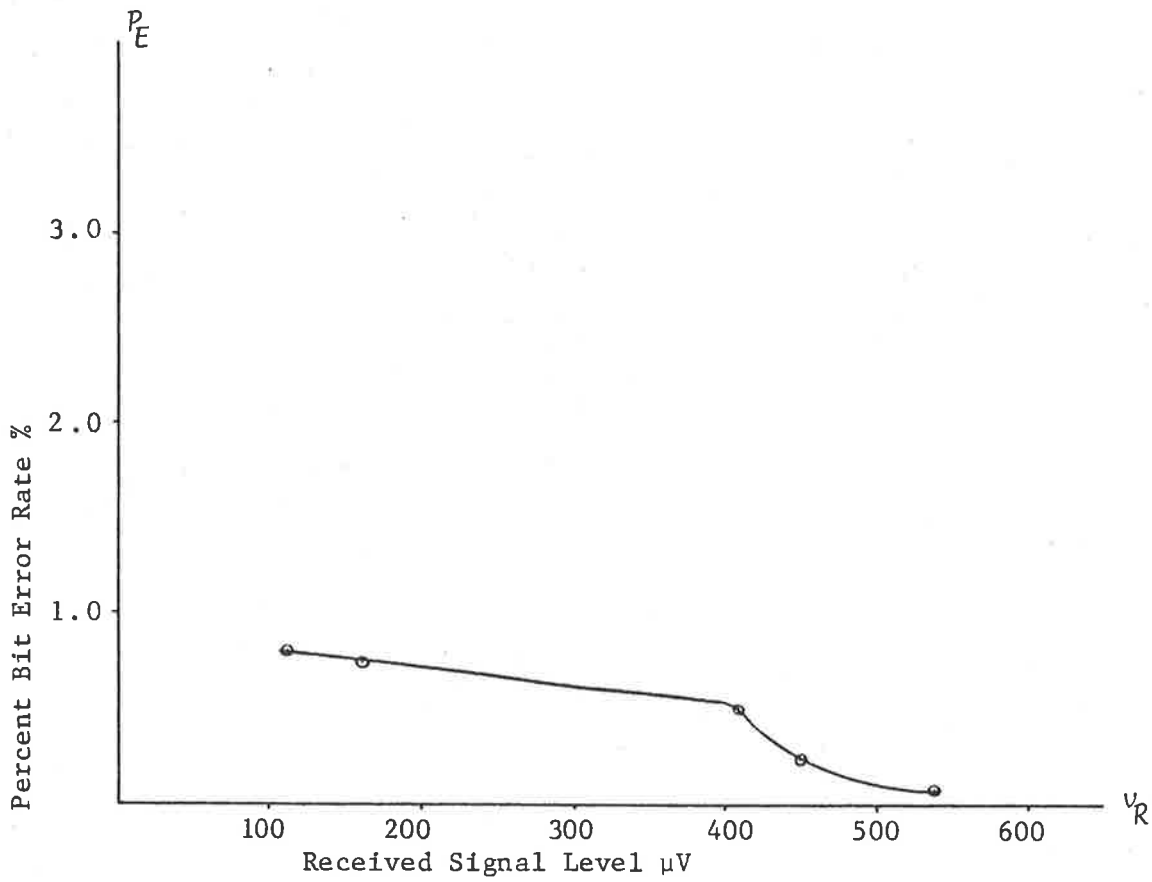


Figure 4.23a Bit Error Rate for 6-Cylinder Vehicle Noise at Idling Speed

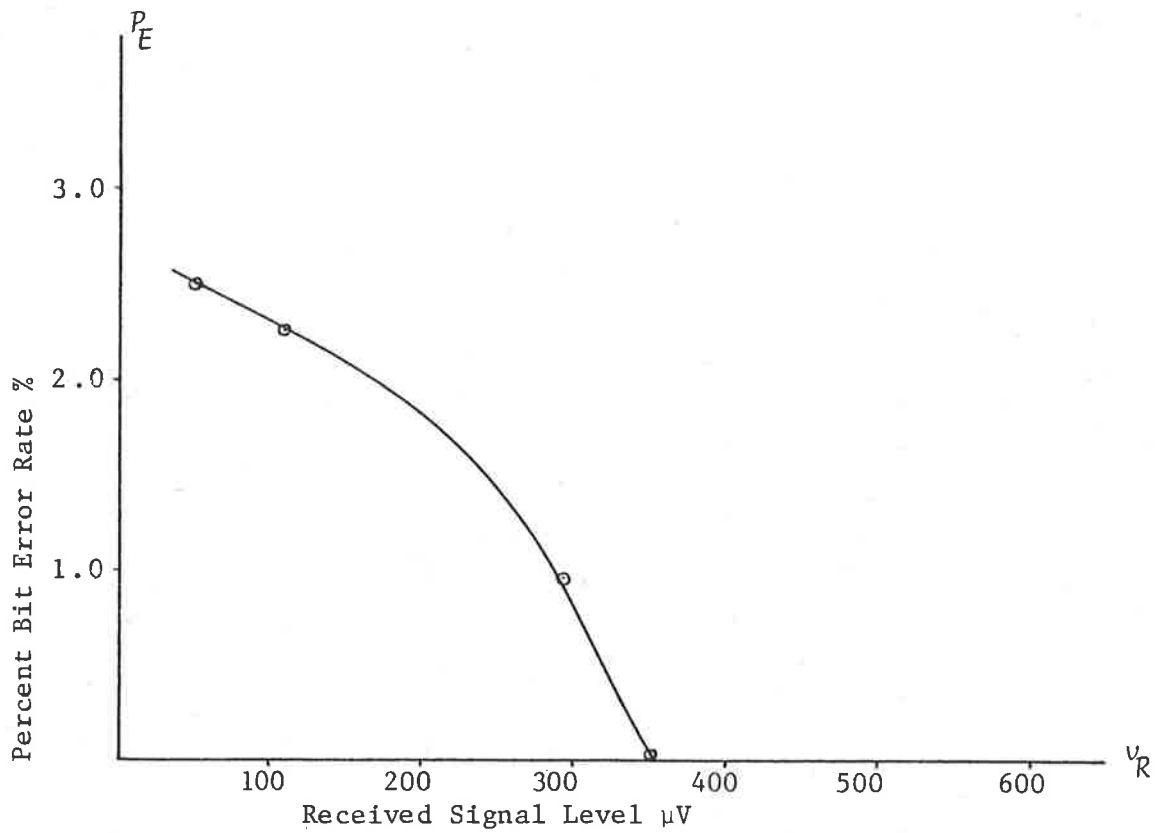


Figure 4.23b Bit Error Rate for 6-Cylinder Vehicle Noise at Cruising Speed

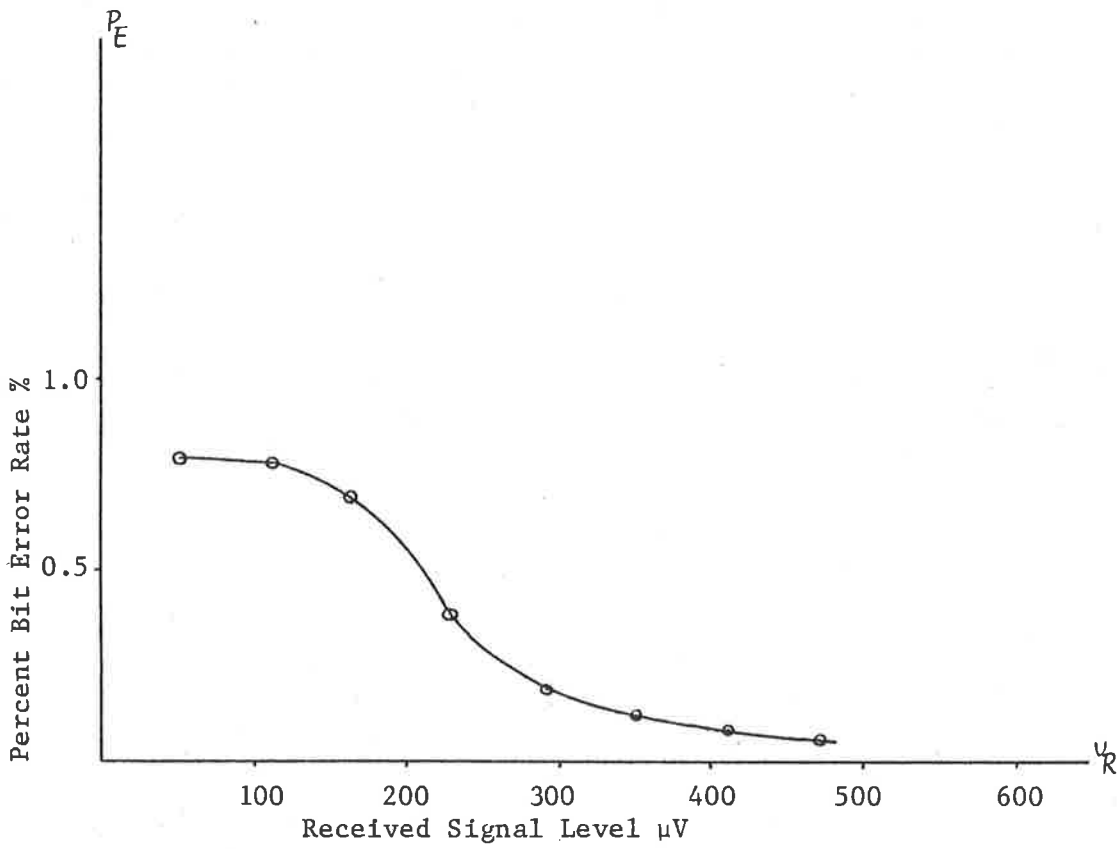


Figure 4.23c Bit Error Rate for V-8 Vehicle Noise at Idling Speed

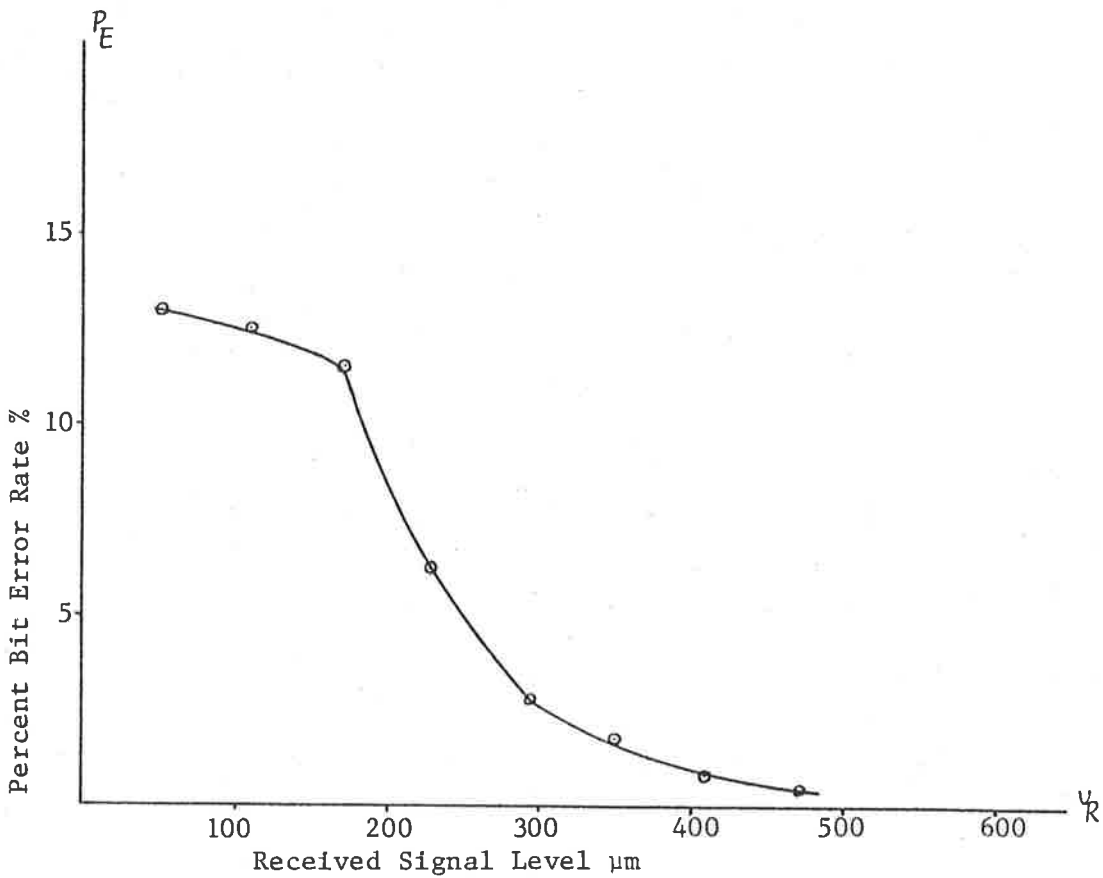


Figure 4.23d Bit Error Rate for V-8 Vehicle Noise at 4.83 Cruising Speed

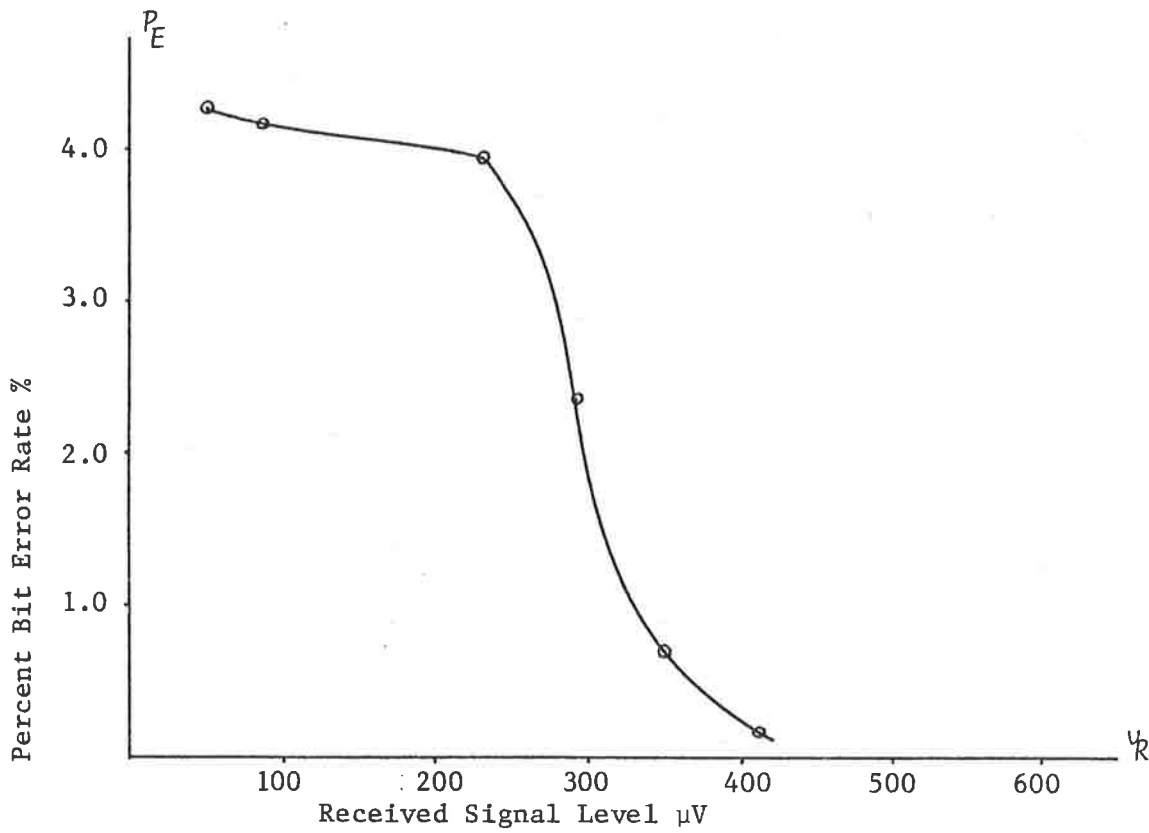


Figure 4.23e Bit Error Rate for Multiple Vehicles Noise at Idling Speed

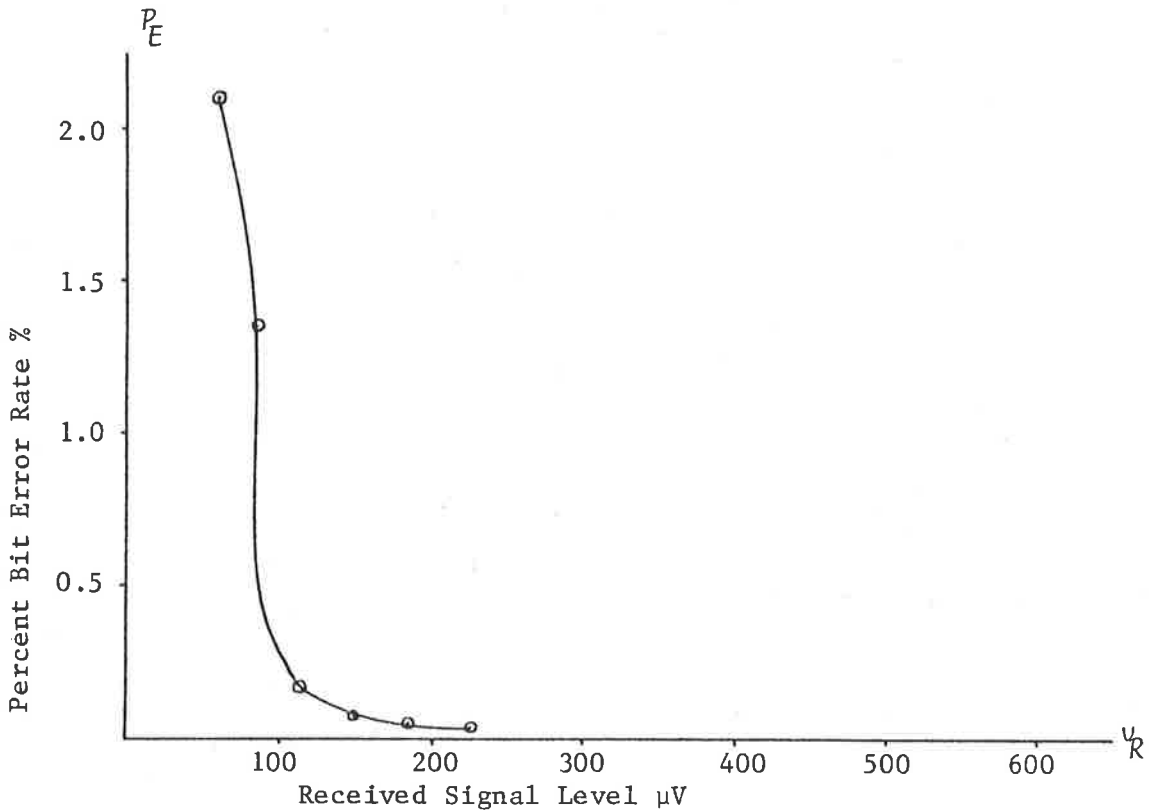


Figure 4.23f Bit Error Rate for ULOS Computer Laboratory Noise

4.7 Conclusions

The study of environmental noise constitutes an important part of the work presented in this thesis. For the evaluation of the performance of a communication system, it is important to study the characteristics of the environmental noise. The measurement of the automotive ignition noise and the error prediction based on these measurements indicate that, in order to obtain error probabilities acceptable in practical systems, error detection and correction schemes should be employed in object identification systems using passive transponders. Some suitable error control strategies are discussed in Chapter 5.

CHAPTER V

ERROR CONTROL

5.1 Introduction

Error control is one of the most important problems in communication engineering. Much research has been done in the field of error control alone over the past two decades (Golay, 1949; Hamming, 1950; Sacks, 1958; Peterson, 1960; Williard, 1978; Scott et al, 1981), and major developments have contributed to the rapid emergence of the field of error detection and correction during this period (Bose and Chaudhari, 1960; Hocquengham, 1959). There have been significant developments in the mathematical theory of encoding and decoding for error control (Peterson and Weldon, 1972). Growth in the use of data communication systems and their intolerance to errors has resulted in a critical look at error control techniques being taken by many authors. The reduction in cost and size of electronic devices and improvements in their performance have resulted in hardware imple-

mentation of error detection and correction strategies becoming an economically viable proposition. This chapter reviews, first, the work done by other researchers in this field to develop the background for the work presented later in this chapter.

The objective in an ideal data communication system is the delivery of error-free messages between a data source and a data sink. A more practical objective has been chosen for the object identification application, that is, to limit the average error rate below a predetermined threshold, firstly by means of suitable encoding of the message at the source and decoding the received signal at the sink and, secondly, by an appropriate selection of other system parameters. The purpose of encoding and decoding is not to correct every conceivable pattern of error, but to correct the most likely patterns.

The error performance of the object identification system is affected by the transmitter power, the various antenna dimensions, the electromagnetic propagation loss in the complete reply link, the modulation parameters, the error control strategy employed, the channel characteristics, the interrogation time, and the signal processing at the

receiver end. The complex interdependence of these system parameters requires a careful consideration of the options to design an optimum system. The effect of some of these system parameters is studied in this chapter, with a view to developing an optimum encoding and decoding strategy.

5.1.1 Transponder Communication System

The passive transponder system seen as a data communication link is shown in Figure 5.1. In an ideal data communication system the symbol received at the data sink matches the symbol transmitted from the data source. In a practical system, however, the received signal is not only corrupted by the noise in the channel and the receiver circuitry, but also distorted due to the nonlinear transfer characteristics of the channel and the processing blocks between the source and the sink. Errors are thus possible at the decision circuitry in the receiver. The error rate depends upon many system parameters, of which the signal level at the source, reply link electromagnetic propagation loss, nonlinear distortion of the various building blocks, amplitude and phase response of the channel, and the noise characteristics are but a few.

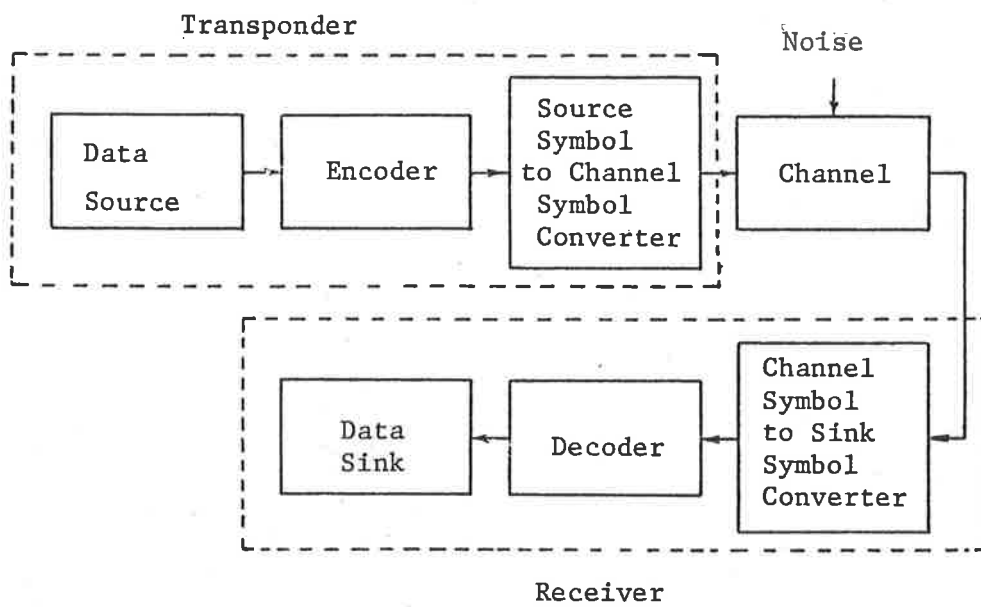


Figure 5.1 Passive Transponder Communication System

It is possible to improve the error performance of the object identification system by increasing the signal-to-noise ratio at the receiver. This may be achieved by increasing the reply signal strength, the antenna dimensions, interrogation time and optimal selection of several other parameters; however, as a result of various constraints involved in a practical system of this nature it is not possible to decrease the probability of error beyond a certain limit. Furthermore, as shown by Shannon's theorem, once the channel is set up it has a definite information capacity. Shannon's theorem gives the capacity of a band-limited channel with fixed average signal power in the presence of additive white-Gaussian noise.

$$C = W \log_2 \left(1 + \frac{S}{N} \right) \quad (5.1)$$

Where the capacity C is the maximum data transfer rate at which the information can be received with arbitrarily small probability of error, the bandwidth W is defined for a normalized system which has unity gain and linear phaseshift over the bandwidth and zero gain outside the band, the average signal power is S and the average noise power is N . The value of C in the presence of non-Gaussian noise or coloured-Gaussian noise is normally higher than that in the presence of white-Gaussian noise with other parameters remaining unchanged.

Another technique of improving the reliability of the received information is to utilize a part of the channel capacity to carry redundant information in place of additional information. In this technique, the reliability of the information transfer increases in proportion to the ratio of redundancy to new information. The redundancy in the transmitted message enables the receiver to detect, and subsequently correct, the errors in the reply word, which procedure results in the improved error performance of the data communication system.

Most of the coding theory outlined in the literature is based on the assumption that each symbol is affected independently by the noise, so that the probability of a given error pattern depends only on the number of errors. Thus, codes that detect any pattern of m or fewer errors in a block of n symbols may be designed. While this may be an appropriate error detecting strategy for some channels, it may not be appropriate for others where errors occur predominantly in bursts. These bursts may last longer than the transmission time for one symbol. Consequently, codes for correcting bursts are required.

Sometimes communication systems employ two-way channels, in which error detecting codes may be used

fairly effectively. When an error is detected at one terminal, a retransmission is requested and, thus, errors can be effectively eliminated. In the object identification applications, it is not possible to employ this strategy. In the case of one-way channels, such as the one studied in this thesis, error-correcting codes may be used to reduce the probability of errors.

Encoding for error correcting codes is no more complex than for error detecting codes; it is the decoding which requires complex signal processing. Error detection is by nature a much simpler task than error correction and, therefore, requires much simpler signal processing to implement than does decoding. Also, error detection with retransmission increases the redundancy when errors occur. Generally, better performance is obtained with this kind of system than with systems employing error correction alone. However, there is a definite limit to the efficiency of a system that uses error detection and retransmission but omits error correction. Short error detecting codes cannot detect errors efficiently and, on the other hand, extremely long codes require frequent retransmissions. A combination of correction of the most frequent error patterns and detection coupled with retransmission for less frequent error patterns is likely to be more

efficient than either error correction or detection alone.

5.1.2 Real Data Channel Models

The system designer has some control over the characteristics of all the building blocks of the system, as shown in Figure 5.1, except the channel which, perhaps, is the most important component. It is necessary to have precise information about the channel, in order to predict the performance of a communication system, and to evaluate improvements due to any error control strategy. The binary symmetric channel (BSC) is the most frequently used mathematical model. In this model of a noisy channel, the following assumptions are made:

- i) The probability of a bit error equals E , where E is less than 0.5. Hence the probability of a bit being received error-free is $1-E$.
- ii) The probability of an error in any bit location is independent of the probability of error in any other bit. In other words, the channel is memoryless.

This model represents a purely random distribution of errors due to noise. Usually it is convenient to approximate the real data channel to the binary symmetric channel model. The probability of

error in an N -bit message when transmitted through such a channel is given by the well-known binomial probability distribution (Kreyszig, 1962)

$$P(E) = (1-E)^N + \frac{N(1-E)^{N-1}E}{1!} + \frac{N(N-1)(1-E)^{N-2}E^2}{2!} + \dots \quad (5.2)$$

The first term in the series is the probability of error-free reception of an N -bit message. The second term in the series denotes the probability of exactly one bit being in error, and so on. The probability of receiving an N -bit message with t or fewer errors may be found by adding the first $(t+1)$ terms of the above expression. In a random error environment, the probability of receiving a message sequence with t or fewer errors is higher than the probability of receiving the message error-free.

Although most data communication channels may be reasonably accurately described by the binary symmetric channel, some channels have characteristics which cannot be modelled by the binary symmetric channel representation. One such channel, called the burst error channel, contains bursts of high error rates interspersed with periods of low error rates. An idealized burst error channel model which consisted of burst periods followed by gap in-

tervals resulted from the efforts to mathematically model burst noise environment (Gilbert, 1960). During the burst period, the channel was modelled as a binary symmetric channel with the probability of error E_b and, similarly, during the gap interval, the channel was modelled as another binary symmetric channel with the probability of bit error E_g , where $E_g < E_b$. The Gilbert burst error channel model also assumed that the burst and gap intervals were independent of each other, and that the burst frequency was randomly distributed. The benefits of the Gilbert model are that it introduces order and simplicity into a complex problem.

Another channel model which has been studied extensively is the binary erasure channel (Gallager, 1968). The models for the binary symmetric channel and the binary erasure channel are shown in Figure 5.2. The source-symbol to channel-symbol converter and the channel-symbol to sink-symbol converter are included in the channel model for ease of representation. The binary erasure channel model is based on the following assumptions (Peterson and Weldon, 1972):

- i) The probability of an erasure equals E , where E is less than 0.5. Hence, the probability that the same bit is received as was transmitted is $1-E$.
- ii) The binary symbols are assumed to be affected independently.

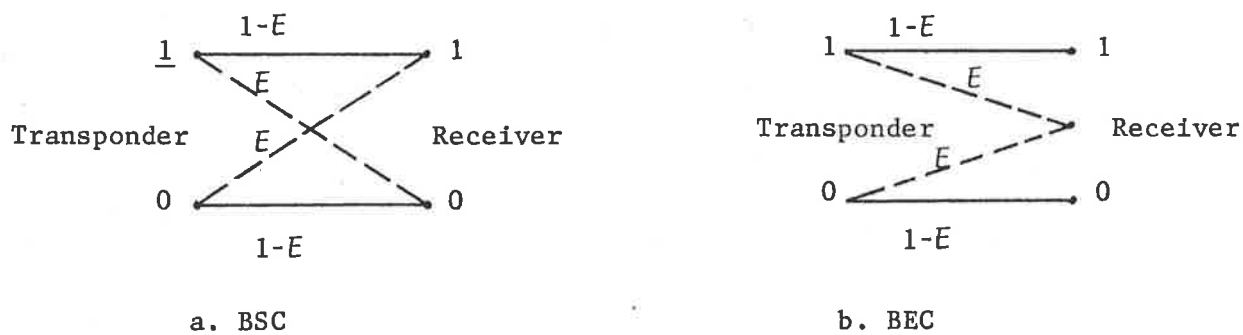


Figure 5.2 Comparison of Channel Models

With a binary erasure channel, the locations of the perturbed symbols are known, whereas it is impossible to know the location of the erroneous bits in a binary symmetric channel when there is no error correcting strategy employed. This property makes correction of erasures easier than correction of errors. The erasure channel results when the channel-symbol to sink-symbol converter

is designed to deliver an erasure or "don't know" symbol when an error occurs due to the channel noise.

A generalized data channel which includes both error and erasure channels is shown in Figure 5.3.

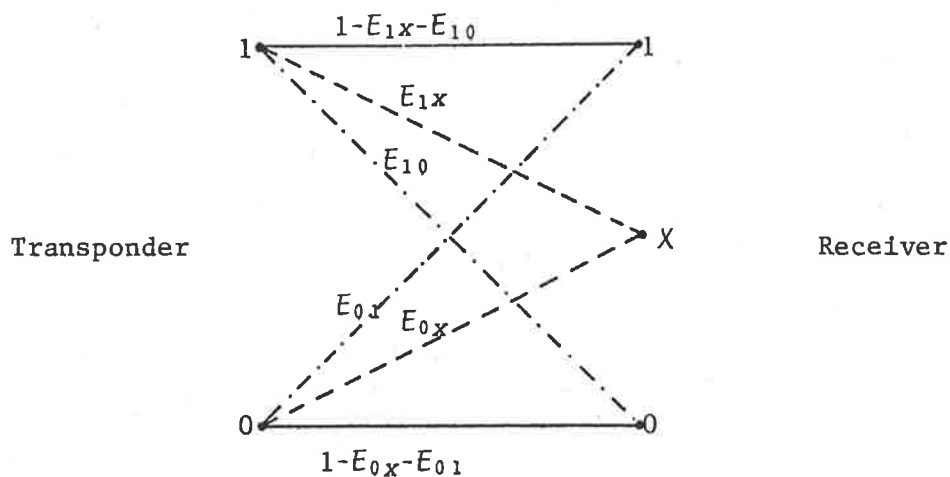


Figure 5.3 A Generalized Binary Channel Model

One assumption which holds for the majority of real data channels is that the channel, apart from being memory-less, does not distinguish between 1's and 0's and, therefore, the probability of error and erasure for both 1's and 0's is equal, as given by Equations (5.3a) and (5.3b).

$$E_{1x} = E_{0x} \quad (5.3a)$$

$$E_{10} = E_{01} \quad (5.3b)$$

The probability of errors in a message length of N bits may be calculated by adding the appropriate terms in the expression

$$\begin{aligned}
 P(E) = (1-E)^N &+ \frac{N(1-E)^{N-1} E_x}{1!} + \frac{N(1-E)^{N-1} E_x}{1!} \\
 &+ \frac{N(N-1)(1-E)^{N-2} E_x E_y}{1!} \\
 &+ \frac{N(N-1)(1-E)^{N-2} E_x^2}{2!} + \frac{N(N-1)(1-E)^{N-2} E_y^2}{2!} \\
 &+ \frac{N(N-1)(N-2)(1-E)^{N-3} E_x^2 E_y}{2!} \\
 &+ \frac{N(N-1)(N-2)(1-E)^{N-3} E_x E_y^2}{2!} + \dots
 \end{aligned} \quad (5.4)$$

where

E_x = probability that a bit is erased

E_y = probability that a bit is in error

and $E = E_x + E_y$

The usefulness of expression for probability of error as given by Equation (5.4) is somewhat limited due to the difficulty in estimation of E_x and E_y . Alternative techniques for the prediction of the error performance of a data communication

system in a given channel have already been discussed in Chapter four.

5.2 Types of Code

An encoder accepts at its input a continuous sequence of information digits, and produces another sequence with somewhat more digits at its output. The encoder may be incorporated within the data source. The encoded message is then transformed into signals suitable for transmission by the source-symbol to channel-symbol converter and fed to the channel. At the receiver end, a reverse transformation is applied to retrieve the encoded message corrupted by noise. The decoder normally translates these symbols back to the information symbols but, sometimes, the error patterns are such that the decoder is unable to detect and correct the errors.

The problem of coding can be best described by assuming all the blocks between the encoder and decoder to be part of the channel. The encoder adds redundancy to the information sequence, while the decoder uses this redundancy to detect and correct the errors which may have resulted during transmission of the message through a hostile channel. The

error correcting capability of a code increases with the redundancy. Conceptually, the encoding and decoding may be performed by means of look up tables, whereby the encoder transforms the information space to a code space and the decoder transforms the code space to a reply message space. The problem associated with such techniques is that it becomes practically impossible to decode codes of very large block lengths; however, rapid developments in VLSI technology may make this approach feasible for modest block lengths in the near future.

The problem can be simplified to a great extent by a proper understanding of the properties of different codes. Even more importantly, the mathematical structures can make the hardware implementation feasible. The main aspects of the coding problem for a data communication system are:

- i) to find codes with necessary and sufficient error-correcting capability
- ii) to find a practical method of encoding
- iii) to find a practical method of error correction.

A useful strategy employed by many workers in this field has been to find codes that could be proved mathematically to satisfy the required error-correcting ability. The mathematical structure of

these codes is then exploited to meet the requirements of the encoder and the decoder implementation (Peterson and Weldon, 1972).

To begin our study of error detection and correction coding theory, it should be recognized that there are two fundamentally different types of such codes. The encoder for a block code breaks a digital data stream into k -digit segments. It then operates on these segments independently, according to the particular code to be employed, expanding the number of digits as appropriate to provide for the error detection or correction. For each k -digit information sequence, the encoder generates an n -digit channel sequence, where $n > k$. The result, now called a code word, is transmitted, corrupted by noise, and decoded independently of all other code words. The quantity n is referred to as the code length or block length.

In the other type of code, called a tree code, the encoder operates on the information sequence without breaking it up into independent segments. Rather, it processes the information continuously, and associates each complete information sequence with a code sequence containing somewhat more digits. The encoder generates an output code sequence by processing in serial fashion the input

information sequence such that the output at each stage depends on the value of the present and all the preceding information digits. The name tree code is derived from the fact that the encoding rules for this type of code are most conveniently described by means of a tree graph. A subset of the tree codes, known as the convolutional codes, have similar error-correcting capabilities and limitations to those of block codes.

The encoding and decoding for tree codes require complex signal processing. For object identification applications, the use of tree codes does not offer any cost or size advantage over that of block codes. The tree codes are, therefore, not considered further in this work.

While the principal purpose of this chapter is to review the work done in the field of error-correcting codes from the point of view of selection of a suitable error control strategy for object identification applications, most of the error correcting codes developed over the last two decades are not discussed in this thesis because of the fact that these coding techniques were not employed in the applications considered in this thesis. During the search for an optimum error control strategy, Bose-Chaudhuri-Hocquenghem (BCH) codes were found

to be most suitable for this purpose. It is for this reason that BCH codes are discussed in greater detail in this chapter. A thorough review of the mathematical structure of the BCH codes will, therefore, be provided to establish a basis for the selection of encoding and decoding parameters, and the design of hardware suitable for use in the object identification applications.

5.3 Review of Linear Algebra

A group G is a set of objects, or elements, for which an operation "+" is defined and for which the following axioms hold.

Axiom G1 - Closure: The operation "+" can be applied to any two group elements to result in a third group element.

$$a + b = c$$

$$\text{where } a, b, c \in G$$

Axiom G2 - Associative Law: For any three elements a, b, c of the group

$$(a+b) + c = a + (b+c) = a + b + c$$

Axiom G3 - Identity: There exists a unique identity element i of the group such that

$$a + i = i + a = a$$

for every element a of the group.

Axiom G4 - Inverse: Every element a of the group has a unique inverse element a' such that

$$a + a' = a' + a = i$$

In addition to the above axioms, if

$$a + b = b + a$$

for all a, b belonging to the group, the group is called Abelian or commutative. The number of elements in a group is called the order of the group.

A ring R is a set of elements for which two operations are defined. One is called addition and denoted by "+", the other is called multiplication and denoted by "•", even though these operations may or may not be ordinary addition or multiplication. For R to be a ring, the following axioms must also hold:

Axiom R1 - The set R is an Abelian group under addition.

Axiom R2 - Closure: For any two elements a and b of R , the product $a \cdot b$ is defined and is an element of R .

Axiom R3 - Associative Law: For any three elements a, b, c of R

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c$$

Axiom R4 - Distributive Law: For any three elements a, b, c of R

$$a \cdot (b+c) = a \cdot b + a \cdot c$$

and

$$(b+c) \cdot a = b \cdot a + c \cdot a$$

In addition, if the multiplication operation is commutative, the ring is called commutative.

A field F is a commutative ring with a multiplicative identity m such that

$$a \cdot m = m \cdot a = a$$

where a is an element of the field F .

Further, every non-zero element of the field has a multiplicative inverse

$$a \cdot a' = a' \cdot a = m$$

where a is an element of the field F

a' is the inverse of a and also belongs to the field F

m is the multiplicative identity of the field F

A non-commutative ring in which every non-zero element has an inverse is called a skew field. The non-zero elements of a field satisfy all the axioms for a group and thus form a group under the operation multiplication.

It may be shown that a finite field of p elements may be obtained by taking integers $(0, 1, \dots, p-1)$ as the field elements and the ordinary rules of integer arithmetic with the results expressed modulo p as defining the properties of the conjunctions addition and multiplication, provided p is a prime. It may be further shown that integers modulo q do not form a field if q is not prime. It can, however, be shown that fields with $q=p^m$ elements, where p is prime and m is an integer can be formed. The rules for arithmetic manipulations will not be, in view of the above statement, the rules of simple arithmetic modulo q . The details of these rules will be discussed later.

A set V of elements is called a vector space over a field F if it satisfies the following axioms:

Axiom V1: The set V is an Abelian group under addition. The notation \oplus is used for vector addition.

Axiom V2: For any vector y in V and any field element c , a product $c \odot y$, which is a vector in V , is

defined. Further

$$1 \odot v = v$$

where 1 is the multiplicative identity of the field F .

Axiom V3 - Distributive Law: If u and v are vectors in V and c is a scalar in F .

$$c \odot (u \oplus v) = c \odot u \oplus c \odot v$$

Axiom V4 - Distributive Law: If v is a vector in V and c and d are scalars in F

$$(c+d) \odot v = c \odot v + d \odot v$$

Axiom V5 - Associative Law: If v is a vector in V and c and d are scalars in F

$$(c \cdot d) \odot v = c \odot (d \odot v)$$

A subset of a vector space is called a subspace if it satisfies all the axioms of a vector space. In order to check whether a subset of a vector space is a subspace, it is necessary only to check for closure under vector addition and multiplication by scalars.

A linear combination of k vectors $v_1 \dots v_k$ is a sum of the form

$$u = a_1 \odot v_1 \oplus a_2 \odot v_2 \oplus \dots \oplus a_k \odot v_k$$

where a_i are scalars belonging to field F .

It may be easily shown that the set of all linear combinations of a set of vectors $v_1 \dots v_k$ of a vector space V is a subspace of V . A set of vectors $v_1 \dots v_k$ is linearly independent if and only if there are scalars $c_1, c_2 \dots c_k$ not all zero, such that

$$c_1 \odot v_1 \oplus c_2 \odot v_2 \oplus \dots \oplus c_k \odot v_k = \underline{0}$$

A set of vectors is said to span a vector space if every vector in the vector space equals a linear combination of vectors in the set. The number of linearly independent vectors that span a vector space is called the dimension of the space.

In most useful applications, the vectors consist of several components. Such a vector has the form

$$v = (v_1, v_2 \dots v_k)$$

where v_i are elements of the field F over which the vector space V is defined.

An $m \times n$ matrix \underline{M} over a field F is defined as an ordered set of $m \cdot n$ components, a_{ij} , where each a_{ij} is an element of the field F , arranged in a rectangular array of m rows and n columns.

$$\underline{M} = [a_{ij}]_{mn} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \\ \vdots & & & \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix} \quad (5.5)$$

The addition of two matrices $\underline{\underline{A}} \oplus \underline{\underline{B}}$, where the two matrices have the same number of rows and the same number of columns, is defined and is performed by adding the corresponding components.

$$\text{If } \underline{\underline{A}} = [a_{ij}]_{mn} \quad \text{and} \quad \underline{\underline{B}} = [b_{ij}]_{mn}$$

$$\text{then } \underline{\underline{C}} = \underline{\underline{A}} \oplus \underline{\underline{B}} = [a_{ij} + b_{ij}]_{mn}$$

The product of a matrix $\underline{\underline{M}}$ with an element c of the field F is defined by the following rule

$$\text{If } \underline{\underline{M}} = [a_{ij}]_{mn}$$

$$\text{then } c \odot \underline{\underline{M}} = [c \cdot a_{ij}]_{mn}$$

The $\underline{\underline{0}}$ matrix is defined as the one with 0's as the components, where 0 is the additive identity of the field F .

The m rows of the matrix may be thought of as m n -tuples or vectors and, similarly, n columns of the matrix may be thought of as n m -tuples or vectors.

The row space of an $m \times n$ matrix $\underline{\underline{M}}$ is defined as the set of all linear combinations of row vectors of $\underline{\underline{M}}$. The row space of a matrix is a subspace of the vector space of n -tuples. The dimension of a vector space is defined as the number of linearly independent vectors that span the vector space.

The dimension of the row space is called the row rank. Similarly, all linear combinations of column vectors form the column space and its dimension is called the column rank. It may be shown that the row rank and the column rank of a matrix are equal and, therefore, called the rank of the matrix. A row vector may be thought of as a $1 \times n$ matrix and a column vector may be thought of as an $m \times 1$ matrix.

A non-commutative product of an $m \times k$ matrix $\underline{A} = [a_{ij}]_{mk}$ and a $k \times n$ matrix $\underline{B} = [b_{ij}]_{kn}$ denoted $\underline{A} \times \underline{B}$, is defined if and only if the number of columns of the matrix \underline{A} equals the number of rows of the matrix \underline{B} . The $m \times n$ product matrix $\underline{C} = [c_{ij}]_{mn}$ is defined by the rule

$$\underline{C} = \underline{A} \times \underline{B} \quad (5.6a)$$

and

$$c_{ij} = \sum_{l=1}^k a_{il} \cdot b_{lk} \quad (5.6b)$$

A vector \underline{v} is said to belong to the null space of a matrix \underline{M} if it is orthogonal to each row of the matrix. The orthogonality is expressed by the relation

$$\underline{v} \times \underline{M}^T = 0 \quad (5.7)$$

where \underline{M}^T is the transpose of the matrix \underline{M} .

The vector space containing all the vectors that satisfy the above equation forms the null space of matrix \underline{M} .

A polynomial, $P(X)$ in a single indeterminate X over a field F is an expression of the form

$$P(X) = a_0X^0 \boxplus a_1X^1 \boxplus a_2X^2 \dots \boxplus a_nX^n$$

where a_i , called the coefficients of the polynomial, are all elements of the field F .

At this stage of the exposition, no properties of the X^i 's are asserted, and the notation X^i merely distinguishes potentially different objects, and the notation \boxplus merely serves as a separator for these objects. It is implied that multiplication of the form cX^i between the element c of field F and X^i exists and the associative law $a(bX^i) = abX^i$ holds. The addition $P(X) \boxplus Q(X)$ of two polynomials $P(X)$ and $Q(X)$ is performed by term by term addition according to the rule:

$$aX^i \boxplus bX^i = (a+b)X^i$$

where "+" is the addition conjunction of the field F .

The equality of two polynomials $P(X)$ and $Q(X)$ implies term by term equality of the polynomials. The 0 polynomial is defined as the polynomial with all coefficients as 0, the additive identity of the field F .

The degree of a polynomial is defined as the largest power of X in a term with non-zero coefficient. The degree of a polynomial is 0, by definition.

The open multiplication of two polynomials $P(X) \cdot Q(X)$, which uses the addition operator $+$ and separator \boxplus , is defined by the following rule

$$R(X) = P(X) \diamond Q(X)$$

and $a_i X^i \diamond b_j X^j = a_i \cdot b_j X^{i+j}$

With the rules of multiplication defined by the above equations, the polynomial $R(X)$ is said to be divisible by either $P(X)$ or $Q(X)$. It may be easily shown that, with these rules addition and multiplication, polynomials form a ring, but such a ring is not a finite one. The additive and multiplicative identities of such a ring are the 0 polynomial, which consists of all 0's and the 1 polynomial, which consists of a 1 followed by all 0's.

Polynomials are a convenient way of expressing vectors. A vector \underline{v} with components from a field F may be expressed as a polynomial $P(X)$ with coefficients equal to the components of \underline{v} . If

$$\underline{v} = (a_0, a_1, \dots, a_{n-1})$$

the corresponding polynomial is

$$P(X) = a_0 X^0 \oplus a_1 X^1 \oplus a_2 X^2 \oplus \dots \oplus a_{n-1} X^{n-1}$$

The 0 and 1 polynomials have the vector representations $(0 \ 0 \ 0 \ \dots \ 0)$ and $(1 \ 0 \ 0 \ 0 \ \dots \ 0)$ respectively or in short $\underline{0}$ and $\underline{1}$.

An irreducible polynomial $\mathcal{D}(X)$ of degree m with coefficients from field F is one which is not divisible by any polynomial of degree less than m and greater than 0 with coefficients from the field F .

The roots, b_j , $j=1, \dots, n$ of a polynomial $P(X)$ are defined when the polynomial is divisible by linear factors of the form $(b_j X^0 \oplus X^1)$, where all b_j are elements of the field F over which the polynomial $P(X)$ was defined. Therefore, under the rules of open multiplication, the irreducible polynomial $\mathcal{D}(X)$ over a field F does not have a root in the field F .

For further useful development, we now introduce the concept of closed multiplication of polynomials using an irreducible polynomial. The closed multiplication of two polynomials $P(X)$ and $Q(X)$, both of degree $m-1$ or less, using an irreduc-

ible polynomial $\mathcal{D}(X)$ of degree m is obtained by first using the open multiplication \cdot to form a polynomial $R(X)$ and then reducing $R(X)$ modulo $\mathcal{D}(X)$. Reduction of a polynomial $R(X)$ using an irreducible polynomial $\mathcal{D}(X)$ requires elimination of terms of $R(X)$ of degree m or greater by addition to $R(X)$ of suitable polynomial multiples of $\mathcal{D}(X)$.

If F is a field of finite elements, then it may be shown that polynomials using closed multiplication form a finite field.

A form of vector multiplications may be defined by first expressing vectors as the corresponding polynomials and then employing either the open or the closed rules of polynomial multiplication. In the analysis to follow, the term multiplication is used to describe both open and closed types of multiplication. The type of multiplication used in a given situation should be identified from the context.

In the above development, to clearly distinguish the kind of objects being added, various addition conjunctions have been denoted by different symbols. In the text to follow only a single notation "+" will be used for all addition conjunctions. The particular type of addition employed should be

identified from the context. This remark is also applicable for the polynomial separator. Similarly, all multiplication conjunctions will be denoted by a single notation " \cdot ".

A finite field with p elements operating under a particular addition and multiplication rule has been mentioned earlier in this section. A generalization of this field is the Galois field $GF(p^m)$ containing p^m elements where p is a prime and m is an integer. Although it is possible to use as labels for the field elements the integers from 0 to p^m-1 , such a representation is, for values of $m>1$, not useful because definition of the conjunctions "+" and " \cdot " in such a representation becomes difficult. Useful representations of the field $GF(p^m)$ for $m>1$ will be discussed later.

The Galois field with the fewest elements is $GF(2)$ and consists of two elements commonly designated as 0 and 1 with the following rules of combination:

$$0 + 0 = 1 + 1 = 0$$

$$0 + 1 = 1 + 0 = 1$$

$$0 \cdot 0 = 1 \cdot 0 = 0 \cdot 1 = 0$$

$$1 \cdot 1 = 1$$

It may be pointed out that these are the ordinary rules for binary arithmetic without carry.

As has been indicated earlier in this section, a finite field with a prime number of elements, p , can be represented by integers modulo p , e.g. $GF(2)$ is represented by integers modulo 2. The associated finite fields $GF(p^m)$ with p^m elements, $m > 1$ are called the extensions of $GF(p)$. It can be shown that elements of $GF(p^m)$ can be represented by the set of polynomials of degree less than m in a single indeterminate with coefficients in $GF(p)$, with polynomial multiplication being defined such that the results are reduced modulo p and modulo an irreducible polynomial of degree m with coefficients in $GF(p)$. These ideas are illustrated by the following example.

Example: Addition and Multiplication in $GF(2^3)$

The Galois field $GF(2^3)$ has 8 elements shown as the below set of binary vectors.

000, 100, 010, 110, 001, 101, 011, and 111

These elements are denoted $\alpha_0, \alpha_1, \dots, \alpha_7$ respectively. The addition of the field elements, performed component by component modulo 2, yields the below addition table.

+	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
α_0	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
α_1	α_1	α_0	α_3	α_2	α_5	α_4	α_7	α_6
α_2	α_2	α_3	α_0	α_1	α_6	α_7	α_4	α_5
α_3	α_3	α_2	α_1	α_0	α_7	α_6	α_5	α_4
α_4	α_4	α_5	α_6	α_7	α_0	α_1	α_2	α_3
α_5	α_5	α_4	α_7	α_6	α_1	α_0	α_3	α_2
α_6	α_6	α_7	α_4	α_5	α_2	α_3	α_0	α_1
α_7	α_7	α_6	α_5	α_4	α_3	α_2	α_1	α_0

It may be noticed from the above table that α_0 is the additive identity and each element is the additive inverse of itself.

The elements of the field $GF(2^3)$ may also be expressed as polynomials with coefficients from the field $GF(2)$ as shown below.

$$\begin{aligned} \alpha_0 &= 000 = 0 + 0X + 0X^2 \\ \alpha_1 &= 100 = 1 + 0X + 0X^2 \\ \alpha_2 &= 010 = 0 + 1X + 0X^2 \\ \alpha_3 &= 110 = 1 + 1X + 0X^2 \\ \alpha_4 &= 001 = 0 + 0X + 0X^2 \\ \alpha_5 &= 101 = 1 + 0X + 1X^2 \\ \alpha_6 &= 011 = 0 + 1X + 1X^2 \\ \alpha_7 &= 111 = 1 + 1X + 1X^2 \end{aligned}$$

It may be shown that a set of polynomials of degree less than m with coefficients from $GF(P)$ form a field when multiplication is defined using an irreducible polynomial of degree m . It may further be shown that such polynomials can be represented as elements of the Galois field $GF(p^m)$. In the case $m=3$ there is only a single irreducible polynomial of degree 3 with coefficients in the field $GF(2)$. This polynomial is (Peterson and Weldon, 1972):

$$D(X) = 1 + X + X^3 \quad (5.8)$$

Using this irreducible polynomial, an example of multiplication of two field elements is illustrated below.

$$\begin{array}{r} \alpha_6 = X^2 + X \\ \alpha_1 = \frac{X + 1}{X^2 + X} \\ \hline X^3 + X^2 \\ \hline X^3 + 0 + X \end{array}$$

Reducing this polynomial modulo $D(X)$ implies adding polynomial multiples of $D(X)$ to eliminate terms of degree greater than or equal to m .

$$\alpha_6 \cdot \alpha_3 = (X^2 + X)(X + 1) = X^3 + X$$

or $\alpha_6 \cdot \alpha_3 = (X^3 + X) + (X^3 + X + 1) = 1 + 0X + 0X^2 = (100) = \alpha_1$

It may be noted here that in modulo 2 arithmetic addition and subtraction are equivalent. A suitable multiplication table for the elements of $GF(2^3)$ may be easily generated using the rules illustrated above. The result is:

•	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
α_0	α_0	α_0	α_0	α_0	α_0	α_0	α_0	α_0
α_1	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
α_2	α_0	α_2	α_4	α_6	α_3	α_1	α_7	α_5
α_3	α_0	α_3	α_6	α_5	α_7	α_4	α_1	α_2
α_4	α_0	α_4	α_3	α_7	α_6	α_2	α_5	α_1
α_5	α_0	α_5	α_1	α_4	α_2	α_7	α_3	α_6
α_6	α_0	α_6	α_7	α_1	α_5	α_3	α_2	α_4
α_7	α_0	α_7	α_5	α_2	α_1	α_6	α_4	α_3

The symmetry of the above table illustrates the commutative property of the multiplication. The fact that the column under α_1 is the multiplicative identity. The fact that every column other than that of α_0 contains α_1 implies that each non-zero field element has a multiplicative inverse. Thus it may be seen that the elements of $GF(2^3)$, and the corresponding polynomials, form a field.

It may be pointed out that the fields $GF(2)$ and $GF(2^m)$ are most useful for the work presented in this thesis and therefore only these finite fields are considered in the remainder of this chapter.

An important property of the field $GF(2^m)$ leading from the closed multiplication is that

$$\alpha^i \cdot \alpha^j = \alpha^{i+j}$$

where α^i and α^j are two field elements and α^{i+j} is also a field element

There are major consequences which flow from this result. First of all, the X^i 's in a polynomial, which were previously called the identifiers in a polynomial representation of a vector, can themselves be taken as vectors. These vectors may have a polynomial representation using another set of identifiers Y^l . A property of these new vectors is that they are represented by polynomials of degree $(m-1)$ in Y with coefficients taken from the field $GF(2)$. Secondly, the separator \boxplus takes the meaning of the addition conjunction "+" of the vectors. It can then be seen that an equation of the form:

$$P(X) = 0 \tag{5.9}$$

has a meaning and that roots of the polynomial may be found by substituting vectors in the polynomial which satisfy the above equation. It can be seen that the definition of a root α of the polynomial derived from the Equation (5.9) and the earlier definition where the root was defined in terms of linear factors of the form $(X+\alpha)$ which now becomes $(X+\alpha)$, are equivalent.

It may be shown that with an appropriate choice of irreducible polynomial, certain of the field elements of $GF(2^m)$ called the primitive elements have the property that, firstly, they are the roots of the irreducible polynomial and, secondly, the 2^m-1 non-zero field elements are obtained by taking the first 2^m-1 powers of that primitive element. The irreducible polynomials with this property are called the primitive polynomials. Further, it may be shown that all roots of a primitive polynomial are primitive elements.

It should be noted at this point that the irreducible polynomial used to define the elements of $GF(2^m)$ does not have roots in $GF(2)$; instead, the roots are elements of $GF(2^m)$.

It should also be noted that an irreducible polynomial $\mathcal{D}(X)$ which is used for the definition of the closed multiplication of vectors, itself exists under open multiplication. Alternatively, the factors of $\mathcal{D}(\alpha)$ when multiplied using the rules of open multiplication result in $\mathcal{D}(\alpha)$. It is obvious that the same factors when multiplied using closed multiplication result in the \mathcal{Q} vector.

It may be shown that every non-zero field element α of $GF(2^m)$ is, in the vector representation, a root of the equation

$$\alpha^{2^m-1} = \underline{1} \quad (5.10)$$

where $\underline{1}$ is the vector representation of the multiplicative identity of the field $GF(2^m)$ and the actual components of $\underline{1}$ depend upon the particular irreducible polynomial $\mathcal{D}(X)$ chosen to establish the multiplication rule.

Thus, if α is a non-zero field element, then $\alpha^{-1} = \alpha^{2^m-1}$ is also a field element.

The seven non-zero elements of the field $GF(2^3)$ are calculated in the example below.

Example: Calculation of Non-zero Field Elements of $GF(2^3)$

The primitive polynomial of degree 3 with coefficients in $GF(2)$ is

$$\mathcal{D}(X) = 1 + X + X^3$$

Let α be one of the primitive elements; it therefore must satisfy the equation $\mathcal{D}(\alpha) = 0$, i.e.

$$1 + \alpha + \alpha^3 = 0$$

or
$$\alpha^3 = \underline{1} + \alpha$$

By definition $\alpha^0 = \underline{1}$ and $\alpha^1 = \alpha$. The vector representation of the multiplicative identity of this field is (1 0 0). It may be shown that one of

the roots of the irreducible polynomial $\mathcal{D}(X)$, which is a primitive element, has representation (0 1 0). From these results and the equation $\alpha^3 = 1 + \alpha$, one may generate the remaining non-zero elements of $GF(2^3)$, as shown below:

Representation of $GF(2^3)$

FIELD ELEMENT	REPRESENTATION IN TERMS OF		
	PRIMITIVE ELEMENT	POLYNOMIAL	VECTOR
α^0	1	$1 + 0X + 0X^2$	1 0 0*
α^1	α	$0 + X + 0X^2$	0 1 0†
α^2	α^2	$0 + 0X + X^2$	0 0 1
α^3	$1 + \alpha$	$1 + X + 0X^2$	1 1 0
α^4	$\alpha + \alpha^2$	$0 + X + X^2$	0 1 1
α^5	$1 + \alpha + \alpha^2$	$1 + X + X^2$	1 1 1
α^6	$1 + \alpha^2$	$1 + 0X + X^2$	1 0 1
α^7	1	1	1 0 0

* There is only one representation of the multiplicative identity.

† This is but one solution of several which can be found for the equation $1 + \alpha + \alpha^3 = 0$. The other two are (0 0 1) and (0 1 1), as may be confirmed by substitution.

5.4 Linear Codes

The linear codes form an important class of codes to which all block and tree codes belong. One important property of linear codes is that they can be defined with symbols chosen from a set of arbitrary size. In the study of the codes suitable for object identification application elements from the Galois field $GF(2^m)$ are used.

5.4.1 Definitions

Some of the concepts to follow are useful in discussion of the error detection and correction capabilities of the codes.

A v -letter n -bit binary signalling alphabet A_n^v , containing check bits as appropriate and used for transmission, is defined as a set v distinct sequences $\alpha_0, \alpha_1 \dots \alpha_{v-1}$ of n binary digits. The individual sequences are called the letters of the alphabet or the code words. The alphabet A_n^v is also called the code.

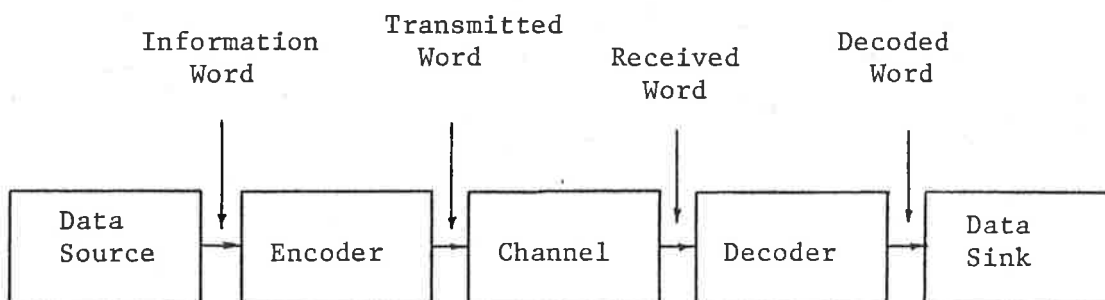
An encoder E_n^v sets up a one-to-one correspondence between the input messages to the encoder and the output letters of the signalling alphabet. To transmit a message over the channel, the n individual digits of the corresponding letter of the

alphabet are presented to the channel in a serial fashion. The output of the channel is an n -bit sequence belonging to the set B_n of all possible binary sequences of length n .

A decoder D_n^v partitions B_n into $v+u$ disjoint subsets $s_0, s_1 \dots s_v, s_{v+1} \dots s_{v+u-1}$, sets up a correspondence between these subsets and the v letters of the alphabet A_n^u , the remaining u subsets being used for decoding failure conditions, and defines probabilities $P_0, P_1 \dots P_{v+u-1}$ of erroneous detection of each of the v letters and the u decoding failure conditions as illustrated by the example given below.

Example: Encoding and Decoding for Error Control

Consider a binary symmetric channel with bit error probability E . Let the single bit information be represented by either a 1 or a 0. The information flow between the data source and the data sink is shown by the following diagram. The encoder in



Information Flow Diagram

this example repeats the information bit $(n-1)$ times and transmits the resultant code word. The information, transmitted, received and decoded words and probability of error for each word for three values of n are shown in the table below.

INFORM. WORDS	TRANS. WORDS	RCVD WORDS	DECODED WORDS	PROB. OF ERROR	REMARKS
1	1	1	1	E	No error Detection or Correction Possible
0	0	0	0	E	
1	11	11	1	E^2	Single Error Detection but no Correction Possible
		10	Decoding Error	$4E$	
		01			
0	00	00	0	E^2	
1	111	111	1	E^3+3E^2+3E	Single Error Correction Possible
		110			
		101			
		011			
		001	0	E^3+3E^2+3E	
010					
100					
0	000	000			

The effectiveness of an error control strategy may be determined by comparison of these probabilities. Once a sequence belonging to a subset S_i is received, it is either decoded as the letter α_i and interpreted as the corresponding message or a message indicating the decoding failure condition is issued.

To introduce the definition of a block, we first observe that the set of all n -tuples with entries chosen from the field of q elements is a vector space. A set of such vectors of length n is called a linear block code if and only if it is a subspace of the vector space of n -tuples. It may be recalled that the basic property of a subspace is its closure under addition. It will be shown later that this property of the linear block codes simplifies encoding and decoding operations.

Useful properties of vectors used in coding theory are derived in terms of Hamming and Lee weight of the vectors. The Hamming weight $w(\mathbf{v})$ of a vector \mathbf{v} is defined as the number of non-zero components in the vector. Since the Hamming distance between two vectors \mathbf{v}_1 and \mathbf{v}_2 is the number of positions in which they differ, the distance between \mathbf{v}_1 and \mathbf{v}_2 is equal to $w(\mathbf{v}_1 - \mathbf{v}_2)$. If \mathbf{v}_1

and \mathcal{U}_2 are code words of a linear block code, then $\mathcal{U}_1 - \mathcal{U}_2$ must also be a code word, since the set of all code words is a vector space. Therefore, the distance between any two code vectors equals the weight of another code vector, and the minimum distance for a linear code equals the minimum weight of its non-zero vectors.

The Lee weight of an n -tuple $(a_{n-1}, a_{n-2}, \dots, a_1, a_0)$ with a_i chosen from the set $(0, 1, \dots, q-1)$ where q is an arbitrary positive integer greater than one, is defined as

$$w_L = \sum_{i=0}^{n-1} |a_i| \quad (5.11)$$

where

$$\begin{aligned} |a_i| &= a_i & 0 \leq a_i \leq \frac{q}{2} \\ &= q - a_i & \frac{q}{2} < a_i \leq q-1 \end{aligned}$$

The Lee distance between two n -tuples is defined as the Lee weight of their difference. It can be easily shown that in the case of binary vectors, Hamming distance and Lee distance are the same. The block codes reviewed in this chapter are based on the concept of Hamming distance. The error correcting properties of block codes are re-

lated to the minimum Hamming distance of the codes. The concept of Lee distance and Lee weight are very useful in the study of codes employing m -ary signalling with $m > 2$. For the object identification application m -ary signalling does not appear to offer any advantage over binary signalling and, therefore, is not considered further.

5.4.2 Cyclic Codes

A subspace V of n -tuples is called a cyclic code if for each vector $\mathcal{V} = (a_0, a_1, \dots, a_{n-1})$ in V , the vector $\mathcal{V} = (a_{n-1}, a_0, \dots, a_{n-2})$, obtained by shifting the components of \mathcal{V} cyclically one unit to the right, is also in V . It will be shown that the process of applying cyclic encoding to R information digits results in $n-R$ check digits being added to the information digits at the end to give an n digit code word.

It may be shown that a cyclic code may alternatively be defined in terms of a generator polynomial $G(X)$ of degree $n-k$, i.e. of degree equal to the number of check digits. This and all subsequent polynomials are assumed to have coefficients from the field $GF(2)$ and the rules of addition and multiplication for such polynomials are that the results are reduced modulo two. In this definition,

a polynomial of degree less than n is a code polynomial if and only if it is divisible by the generator polynomial $G(X)$. With this definition, the sum of two code polynomials is also a code polynomial.

To encode a message polynomial $P(X)$, $X^{n-k} P(X)$ is firstly divided by $G(X)$. The quotient and remainder polynomials of this division being $Q(X)$ and $R(X)$ respectively. This division may be expressed by the following equation

$$X^{n-k} P(X) = Q(X)G(X) + R(X) \quad (5.12)$$

The remainder $R(X)$ is then added to $X^{n-k} P(X)$ to result in the code polynomial $F(X)$ given by

$$F(X) = X^{n-k} P(X) + R(X) = Q(X)G(X) \quad (5.13)$$

where the properties of addition modulo have been used.

$F(X)$ is a code polynomial by virtue of the fact that, as shown by Equation (5.13), it is divisible by the generator polynomial $G(X)$. The advantage of this scheme is that it results in a code polynomial $F(X)$ in which the high-order coefficients are the same as the message digits and the low-order coefficients are the same as the check digits. This is because of the fact that the division of $X^{n-k} P(X)$

by $G(X)$ results in a remainder polynomial $R(X)$ with degree less than $n-k$ and $X^{n-k}P(X)$ has zero coefficients in the $n-k$ low-order terms. This encoding technique is illustrated by the example below.

Example: Finding a Code Word for a Given Generator Polynomial

Consider a code for which block length n is 15 and the number of information digits k is 10 and which uses the generator polynomial $G(X) = 1 + X^2 + X^4 + X^5$. To encode the message 1 0 1 0 0 1 0 0 0 1 the corresponding polynomial $P(X)$ is written

$$P(X) = 1 + X^2 + X^5 + X^9$$

$X^5P(X)$ is then divided by $G(X)$. The result of the division is expressed as

$$(X^5+X^7+X^{10}+X^{14}) = (1+X^2+X^4+X^5)(1+X+X^2+X^3+X^7+X^8+X^9) + (1+X)$$

The quotient polynomial $Q(X)$ and remainder polynomial $R(X)$ are expressed below

$$Q(X) = 1 + X + X^2 + X^3 + X^7 + X^8 + X^9$$

$$R(X) = 1 + X$$

The remainder polynomial $R(X)$ is added to $X^5P(X)$ to give the code polynomial $F(X)$, which is given below

$$F(X) = (1+X) + (X^5+X^7+X^{10}+X^{14}) \quad (5.14)$$

The corresponding code vector is then expressed as

$$\underline{E} = (1\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 1)$$

Check	Information
Digits	Digits

5.4.2.1 Error Detection and Correction for Cyclic Codes

An encoded message containing errors can be represented by the polynomial

$$H(X) = F(X) + E(X) \quad (5.15)$$

where $F(X)$ is the transmitted code polynomial and $E(X)$ is a polynomial having a non-zero element in each erroneous position. Because of the modulo two addition $H(X)$ is the true encoded message with the erroneous positions changed.

If the received polynomial $H(X)$ is not divisible by the generator polynomial $G(X)$, then an error has occurred. If, on the other hand, $H(X)$ is divisible by $G(X)$, then either no error or an undetectable error has occurred. Since $F(X)$ is constructed such that it is divisible by $G(X)$, $H(X)$ is divisible by $G(X)$ if and only if $E(X)$ is divisible by $G(X)$. Therefore, if the generator polynomial is chosen such that no error polynomial $E(X)$ which is to be detected is divisible by $G(X)$, then $H(X)$ will not be divisible by $G(X)$ when such errors occur.

The ability of a code to correct errors is related to its ability to detect errors. A code which detects all combinations of $2t$ errors can correct any combination of t errors, since if t or fewer errors occur, changing all combinations of t or fewer positions results in a code polynomial only if all the erroneous positions are changed; otherwise it results in a code polynomial with $2t$ or fewer errors, which are detectable. A similar argument may be used to show that greater than t errors cannot be corrected. Hence a $2t$ error-detecting code is also a t error-correcting code. Once again, the same argument may be used to show that any code capable of detecting any two error bursts of length b or less can correct any single burst of length b or less. The converse of these statements is also true: any t error-correcting code can detect any combination of $2t$ or fewer errors and any code capable of correcting any single burst of length b can be used instead to detect any combination of two bursts of length b each.

5.4.3 BCH Codes

Bose-Chaudhuri-Hocquenghem (BCH) Codes belong to the class of linear block codes. The mathematical structure of these codes make the implementation of encoding and decoding comparatively simple.

Furthermore, the size and complexity of the hardware and software for error correction increases as a small power of the error-correcting capability of the code.

The BCH codes have some remarkable properties which make them a very powerful tool for error control in data communication systems (Bose and Chaudhuri, 1960; Hocquenghem, 1959). For positive integers m and t , a BCH code with block length 2^m-1 , that corrects all combinations of t or fewer errors and requires at most $m \cdot t$ parity check digits can be formed. It has been noticed that in some cases, depending upon the values of m and t , fewer check digits may be required.

BCH codes are a generalization of the Hamming Codes. One of the very useful properties of the BCH codes is that they are cyclic block codes (Peterson and Brown, 1961). It can be shown that the encoding process for cyclic block codes can be performed very efficiently by means of feedback shift-registers. In general, BCH codes with block lengths other than 2^m-1 may be formed, but BCH codes are most efficient, as determined by the ratio of information digits to the block length in a code word, when the codes have a block length of 2^m-1 .

5.4.3.1 Construction of BCH Codes

Consider a primitive polynomial $\mathcal{D}(X)$ of degree m with coefficients from $GF(2)$. Let α be a root of the polynomial $\mathcal{D}(X)$ and thus a primitive element of the field $GF(2^m)$. It is therefore possible to represent all the non-zero elements of $GF(2^m)$ as powers of α . A BCH code may be defined by using the primitive element α .

A BCH code of block length $n = 2^m - 1$, capable of correcting all t or fewer errors called the $BCH(n, k, t)$ with k information digits is defined by the parity check matrix \underline{M}

$$\underline{M} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \alpha & \alpha^3 & \dots & \alpha^{2t-1} \\ \alpha^2 & (\alpha^2)^3 & \dots & (\alpha^2)^{2t-1} \\ \alpha^{2^{m-1}} & (\alpha^{2^{m-2}})^3 & \dots & (\alpha^{2^{m-2}})^{2t-1} \end{bmatrix}$$

$$\equiv \begin{bmatrix} 1 & 1 & \dots & 1 \\ \alpha & \alpha^3 & \dots & \alpha^{2t-1} \\ \alpha^2 & (\alpha^3)^2 & \dots & (\alpha^{2t-1})^2 \\ \alpha^{2^{m-1}} & (\alpha^3)^{2^{m-2}} & \dots & (\alpha^{2t-1})^{2^{m-1}} \end{bmatrix} \quad (5.16)$$

The matrix \underline{M} is a $(2^m - 1) \times t$ matrix of elements from $GF(2^m)$, where each element of $GF(2^m)$ is an m -bit row vector with components from the field $GF(2)$. Therefore, the matrix \underline{M} has $(2^m - 1) \times m \times t$ components which are elements of the field $GF(2)$. The BCH

code described by the parity check matrix \underline{M} is the left null space of this matrix. In other words, a (2^m-1) bit vector is a code word if the product of that vector and the matrix is the zero vector.

The steps involved in the generation of the parity check matrix \underline{M} are:

- i) Find an irreducible polynomial $\mathcal{D}(X)$ of degree m with coefficients in $GF(2)$.
- ii) Denote α to be a root of the equation $\mathcal{D}(\alpha) = 0$.
- iii) Find all non-zero elements of the field $GF(2^m)$ by taking the first 2^m-1 powers of α .
- iv) The non-zero elements of $GF(2^m)$ form the first column of the matrix \underline{M} . The other columns of the matrix are formed by raising the elements of the first column to an appropriate power as given by Equation (5.16) and reducing the resultant polynomial using the equations

$$\mathcal{D}(\alpha) = 0 \quad (5.17a)$$

$$\alpha^{2^m-1} = 1 \quad (5.17b)$$

- v) At this stage the matrix \underline{M} is a $(2^m-1) \times m \times t$ matrix with components from $GF(2)$, but in some cases, depending on the values of m and t , a smaller but equally useful parity check matrix may be found by removing linearly dependent columns. The number of columns in this reduced matrix, which number is less than or equal to $m \cdot t$, represents the number of

check bits in the code. Hence, the upper limit on the number of check bits is $m \cdot t$. For values of m greater than four, several irreducible polynomials of degree m with coefficients in $GF(2)$ may be found. Using different irreducible polynomials results in different codes, each having the same error correcting ability, but possibly requiring different numbers of check digits to achieve the same result (Chen and Lin, 1969).

For object identification purposes, a BCH code of block length of 63 bits capable of correcting up to 3 errors is proposed. The choice of error-correcting capability of the code is largely based on the study of the environmental noise statistics in Chapter 4. The trade-off involved in selection of the error-correcting ability of the code and other system parameters is considered later in this chapter. The BCH(63,43,3) code is used in the remainder of the chapter for illustration of various mathematical operations required in the construction of the code, and for the design and implementation of encoding and decoding schemes for the BCH codes.

For the BCH(63,43,3) code,

the block length = $n = 2^m - 1 = 63$

$$\therefore m = 6$$

and the error-correction desired is $t = 3$.

\therefore the number of parity check digits = $m \cdot t = 18$
and the number of information digits = $n - m \cdot t = 45$

It has been stated that all non-zero elements of $GF(2^m)$ may be represented as powers of α . It may be shown that each element α^j of $GF(2^m)$ is a root of at least one irreducible polynomial $\mathcal{D}(X)$ of degree m or less. It then follows that for each element α^j an irreducible polynomial of the minimum degree, called the minimum polynomial, may be found for which it is a root. It may be easily shown, using modulo 2 algebra, that for any polynomial in α

$$P(\alpha^2) = [P(\alpha)]^2 \quad (5.18)$$

Therefore, if α^j is a root of an irreducible polynomial $\mathcal{D}(X)$, then $\alpha^{2j}, \alpha^{4j}, \alpha^{8j} \dots$ are all roots of $\mathcal{D}(X)$. The irreducible polynomials of degree 6 or less which have roots in $GF(2^6)$ are available from the table of irreducible polynomials (Peterson and Weldon, 1972) and are listed in Table 5.1. The roots of these polynomials are also listed. $\mathcal{D}_1(X)$ listed in the table is used in the definition of multiplication and α , which is the root of $\mathcal{D}_1(X)$ and hence a primitive element, is used in the definition of the non-zero field elements of $GF(2^6)$.

Table 5.1 Irreducible Polynomials and Their Roots
for $m=6$

DESIGNATION	POLYNOMIAL	ROOTS
$\mathcal{D}_1(X)$	$1 + X + X^6$	$\alpha, \alpha^2, \alpha^4, \alpha^8, \alpha^{16}, \alpha^{32}$
$\mathcal{D}_3(X)$	$1 + X + X^2 + X^4 + X^6$	$\alpha^3, \alpha^6, \alpha^{12}, \alpha^{24}, \alpha^{48}, \alpha^{96} = \alpha^{33}$
$\mathcal{D}_5(X)$	$1 + X + X^2 + X^5 + X^6$	$\alpha^5, \alpha^{10}, \alpha^{20}, \alpha^{40}, \alpha^{17}, \alpha^{34}$
$\mathcal{D}_7(X)$	$1 + X^3 + X^6$	$\alpha^7, \alpha^{14}, \alpha^{28}, \alpha^{56}, \alpha^{49}, \alpha^{35}$
$\mathcal{D}_9(X)$	$1 + X^2 + X^3$	$\alpha^9, \alpha^{18}, \alpha^{36}$
$\mathcal{D}_{11}(X)$	$1 + X^2 + X^3 + X^5 + X^6$	$\alpha^{11}, \alpha^{22}, \alpha^{44}, \alpha^{25}, \alpha^{50}, \alpha^{37}$
$\mathcal{D}_{21}(X)$	$1 + X + X^2$	α^{21}, α^{42}

The subscripts used in the designation of the polynomials have the significance that the polynomial $\mathcal{D}_j(X)$ is the polynomial of minimum degree which has α_j as one of the roots. The other roots are found by using the fact that if α is a root, then $\alpha^2, \alpha^4, \alpha^8, \dots$ are all roots and also by using the relationship given by Equation (5.10). Among the polynomials listed in the above table $\mathcal{D}_1(X), \mathcal{D}_3(X), \mathcal{D}_5(X), \mathcal{D}_7(X)$ and $\mathcal{D}_{11}(X)$ are of degree 6 and $\mathcal{D}_1(X), \mathcal{D}_5(X)$ and $\mathcal{D}_{11}(X)$ are primitive. Irreducible polynomials with subscripts other than those shown are not listed because these polynomials have already appeared with a subscript higher in the table.

The representation of the non-zero field elements of $GF(2^6)$ using the irreducible polynomial $\mathcal{D}_1(X)$ given in the above table and the solution

$\alpha = (0.10000)^*$ are listed in Table 5.2. The complete set of non-zero field elements and the parity check matrix \underline{M} for the BCH(63,45,3) code have been computed and listed in Table 5.3.

5.4.3.2 Encoding of BCH Codes

It may be shown that a basic property of the BCH codes is that for a BCH(n, k, t) code, a code polynomial $F(X)$ is a code polynomial if and only if $(\alpha, \alpha^2, \alpha^3 \dots \alpha^{2t})$ are all roots of the polynomial $F(X)$. As described earlier, every even power of α is a root of the same minimum polynomial as some lower odd powers of α . Therefore, an equivalent requirement is that $F(X)$ is a code polynomial if and only if $(\alpha, \alpha^3, \dots \alpha^{2t-1})$ are all roots of the polynomial. This definition of a code polynomial may also be derived using the property of BCH code that the code polynomials form the left null space of the parity check matrix \underline{M} . Let the code vector be $\underline{L} = (\delta_0, \delta_1, \dots \delta_{n-1})$, which is represented by the code polynomial $F(X)$,

$$F(X) = \delta_0 + \delta_1 X + \dots + \delta_{n-1} X^{n-1} \quad (5.19)$$

where each δ_i is an element of the field $GF(2)$

*Other solutions for the primitive element α are possible and result in different representations which turn out to be mere permutations of the one given.

Table 5.2 Representation of $GF(2^6)$

α^0	$= 1$	$= 1 0 0 0 0 0$
α^1	$= \alpha$	$= 0 1 0 0 0 0$
α^2	$= \alpha^2$	$= 0 0 1 0 0 0$
α^3	$= \alpha^3$	$= 0 0 0 1 0 0$
α^4	$= \alpha^4$	$= 0 0 0 0 1 0$
α^5	$= \alpha^5$	$= 0 0 0 0 0 1$
α^6	$= 1 + \alpha$	$= 1 1 0 0 0 0$
α^7	$= \alpha + \alpha^2$	$= 0 1 1 0 0 0$
α^8	$= \alpha^2 + \alpha^3$	$= 0 0 1 1 0 0$
α^9	$= \alpha^3 + \alpha^4$	$= 0 0 0 1 1 0$
α^{10}	$= \alpha^4 + \alpha^5$	$= 0 0 0 0 1 1$
α^{11}	$= 1 + \alpha + \alpha^5$	$= 1 1 0 0 0 1$
α^{12}	$= 1 + \alpha^2$	$= 1 0 1 0 0 0$
.		
.		
.		
α^{60}	$= 1 + \alpha^3 + \alpha^4 + \alpha^5$	$= 1 0 0 1 1 1$
α^{61}	$= 1 + \alpha^4 + \alpha^5$	$= 1 0 0 0 1 1$
α^{62}	$= 1 + \alpha^5$	$= 1 0 0 0 0 1$
α^{63}	$= 1$	$= \alpha^0$

Table 5.3 Parity Check Matrix for BCH(63,45,3) Code

```

10000010000001000000
0100000001000000001
001000110000000011
0001000000100000101
0000101010000001111
0000010001010100001
110000111100110011
011000110111110100
001100100010111101
000110011100100110
110001010010011101
10100001010100111
0101001011101100
000101100010100011
110101011100000100
101010110011001100
010101010010010100
110101011010111100
011101011011100101
111110010111001110
0111110010110010010
111111101100110110
101111110101110111
100111111010101100
100011011111010101
100001100111111111

```

Because \underline{F} is a code vector

$$\underline{F} \cdot \underline{M} = Q \quad (5.20)$$

Using the definition of \underline{M} given in Equation (5.16), the above equation yields

$$F(\alpha) = F(\alpha^3) = \dots = F(\alpha^{2^t-1}) = Q \quad (5.21)$$

It is also clear that if α satisfies all of Equation (5.21), then Equation (5.20) is also satisfied. The above result implies that the BCH(n, k, t) code, defined by the parity check matrix \underline{M} , consists of all polynomials of degree $n = 2^m - 1$ which have all of $\alpha, \alpha^3, \dots, \alpha^{2^t-1}$ as roots. Using the property of cyclic codes given by Equation (5.13), it may be shown that a polynomial $G(X)$ is a generator polynomial for the BCH code defined by the parity check matrix \underline{M} if and only if $\alpha, \alpha^3, \dots, \alpha^{2^t-1}$ are roots of $G(X)$. As described earlier, if α is a root of a polynomial, then $\alpha^2, \alpha^4, \alpha^8, \dots$ are also roots of the same polynomial. It may also be recalled that each element α^j of the field $GF(2^m)$ is a root of a unique minimum polynomial $D_j(\alpha)$. Using these properties, it can be seen that the generator polynomial must be divisible by each of the irreducible polynomials $D_1(X), D_3(X), \dots, D_{2^t-1}(X)$ and, hence, by their least common multiple:

$$G(X) = \text{LCM}[\mathcal{D}_j(X)] \quad (5.22)$$

$$j=1,3,\dots,2t-1$$

It has been described earlier that each $\mathcal{D}_j(\alpha)$ has roots in $GF(2^m)$ and therefore may be expressed as a product of the form

$$\mathcal{D}_j(\alpha) = (\alpha - \beta_1)(\alpha - \beta_2) \dots (\alpha - \beta_d) \quad (5.23)$$

The number of factors in the above equation depends upon the degree of $\mathcal{D}_j(\alpha)$.

When the generator polynomial given by Equation (5.22) is expressed as the product of linear factors, over $GF(2^m)$, duplicate factors may occur and some of the columns of the parity check matrix are linearly dependent. For any two elements of $GF(2^m)$ α^i and α^j that are roots of the same irreducible polynomial $\mathcal{D}_i(\alpha)$, duplication occurs because α^j is also a root of the polynomial $\mathcal{D}_j(\alpha)$. In a situation like this, the parity checks produced by the column of powers of α^j in matrix \underline{M} will be satisfied if and only if the parity checks produced by the column of powers of α^i are satisfied, and thus one of the parity check equations is unnecessary. The number of check digits required to provide t error correction, in the situation described above, for a code with block length 2^m-1 , is less than $m \cdot t$.

The BCH(63,45,3) code has a block length of $n = 63$, number of information digits $k = 45$, and error correcting capability $t = 3$ and, thus, requires $m \cdot t = 18$ check digits. The generator polynomial for such a code is written as

$$G(X) = \mathcal{D}_1(X) \cdot \mathcal{D}_3(X) \cdot \mathcal{D}_5(X) \quad (5.24)$$

where $\mathcal{D}_1(X)$, $\mathcal{D}_3(X)$ and $\mathcal{D}_5(X)$ are the irreducible polynomial of degree 6 over $GF(2)$ with roots listed in Table 5.1. In this situation the polynomials $\mathcal{D}_1(X)$, $\mathcal{D}_3(X)$ and $\mathcal{D}_5(X)$ have all independent roots. Therefore, the LCM in Equation (5.22) yields the product given in Equation (5.24).

The number of parity check digits in a code is given by the total number of α 's that are roots of $G(X)$. In the case of the BCH(63,45,3) code, the code generator polynomial has 18 roots; therefore, the number of parity check bits for this code is 18.

The code generator polynomial $G(X)$ may be formed by first selecting an irreducible polynomial of appropriate degree such that α is a root of this polynomial and the first $2^m - 1$ powers of α define all the non-zero elements of the field $GF(2^m)$. For the BCH(63,45,3) code $m = 6$, and the irreducible polynomial of degree 6 with coefficients from $GF(2)$ is selected to be

$$D_1(X) = 1 + X + X^6 \quad (5.25)$$

Although it is possible to use the tables of irreducible polynomials such as the one given in Table 5.1, it is also possible to find the other polynomial as described below.

The irreducible polynomial $D_2(X)$ has the general form

$$D_3(X) = \sum_{i=0}^6 d_i X^i \quad (5.26)$$

where each d_i belongs to $GF(2)$

And all roots of $D_3(X)$ described earlier must satisfy the equation

$$D_3(\alpha) = 0$$

The set of simultaneous equations formed by substituting the roots $\alpha^3, \alpha^6, \alpha^{12}, \alpha^{24}, \alpha^{48}$, and α^{33} in the equation

$$\sum_{i=0}^6 d_i X^i = 0 \quad (5.27)$$

may be solved to find the coefficients d_i and, hence, $D_3(X)$. $D_5(X)$ may be found by following a similar procedure. $D_3(X)$ and $D_5(X)$ are thus found to be

$$D_3(X) = 1 + X + X^2 + X^4 + X^6 \quad (5.28)$$

$$\text{and } D_3(X) = 1 + X + X^2 + X^5 + X^6 \quad (5.29)$$

Hence, the code generator polynomial may be written as

$$G(X) = (1+X+X^6)(1+X+X^2+X^4+X^6)(1+X+X^2+X^5+X^6)$$

or $G(X) = 1+X+X^2+X^3+X^6+X^7+X^9+X^{15}+X^{16}+X^{17}+X^{18}$

(5.30)

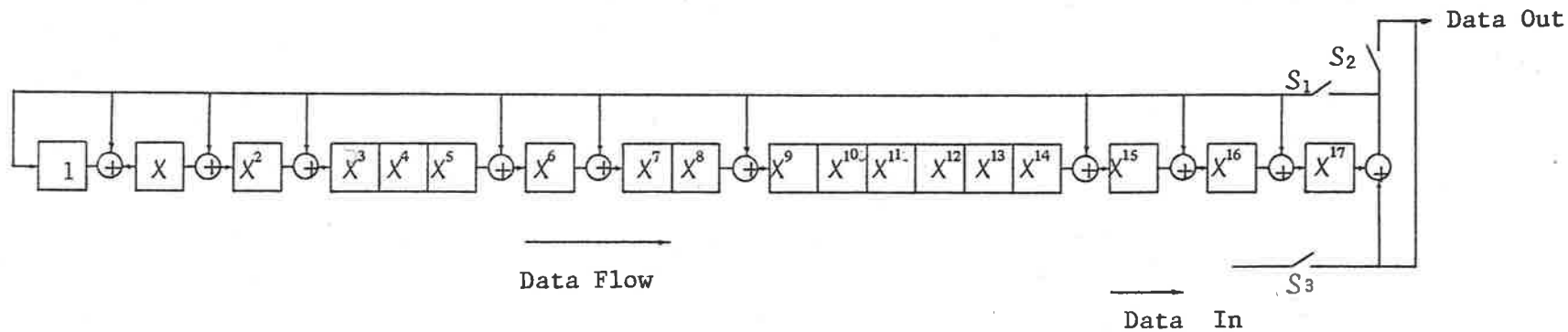
This code generator polynomial may alternatively be written as the vector \mathcal{G} comprising the coefficients of the polynomial $G(X)$

$$\mathcal{G} = 1111000001011001111 \quad (5.31)$$

5.4.3.3 Encoder Design

The implementation of an encoder for the BCH(63,45,3) code described by the code generator polynomial given by Equation (5.30) is based on the definition of a cyclic code given by Equation (5.13). The encoder for the BCH(63,45,3) code using an 18-stage feedback shift-register arrangement is shown in Figure 5.4.

The effectiveness of any error detection and correction strategy depends critically upon the interconnection of the stages. If there are no feedback loops, the shift register functions as



Switch Position: Parity Generation - S_1, S_3 closed and S_2 open
 Parity Output - S_1, S_3 open and S_2 closed

Figure 5.4 A Shift Register Encoder for the BCH(63,45,3) Code Using the Code Generator Polynomial $G(X) = 1 + X^2 + X^3 + X^6 + X^7 + X^9 + X^{15} + X^{16} + X^{17} + X^{18}$

merely a storage device and repeats a part of the message. This encoding strategy does not utilise the available hardware optimally. A code word in which only a portion of the message is repeated is capable of detecting errors in the portion of the message which has been repeated and not capable of detecting errors in the rest of the message. The use of appropriate feedback strategy enables detection of multiple errors independent of their bit positions. The encoder shown in Figure 5.4 uses a 45-bit information sequence to generate a 63-bit code word. Initially, the data is fed into the shift-register and the output line by closing the switches S_1 and S_3 and opening the switch S_2 . Once the message is complete, the parity check bits generated by the shift-register are shifted out of the shift register by closing the switch S_2 and opening the switches S_1 and S_3 . It may be shown that the code word thus generated has a Hamming distance of seven and, therefore, as explained in Section 5.4.1, it is capable of correcting up to three errors.

An alternative approach to the design of the encoder, in which the number of storage elements required is equal to the number of information digits, is also possible. A comparison of the number of information digits with the number of parity

check digits gives an indication of the efficiency of the two encoders described above. For the object identification application, the number of check bits is less than the number of information bits and, hence, the implementation of the encoder shown in Figure 5.4 offers an economical solution to the encoding problem for the BCH(63,45,3) code.

5.4.3.4 Decoding of BCH Codes

The decoding of BCH codes is much more complex than the encoding. The complexity of the problem, the nature of outputs required from a decoder, and the desired simplicity of the decoder require that the decoding problem be divided into several smaller problems, as described below.

Given an n -bit received vector \underline{R} and corresponding polynomial $R(X)$, the task of the decoding equipment involves, first of all, finding whether the received word has any errors or not. If the received word has a number of errors, then the decoder must find out how many errors have occurred. When the number of errors in the received word is greater than the error-detecting and correcting ability of the code, as illustrated in the example in Section 5.4.1, the decoding strategy fails. An indication of such a condition should be provided at the decoder output.

Finally, if error correction is also required, then the decoder must find the error-bit positions.

The decoding strategy developed here is based on the definitions of the BCH codes given by Equations (5.13) and (5.20). Let $F(X)$, $E(X)$ and $R(X)$ be the transmitted, error and received polynomials, respectively. The corresponding n -bit row vectors are \underline{F} , \underline{E} and \underline{R} . The received polynomial is the sum of the code and error polynomials:

$$\underline{R} = \underline{F} + \underline{E} \quad (5.32)$$

where the error vector \underline{E} has non-zero coefficients at the bit positions where errors have occurred. The first step in decoding is to generate an error-syndrome polynomial $S(X)$ by generating the corresponding syndrome vector \underline{S} which is the product of the received vector and the parity check matrix.

$$\underline{S} = \underline{R} \cdot \underline{M} \quad (5.33)$$

Substituting the value of R from Equation (5.32) and using the definition of the code vector in Equation (5.20) results in

$$\underline{S} = \underline{E} \cdot \underline{M} \quad (5.34)$$

It is apparent from the above equation that the syndrome vector is a zero vector when no errors occur, i.e. $\underline{S} = \underline{0}$ when $\underline{E} = \underline{0}$.

This property of the syndrome vector is used in the first step of decoding to determine whether or not any errors have occurred during the transmission of the coded message. The occurrence of errors is detected by a non-zero syndrome vector. It is necessary to understand the characteristics of the error-syndrome vector in order to formulate a decoding strategy, namely, to extract the redundant information provided by the parity check bits. This is explained below.

If the errors have occurred in bit positions (i_1, i_2, \dots, i_v) , then the error vector \underline{E}

$$\underline{E} = (e_1, e_2, \dots, e_n) \quad (5.35)$$

$$\begin{aligned} \text{where } e_i &= 1 && \text{for } i=i_1, i_2, \dots, i_v \\ &= 0 && \text{otherwise} \end{aligned}$$

A comparison of the right-hand expression of Equation (5.34) using the error vector given in Equation (5.35) and the elements of $GF(2^m)$ which constitute the first column of the parity check matrix \underline{M} given by equation (5.16) reveals a one-to-one correspondance between the error positions and

the elements of $GF(2^m)$. This correspondance leads to the definition of error positions under i_j and the corresponding error position vector X_j^* is defined as the element α^{i_j} of the field $GF(2^m)$ as this element results in contributions to a non-zero syndrome vector when an error at position i_j occurs.

For the further useful development of the properties of the syndrome vector for use in decoding of BCH codes, symmetric polynomial and elementary symmetric functions are defined below.

Consider a commutative ring of polynomials. A polynomial $P(X_1, X_2 \dots X_n)$ in n indeterminates X_i is called a symmetric polynomial if the polynomial remains unchanged by any permutation of the indeterminates $(X_1, X_2 \dots X_n)$. Examples of such symmetric polynomials include sum, product, sum of powers and many others.

Consider a symmetric polynomial expressed in the product form with the expansion given by

$$\begin{aligned} P(X) &= (X-X_1)(X-X_2) \dots (X-X_n) \\ &= X^n - \sigma_1 X^{n-1} + \sigma_2 X^{n-2} - \dots + (-1)^n \sigma_n \end{aligned}$$

*In literature, X_j 's are somewhat inappropriately defined as the error position numbers (Peterson and Weldon, 1972).

The above equation may also be written as

$$P(X) = \prod_{k=1}^n (X - X_k) = X^n + \sum_{k=1}^n (-1)^k \sigma_k X^{n-k}$$

The coefficients of the powers of X in the above equation are given by

$$\sigma_1 = \sum_i X_i$$

$$\sigma_2 = \sum_{i < j} X_i X_j$$

$$\sigma_3 = \sum_{i < j < k} X_i X_j X_k$$

$$\sigma_n = X_1 X_2 \dots X_n$$

The function σ_i 's are called the elementary symmetric functions of (X_1, X_2, \dots, X_n) . Some of the most useful properties of elementary symmetric functions are derived from the so-called Fundamental Theorem on Symmetric Polynomials, which states: Any symmetric polynomial $P(X_1, \dots, X_n)$ in n indeterminates can be expressed as a polynomial in the elementary symmetric functions. It may be shown that for a given symmetric polynomial there is a unique representation in terms of elementary symmetric functions.

For decoding of BCH codes used in the object identification system described in this thesis, X, X_i 's and σ_i 's are all elements of $GF(2^m)$.

Returning to the development of properties of a syndrome vector, we note that when the error vector has several non-zero elements, as defined in Equation (5.35), the $m \cdot t$ -bit syndrome vector consists of t m -bit vectors of the form

$$\underline{S} = (S_1, S_2, \dots, S_{2t-1}) \quad (5.36)$$

$$\text{where } S_\ell = \sum_{j=1}^v (\alpha^j)^{\ell} \quad (5.37)$$

It may be noted that only odd index S_ℓ occur in \underline{S} . For the further development of the theory, the even index S_ℓ will be needed. Because of the modulo-2 addition in the field $GF(2^m)$,

$$(\alpha + \beta)^2 = \alpha^2 + \beta^2$$

where α, β are elements of $GF(2^m)$.

Using this property, the element S_ℓ for even ℓ may be expressed in terms of S_ℓ with odd indices, as shown below.

$$\begin{aligned} (S_\ell)^2 &= \left[\sum_{j=1}^v (\alpha^j)^{\ell} \right]^2 && ; \ell = 1, 3 \dots 2t-1 \\ S_{2\ell} &= \sum_{j=1}^v (\alpha^j)^{2\ell} = S_{2\ell} && ; \ell = 1, 3 \dots 2t-1 \end{aligned} \quad (5.38)$$

Hence,

$$\begin{aligned} S_2 &= S_1^2 \\ S_4 &= S_2^2 = S_1^4 \\ S_6 &= S_3^2 \quad \text{and so on.} \end{aligned}$$

The j th error position vector \underline{x}_j has been defined as

$$\underline{x}_j = \alpha^{i_j} \quad (5.39)$$

Using this notation, Equations (5.37) and (5.38) are rewritten, for the case when t errors occur, as

$$\underline{s}_\ell = \sum_{j=1}^t \underline{x}_j^\ell \quad (5.40)$$

$$\underline{s}_\ell^2 = \sum_{j=1}^t \underline{x}_j^{2\ell} = \underline{s}_{2\ell} \quad (5.41)$$

where $\ell = 1, 3, \dots, 2t-1$

Equation (5.40) represents a set of t equations in t unknowns, the \underline{x}_j 's. The exhaustive search method in which all combinations of t of 2^m-1 field elements of $GF(2^m)$ are examined as possible solutions requires a large amount of computing time which increases exponentially when either m or t or both increase.

One approach to solve the equation utilises a useful property of the syndrome vector. It may be easily shown that the component vectors \underline{s}_ℓ of the syndrome vector are power sum symmetric functions. Thus, the parity checks resulting in the syndrome vector given by Equation (5.40) give the first t

odd power sum symmetric functions, and the first t even power sum symmetric functions are given by Equation (5.41). It may be shown that the elementary symmetric functions σ_i are related to the power sum symmetric functions S_i by Newton's identities given below.

$$\begin{aligned}
 S_1 - \sigma_1 &= 0 \\
 S_2 - S_1 \sigma_1 + 2\sigma_2 &= 0 \\
 S_3 - S_2 \sigma_1 + S_1 \sigma_2 - 3\sigma_3 &= 0 \\
 S_4 - S_3 \sigma_1 + S_2 \sigma_2 - S_1 \sigma_3 + 3\sigma_4 &= 0
 \end{aligned}
 \tag{5.42}$$

One approach to solve for the error position vectors X_j which requires considerably less computation than the exhaustive search method (Peterson, 1960) is to solve the Newton's identities for the elementary symmetric functions σ_i . The error position numbers i_j and corresponding error position vectors X_j are then found by substituting each non-zero field element of $GF(2^m)$, one at a time, into Equation (5.43) given below. If the equation is satisfied, the bit is erroneous and must be inverted to correct it. If the equation is not satisfied, the bit is correct.

$$X^t - \sigma_1 X^{t-1} + \sigma_2 X^{t-2} - \sigma_3 X^{t-3} \dots + (-1)^t \sigma_t = 0
 \tag{5.43}$$

The technique of error correction outlined above is illustrated with an example of BCH(63,45,3) code. The parity check matrix for the code is given in Table 5.3. This code is capable of correcting up to 3 errors.

Example: Decoding of BCH Code

Let the transmitted vector be \underline{E} , a vector of all zeros.

$$\underline{E} = (0\ 0\ 0\ \dots\ 0)$$

Let the errors to have occurred in bit positions one, four and seven. Then the 63-bit error and the received vectors have three non-zero elements, i.e. at 1st, 4th and 7th bit positions

$$\underline{R} = \underline{E} = (1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ \dots\ 0\ 0\ 0\ 0)$$

Then, using Equation (5.33) and the parity check matrix given in Table (5.3), the 18-bit syndrome vector may be calculated as

$$\begin{aligned} \underline{S} &= (0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0) \\ &= (S_1\ S_3\ S_5) \end{aligned}$$

where

$$\begin{aligned} S_1 &= 0\ 1\ 0\ 1\ 0\ 0 \\ S_3 &= 0\ 1\ 1\ 0\ 1\ 0 \\ S_5 &= 0\ 1\ 0\ 1\ 1\ 0 \end{aligned} \tag{5.44}$$

A comparison of $\mathfrak{S}_1, \mathfrak{S}_3$ and \mathfrak{S}_5 with the elements of $GF(2^6)$ listed in Table 5.3 shows that $\mathfrak{S}_1 = \alpha^{13}$, $\mathfrak{S}_3 = \alpha^{36}$ and $\mathfrak{S}_5 = \alpha^{49}$. These \mathfrak{S}_i are the odd power sum symmetric functions. The even power sum symmetric functions are calculated using Table 5.3, as shown below.

$$\begin{aligned}\mathfrak{S}_2 &= \mathfrak{S}_1^2 = (\alpha^{13})^2 = \alpha^{26} = 111000 \\ \mathfrak{S}_4 &= \mathfrak{S}_1^4 = (\alpha^{13})^4 = \alpha^{52} = 101010 \quad (5.45) \\ \mathfrak{S}_6 &= \mathfrak{S}_3^2 = (\alpha^{36})^2 = \alpha^{72} = \alpha^9 = 000110\end{aligned}$$

The Newton's identities are solved for the elementary symmetric functions. The Newton's identities appropriate for this case are given below

$$\mathfrak{S}_1 - \sigma_1 = 0 \quad (5.46a)$$

$$\mathfrak{S}_2 - \mathfrak{S}_1\sigma_1 + 2\sigma_2 = 0 \quad (5.46b)$$

$$\mathfrak{S}_3 - \mathfrak{S}_2\sigma_1 + \mathfrak{S}_1\sigma_2 - 3\sigma_3 = 0 \quad (5.46c)$$

$$\mathfrak{S}_4 - \mathfrak{S}_3\sigma_1 + \mathfrak{S}_2\sigma_2 - \mathfrak{S}_1\sigma_3 + 4\sigma_4 = 0 \quad (5.46d)$$

$$\mathfrak{S}_5 - \mathfrak{S}_4\sigma_1 + \mathfrak{S}_3\sigma_2 - \mathfrak{S}_2\sigma_3 + \mathfrak{S}_1\sigma_4 - 5\sigma_5 = 0 \quad (5.46e)$$

The general set of equations given in (5.46) can usually be reduced to a smaller set of equations using various properties of elementary symmetric functions and power-sum symmetric functions. First of all, a consequence of the use of modulo-2 algebra is that addition and subtraction in Equation (5.46) are equivalent. Further, by substituting the values of \mathfrak{S}_2 and \mathfrak{S}_4 from Equation (5.45), it may be shown

that Equation (5.46b) and (5.46d) are equivalent to (5.46a) and (5.46c), respectively. Further, it can be shown that if only p errors occur, the elementary symmetric functions of order greater than p are all zero.

The Newton's identities for this example may then be written as

$$\mathfrak{S}_1 + \mathfrak{Q}_1 = 0 \quad (5.47a)$$

$$\mathfrak{S}_3 + \mathfrak{S}_1^2 \mathfrak{Q}_1 + \mathfrak{S}_1 \mathfrak{Q}_2 + \mathfrak{Q}_3 = 0 \quad (5.47b)$$

$$\mathfrak{S}_5 + \mathfrak{S}_1^4 \mathfrak{Q}_1 + \mathfrak{S}_3 \mathfrak{Q}_2 + \mathfrak{S}_1^2 \mathfrak{Q}_3 + \mathfrak{S}_1 \mathfrak{Q}_4 = 0 \quad (5.47c)$$

Solving the above set of simultaneous equations for the situation when three errors have occurred, we obtain

$$\begin{aligned} \mathfrak{Q}_1 &= \mathfrak{S}_1 \\ \mathfrak{Q}_2 &= (\mathfrak{S}_1^2 \mathfrak{S}_3 + \mathfrak{S}_5) / (\mathfrak{S}_1^3 + \mathfrak{S}_3) \\ \mathfrak{Q}_3 &= (\mathfrak{S}_1 \mathfrak{S}_5 + \mathfrak{S}_3^2 + \mathfrak{S}_1^3 \mathfrak{S}_1 + \mathfrak{S}_1^6) / (\mathfrak{S}_1^3 + \mathfrak{S}_3) \end{aligned} \quad (5.48)$$

If, on the other hand, only one error has occurred, then

$$\mathfrak{S}_1^3 + \mathfrak{S}_3 = 0 \quad (5.49)$$

The importance of expressing all elements of $GF(2^m)$ in terms of powers of α may now be apparent. The division of vectors in Equation (5.48) merely

reduces to subtraction of powers of α once the values of S_{λ} 's are substituted from Equations (5.44) and (5.45) into Equation (5.48). The elementary symmetric functions σ_{λ} 's are then calculated to be

$$\begin{aligned} \sigma_1 &= 010100 = \alpha^{13} \\ \sigma_2 &= 110010 = \alpha^{16} \\ \sigma_3 &= 000110 = \alpha^9 \end{aligned} \quad (5.50)$$

The next step in the decoding procedure is to determine the error-bit locations, which is done by substituting the elements of $GF(2^6)$ in the equation

$$X^3 + \sigma_1 X^2 + \sigma_2 X + \sigma_3 = 0 \quad (5.51)$$

The above equation is satisfied for $X = \alpha^0, \alpha^3$ and α^6 . The above equation may therefore be written as

$$(X + X_1)(X + X_2)(X + X_3) = (X + \alpha^0)(X + \alpha^3)(X + \alpha^6) = 0 \quad (5.52)$$

Therefore, the error position vectors X_{λ} 's are α^0, α^3 and α^6 , indicating error-bit position numbers as 1st, 4th and 7th. This technique reduces the maximum number of comparisons required to locate the error bit positions from ${}^n C_3$ to ${}^n C_1$.

5.4.3.5 Decoder Building Blocks

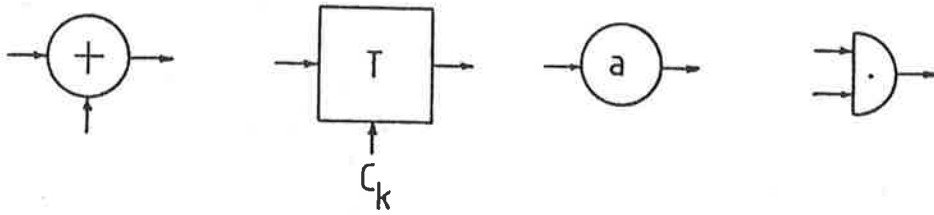
The encoding and decoding equipment for the BCH codes consist of linear finite-state switching cir-

cuits. As explained earlier, the information contained in the BCH codes is usually represented by elements of $GF(2^m)$.

There are four basic building blocks normally used in the decoding hardware. Elementary operations may be performed on the elements of $GF(2)$ by means of these devices. The hardware to perform various arithmetic operations on the elements of $GF(2^m)$ may be built by an appropriate interconnection of a number of these elementary devices.

The first building block is an adder which has two inputs and one output, the output being the sum of two inputs. The second is a delay or storage element which has one input and one output, the output at any instant of time being the same as the input one unit of time earlier. The third type of device is a constant multiplier. This single input single output device multiplies an element of the field $GF(2)$ by a constant which is also an element of $GF(2)$ and, finally, a variable multiplier is a two-input single output device which multiplies two elements of the field $GF(2)$ and results in an element of the field $GF(2)$. The rule of interconnection of these basic building blocks is that any number of inputs may be connected to any output, but two outputs are never connected together. These

basic elements are represented schematically in Figure 5.5 below.



a. Adder b. Storage Device c. Constant Multiplier d. Variable Multiplier

Figure 5.5 Basic Building Blocks for Linear Switching Circuits

A linear finite-state switching circuit is one which consists of a finite number of adders, memory devices and constant multipliers connected according to the rules of interconnections. The field $GF(2)$ consists of only two elements, namely a 1 and a 0. In this situation, the adder is an exclusive-or gate, the delay device is a single stage shift-register and the constant multiplier is a closed switch when the constant is a 1 and an open switch when the constant is a 0.

The elements of higher order fields such as $GF(2^m)$ are conveniently expressed as polynomials of

order m with coefficients from the field $GF(2)$. In a serial processor, the high-order coefficients are transmitted first. On the input-output lines of a serial processor, the elements of the field $GF(2^m)$ or the corresponding polynomial would appear as a sequence of m elements of the field $GF(2)$ with the highest order coefficient appearing first in time, the next highest coefficient appearing on the line one unit of time later, and so on. In a parallel processor, on the other hand, all the coefficients of the polynomial appear on the m -line bus at the same instant of time.

A circuit configuration to perform addition of two elements of $GF(2^m)$ is shown in Figure 5.6 below.

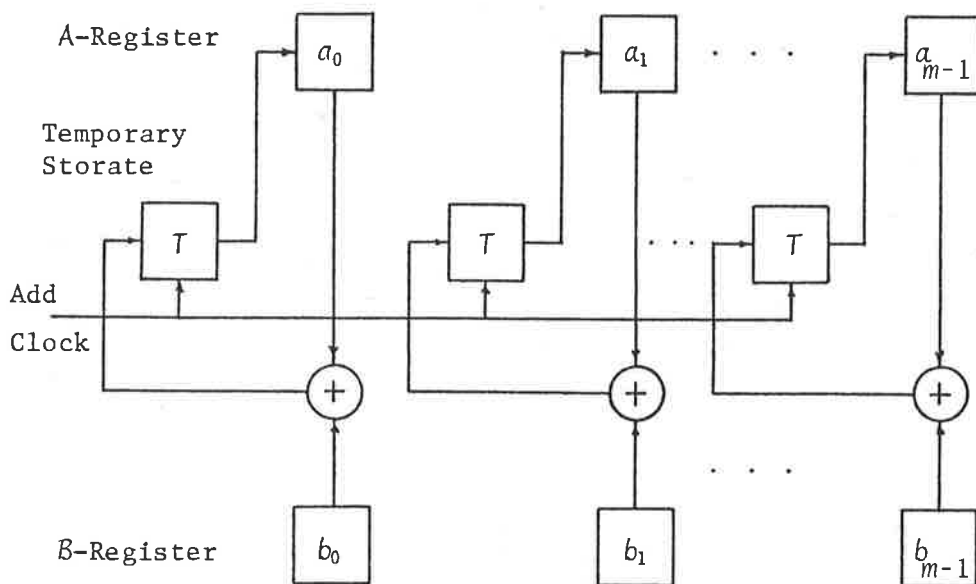


Figure 5.6 Addition of Elements of $GF(2^m)$

The two elements to be added are stored in two registers. The addition is performed by means of an appropriate number of basic adders in $GF(2)$ and the result is temporarily stored in an m -bit delay line. When the add signal appears, the output of the delay line is shifted back to one of the registers. Let

$$A(X) = a_0 + a_1 X + a_2 X^2 + \dots + a_{m-1} X^{m-1} \quad (5.53)$$

and $B(X) = b_0 + b_1 X + b_2 X^2 + \dots + b_{m-1} X^{m-1} \quad (5.54)$

then the sum polynomial $C(X)$ is defined as

$$C(X) = c_0 + c_1 X + c_2 X^2 + \dots + c_{m-1} X^{m-1} \quad (5.55)$$

where $c_i = a_i + b_i \quad i=0,1,\dots,m-1 \quad (5.56)$

There are two types of multiplication which are normally used in the linear algebra. In an open multiplication of two polynomials $A(X)$ and $B(X)$ of degree p and q , respectively, result in a product polynomial $C(X)$ of degree $p+q$. Let

$$A(X) = a_0 + a_1 X + a_2 X^2 + \dots + a_p X^p \quad (5.57)$$

$$B(X) = b_0 + b_1 X + b_2 X^2 + \dots + b_q X^q \quad (5.58)$$

then the product polynomial

$$\begin{aligned} C(X) &= A(X) \cdot B(X) \\ &= c_0 + c_1 X + c_2 X^2 + \dots + c_r X^r \end{aligned} \quad (5.59)$$

where $n = p + q$ and

$$\begin{aligned}
 c_0 &= a_0 \cdot b_0 \\
 c_1 &= (a_0 b_1 + a_1 b_0) \\
 c_2 &= (a_0 b_2 + a_1 b_1 + a_2 b_0) \\
 &\vdots \\
 c_{n-1} &= (a_{p-1} \cdot b_q + a_p \cdot b_{q-1}) \\
 c_n &= a_p \cdot b_q
 \end{aligned}
 \tag{5.60}$$

One circuit configuration for an open multiplier is shown below in Figure 5.7. The coefficients of polynomials $A(X)$ and $B(X)$ are stored in two shift-registers with p and q storage elements, respectively. The coefficients of the polynomial

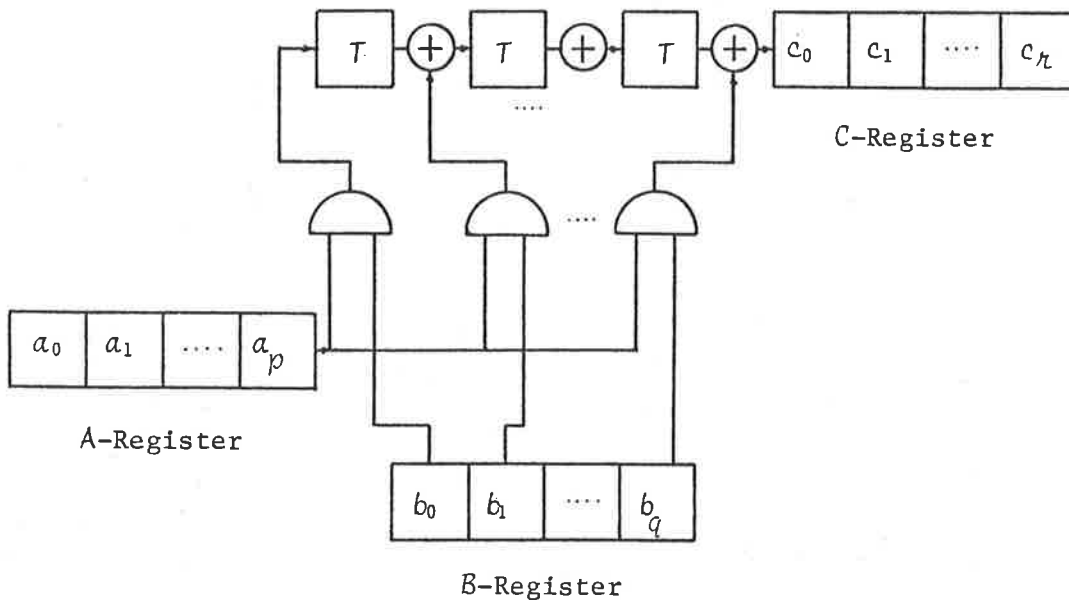


Figure 5.7 Multiplication of Two Polynomials

$A(X)$ are fed to the input of the multiplier serially such that highest order coefficients of $A(X)$ appear on the line first. The coefficients of the polynomial $B(X)$ are fed to the input of the multiplier in parallel. The product polynomial is stored in an n -stage shift-register. Once all the coefficients of the polynomial $B(X)$ have been fed to the line, the serial input is disabled and zeros are fed to that line.

To make the circuit configuration simple, it is assumed that $p \leq q$. Initially, the storage elements in the multiplier are all cleared. With the first clock pulse, the output is $a_p b_0$, and the storage devices in the multiplier contain $a_p b_0$, $a_p b_1 \dots a_p b_{q-1}$. The output after the second clock pulse is $a_{p-1} b_q + a_p b_{q-1}$ and the storage devices contain $a_{p-1} b_0$, $(a_{p-1} b_1 + a_p b_0)$, $(a_{p-1} b_2 + a_p b_1)$, \dots , $(a_{p-1} b_{q-1} + a_p b_{q-1})$. The operation continues in a similar manner, and the output is stored in a shift-register. It is obvious that the result of this multiplication is completely stored in the output register in $p+q$ clock cycles (Peterson and Heldon, 1972).

The multiplication may also be performed in a parallel processor in which multiplication is completed in one clock cycle. Combinatorial logic is

used to generate the coefficients of the product polynomial and all the coefficients are available on the output lines at the same time.

In a closed multiplication, two polynomials of degree less than or equal to n are multiplied to result in a polynomial of degree n . For instance, when two elements of the field $GF(2^m)$ are multiplied, the resultant element is also an element of the field $GF(2^m)$. This multiplication is usually performed in two steps. First, an open multiplication of the two polynomials corresponding to the elements of $GF(2^m)$ is performed and then one of the primitive polynomials associated with the field $GF(2^m)$ is used to convert this product into a polynomial of degree less than m . Closed multiplication is illustrated below by an example in which two elements of $GF(2^6)$ are multiplied.

Example: Closed Multiplication in $GF(2^6)$

$$A(X) = \sum_{i=0}^5 a_i X^i \quad (5.61)$$

$$B(X) = \sum_{i=0}^5 b_i X^i \quad (5.62)$$

then the open product $C(X) = A(X) \cdot B(X)$

or

$$C(X) = \sum_{i=0}^{10} c_i X^i \quad (5.63)$$

where coefficients c_i are given by Equation (5.60)

A primitive polynomial for the field $GF(2^6)$ is

$$\mathcal{D}_1(X) = 1 + X + X^6 \quad (5.64)$$

The method of reducing a polynomial of degree greater than or equal to m using an irreducible polynomial of degree m has already been discussed. Polynomial multiples of $\mathcal{D}_1(X)$ are added to the polynomial of degree less than 6. The polynomial multiples necessary to perform the task in this case are given below

$$\begin{aligned} X \mathcal{D}_1(X) &= X + X^2 + X^7 \\ X^2 \mathcal{D}_1(X) &= X^2 + X^3 + X^8 \\ X^3 \mathcal{D}_1(X) &= X^3 + X^4 + X^9 \\ X^4 \mathcal{D}_1(X) &= X^4 + X^5 + X^{10} \end{aligned} \quad (5.65)$$

Addition of appropriate scalar multiples of the polynomials given in Equations (5.64) and (5.65) to (5.63) eliminates the powers of X greater than 6 and yields

$$\begin{aligned} C(X) &= (C_0 + C_6) + (C_1 + C_6 + C_7)X + (C_2 + C_7 + C_8)X^2 \\ &+ (C_3 + C_8 + C_9)X^3 + (C_4 + C_9 + C_{10})X^4 + (C_5 + C_{10})X^5 \end{aligned} \quad (5.66)$$

or

$$C(X) = \sum_{i=0}^5 d_i X^i \quad (5.67)$$

where

$$\begin{aligned}
 d_0 &= C_0 + C_6 \\
 d_1 &= C_1 + C_6 + C_7 \\
 d_2 &= C_2 + C_7 + C_8 \\
 d_3 &= C_3 + C_8 + C_9 \\
 d_4 &= C_4 + C_9 + C_{10} \\
 d_5 &= C_5 + C_{10}
 \end{aligned}
 \tag{5.68}$$

The product of the elements of the field $GF(2^6)$ is expressed as another element of the field as given by Equation (5.67). A circuit configuration to perform closed multiplication of two elements of the field $GF(2^6)$ is shown in Figure 5.8.

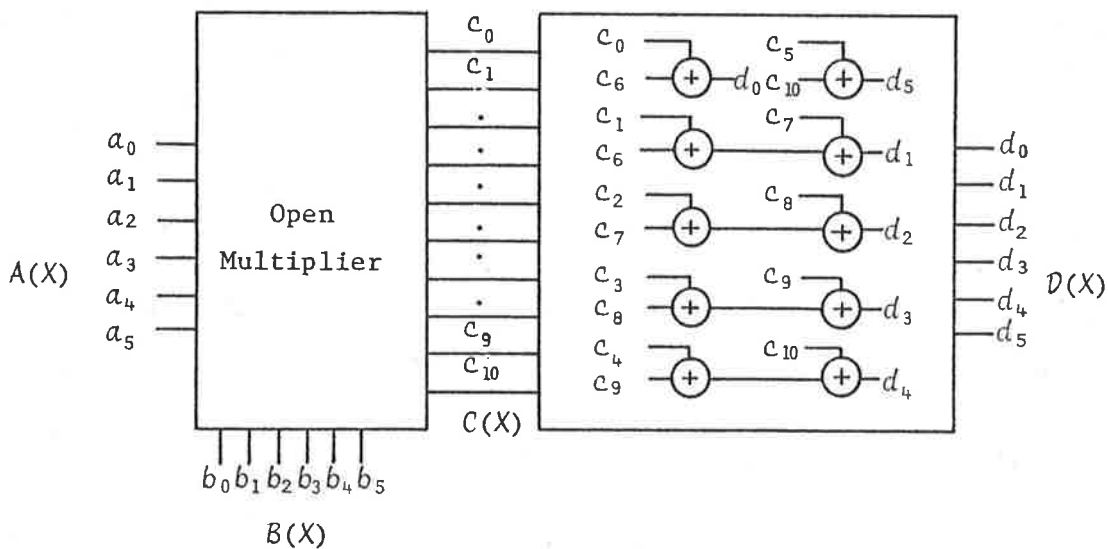


Figure 5.8 Closed Multiplication of Elements of $GF(2^6)$

Another operation which is frequently performed in linear switching circuits is a division of two polynomials. When a polynomial $A(X)$ of degree p is divided by another polynomial $B(X)$ of degree q ($p > q$) the result is a quotient polynomial $C(X)$ of degree $p - q$ and a remainder polynomial $D(X)$ of degree q . These polynomials are then related by the equation

$$A(X) = C(X) \cdot B(X) + D(X) \quad (5.69)$$

The consequence of similarity between the multiplication and division might be expected to be that the implementation of Equation (5.71) would lead to a circuit somewhat similar to the one obtained for open multiplication. Let

$$A(X) = \sum_{i=0}^p a_i X^i \quad (5.70)$$

$$\text{and } B(X) = \sum_{i=0}^q b_i X^i \quad (5.71)$$

The circuit to divide the polynomial $A(X)$ by $B(X)$ is shown in Figure 5.9. The coefficients of these polynomials are stored in two shift-registers of appropriate length. The coefficients of $A(X)$ are fed in serially, while the coefficients of $B(X)$ are fed in in parallel. Initially, the storage elements in the divider are all cleared to 0. The output is 0 for the first q shifts, that is until

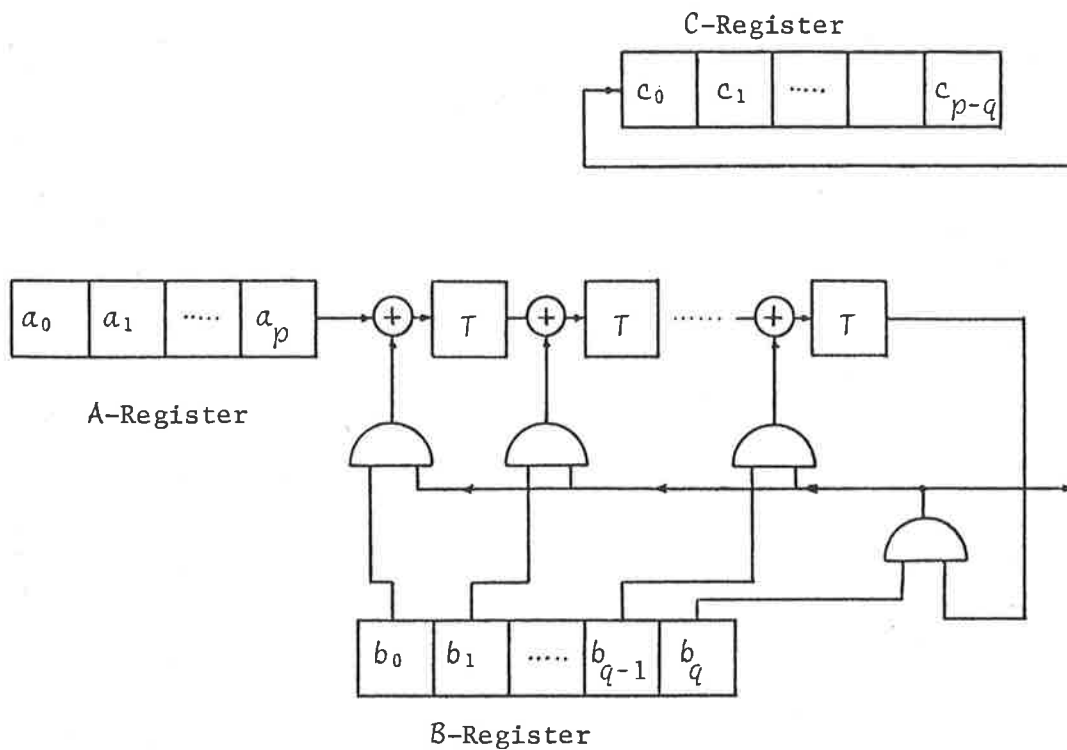


Figure 5.9 Division of Two Polynomials

the first input symbol reaches the end of the shift-register. The next clock pulse results in the first non-zero output, which is $a_p \cdot b_q^{-1}$, the highest order coefficient of the quotient polynomial. Since the coefficients of the polynomials $A(X)$ and $B(X)$ are elements of the field $GF(2)$, modulo-2 arithmetic is used. The following relationships may therefore be used

$$\begin{array}{ll}
 & b_k^{-1} = b_k \quad \text{if } b_k \neq 0 \\
 \text{and} & -b_k = b_k \quad \text{for all } b_k
 \end{array} \tag{5.72}$$

For each coefficient c_j of the quotient polynomial, the polynomial $c_j B(X)$ must be subtracted from the dividend. The feedback is arranged to achieve this end. After a total of p shifts, the entire quotient is stored in the c -register and the remainder polynomial stays in the storage devices in the divider circuit. Simultaneous multiplication and division involving several polynomials may be achieved by appropriate combination of the circuits shown in Figures 5.7 and 5.9 (Peterson and Weldon, 1972).

The decoding of the BCH codes may be accomplished by connecting a number of building blocks described above appropriately. In this approach, the decoding tasks starting from generation of the syndrome vector through to correcting the erroneous bit are implemented in hardware. In a software implementation, the calculations for the decoding are performed in a microcomputer by means of appropriate software.

Another decoding strategy, which can be effectively used for codes of small block length, associates each reply word with an information word.

This is achieved for a BCH(n, k, t) code by storing a k -bit information sequence at each of the 2^n memory locations in a ROM. The n -bit received signal is used to select one of 2^n memory cells and the contents of the memory cell are read as the information sequence. This technique requires a ROM size of $2^n \times k$ bits and, therefore, memory size becomes prohibitive for large n or k .

A still further decoding scheme which generates the syndrome vector in hardware and then performs, using software, a decomposition of the syndrome vector into rows of the parity check matrix to locate the error-bit position is shown in Figure 5.10 below.

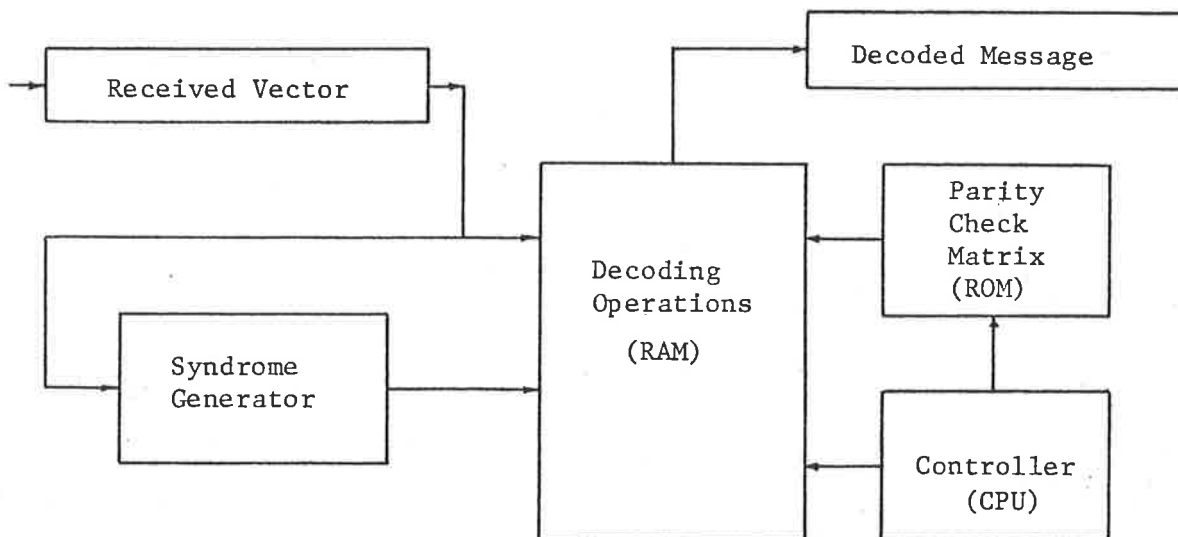


Figure 5.10 A Decoding Strategy for BCH Codes

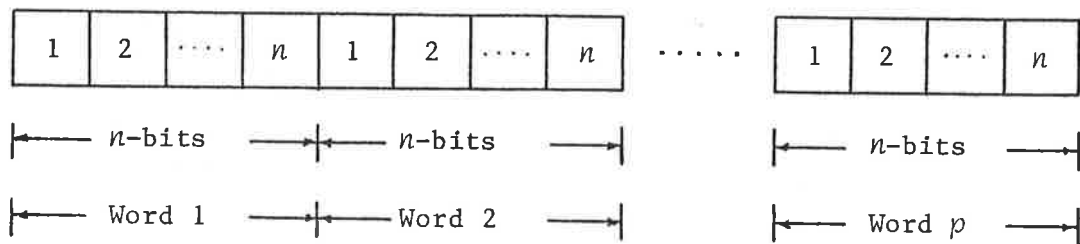
The operation of the decoder shown in Figure 5.10 is now described for the BCH(63,45,3) code. When a 63-bit vector is received over the data lines, it is stored in the RAM of the computer and at the same time it is fed to the syndrome generation logic. Once the syndrome vector is generated, it is also stored in the RAM. The computer then loads the parity check matrix into the RAM and uses the appropriate decoding algorithm to locate the erroneous bits. The received vector is then corrected by inverting the bits in error and the corrected information sequence is stored for further processing.

There are several advantages of this decoding scheme over the complete software or complete hardware decoding. This strategy is time efficient compared with the software implementation, more flexible than the other schemes, and less demanding compared with full hardware implementation. For example, the decoder may be used for another BCH code by merely changing the parity check matrix stored in the ROM and the syndrome generator.

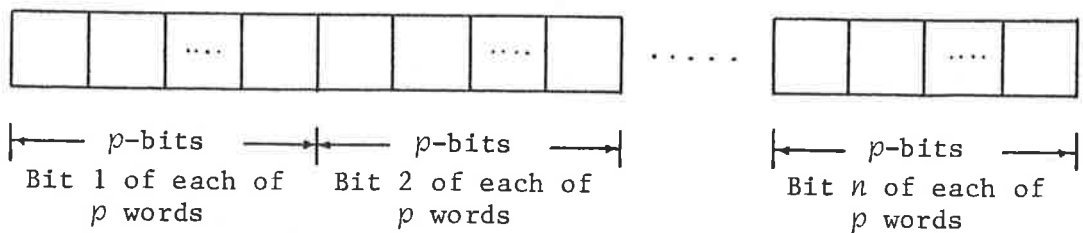
5.4.4 Disadvantages of BCH Codes

Although the BCH codes are highly adaptable and very economical, these codes, too, have their disadvantages. Sometimes, in real channels, the errors appear in bursts. When BCH encoding is employed in a communication system operating over a burst error channel, usually it is possible to detect relatively isolated errors, but correction of an error burst is not possible using the BCH codes unless further signal processing at both transmitter and receiver ends is employed. The correction of the many errors associated with the burst requires special encoding strategies to locate and correct the error bursts. Encoding schemes such as the Hagelbarger Code (Hagelbarger, 1959) and the Fire Code (Fire, 1959) are aimed at correcting errors which occur in bursts with relatively long error-free intervals between the bursts. The principle used in the Hagelbarger code is that the amount of redundancy required to correct a single error in a block length of n bits is smaller than the amount required for multiple errors. This is achieved by defining parity checks on message bits spread out in time. The bit spacing between the message bits, for which parity checks are generated, is selected according to the burst length to be corrected. The advantage of such a scheme is that, when error bursts occur, they are not likely to

corrupt more than one message bit of the group whose parity is being checked. The same principle may be applied to modify the BCH code for burst error correction. When overall average error rates are low but errors occur in bursts, interleaving can be applied to distribute the errors among the data words. This is illustrated in Figure 5.11, where transmission of successive and interleaved words is shown. In order to distribute the errors, the words are temporarily stored and the bits are



a. Successive Data Words



b. Interleaved Data Words

Figure 5.11 Transmission of Successive and Interleaved Data Words

interleaved before transmission; an error burst will then corrupt only a small number of bits in several data words. The use of redundancy encoding provided by the BCH code on the original word enables the correction of errors without a need for increase in redundancy.

A statistical analysis of the performance of a code usually reveals the effectiveness of the code against random errors, but can be deceptive in that weakness against systematic errors may be overlooked. For example, even though a BCH code of moderate length detects practically all error patterns, it is relatively ineffective against the simple fault of slipped synchronism, i.e., a displacement between the word framing at receiver and transmitter (Bennett and Davey, 1965). This is illustrated for the BCH(63,45,3) code in the example given below.

Example: Weakness of BCH Codes Against Slipped Synchronism

Let the information sequence $(a_1, a_2 \dots a_{45})$ in a word generate the check bits $(c_0, c_1 \dots c_{17})$ in the shift-register stages 1, 2 \dots 18, respec-

tively, shown in Figure 5.4. The code word sent on the line then consists of the sequence $(a_1, a_2 \dots a_{45}, c_{17}, c_{16} \dots c_0)$, since the states of the higher order shift-register stages are shifted out to the line first. Assuming that the receiver is delayed by one bit from its normal position relative to the incoming signal, the information bits are read as $(a_2, a_3 \dots a_{44}, a_{45}, c_{17})$ and the check bits as $(c_{16}, c_{15} \dots c_1, c_0, \alpha_1)$, where α_1 is the first information bit of the next code word transmitted. The local shift-register at the receiver end is driven by the erroneous set of information bits $(a_2, a_3 \dots a_{45}, c_{17})$ to generate the fallacious check bits $(b_{17}, b_{16} \dots b_1, b_0)$ which are compared, bit by bit, with the also incorrect received check sequence $(c_{16}, c_{15} \dots c_0, \alpha_1)$.

In order to calculate what actually happens in this example without a tedious enumeration of sequential operations, the algebraic techniques introduced in Section 5.4 are utilized. Assuming that the states of all the shift-register stages are 0 at the beginning of the word, the state of the shift-register stages after the j th message bit is represented by the coefficients of the polynomial

$$P_j(X) = d_{0j} + d_{1j} X + \dots + d_{17j} X^{17} \quad (5.73)$$

where d_{k_j} is the state of stage $(k+1)$ after the information bit is applied. Using modulo-2 addition, the circuit shown in Figure 5.4 may be described by Equation (5.74)

$$\begin{aligned} P_{j+1}(X) &= XP_j(X) + c_j(1+X+X^2+X^3+X^6+X^7+X^9+X^{15}+X^{16}+X^{17}) \\ &= XP_j(X) + c_j X^{18} \end{aligned} \quad (5.74)$$

where the convention

$$X^{18} \equiv 1 + X + X^2 + X^3 + X^6 + X^7 + X^9 + X^{15} + X^{16} + X^{17} \quad (5.75)$$

is applied wherever necessary to reduce all exponents to values less than 18. The state of the shift-register after all the 45 information bits have been applied is then given by

$$\begin{aligned} P_{45}(X) &= a_1 X^{62} + a_2 X^{61} + \dots + a_{44} X^{19} + a_{45} X^{18} \\ &= c_0 + c_1 X + c_2 X^2 + \dots + c_{17} X^{17} \end{aligned} \quad (5.76)$$

where the c_j 's may be expressed in terms of the a_j 's by the use of Equation (5.75). The fallacious check bits generated by the mistimed register at the receiver may be represented by

$$\begin{aligned} Q(X) &= a_2 X^{62} + a_3 X^{61} + \dots + a_{44} X^{20} + a_{45} X^{19} + c_{17} X^{18} \\ &= d_0 + d_1 X + \dots + d_{17} X^{17} \\ &= X(a_1 X^{62} + a_2 X^{61} + \dots + a_{44} X^{19} + a_{45} X^{18}) \\ &\quad + c_{17} X^{18} + a_1 X^{63} \\ &= XP_{45}(X) + c_{17} X^{18} + a_1 X^{63} \\ &= c_0 X + c_1 X^2 + \dots + c_{16} X^{17} + a_1 X^{63} \end{aligned} \quad (5.77)$$

The properties of the BCH code of block length (2^m-1) may now be applied. The primitive element for the BCH(63,45,3) code is α , which satisfies Equation (5.78) given below:

$$\alpha^{2^m-1} = \underline{1} \quad (5.78)$$

which in this example yields

$$\alpha^{63} = \underline{1} \quad (5.79)$$

When the presumed received check bits are compared bit by bit with the locally generated check bits as given by Equation (5.77), all the checks succeed except possibly the last, which compares α_1 and α_1 . These two symbols are the first information bits of two adjacent words and typically have a probability of 0.5 that they are alike. Hence, a one-bit slip in word synchronism has only a 50% likelihood of being detected.

A simple remedy for the problem of clipped synchronism at the receiver is to invert the check bits before transmission. A second inversion of the bits at the presumed parity check bit positions is performed at the receiver before the parity checks are made. In the example discussed, the expression for $Q_{45}(X)$ may be rewritten as

$$Q_{45}(X) = a_2 X^{62} + a_3 X^{61} + \dots + a_{44} X^{20} + a_{45} X^{19} + (1+c_{17}) X^{18} \quad (5.80)$$

where $(1+c_{17})$ is the inverse of c_{17}

Once again, combining Equations (5.75) and (5.80) yields

$$\begin{aligned} Q_{45}(X) = & (1+c_0)X + (1+c_1)X^2 + (1+c_2)X^3 + c_3 X^4 + c_4 X^5 \\ & + (1+c_5)X^6 + (1+c_6)X^7 + c_7 X^8 + (1+c_8)X^9 \\ & + c_9 X^{10} + c_{10} X^{11} + c_{11} X^{12} + c_{12} X^{13} + c_{13} X^{14} \\ & + (1+c_{14})X^{15} + (1+c_{15})X^{16} + (1+c_{16})X^{17} \quad (5.81) \end{aligned}$$

The bit by bit comparison of line and local check bits is shown below:

Line	c_6	c_5	c_4	c_3	c_2	c_1	c_0	c_9	c_8	c_7	c_6
			c_5	c_4	c_3	c_2	c_1	c_0	$(1+c_1)$		
Local	$(1+c_6)$	$(1+c_5)$	$(1+c_4)$	c_3	c_2	c_1	c_0	c_9	$(1+c_8)$	c_7	$(1+c_6)$
			$(1+c_5)$	c_4	c_3	$(1+c_2)$	$(1+c_1)$	$(1+c_0)$	$(1+c_0)$	$(1+c_0)$	$(1+a_1)$

Hence, if there is no error except for the slipped synchronism, nine of the parity checks fail and a tenth one fails half the time. Thus, the errors caused by the slipped timing may be detected using the above technique.

5.5 Optimization of Encoder and Decoder Parameters

The overall optimization of the system parameters in a communication system is a complex task. The conflicting nature of some of the criteria makes the task even more difficult.

In a practical system, often the optimization problem is divided into smaller problems and local optimization is performed. Although the error performance of a communication system such as the object identification system is dependant upon various system parameters such as transmitter power, interrogation distance, interrogation time, propagation loss in the reply line, type of modulation, encoding and decoding strategy employed, it is reasonable to perform a local optimization of the encoding and decoding parameters.

In this optimization, one important parameter is the amount of redundancy provided by the code. The redundancy is a function of the block length of the code and the error detection and correction capability of that code. The message length in a communication system is usually selected on the basis of system requirements other than the encoding-decoding strategy to be used.

The amount of error correction desired in any application is based on the error performance predicted for the communication channel. The object of the optimization proposed here is to reduce hardware complexity and decoding time at the cost of reduced error detection and correction capability of the code (Roy and Cole, 1981). This is illustrated in Table 5.4, where four encoding strategies for codes of block length 64 bits are compared. The number of binary comparisons required in the decoder of Figure 5.10 to locate all possible errors is calculated for each case.

Table 5.4 Comparison of Encoding Schemes

CODE	BCH(63,45,3)	BCH(31,21,2)	BCH(15,11,1)	BCH(7,4,1)
INFO. RATE	45/64	42/64	44/64	32/64
BINARY COMPARISONS	714798	9300	240	168

One encoding strategy is to use the BCH(63,45,3) code which has a block length of 63 bits and is capable of correcting up to 3 errors. Augmenting this code by another parity bit results in a code with a block length of 64 bits and the resultant code is capable of correcting up to 3 errors and detecting

up to 6 errors. The information transfer rate R for the code is given by

$$R = \frac{\text{Number of information bits } k}{\text{Block length of the code } n} = \frac{k}{n} \quad (5.82)$$

The syndrome vector for this code is 18 bits long. If the received vector has up to 3 errors, the last parity bit may be ignored, as it is used to enhance the detection capability of the code. The total number of possible error patterns for a code capable of correcting up to t errors is given by Equation (5.83) below.

$$\text{Total number of possible error patterns} = \sum_{i=1}^t n c_i \quad (5.83)$$

The number of comparisons of the syndrome vector with the parity check matrix is the same as the total number of possible error patterns. In the case of the BCH(63,45,3) code, because the syndrome vector is 18 bits long, each comparison of the syndrome vector implies 18 binary comparisons.

The third encoding strategy considered in this study uses BCH(15,11,1) code four times. Once again, each code block of 15 bits is augmented by another parity bit to enable within each block cor-

rection of single errors and the detection of double errors. Also included in this study is the BCH(7,4,1) code, better known as the Hamming code. The code is augmented by a parity bit which enables correction of single errors and detection of double errors. The Hamming code is used eight times to encode the complete block of 64 bits.

Although, on the basis of the information transfer rate, it first appears that the first three encoding schemes provide similar amounts of redundancy, the resultant effect of the redundancy in enabling error detection and correction is quite different, and in particular substantially different numbers of binary comparisons are required for the error correction. A plot of the number of binary comparisons versus the order of the BCH code used for encoding is shown in Figure 5.12. The advantage of using a lower order BCH code, even when the block length is quite large, is evident.

Practical evaluation of the performance of a specific encoding strategy requires a knowledge of the statistical distribution of errors. Even when the error statistics are known, analytical evaluation of the probabilities of undetected and uncorrected errors is a formidable task. The most effective techniques for such error prediction studies

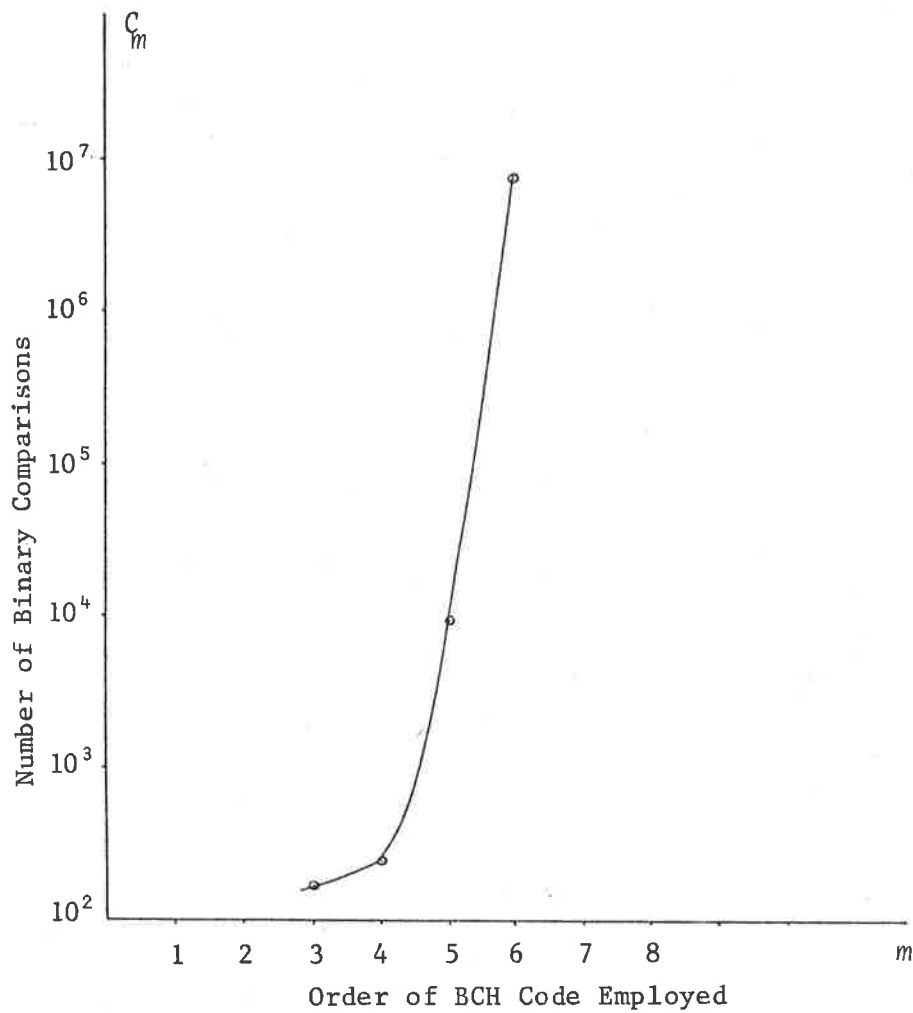


Figure 5.12 A Comparison of BCH Codes

are computer simulation and direct measurements on a real channel. The results of an error prediction study using these methods are provided in Chapter 6. An accurate estimate of the system error rate in a given environment not only helps in designing a better system, but also helps in reducing the cost and complexity.

CHAPTER VI

SYSTEM DESIGN CONSIDERATIONS

6.1 Introduction

The principal aim in any object identification system is to establish communication between a transponding unit which is attached to the object and an interrogation unit which is remote from the object. The requirements in most object identification applications include an interrogation unit transmitting a signal to initiate interrogation of the transponder, a transponder subsequently responding with a code modulated reply signal, and reception and interpretation of this reply such that the object is identified by the operator. In the case of passive transponder systems, the interrogation unit must also supply energy for the normal operation of the transponders. These functional requirements of the identification system can be realized by a transponding unit which includes, in the case of a two-port passive transponder, a receptor antenna, a transponder antenna, a power supply unit

and code generation and modulation circuitry, an interrogation unit which includes a transmitter capable of supplying sufficient power to the transponder for its normal operation, and a receiver capable of interpreting the signal received from the transponder to a form suitable for the operator.

Considerations for the design and satisfactory operation of transponders in object identification situations have been discussed in the previous chapters. The corresponding considerations for the interrogation system are discussed in this chapter.

6.2 Practical System Considerations

The interrogation unit provides the communication interface between the transponder and the outside world in a practical object identification system. The objective of the design is not only to achieve an optimal system performance in a given application, but also to choose the design parameters such that the identification system may be adapted for use in a large number of applications. Another objective is to ensure that performance of the overall system is not degraded due to the interaction between these parameters.

In order to transfer the information over an electromagnetic link, the baseband information signal must be transformed to a frequency band suitable for object identification applications. The required frequency transformation is provided by modulating a carrier frequency by the information-bearing signal. The choice of modulation technique which provides maximum signal-to-noise ratio at the output of the demodulator and is easily implemented in the low-frequency transponders is one of the concerns of this chapter.

The transponder reply code usually consists of a string of binary digits. This information sequence is serially transmitted by the transponder. A digital receiver requires timing information in order to interpret the received signal sequence properly. The timing information may be conveyed either by the data channel or by an auxiliary timing channel. Symbol synchronization is a serious problem in low signal-to-noise communication situations. Methods of providing the necessary synchronization are also considered in this chapter.

The signal-to-noise performance of the identification system can be optimized by providing filters of suitable characteristics. The filtering of the received signal reduces the unwanted interference

and restricts the noise input, thereby improving the system performance.

A block schematic of the interrogation system showing its major subsystems is illustrated in Figure 6.1. Each of these subsystems is considered in the sections to follow.

6.3 Design of the Transmitter

The transmitter is an important part of the interrogation system. The characteristics of the transmitter affect the selection of other system parameters. It is therefore necessary to consider the requirements of the overall system in order to choose the design parameters for the transmitter. The principal components of the transmitter include an oscillator, power amplifier, antenna matching network, and an antenna.

The satisfactory operation of the identification system requires that the oscillator frequency be externally tunable near the desired transmitter frequency. This is to enable on-site tuning of the transmitter to the resonant frequency of the transmitter antenna. It is necessary to provide this

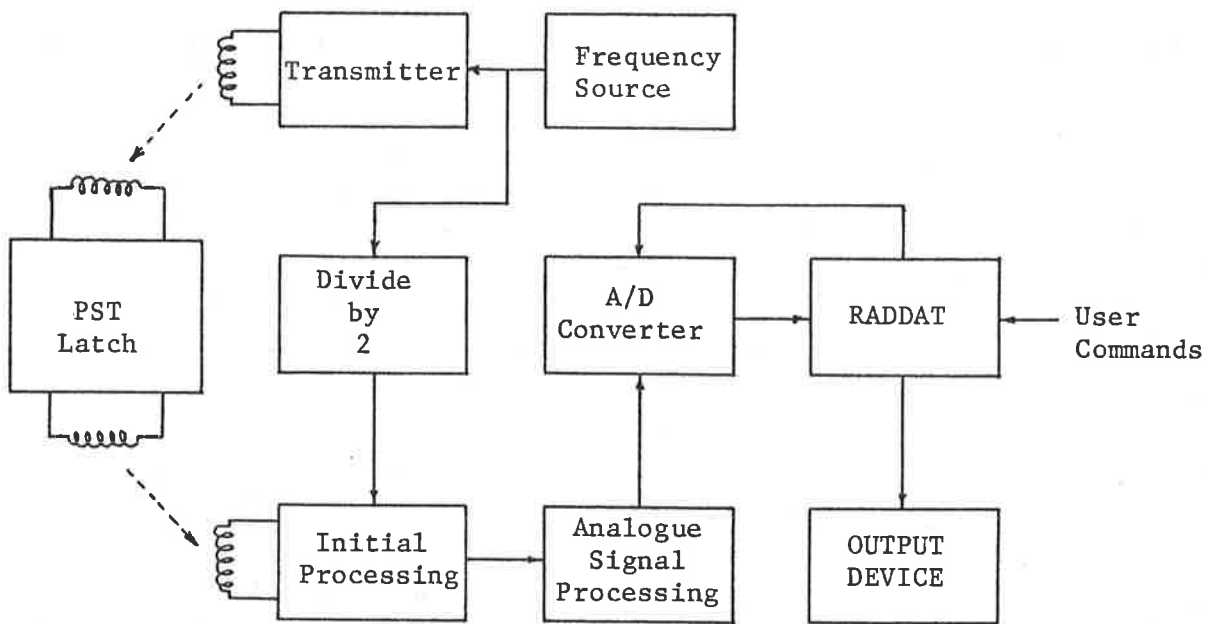


Figure 6.1 Interrogation System for PST System

facility in order to compensate for different environmental effects at different interrogation sites which cause the shift in the resonant frequency of the antenna. The tunable oscillator also allows experimental measurement of the transmitter antenna parameters. The output voltage and frequency of the oscillator should be stable with temperature. It is desirable to have a stable frequency of operation to avoid degradation in system performance due to decrease in load impedance when operated at frequencies other than the resonant frequency. The amplifier should be capable of providing sufficient interrogation energy to transponder located at reasonable distances. In practical object identification applications, this requirement is described in terms of the signal-to-noise ratio at the receiver input. The study of environmental noise in various situations indicates that transmitter power of the order of 10 watts is sufficient in most identification applications. In order to simplify the mechanical design of the transmitter, it is required that the power amplifier should have high efficiency. The cooling requirement in an efficient power amplifier is less stringent and, therefore, results in simple and economical design. A further requirement is that the transmitter amplifier should be capable of driving a range of load impedances and being driven by a 50 ohm source. This requirement increases the

flexibility of the transmitter and simplifies the calibration procedure of the system. The transmitter dynamic impedance is matched to the power amplifier output impedance by means of an appropriate matching network. The matching network should be capable of providing a power match between the amplifier and a range of antenna impedances to allow flexibility.

The circuit diagram of an oscillator using the LM318 OPAMP is shown in Figure 6.2. The oscillator is capable of driving a conventional Class B push-pull amplifier. One implementation of the power amplifier which is capable of providing output power in excess of 50 watts across a load of 6 ohms is shown in Figure 6.3. In later stages of development, a commercially available hybrid power amplifier is used. The internal circuit schematic and connection diagram for the transmitter amplifier realization is shown in Figure 6.4.

The matching between the amplifier output and the transmitter antenna is essential. The amplifier output impedance of the order of 6 ohms must be matched to the transmitter antenna dynamic impedance of 76 ohms. The impedance matching required for efficient operation with minimum of distortion in the output voltage may be achieved either by the

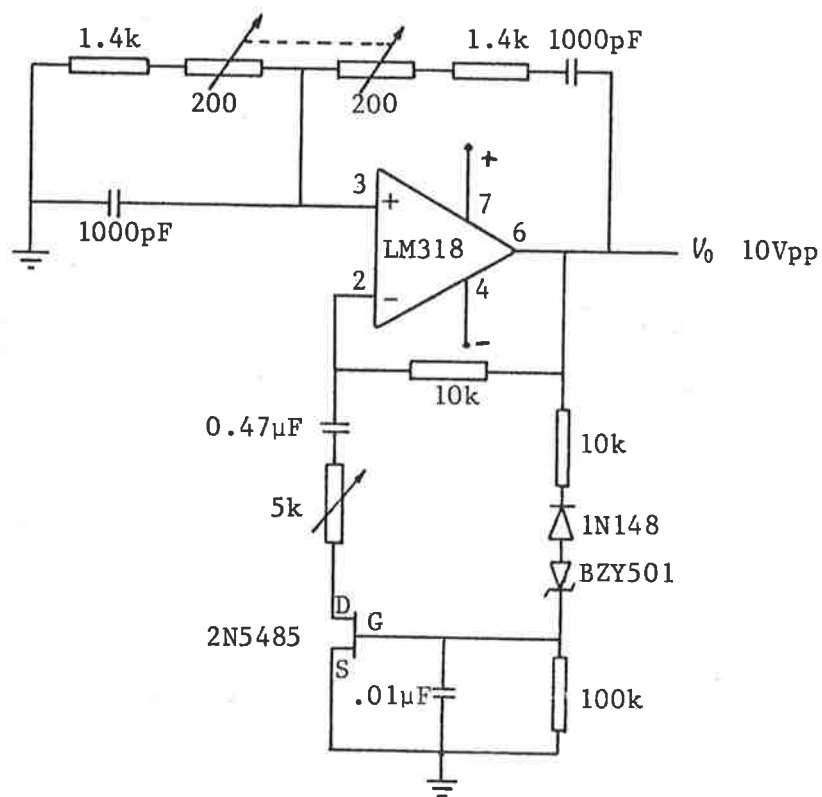


Figure 6.2 An Oscillator for Interrogation System

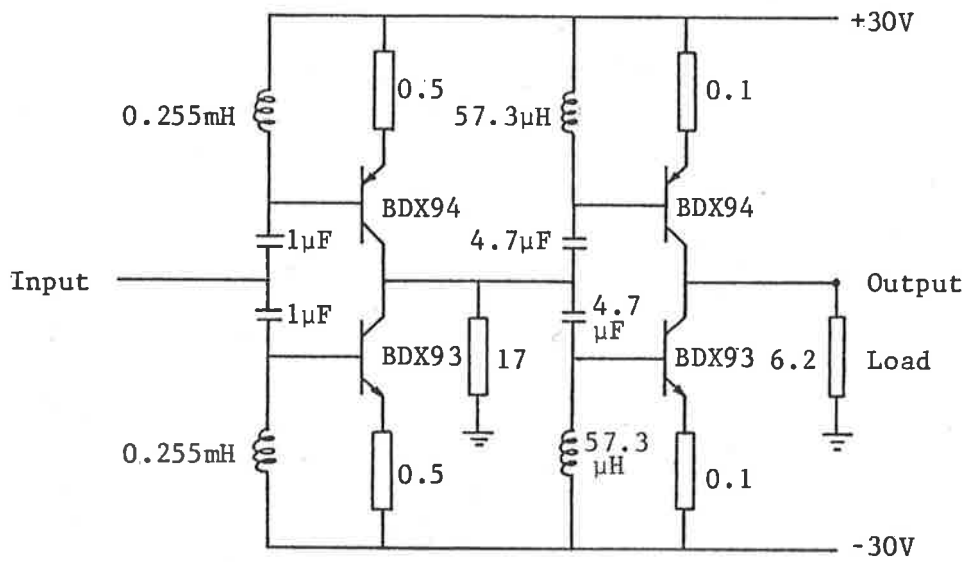


Figure 6.3 A Transmitter Power Amplifier Implementation

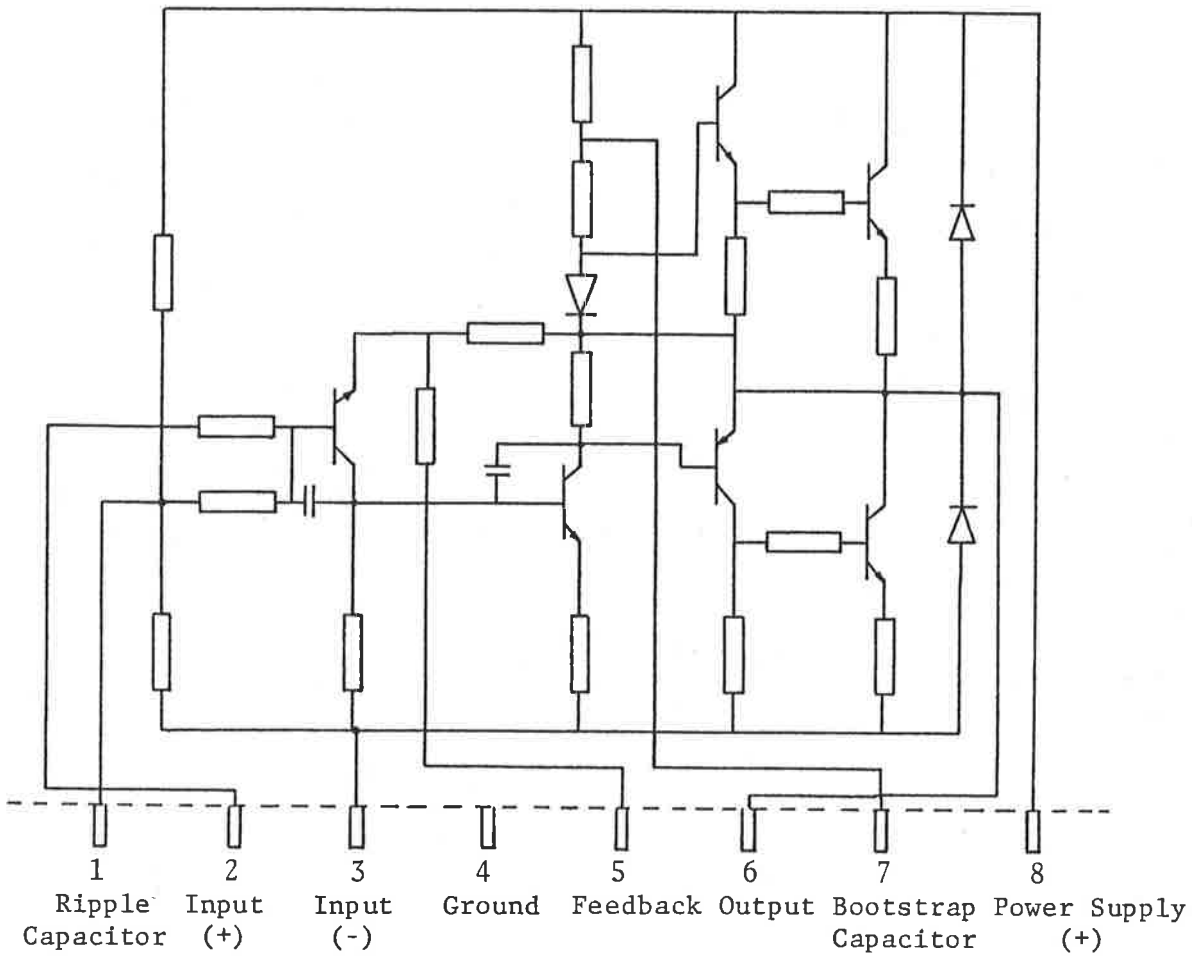


Figure 6.4a Internal Schematic of Hybrid Amplifier SI-1020G

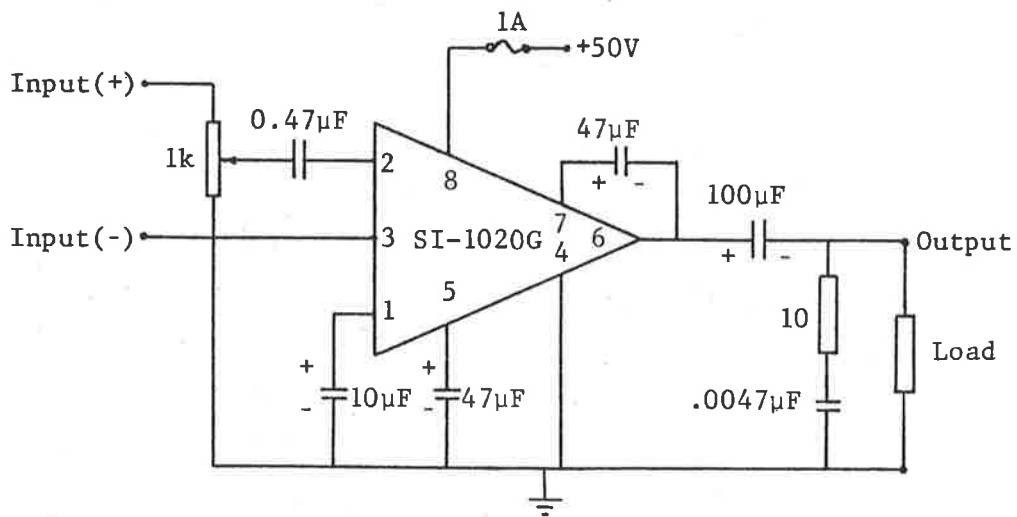


Figure 6.4b Power Amplifier Design Using Hybrid Amplifier SI-1020G

split capacitor tuning or by transformer matching. The tapped resonating capacitor transforms the parallel resonant load impedance, Z_p to the amplifier output impedance, Z_0 by means of the circuit arrangement of Figure 6.5. The impedance transformation is given by Equation (6.1) below

$$Z_0 = \left(\frac{C_1}{C_1 + C_2} \right)^2 \cdot Z_p \quad C_2 \gg C_1 \quad (6.1)$$

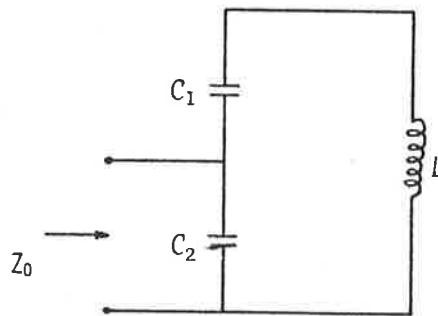


Figure 6.5 Tapped Resonating Capacitor Matching Network

The drawback of this technique arises from the difficulty of conveniently achieving the accurate tapping ratio required in Equation (6.1) with the total series capacitance providing the tuning for the desired resonant frequency. For this reason alone, transformer matching appears to be more suitable in this application. Another advantage of

the transformer matching is that the impedance of the cable which connects the amplifier to the antenna may be considered as a part of the antenna and, therefore, a better matching may be obtained. The transformer using an E-type ferrite core suitable for this transformation is shown in Figure 6.6.

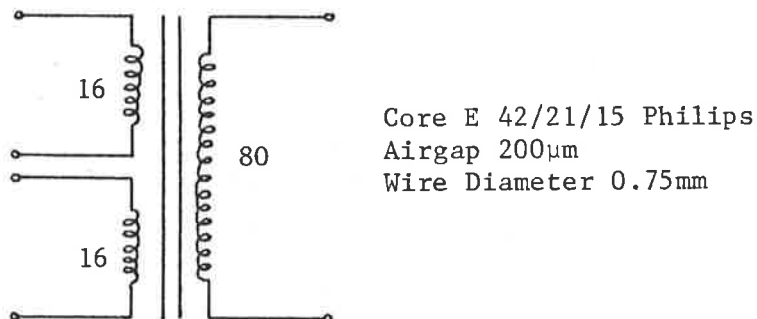


Figure 6.6 Matching Transformer Design

This has dual primary winding which may be in push-pull or parallel configurations to suit various amplifier output circuit arrangements. The physical and electrical characteristics of the transformer are described in Table 6.1.

An important limitation to the level of transmitter power which may be employed arises from the requirement that significant interference to other users of the electromagnetic spectrum does not occur

Table 6.1 Transformer Characteristics

Core	E 42/21/15 Philips
Airgap	200 μm
Wire Diameter	0.75 mm
Turns Ratio	5
Transfer Efficiency	96%

as a result of spreading of the transmitter field outside the interrogation region. The problem is most significant in the applications where the screening of the sensing region is not possible. As shown in Appendix A, the proposed identification system may be operated at the transmitter power of 90 watts, if required. It is, however, expected that in most object identification applications the transmitter would be operated at much lower power levels.

6.4 Selection of Modulation Method

All communication systems consist of three basic building blocks: a data source where information is generated, a channel through which the information is transmitted, and a data sink where the

information is accepted. The data source in the case of the passive transponder system is the code stored in the transponder memory, the channel is the electromagnetic link between the transponder and the receiver, and the sink is the computer where identification information is stored.

In most communication applications, the source generates dc pulses. The transmission channel must provide dc continuity to transmit these pulses accurately. In transponder systems, this is not the case and, therefore, the baseband signal must be translated to a higher frequency band before transmission. The information exchange then requires a modulator at the transmitter end, which translates the data wave to an appropriate frequency band and a demodulator at the receiver end, which transforms the received signal to the baseband signal. Basically, there are three methods of spectrum shifting which may be provided by amplitude, frequency or phase modulation of a sinusoidal carrier by a data signal. In the case of a digital modulating signal, these methods are called Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK) and Phase Shift Keying (PSK). Idealized waveforms for these modulation methods are shown in Figure 6.7 for a binary modulating waveform. As illustrated in the figure, the carrier amplitude has one of the two levels in

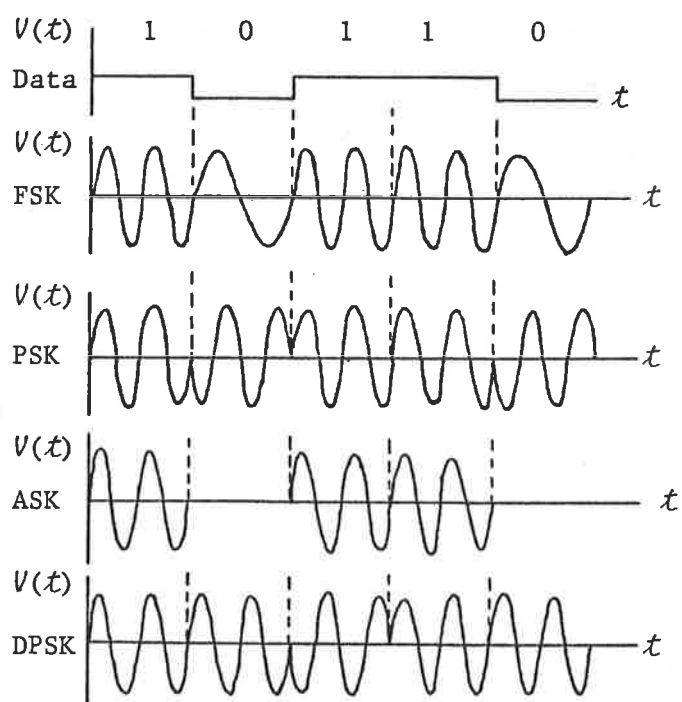


Figure 6.7 Comparison of FSK, PSK, ASK and DPSK Waveforms

the ASK signal, the FSK carrier frequency has one of the two values and the PSK signal phase of the carrier has one of the alternative values. In a practical system, the term "shift-keyed" is somewhat misleading as the amplitude, frequency or phase transition is gradual rather than that illustrated in the idealized figure. These basic modulation techniques for digital signals are briefly described, and then compared to enable selection of a suitable scheme for object identification applications.

6.4.1 Binary Amplitude Shift Keying

The block diagram of a typical data communication system is shown in Figure 6.8. The output of the data source is usually a train of rectangular on-off pulses. These pulses are passed through a low-pass shaping filter before modulating the carrier so that the sidebands produced by the modulator are restricted. The modulator varies the carrier amplitude in accordance with the baseband waveshape. The transmitter output bandpass filter further restricts the extent of the sidebands, and thereby controls interchannel interference. At the receiving end, another bandpass filter is used to attenuate noise and interference outside the band of interest. The detector recovers the baseband wave by means of a product modulator or by simple rectifi-

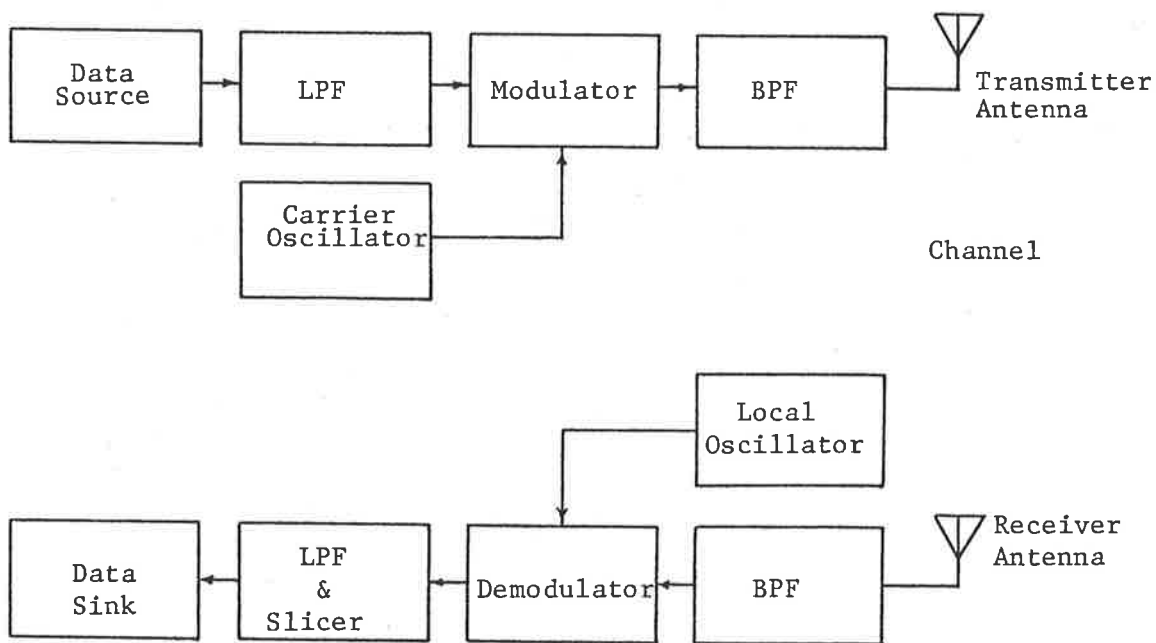


Figure 6.8 A Typical Data Communication System

cation. The low-pass filter filter then separates the baseband signal from higher frequency components and the data pulses are recovered by slicing.

The modulator performs the function of multiplying the sine wave $\cos(\omega_c t + \theta)$ from the carrier source by the data wave $s(t)$. The resulting ASK signal may be written as

$$E(t) = s(t) \cos(\omega_c t + \theta) \quad (6.2)$$

Let $s(t)$ be representable by a Fourier integral as given by Equation (6.3)

$$s(t) = \frac{1}{\pi} \int_0^{\infty} S(\omega) \cos[\omega t + \phi(\omega)] d\omega \quad (6.3)$$

$S(\omega) d\omega/\pi$ and $\phi(\omega)$ may be regarded as the amplitude and phase of the spectral component at frequency ω .

Then the ASK signal may be expressed as

$$\begin{aligned} E(t) &= \frac{1}{2\pi} \int_0^{\infty} S(\omega) \cos[(\omega_c + \omega)t + \theta + \phi(\omega)] d\omega \\ &+ \frac{1}{2\pi} \int_0^{\omega_c} S(\omega) \cos[(\omega_c - \omega)t + \theta - \phi(\omega)] d\omega \\ &+ \frac{1}{2\pi} \int_{\omega_c}^{\infty} S(\omega) \cos[(\omega - \omega_c)t - \theta + \phi(\omega)] d\omega \end{aligned} \quad (6.4)$$

The last term of Equation (6.4) represents foldover distortion, and is eliminated by the low-pass shaping filter by restricting the frequency content of the data wave below the carrier frequency. The contribution of the third integral then becomes negligibly small, and the upper limit of the first integral may be replaced by ω_c . It may be easily

shown, using the expression for $E(t)$, that any dc component of $S(\omega)$ will appear in the transmitted signal as a carrier frequency component.

The demodulation may be performed either by synchronous detection or envelope detection. In synchronous detection, the received ASK signal is multiplied by a local carrier signal which has the same frequency and phase as that associated with the received signal. The output of the detector may be calculated if the amplitude and phase characteristics of the two band pass filters and the channel are known. Let the overall amplitude and phase characteristics of the band pass filters and the transmission channel be $A(\omega)$ and $B(\omega)$, respectively. Then the input to the detector is given by

$$E(t) = \frac{1}{2\pi} \int_0^{\omega_c} S(\omega) A(\omega_c + \omega) \cos [(\omega_c + \omega)t + \theta + \phi(\omega) + B(\omega_c + \omega)] d\omega$$

$$+ \frac{1}{2\pi} \int_0^{\omega_c} S(\omega) A(\omega_c - \omega) \cos [(\omega_c - \omega)t + \theta - \phi(\omega) + B(\omega_c - \omega)] d\omega$$

The local carrier used in synchronous detection is $\cos [\omega_c t + \theta + B(\omega_c)]$. The detector output may then be expressed as

$$\begin{aligned}
E_D(t) &= \frac{1}{4\pi} \int_0^{\omega_c} S(\omega) A(\omega_c + \omega) \cos [(2\omega_c + \omega)t + 2\theta + \phi(\omega) + \\
&\quad B(\omega_c + \omega) + B(\omega_c)] d\omega \\
&+ \frac{1}{4\pi} \int_0^{\omega_c} S(\omega) A(\omega_c + \omega) \cos [\omega t + \phi(\omega) + B(\omega_c + \omega) - B(\omega_c)] d\omega \\
&+ \frac{1}{4\pi} \int_0^{\omega_c} S(\omega) A(\omega_c - \omega) \cos [(2\omega_c - \omega) + 2\theta - \phi(\omega) + \\
&\quad B(\omega_c - \omega) + B(\omega_c)] d\omega \\
&+ \frac{1}{4\pi} \int_0^{\omega_c} S(\omega) A(\omega_c - \omega) \cos [\omega t + \phi(\omega) - B(\omega_c - \omega) + B(\omega_c)] d\omega
\end{aligned} \tag{6.6}$$

The first and third integrals are filtered out by the receiver low pass filter. If the amplitude characteristic $A(\omega)$ has even symmetry about ω_c and phase characteristic $B(\omega)$ has odd symmetry about ω_c as given by Equations (6.7) and (6.8)

$$A(\omega_c + \omega) = A(\omega_c - \omega) = \alpha(\omega) \tag{6.7}$$

$$B(\omega_c + \omega) - B(\omega_c) = B(\omega_c) - B(\omega_c - \omega) = \beta(\omega) \tag{6.8}$$

Then, the detector output signal simplifies to

$$E_D(t) = \frac{1}{2\pi} \int_0^{\omega_c} S(\omega) \alpha(\omega) \cos [\omega t + \phi(\omega) + \beta(\omega)] d\omega \tag{6.9}$$

A comparison of Equations (6.9) and (6.3) shows that the recovered baseband signal is the same as the original baseband signal except that it is modified by the channel transmission characteristics.

An alternative approach to synchronous detection is envelope detection, where full wave rectification of the received signal is performed to obtain the baseband signal. If the characteristics of the transmission medium have the symmetry defined by Equations (6.7) and (6.8), then the received signal may be expressed as

$$\begin{aligned}
 E_D(t) &= \frac{1}{2\pi} \int_0^{\omega_c} S(\omega) \alpha(\omega) \cos [(\omega_c + \omega)t + \theta + \phi(\omega) + \beta(\omega) + B(\omega_c)] d\omega \\
 &+ \frac{1}{2\pi} \int_0^{\omega_c} S(\omega) \alpha(\omega) \cos [(\omega_c - \omega)t + \theta - \phi(\omega) - \beta(\omega) + B(\omega_c)] d\omega \\
 &= \frac{1}{\pi} \cos [\omega_c t + \theta + B(\omega_c)] \int_0^{\omega_c} S(\omega) \alpha(\omega) \cos [\omega t + \phi(\omega) + \beta(\omega)] d\omega
 \end{aligned} \tag{6.10}$$

The received signal therefore has zero crossings coincident with the synchronous detection carrier $\cos [\omega_c t + \theta + B(\omega_c)]$. The integral in the above equation does not change sign and, therefore, does not contribute extra zero crossings if the transmitted signal is not more than 100 percent amplitude

modulated. It can be shown that, under these conditions, the envelope detector using an ideal full wave rectifier yields the same output as the ideal synchronous detector.

One implementation of a synchronous detector using a Phase Lock Loop(PLL) is shown in Figure 6.9.

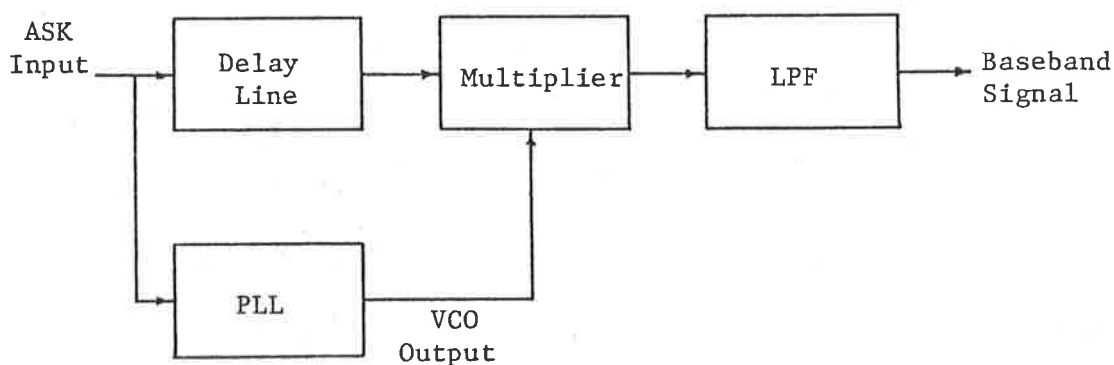


Figure 6.9 An ASK Demodulator

In this mode of operation, the PLL locks on to the carrier of the ASK signal and the output of the voltage controlled oscillator and, hence, the output of PLL is a constant amplitude carrier signal. The multiplier produces the baseband signal and some high frequency components which are filtered out by the low pass filter. The delay is controlled to obtain the maximum output level. If the loop band-

width of the PLL is sufficiently narrow, the signal-to-noise ratio at the output of PLL can be much higher than that at the input. The PLL, therefore, may be used to regenerate weak signals buried in noise.

6.4.2 Binary Frequency Shift Keying

The essential parts of an FSK data transmission system are the same as the ASK system described above. The functions performed by some of the blocks, however, is quite different. As in the ASK system, frequency limiting of the data signal is useful from the point of view of controlling fold-over distortion. The output of the frequency modulated carrier oscillator varies in frequency in accordance with the amplitude of the modulating signal. The resulting spectrum is band limited by the transmitter band-pass filter to fit the allocated band. The receiver band-pass filter helps in restricting the noise and interference outside the band of interest. A limiter often follows the band-pass filter in the frequency modulation systems. The function of the limiter is to remove the amplitude modulation caused by band limiting and in-band noise. Input to the FSK demodulator is, therefore, a constant amplitude signal, and the output is a base-band signal proportional to the instantaneous fre-

quency of the received wave. The low pass filter and slicer have similar functions to those in the ASK system.

The frequency modulated oscillator generates a carrier wave with instantaneous frequency linearly varying with with the baseband signal. The resulting FSK signal may be expressed as

$$E(t) = A_0 \cos \left[\omega_c t + \Theta + \lambda \int_{t_0}^t s(t) dt \right] \quad (6.11)$$

where

A_0 is the amplitude of the unmodulated carrier wave

ω_c is the frequency of the unmodulated carrier wave

Θ is the carrier phase with respect to a particular time origin

t_0 is an arbitrary reference time

$s(t)$ is the modulating signal

λ is a constant of proportionality

The instantaneous frequency of the modulated carrier is defined as the time rate of change of phase and is given by

$$\omega_i = \omega_c + \lambda s(t) \quad (6.12)$$

The basic understanding of the amplitude modulation system is obtained due to a relatively simple Fourier-integral expression of the received signal. The Fourier-integral expression of the FSK signal is much more complex, due to the non-linear relationship of the baseband signal and corresponding FSK signal. The problem related to the non-linearity of the FSK signal can be circumvented, in the case of a binary modulating signal, by considering the FSK signal as the sum of two ASK signals at different frequencies. This approach provides an easy understanding of the FSK system.

A simple method of generating FSK signal is by means of a phase locked frequency control loop. The FSK signal, which is the output of a Voltage Controlled Oscillator (VCO), is phase compared against a reference and an averaged correction voltage is generated. In this way, the carrier frequency can be accurately controlled. The low pass filter prevents the loop from responding to variation in frequency due to modulation. A modification of this approach, where the phase comparator is run at a frequency substantially smaller than the carrier frequency, is illustrated in Figure 6.10. A programmable divider is used to bring the carrier frequency down to the reference frequency. By controlling the division ratio, the carrier frequency can be altered.

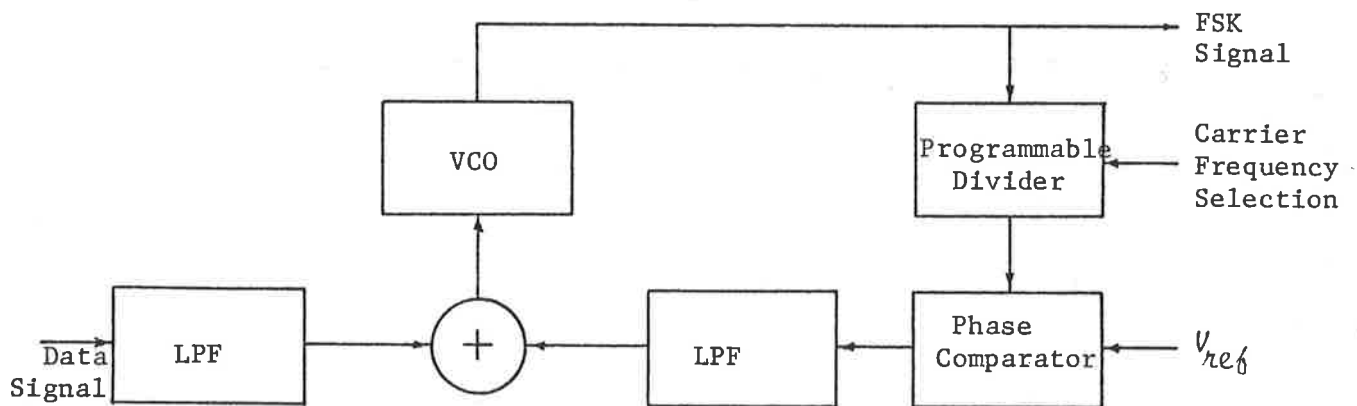


Figure 6.10 An FSK Modulator Using PLL

Detection of FSK signal usually requires filters tuned to each of the possible frequencies. In non-coherent binary FSK detection, the decision whether a 1 or 0 is more likely to have been transmitted is based on which filter has the larger output at the instant of sampling. In order to get the lowest error rate in regeneration of the data, it is

essential to sample the baseband signal with its noise, at a time within each symbol period when the instantaneous signal-to-noise ratio has reached a maximum. This may be achieved by a timing regeneration circuit, which is usually a very narrow-band filter tuned to the symbol rate frequency. Because of the narrow bandwidth, it is capable of regenerating symbol timing rates even at poor signal-to-noise ratios. The block schematic of a non-coherent FSK detector is shown in Figure 6.11.

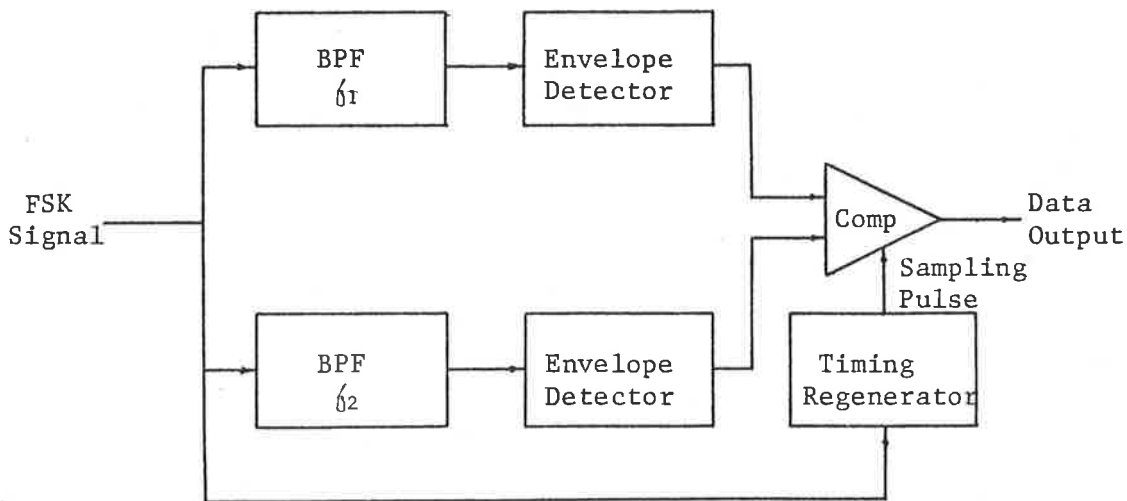


Figure 6.11 Block Schematic of a Non-coherent FSK Detector

Coherent FSK detection is capable of achieving lower bit error rates than the non-coherent FSK detection described above. A block schematic of such a system is shown in Figure 6.12. The phase synchronous carrier signals are generated by means of two PLL blocks. The PLL, being sensitive to the phase of the input signal, can reject noise components that are in phase quadrature with the desired signal, thus resulting in an improved signal-to-noise ratio. In the coherent detector implementation of Figure 6.12, the output of the first band-pass filter is multiplied by a phase synchronous

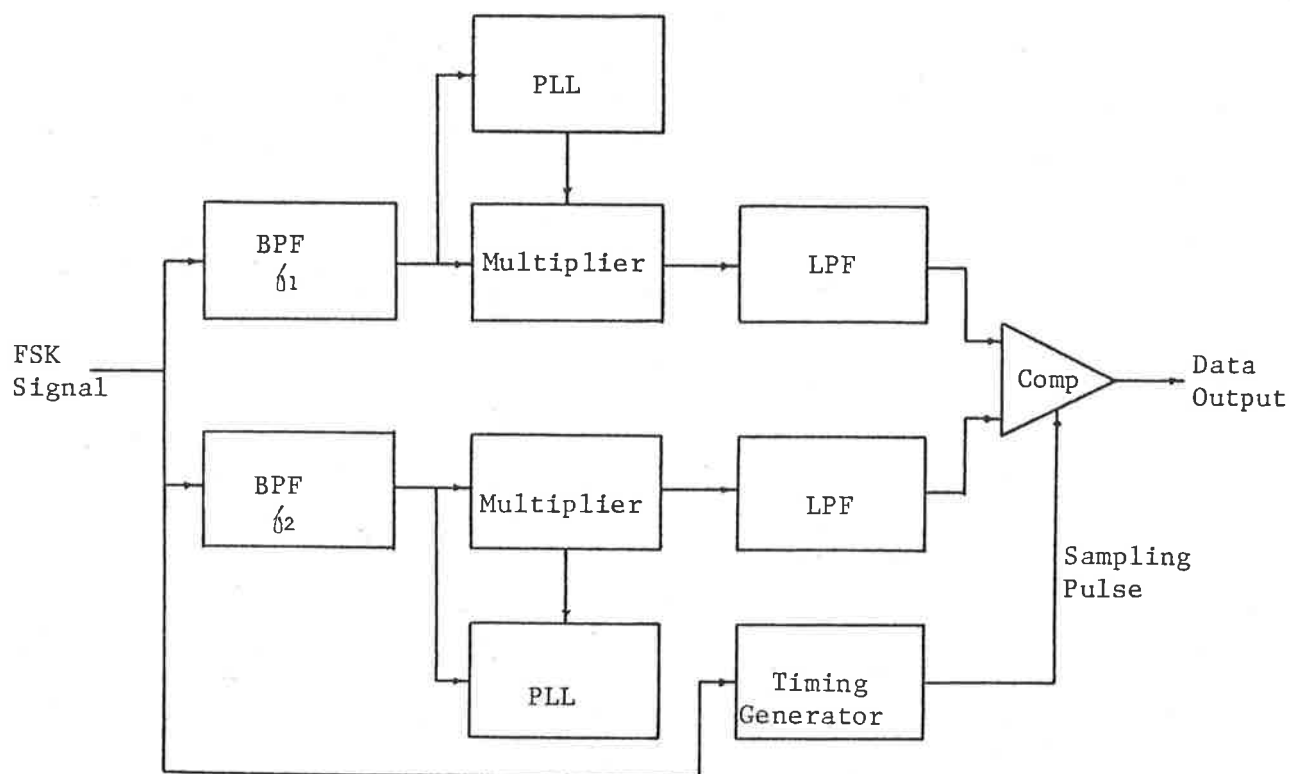


Figure 6.12 A Coherent FSK Detector Using PLL

carrier of frequency f_1 , and the output of the second bandpass filter is multiplied by a phase synchronous carrier of frequency f_2 . The output signals of these multipliers are filtered to eliminate the higher frequency components, and the resulting signals are compared at the sampling instant to decide whether a 1 or a 0 was transmitted.

6.4.3 Binary Phase Shift Keying

Although phase modulation is a type of angle modulation, PSK systems have more resemblance to ASK systems than FSK systems. This can be easily visualized if one compares a binary ASK signal, where a 1 is represented by a positive voltage and 0 is represented by a negative voltage, with a binary PSK signal, where 1's and 0's are represented by 0° or 180° phase-shift, respectively.

The phase modulator generates a carrier wave with instantaneous phase linearly varying with the baseband signal. The resultant PSK signal may be expressed as

$$E(t) = A_0 \cos [\omega_c t + \theta + \lambda s(t)] \quad (6.13)$$

$$\text{and} \quad \phi_i = \theta + \lambda s(t) \quad (6.14)$$

where

A_0 is the amplitude of the unmodulated carrier

ω_c is the frequency of the unmodulated carrier

θ is the carrier phase with respect to a particular time origin

$s(t)$ is the modulating signal

λ is a constant of proportionality

ϕ_i is the instantaneous phase

As in FSK systems, it is desirable that the width of the transmitted spectrum be limited. The two approaches to this problem are known as the Phase Shift Keying (PSK) and Phase Exchange Keying (PEK) (Thompson and Clouting, 1974). Premodulation shaping of the baseband signal is employed in PSK systems to result in a constant amplitude phase modulated signal. The disadvantage of this system is that the resultant signal is not strictly band limited and a finite amount of energy spreads over a wide frequency band and causes interchannel interference. In the PEK system, the spectrum control is achieved by first filtering the baseband signal and then employing linear modulation. The disadvantage of the band limiting is that it causes amplitude modulation to be introduced in the transmitted signal. Consequently, if this signal is passed

through a power amplifier, the non-linearities of the amplifier cause spectrum spreading. In practice, therefore, it is not possible to achieve a strictly band limited signal. For this reason, the term PSK has been used in the text, regardless of whether a PSK or a PEK system is employed.

One of the simplest methods of generating a binary PSK signal is to convert the binary baseband signal to a bipolar signal by addition of a dc level and then to use the bipolar signal as a modulation signal in an ASK system.

Coherent detection of PSK signals can be achieved by multiplying the received signal with a phase-synchronous carrier, as shown in Figure 6.13. One of the problems faced in the PSK system is that the phase of the carrier is defined with respect to a time origin. Unavailability of this time reference causes an ambiguity in the phase origin at the receiver end. The coherent PSK receiver is, therefore, not capable of distinguishing between a binary string and its complement. This problem is overcome by detecting phase difference between consecutive pulses rather than their absolute phases. This scheme, known as Differential PSK (DPSK), is illustrated in Figure 6.14. Detection of DPSK signal is simpler because only a time delay is required instead of a carrier recovery circuit.

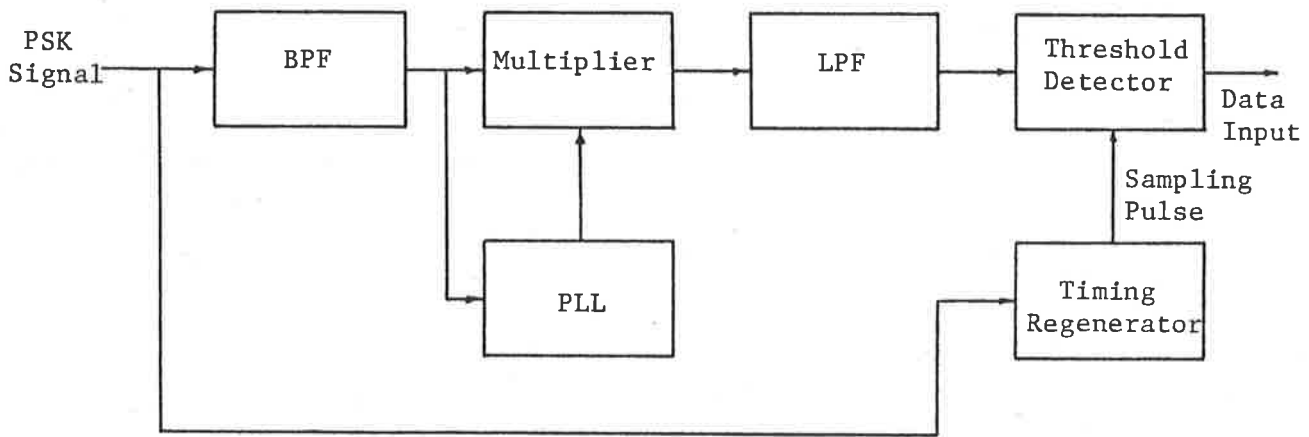
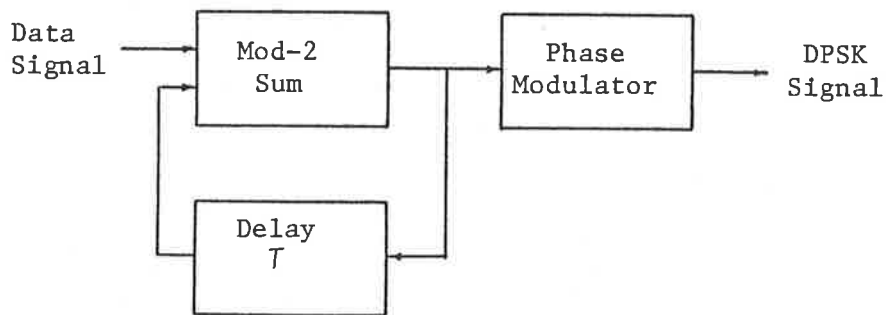
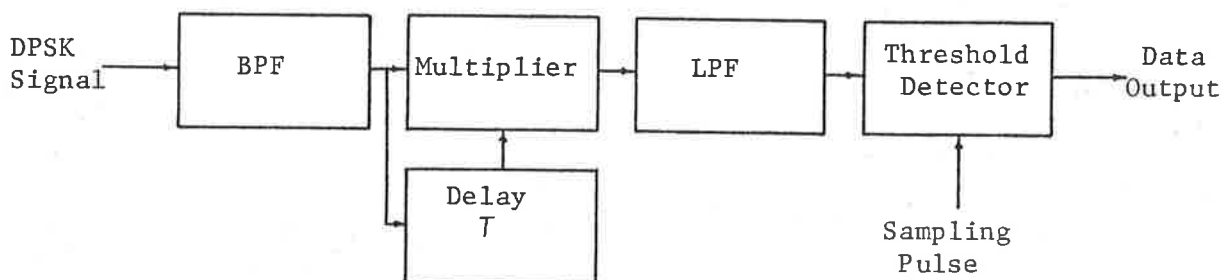


Figure 6.13 Block Schematic of Coherent PSK Detector



a. Generation of DPSK Signal



b. Detection of DPSK Signal

Figure 6.14 DPSK Modulation and Detection

6.4.4 Comparison of Modulation Methods

The decision as to which modulation method is most suitable depends upon the specific circumstances. Many factors must be considered when selecting a modulation method. The most important of these are equipment complexity, noise performance, spectrum occupancy, propagation characteristics of the transmission medium, and sensitivity of performance to equipment design tolerances.

The equipment complexity of a modulation system is an important factor for consideration, as it has a significant effect upon the cost of the system. The signal-to-noise performance of a system determines the bit-error rate and, consequently, the operational range of the communication system. Alternatively, for a given receiver sensitivity, the noise performance is the factor which will determine the transmitter power required. The available spectrum for new communication systems is severely restricted due to ever increasing numbers of users of the frequency spectrum. Spectrum occupancy and interchannel interference are likely to be major issues for consideration in all future systems. In general, spectrum economy is achieved at the cost of degradation in noise performance. The modulation method's susceptibility to external interference and operating environment is difficult to quantify.

However, most multi-level systems are more susceptible to interference than binary systems. As a general rule, the equipment design tolerance must be as wide as possible to make the system economically viable without sacrificing performance. The modulation systems described earlier are now compared for binary baseband signals (Thompson and Clouting, 1975).

6.4.4.1 Noise Performance

The environmental noise encountered in some object identification applications is of impulsive nature. The analysis of the performance of modulation systems in the presence of impulsive noise requires complex mathematical manipulations. The approach used in this study was to make an initial comparison of modulation methods on the basis of the work published by various researchers. The results, based on the assumption that only white-Gaussian noise is present, are plotted later in this section. The system performance in the presence of impulsive noise was studied by means of computer simulation of a receiver using a particular modulation method in conjunction with experimental noise data. Results of these simulations are presented in Section 6.6.3.

Suppression of intersymbol interference in a band-limited communication system such as the one considered in this thesis has major consequences for the design of the transmitter and receiver filters. It has been shown that the intersymbol interference is minimized if the transmitted data signal is shaped to give a raised-cosine spectrum. It has also been shown that for a given average transmitter power the bit error rate in the presence of white-Gaussian noise is minimized by the use of a cosine-shaped receiver bandpass filter with the transmitted signal shaped to give a raised-cosine spectrum at the receiver input (Bennett and Davey, 1965). In the presence of other types of noise, the optimum receiver filter transmittance function is of the form

$$Y(\omega) = \kappa \left[\cos \left(\frac{\pi \omega}{2\omega_s} \right) \right]^\kappa \quad \omega < \omega_s$$

where

κ is a constant

and $\omega_s = 2\pi f_s$

f_s is the signalling frequency

It has been shown that the optimum value of κ for impulsive noise situations is 4/3. The plots of probability of bit error rate at various Signal-to-Noise Ratios (SNR) for the binary modulation methods in the presence of white-Gaussian noise with optimum

filtering employed at the transmitter and receiver ends are shown in Figure 6.15. The signal-to-noise ratio for these plots is defined as the ratio of the peak signal power to the rms noise power at the detector input when the bandwidth of the receiver input filter equals the bit rate of the information signal. In practice, the performance expected from a modulation system is likely to be worse than the theoretical result presented above. In the calculation for these plots it is assumed that appropriate

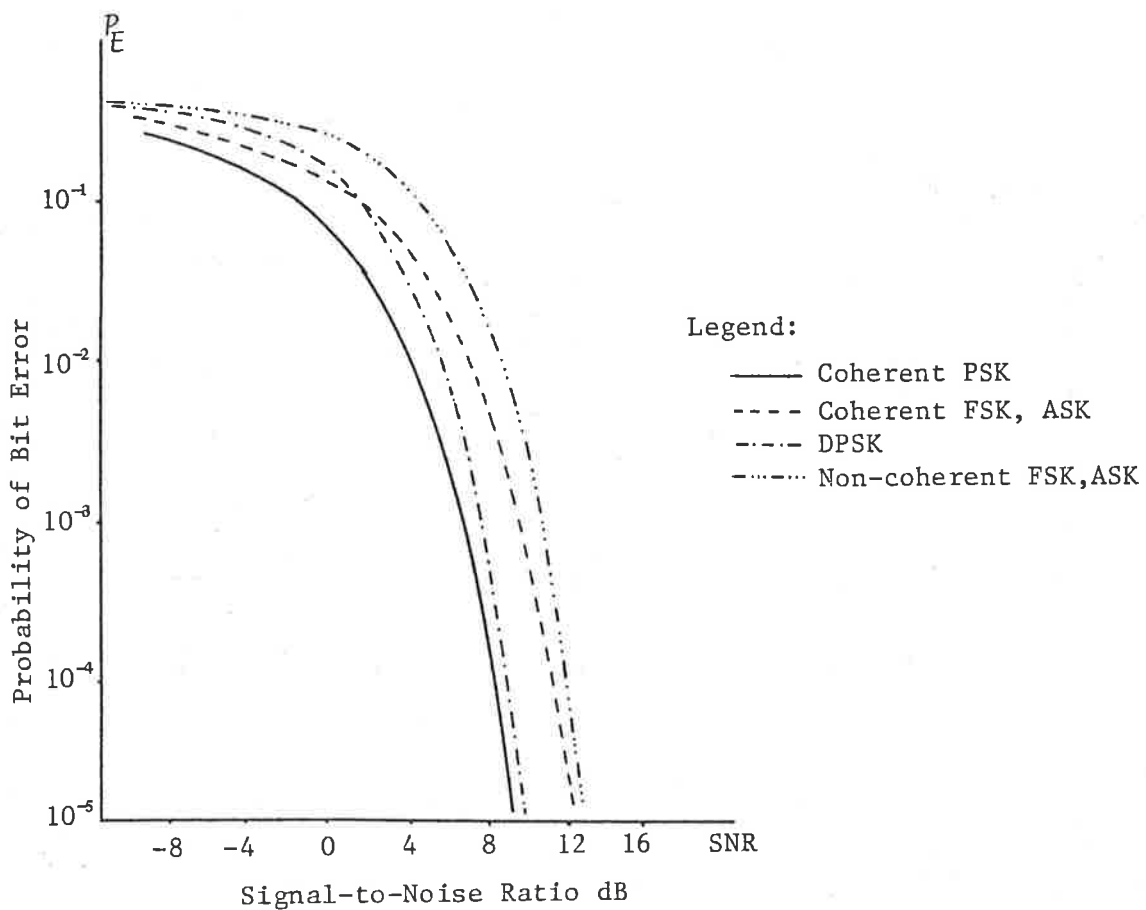


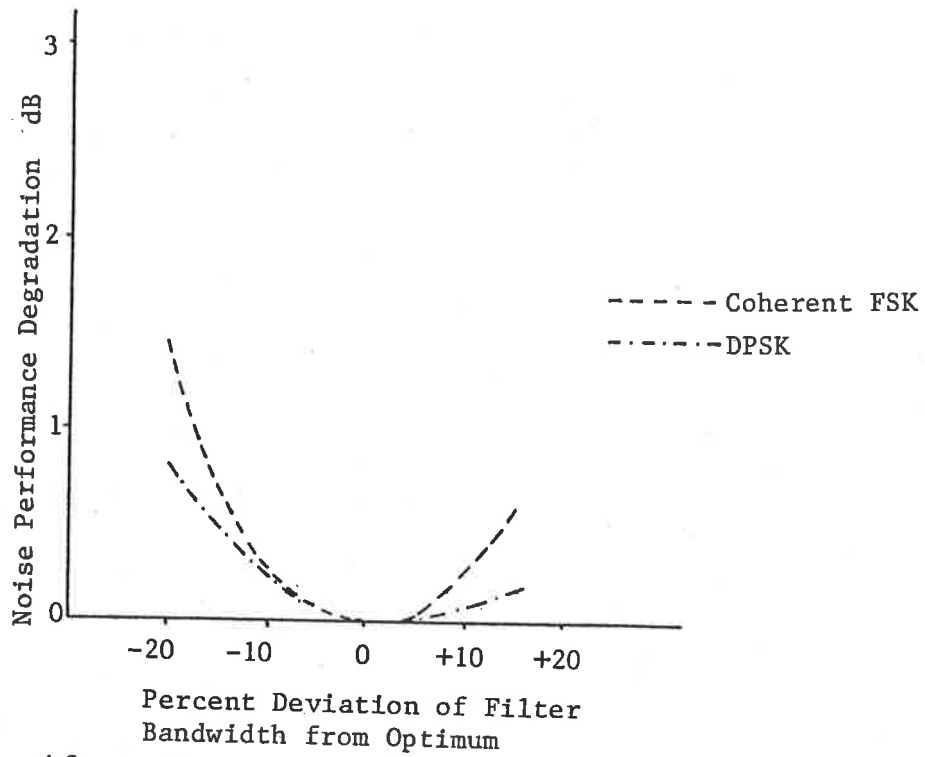
Figure 6.15 Probability of Bit Error Rate for Different Modulation Methods

filtering is provided so that intersymbol interference is negligibly small. Further, it is assumed that the receiver is in perfect synchronism with the transmitter, and the decision logic operates such that sampling is performed at optimum times and decisions are made by comparing it against the optimum threshold voltage.

It can be seen that, for a given signal-to-noise ratio, coherent PSK provides the lowest error rate and non-coherent FSK and ASK provide the highest error rate. Although the DPSK system does not provide the lowest error rate under the ideal conditions, it has a lot of advantages in a practical system. The DPSK system does not suffer from the problem of phase ambiguity of the PSK system. The detection of DPSK signal does not require a phase synchronous local carrier, which is usually very difficult to achieve.

Intersymbol interference occurs when pulses are spread in time so as to affect the adjacent bits at the sampling instants. Intersymbol interference is usually caused by multi-path propagation, attenuation characteristics of the channel, or envelope delay of various sections of transmission path. The filters used to restrict the bandwidth of the transmitted signal usually distort the transmitted wave-

form and, therefore, can cause intersymbol interference. One method of determining the quality of signal transmission is by means of eye patterns, which are formed when the received waveform is displayed on an oscilloscope that is triggered at the bit rate. The smaller the eye, the higher the bit error rate. The FSK system has lower intersymbol interference than the PSK system when the transmitted signal is band-limited to the information bit rate. Furthermore, for this bandwidth the FSK system with a peak-to-peak deviation of 0.7 of the bit rate frequency has slightly lower bit error rate (Tjhung and Wittke, 1970). A typical curve for noise performance degradation with receiver filter bandwidth is shown in Figure 6.16. For narrow bandwidths, errors due to intersymbol interference are predominant, and for wide bandwidths, error rate increases due to increased noise. It may be pointed out that no modulation method results in an error-free transmission, which may be possible by employing a suitable encoding and decoding strategy.



After Tjhung and Wittle, (1980).

Figure 6.16 Effect of Filter Bandwidth on Noise Performance

6.4.4.2 Spectral Occupancy

The spectrum occupancy of a modulation method affects the adjacent channel rejection factor and signal-to-interference ratio. The adjacent channel rejection factor is defined as the proportion of the transmitted power which falls in the passband of a receiver operating at an adjacent channel. It is, therefore, only concerned with the performance of the modulator and not that of the receiver. As illustrated by the plots of Figure 6.17, the more the channels are spaced apart, the less is the adjacent channel rejection and interference (Thompson and

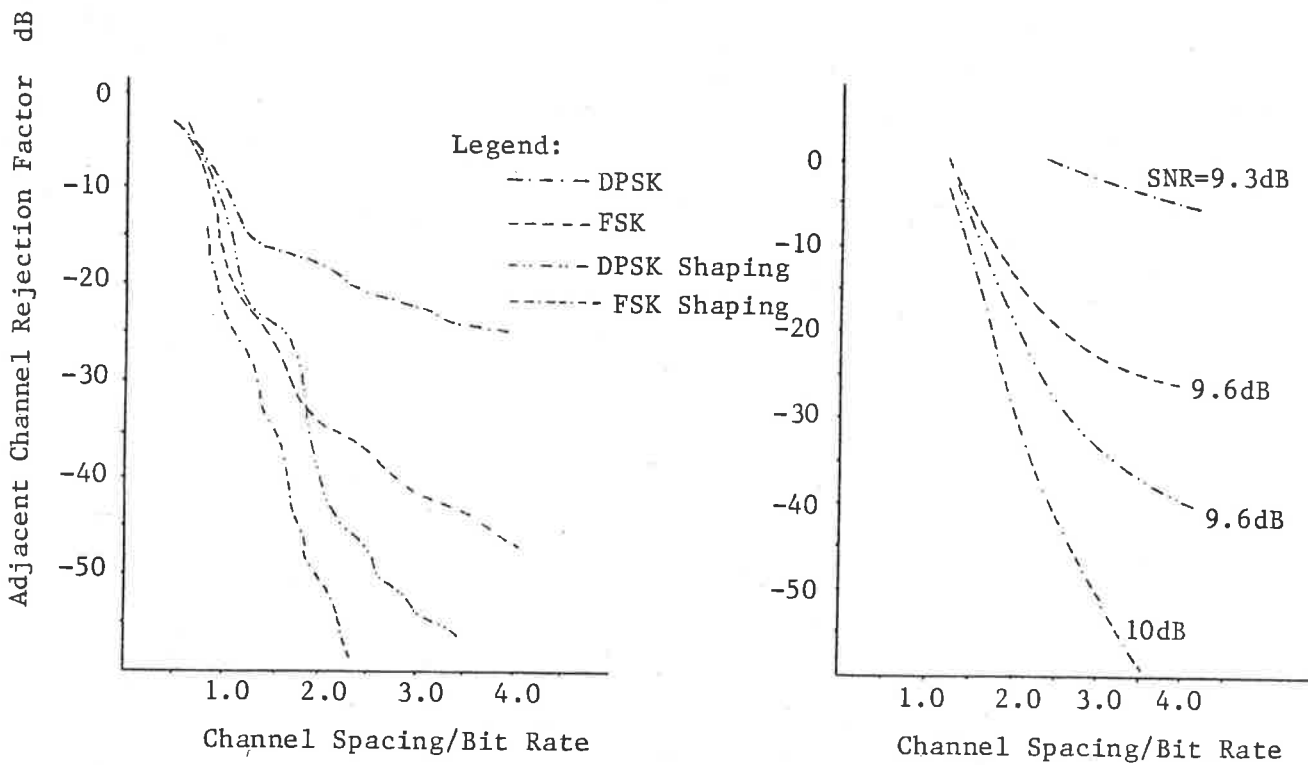


Figure 6.17 Effects of Channel Spacing on the Quality of Communication

Clouting, 1975). It seems, from these curves, that premodulation shaping helps to reduce spectrum spreading and that FSK signals spread less than DPSK signals.

6.4.4.3 Equipment Design Tolerance

The comparison of various modulation schemes under more or less idealized conditions has been made in the above discussion. In practice, however, functional behaviour of most equipment is far from ideal. It is, therefore, expected that performance of the modulation methods would degrade in a practical application.

The noise performance of a modulation system is degraded as the receiver filter bandwidth deviates from the optimum value. The overall group delay distortion introduced by the transmitting and receiving equipment can seriously affect any modulation method. The effects of group delay distortion are more pronounced in binary DPSK and multi-level systems than in binary FSK (Thompson and Clouting, 1975). The noise penalty due to demodulation timing error increases rapidly as the output of the timing regeneration circuit drifts from the optimum value. This degradation can, however, be kept under control by the use of phase lock techniques available.

6.4.4.4 Effects of Transmission Medium

Communication systems using frequency or phase modulation are inherently insensitive to variations in the propagation loss. If amplitude limiters are used, only the angle-modulation information reaches the detector. The ASK system, on the other hand, is very sensitive to level variations. With the decision threshold at halfway between 1 and 0 levels, complete failure occurs with a decrease in level of 6dB. Automatic Gain Control (AGC) is usually employed to maintain accurate decisions.

The amplitude modulation systems are very insensitive to frequency offset. If the channel has a small roll-off, the envelope shape is not materially affected by these small changes in frequency. In the case of synchronous detection of such signals local carrier must follow the frequency variations. The PSK systems with synchronous detection are similarly tolerant of carrier frequency variations. Carrier frequency variations in FSK systems result in a slowly varying dc displacement of the detected signal and, consequently, all decision thresholds are affected. The DPSK systems are also somewhat affected by frequency offset, since the amount of phase change between the adjacent bits varies with the instantaneous carrier frequency.

A sudden phase change of carrier causes problems in all of the carrier type modulation systems. The effect of a sudden phase change, in a band-limited system, is a change in amplitude. In the case of synchronous ASK and PSK systems, the effect can last for a considerable duration, depending upon the response speed of the carrier recovery circuit. The FSK and DPSK systems, on the other hand, are not affected by such transients.

A quantitative comparison of various modulation schemes is provided by Oetting (1979). Oetting's study also includes an extensive guide to the literature available on the subject and translation of the performance figures to a common reference measure. The reference measure used for this purpose is the baseband equivalent $\frac{E_b}{N_0}$, which is the ratio of average signal energy per bit to noise power spectral density at the receiver input required to achieve a bit error rate of 10^{-4} .

6.4.5 Modulation System for PST Applications

The choice of modulation system among the various types available not only depends upon the characteristics of the modulation method, but also upon the requirements of the communication system. Among the carrier modulation methods, the ASK is the

simplest and provides satisfactory communication if the amplitude characteristic of the channel is stable, and the level of environmental noise and interchannel interference is low. The FSK system is very versatile because the system can be easily configured according to requirements. The FSK system provides stable operation over a wide variety of channels, and asynchronous operation is possible. Binary PSK provides the best tolerance to noise. When coherent detection is used, asynchronous operation is possible in PSK systems. The DPSK system is, however, restricted to synchronous operation.

In object identification applications, the system requirements include simple modulator implementation as the modulation circuitry is required to be on board the transponder, inexpensive detection hardware due to the large number of receivers required in some applications, and capability for operating in very weak signal conditions in harsh environmental conditions. For most communication applications, the FSK and DPSK systems are equally good options. In the case of object identification, the DPSK modulation scheme appears to be the optimum choice.

The suitability of the DPSK system for object identification applications can be illustrated by

considering how carrier modulation is provided in passive transponder systems. Let the transponder resonant circuit shown in Figure 6.18a be energized with sinusoidal voltage across the capacitor in one phase corresponding to a particular information bit. When the information bit changes value, it is possible to change the phase of oscillation by changing the position of the switch so that the transponder resonant circuit is connected as shown in Figure 6.18b. The phase reversal of the carrier signal which is the essence of DPSK modulation schemes can be provided in passive transponders simply by means of semiconductor switches, and some control logic to perform the switching with minimum loss in the signal energy due to switching transients. The results of computer simulation of such a modulation system when operated under a realistic noise environment are provided in Section 6.6.3.

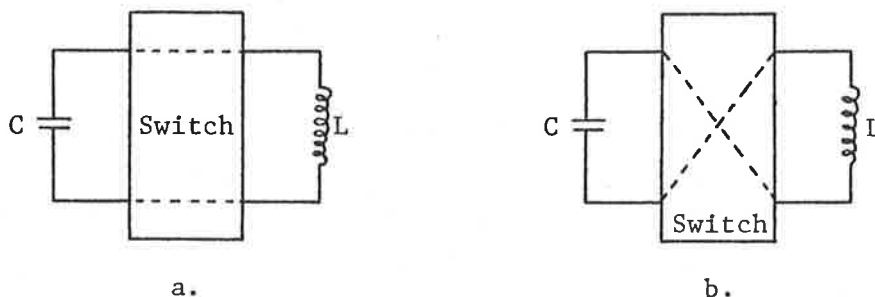


Figure 6.18 Transponder Output Circuit

6.5 Signal Synchronization

In a digital data communication system there is a hierarchy of synchronization problems that must be considered. For the object identification system where a cyclic data wave phase modulates a carrier, the synchronization problems of interest include carrier synchronization, bit synchronization, and word synchronization. Coherent detection of the received signal requires a reference carrier. The aim of carrier synchronization is to generate a reference carrier with phase closely matching that of the received signal. The output of the coherent detector is a baseband signal which must be sampled at the signalling rate. The bit synchronization concerns the generation of the timing wave used in the sample and decision logic. The next hierarchy of synchronization concerns the word timing. In a serial type system, groups of binary symbols are used to represent a word. It is therefore necessary to find the start and end of the words so that words may be interpreted properly. A feature that distinguishes word synchronization from carrier and bit synchronization is that the former is a software problem, while the latter are hardware problems. Word synchronization in a data communication system is usually provided through special design of the information format such as the introduction of redundancy. The optimum usage of the channel capacity

requires, on the other hand, that carrier and bit synchronization be provided without further multiplexing of special timing signals on to the data signal. The recovery of reference carrier and bit timing wave are, therefore, circuit design problems.

6.5.1 Carrier Synchronization

Modulation systems employing synchronous detection not only achieve a better performance but, with the availability of a variety of communication ICs, system complexity and cost have been reduced below that of some other systems. For these reasons, synchronous detection has been studied in this thesis. Methods of establishing and maintaining a reference carrier for this purpose are considered here. When an unmodulated component is present along with the modulated signal, the standard approach to carrier synchronization is to use a phase-locked loop (PLL) which locks on to the carrier component. The loop bandwidth of the PLL is selected to be narrow so that the loop is not perturbed by the sideband components. The recovery of the reference carrier under these circumstances has been studied by various authors (Gardner, 1979; Lindsey, 1972; Stiffler, 1971).

The suppressed-carrier systems, where there is no carrier component in the signal during a random sequence, are of greater interest. The output of a narrow band filter centred at the carrier frequency, when the input signal consists of equal numbers of 1's and 0's, is zero. One way to overcome this problem is to rectify the received wave. The full-wave rectification of the received wave results in a strong second harmonic of the carrier. The frequency can easily be divided to provide the desired reference carrier. A scheme which incorporates this idea is illustrated in Figure 6.19. The carrier

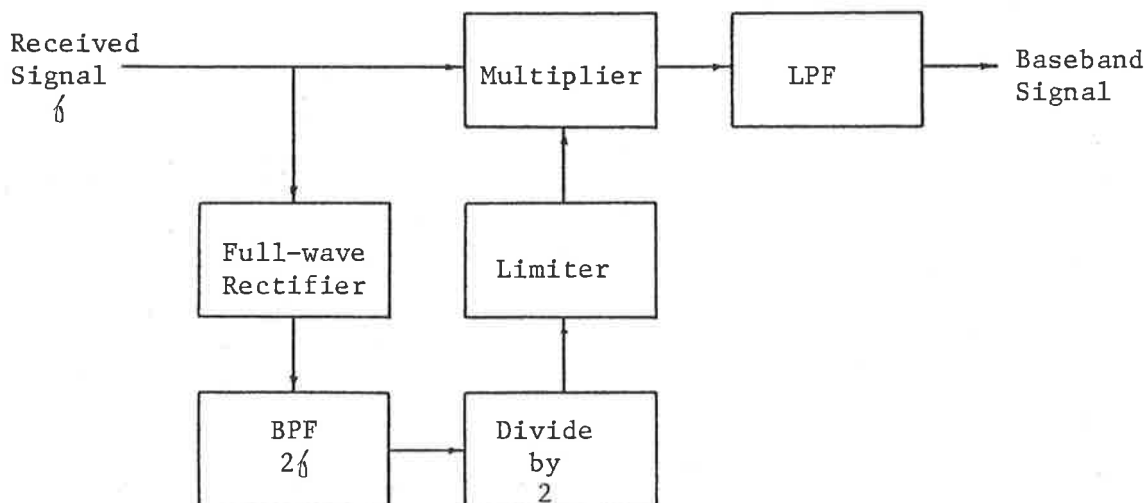


Figure 6.19 Synchronization by Generation of Second Harmonic

generated by this scheme can be either in phase or out of phase with the suppressed carrier and, therefore, the detected baseband signal has a phase ambiguity. If, however, the information is differentially encoded, then the absolute phase of the signal is not important. It is for this reason that the DPSK modulation scheme is selected for the object identification application.

Another method of recovering the reference carrier involves modulation of the received signal by the detected baseband signal in a closed loop system. The received signal in a binary PSK system has two distinct phases corresponding to the two information bits. If the received signal is passed through a binary PSK modulator where the phase of the input signal changes depending upon the polarity of the detected baseband signal, then the output of this modulator is the desired carrier frequency. A circuit configuration using this scheme is shown in Figure 6.20. The time delay in the loop is required to compensate for the delay caused by post-detection low pass filtering so that the reference carrier is phase synchronous with the received signal. The Costas loop, where a phase-locked loop is used instead of the bandpass filter, and the phase reversal feedback from the detected output is introduced into the filter, has a somewhat better performance than the system described here (Costas, 1956).

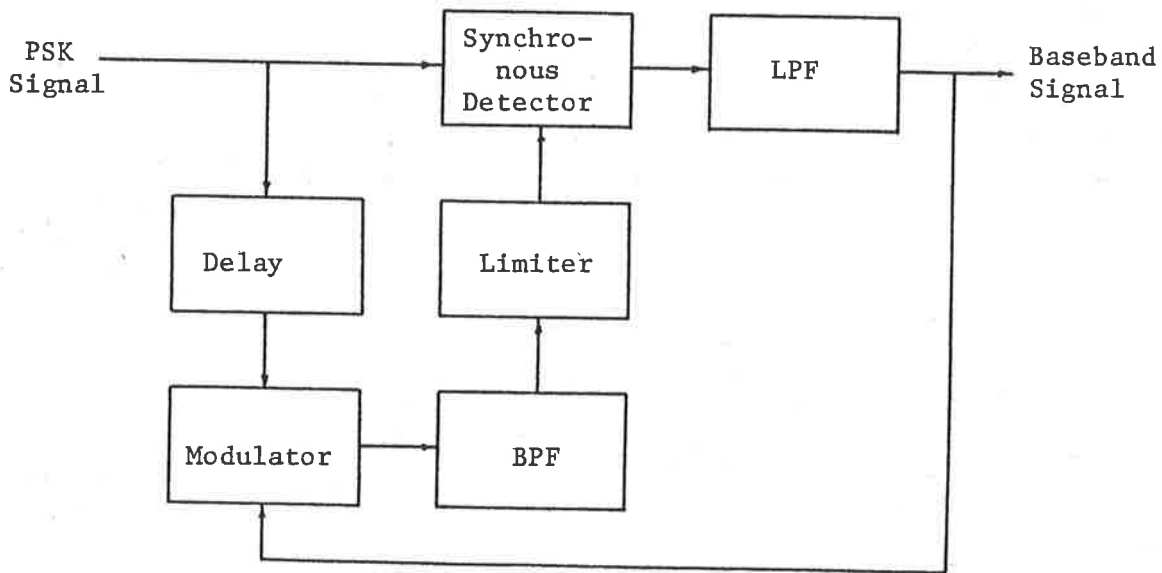


Figure 6.20 Reference Carrier Generation by Modulation of Received Signal

6.5.2 Bit Synchronization

A digital receiver requires timing information to interpret the detected baseband signal sequence properly. Each symbol at the detector output must be sampled at a time when its value has become fully established and is not in a state of transition. The timing which accomplishes this task, in the case

of the binary baseband signal, is called the bit timing, and is provided by the bit synchronization circuit.

The baseband signal transitions, in a synchronous binary channel, occur at integral multiples of unit symbol length. If the bit transitions occur reasonably frequently, then the bit timing can be established from the transitions. One technique which may be employed in this case is the use of a very narrow-band filter tuned to the bit rate frequency. The output of such a filter would be a sinusoidal signal at the bit rate frequency. This signal may then be suitably amplified and passed through a limiting circuit to produce the desired timing wave. One of the drawbacks of the scheme is that, when the baseband signal has very few or no transitions, the output of the bandpass filter starts to diminish and, consequently, without any automatic gain control arrangement, the circuit has limited practical usage. A further problem is the synchronization of the phase of the timing wave with that of the detected baseband signal. Means for synchronizing these two phases must be provided before the system can be employed in a practical situation. Usually, the requirement is to place either the positive or negative going transitions of the timing wave at the centres of the received bit intervals.

A better approach for deriving the timing wave, under these circumstances, is to control the phase of a bit rate local oscillator by means of the transitions in the data signal. The amount of phase correction required in the timing wave may be obtained by comparing the time positions of the data wave transitions with those of the output of the synchronizer. The correction may be provided in one of several ways. One possible technique is to sample a sawtooth timing wave at the instants of transition to derive an error signal which is proportional to the amount of correction required. The error signal can then act on the phase or frequency of the timing oscillator to bring about the correction. It is usually desirable to have the corrections take place slowly. If there are no transitions in the data wave for a certain period of time, then there is no correction made. It is for this reason necessary to operate the timing oscillator at very close to the bit rate frequency so that the synchronism is maintained for a longer period. A digital method of controlling the phase of the timing wave is shown in Figure 6.21 (Bennett and Davey, 1965). A bit timing wave is generated by frequency dividing the output of an oscillator running at N times the bit rate frequency, where N is selected to be large. If the timing wave is found to be leading with respect to the time position of

of the transitions, one of the count pulses from the oscillator is deleted, thereby retarding the phase by $1/N$ of a bit interval. Similarly, if the bit timing wave is found to be lagging compared with the transitions, an extra pulse is fed to the divider circuit, thereby advancing the phase by $1/N$ of the bit interval. In the absence of transitions, the system returns to the bit rate frequency.

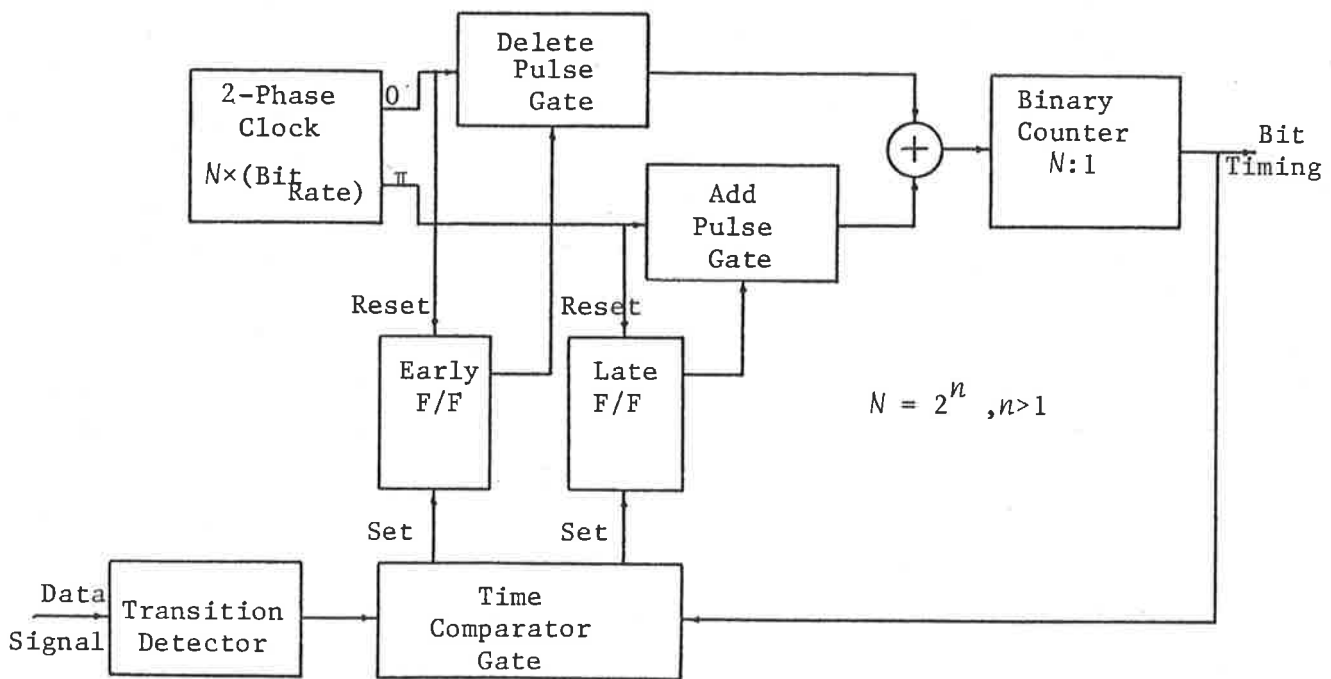


Figure 6.21 Digital Bit Synchronizer Schematic

A simpler technique is employed for the object identification application. The channel characteristics for a number of identification applications are such that they introduce only a negligible amount of phase distortion and frequency offset. The bit rate frequency is chosen to be integrally related to the carrier frequency. Once the reference carrier is generated, a counter is used to derive the bit timing wave. This technique is not useful in channels with phase jitter.

6.5.3 Word Synchronization

Efficient use of codes to transfer information over a communication channel is dependent upon the knowledge of the positions in a long string of received bits at which a code word starts and finishes. If fixed length codes are used for information exchange, then it is only necessary for the receiver to know the starting bit position and the number of bits in the word being transmitted. There are two well-known approaches to the problem of word synchronization, and both utilize redundancy in the code word. The first method relies upon the recognition of synchronizing bits periodically within the information stream by the receiver (Gilbert, 1960; Guibas and Odlyzko, 1978; Scholtz, 1980). The second method eliminates the need for prefixes

and relies, instead, upon properties of certain codes to enable word synchronization (Stiffler, 1962).

A class of comma-free codes called the prefix-synchronized codes are briefly considered here (Gilbert, 1960). The codes of this class have the property that every codeword starts with a fixed prefix $P = (a_1, \dots, a_p)$ of length, p being smaller than the block length, N of the code. The prefix P is selected such that it does not appear as p consecutive bits in the sequence $(a_2, \dots, a_p, b_1, \dots, b_{N-p}, a_1, \dots, a_{p-1})$, where $C = (a_1, \dots, a_p, b_1, \dots, b_{N-p})$ is a codeword if the above condition is satisfied. The procedure in finding a suitable prefix for a given block length is first of all to find the prefix which will maximize the valid codewords called the dictionary size $G_p(N)$ and secondly to find the optimum size of the prefix required to achieve this.

Gilbert showed that, in the case of the binary alphabet, the prefix $P = 11 \dots 10$ of length $p = \lceil \log_2(N \log_2 e) \rceil$ gives a prefix-synchronized code with number of codewords, $G_p(N)$ within a small constant factor of the maximal size for any comma-free code of length N . Guibas and Odlyzko (1978) have shown that Gilbert's choice of optimal prefix is

only true for m -ary signalling systems for values of m being 2, 3 or 4. For passive transponder applications, only binary codes are considered; therefore, the optimum prefix can easily be found.

6.6 Receiver Design Considerations

The tasks of the receiver in the Passive, Subharmonic Transponder (PST) applications include recovery of the transmitted signal and its interpretation, and subsequent display of the message. The receiver is required to perform a large number of operations between reception and eventual display of the information. The operations required to be performed by the receiver are dependent upon the available inputs and desired outputs. These operations, in the case of identification applications, may be performed by analogue and digital signal processing techniques.

6.6.1 Analogue Processing of the Reply Signal

The signals appearing at the PST receiver input include a weak reply signal from a PST label, a strong interfering signal from the interrogation equipment, and environmental noise. It is essential to separate the reply signal from the interference.

Although separation is provided between the transmitter and reply signal in the frequency domain, the interfering signal must be further attenuated in order to avoid saturation of the amplifiers and other circuitry within the receiver. The signal-to-noise ratio at the receiver input is usually low, due to absence of source of energy in the PST labels. It is, therefore, necessary to eliminate the out-of-band noise by appropriate filtering and enhance the signal by means of low-noise amplification. A block schematic of the receiver front-end which performs these functions is shown in Figure 6.22.

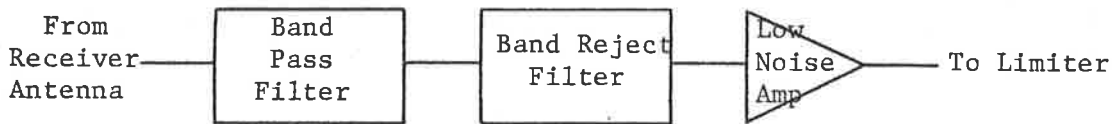


Figure 6.22 The PST Receiver Front End Signal Processing

For the PST receiver, where the reply carrier is at the first subharmonic of the transmitter signal, the normalized transfer function of the passive receiver filter, including the receiver antenna, is of the form given by Equation 6.15.

$$H(s) = \frac{(s^2 + \alpha)}{(s^2 + \beta s + 1)(s^2 + \gamma s + \alpha)} \quad (6.15)$$

where

$$\alpha = 4$$

$$\beta = 1/Q_4$$

$$\gamma = 4/Q_1'$$

The quality factor Q_4 is chosen to provide the desired information bandwidth for the PST system and the quality factor Q_1' is selected to provide a deep notch at the interfering frequency on the one hand and yet maintain the rejection characteristics for small variations in the transmitter signal.

Suitable values for Q_4 and Q_1' are 8 each for the PST applications, and the resultant filter transfer function is given by Equation (6.16)

$$H(s) = \frac{s^2 + 4}{s^4 + 0.625s^3 + 5.0625s^2 + s + 4} \quad (6.16)$$

This filter for the transmitter frequency of $f_1 = 100\text{kHz}$ and receiver centre frequency of $f_2 = 50\text{kHz}$ takes the form shown in Figure 6.23. The overall receiver front end design and the amplitude plot of the overall transfer function is shown in Figure 6.24. A separation of 99dB between the reply and the interference is achieved using commercially available low-noise operational amplifiers. Some of

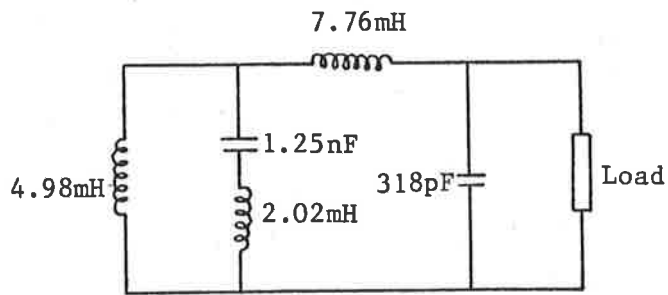


Figure 6.23 Receiver Filter

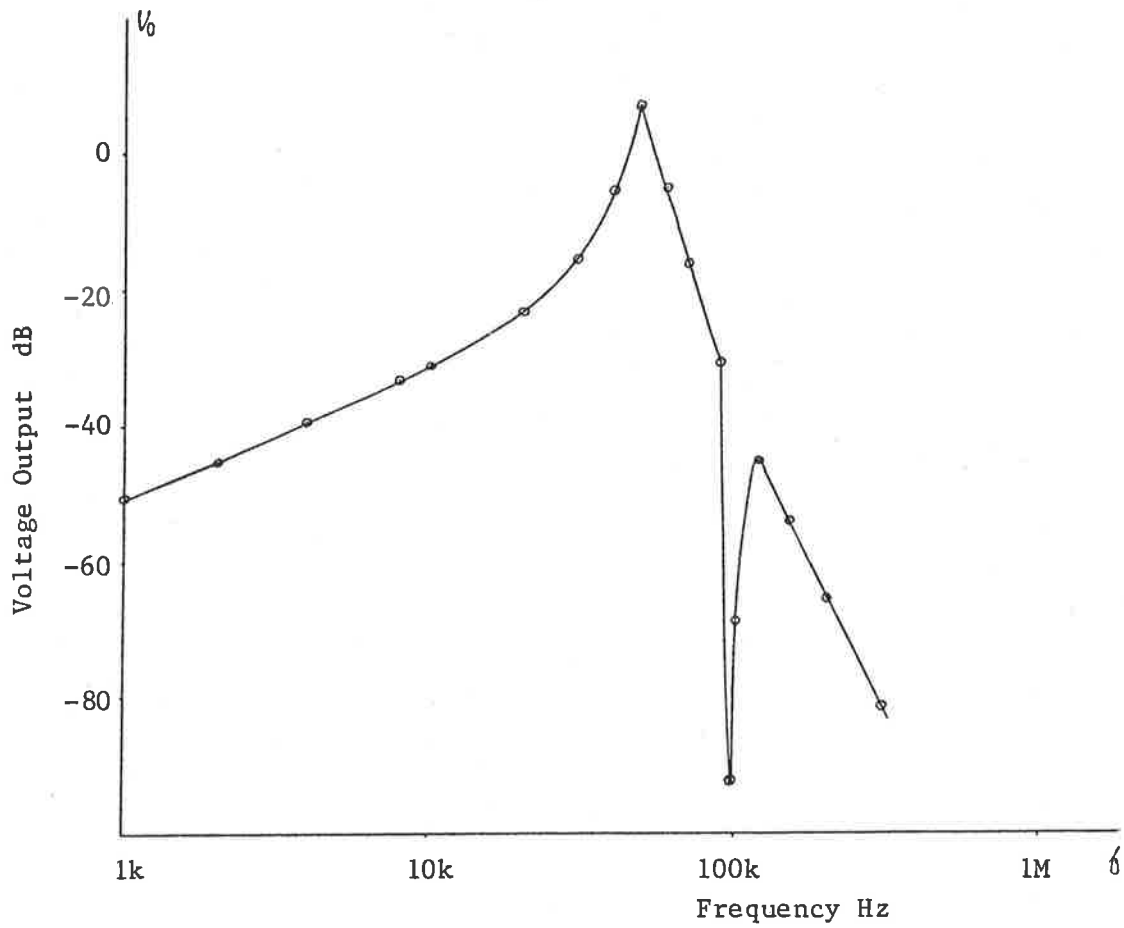


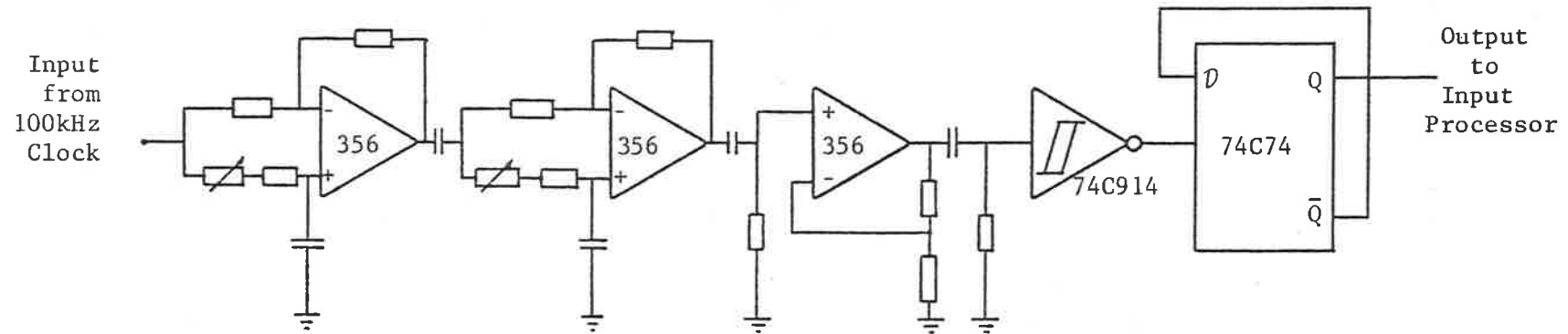
Figure 6.24 Amplitude Plot of the PST Receiver Filter-Amplifier

the other signal processing circuitry is shown in Figure 6.25.

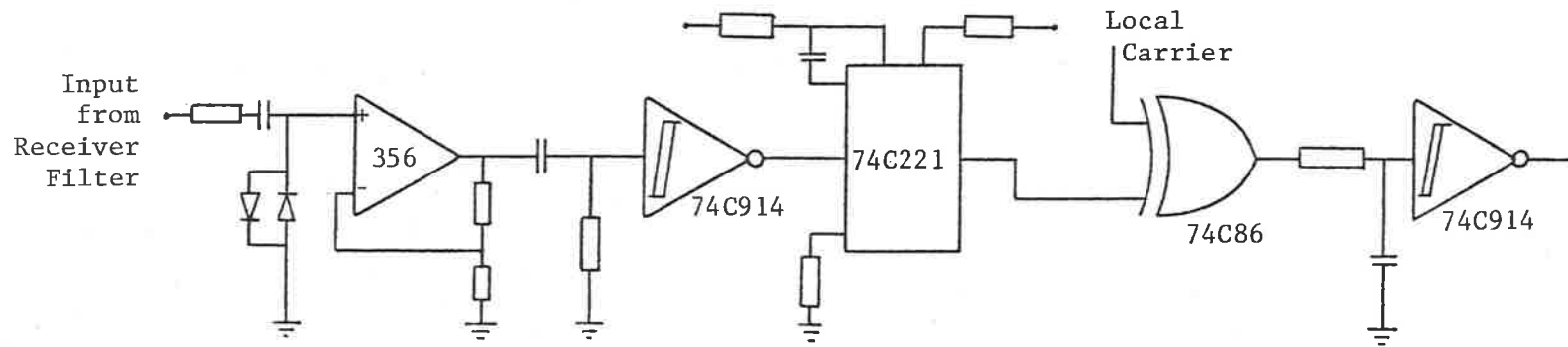
Although commercial OPAMPS are used for the prototype design, an extremely low-noise bipolar amplifier designed by the author for another application was found to be suitable for this application as well. Such an amplifier is expected to provide a much improved signal-to-noise performance. The major contributor of noise in bipolar amplifiers is the $r_{bb'}$ of the transistors. Special design techniques are employed to minimize the $r_{bb'}$ component. Biasing techniques are employed to minimize the drift in quiescent operating conditions. The circuit diagram and the chip layout of the amplifier are shown in Figures 6.26 and 6.27, respectively.

6.6.2 Digital Signal Processing

The principal objective in the implementation of the receiver being described here is to provide the functional requirements appropriate to a PST receiver system and, at the same time, to provide additional flexibility, which is desirable in an experimental programme. One such additional requirement is that the system be suitable for use in the measurement of environmental noise. In one configuration, because of the availability of the



a. Local Carrier Generation



b. Signal Conditioning

Figure 6.25 Receiver Signal Processing

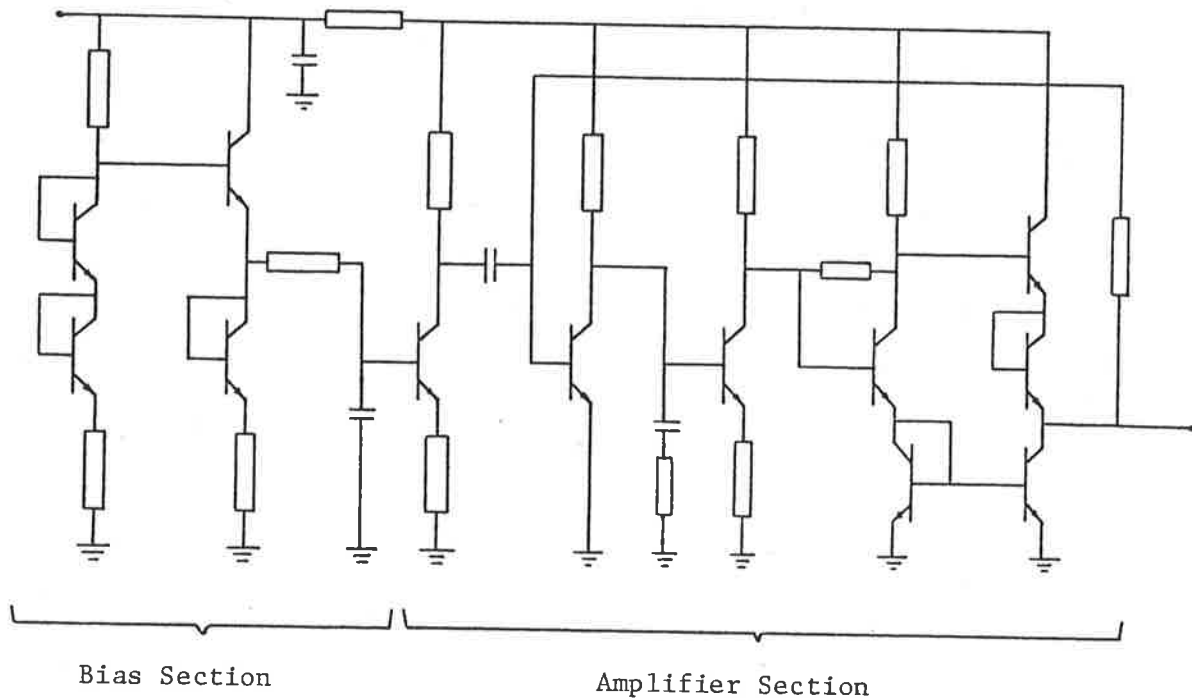
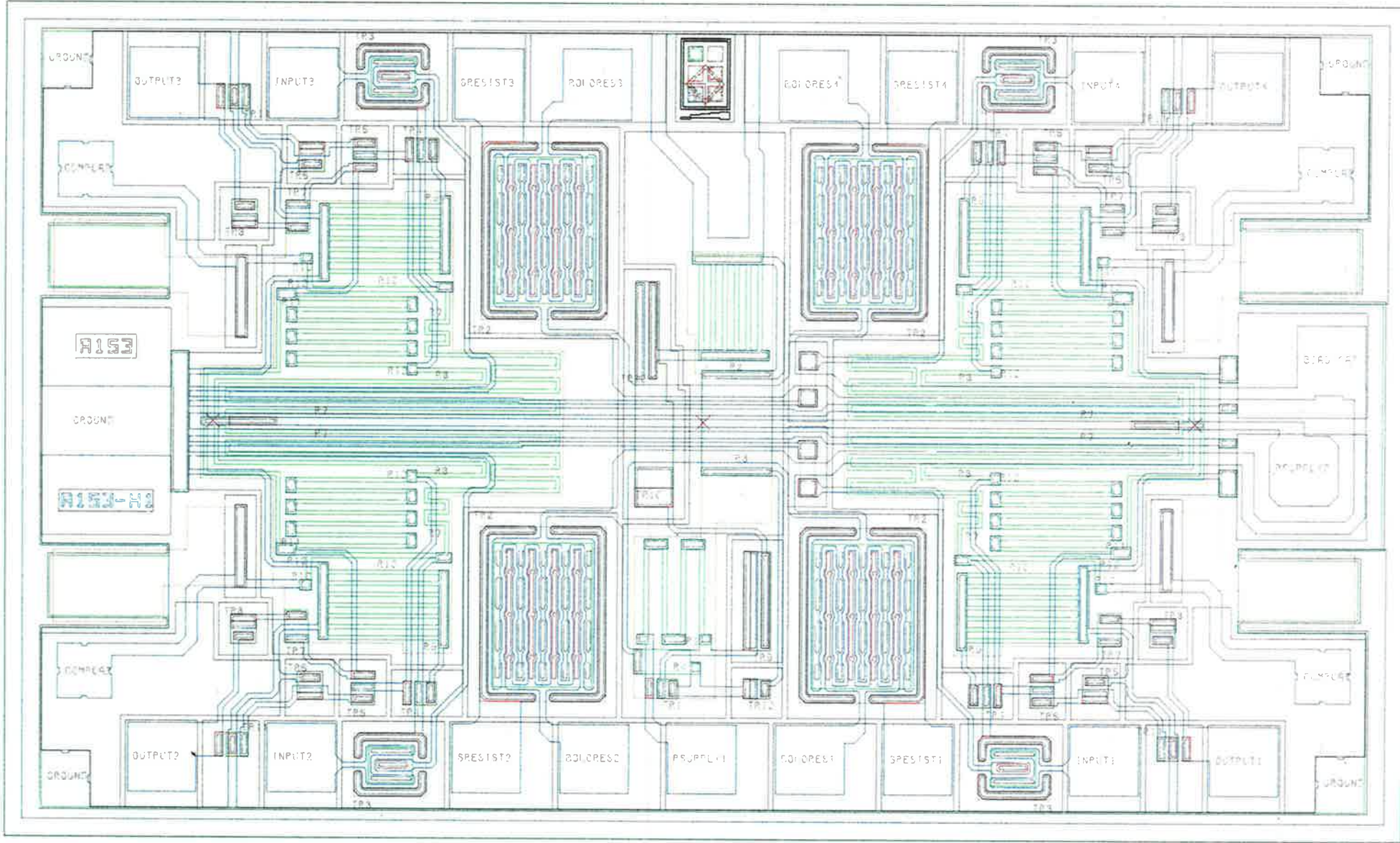


Figure 6.26 A Low Noise Amplifier Design

reference carrier directly from the transmitter, the reply signal is synchronously detected and then sampled and stored in the microcomputer memory for further processing. In another configuration, the rf signal is sampled at a rate sufficient to enable later reconstruction of the signal. Once the reply is demodulated, detection of the message starting bit and decoding of the message is performed under software control. The block schematic of the microcomputer based digital signal processor, called the RADDAT system, developed by the author and built with the assistance of Mr. N. R. Blockley, is shown

Figure 6.27 Layout of Low-Noise Amplifier



0-0000-0

0-0

0-0000-0

in Figure 6.28. The RADDAT has a very large virtual memory because of its capability to transfer data to the mainframe computer and receive it back when necessary. A summary of useful RADDAT commands is given in Table 6.2, and the system specifications are given in Table 6.3.

The RADDAT is, therefore, capable of sampling the reply signal before detection at a predetermined rate set by the capabilities of the A-to-D converter. The software is then used to detect the information sequence, find the synchronizing prefix and find the codeword. The codeword is usually encoded in order to minimize the bit error rate. The BCH encoding is employed at the transponder to achieve this. The received information sequence is then decoded to detect and correct errors which might have occurred during transmission.

6.6.3 Receiver Performance Evaluation

The effectiveness of the receiver described above has been evaluated in the presence of impulsive noise by means of simulation of the receiver functions. The inputs used in these simulations include an assumed signal level and the measured values of environmental noise.

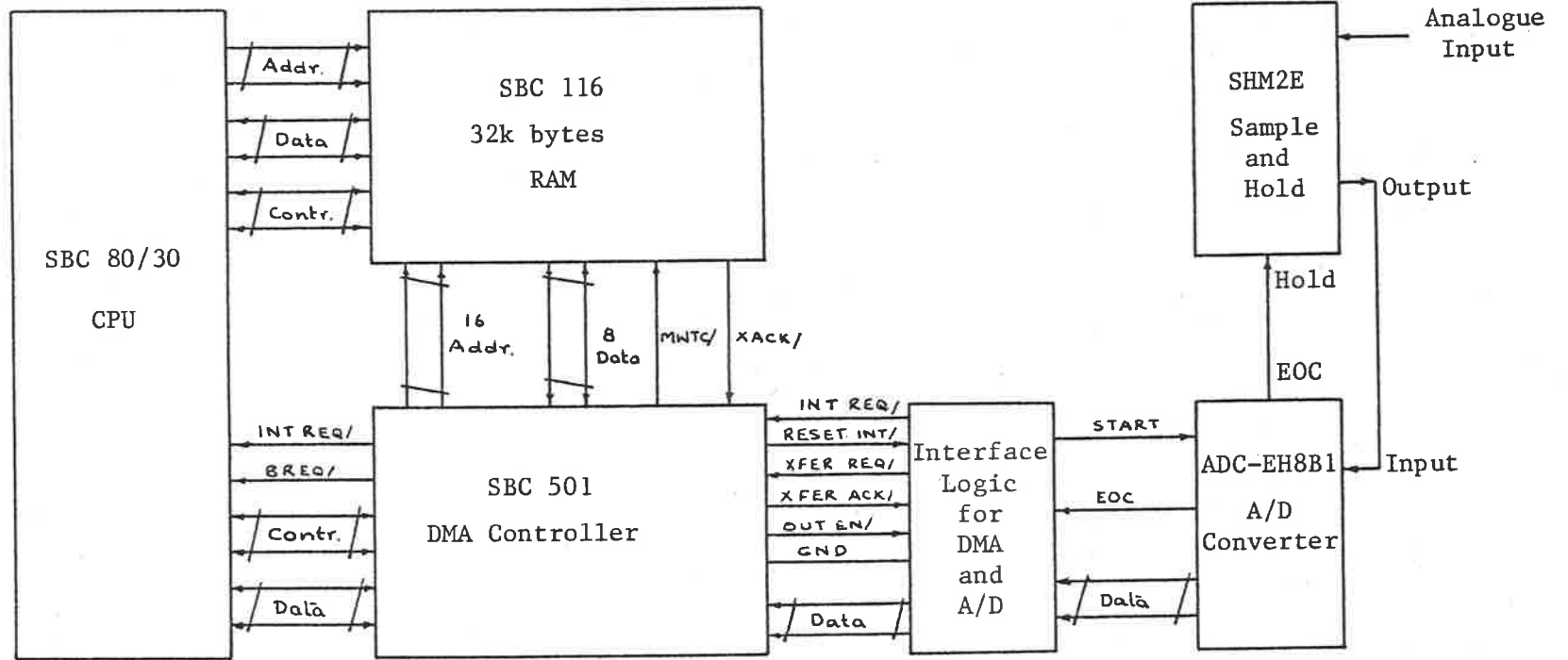


Figure 6.28a RADDAT Block Schematic

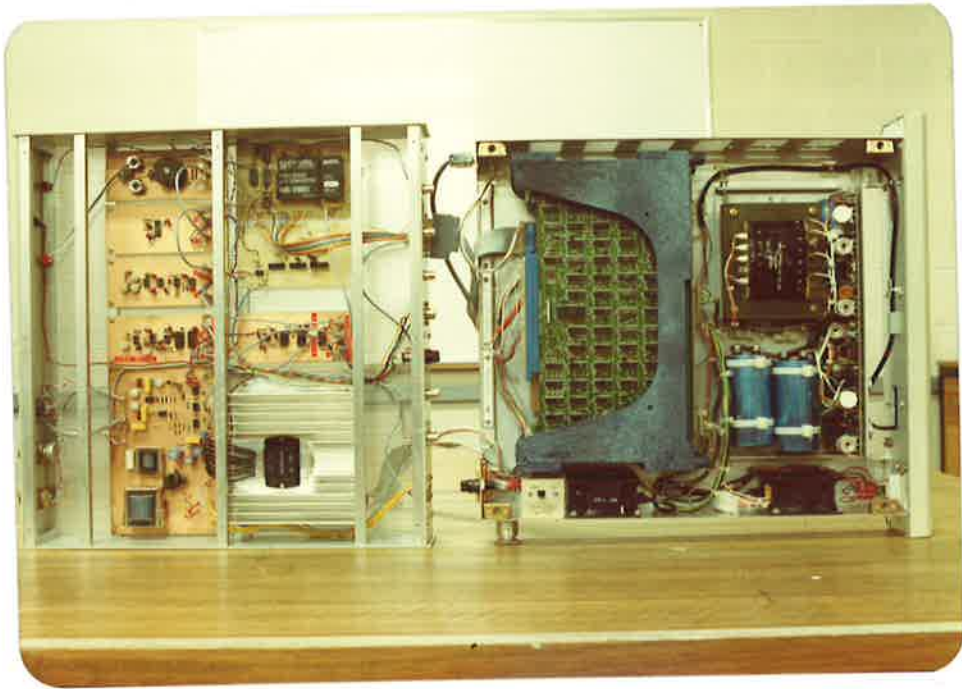


Figure 6.28b The RADDAT System

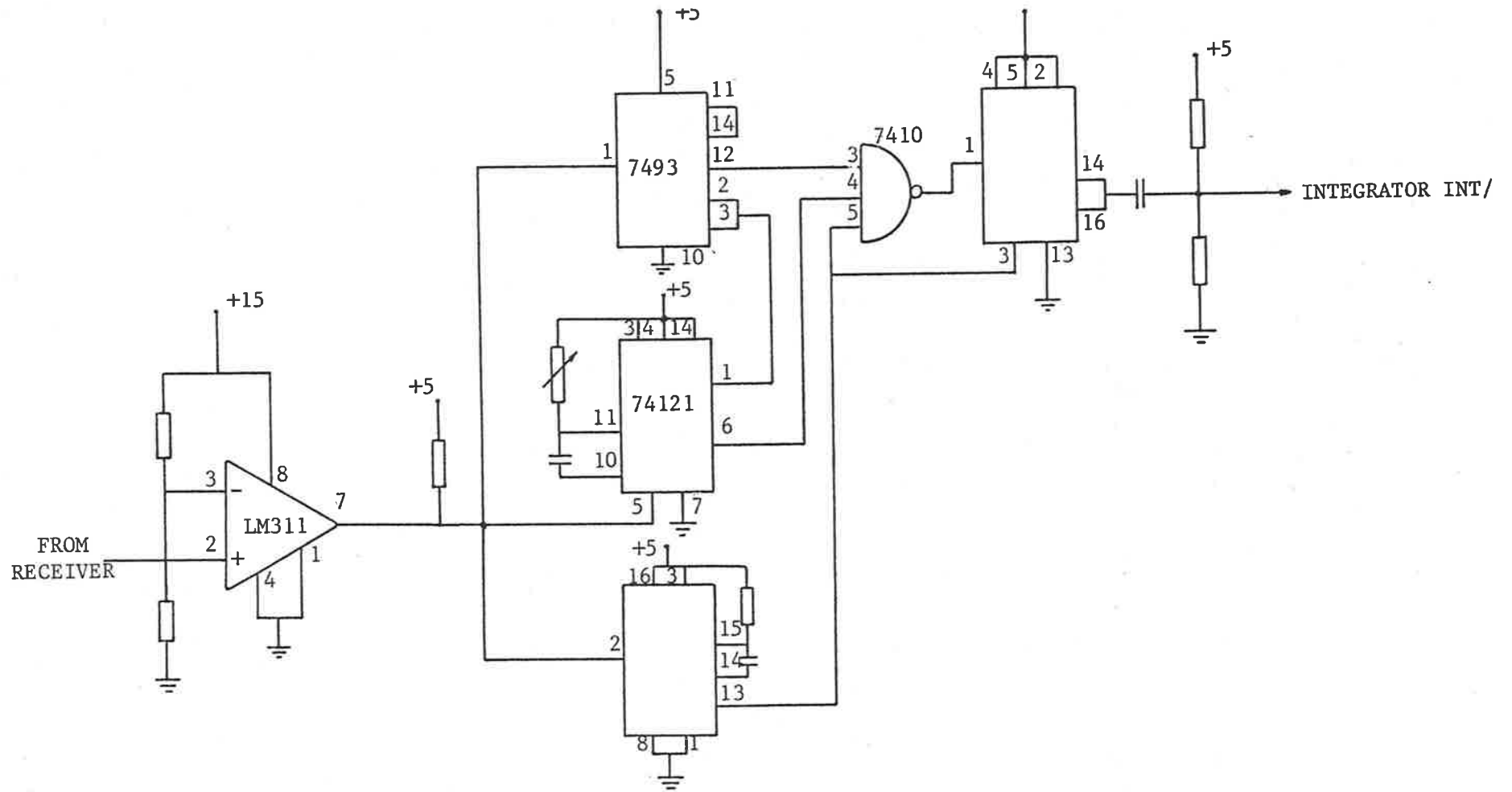


Figure 6.28d Automatic Interrupt from Integrator

Table 6.2 RADDAT Command Summary

COMMAND	ACTION (EXPLANATION)
A<address>	Insert Address Command
C	CYBER Communication Mode Command
D	Display Data Command
F	Find Code Command
I<address>	Insert Sample Data Command
J<prefix pattern>	Insert Code Prefix Command
L	Load Sample Data Command
M	Return to MONITOR Command
N<sample length>	Set Sample Length Command
P	Process Data Command
R<lo address>,<hi address>	Receive Data from CYBER Command
S	Start Sampling Command
T<lo address>,<hi address>	Transmit Data to CYBER Command
V	Load Test Values Command
W	Display Status Word Command

Table 6.3 RADDAT System Specifications

CPU	8085-A
Clock Rate	2.7648 MHz
Cycle Time	1.44 microsecs
Data Length	8-bits
Address Length	16-bits
Address Capacity	65536 bytes
RAM Available	32768 bytes
ROM MONITOR	2048 bytes
ROM RADDAT	2048 bytes
ROM Sockets	8k bytes
Baud Rate	2400
Serial I/O Ports	2
Parallel I/O Ports	9
Interrupts	12 vectored
DMA Transfer Rate	330k bytes/sec
A/D Sampling Interval	4.7 microsecs

The effect of impulsive noise on a binary DPSK system is evaluated by simulating a coherent DPSK detector. The bit error rate for various signal-to-noise ratios is plotted in Figure 6.29. The degradation in performance due to bit synchronization error is also illustrated. The signal-to-noise ratio is defined by calculating signal power and peak noise power across the receiver antenna dynamic impedance.

The effect of employing BCH codes on bit error rate is also investigated by means of simulation. In this case, it is assumed that there is no bit synchronization error. A BCH decoder capable of detecting up to two errors and correcting one error in a 64-bit information sequence is used in this study. The bit error rate after correction of all single bit errors in a complete message and the resultant plot for two noise situations are shown in Figure 6.30.

6.7 Conclusions

These studies of the receiver characteristics indicate that the proposed receiver is suitable for PST applications. It also appears, from the simulation, that a BCH code capable of detecting double

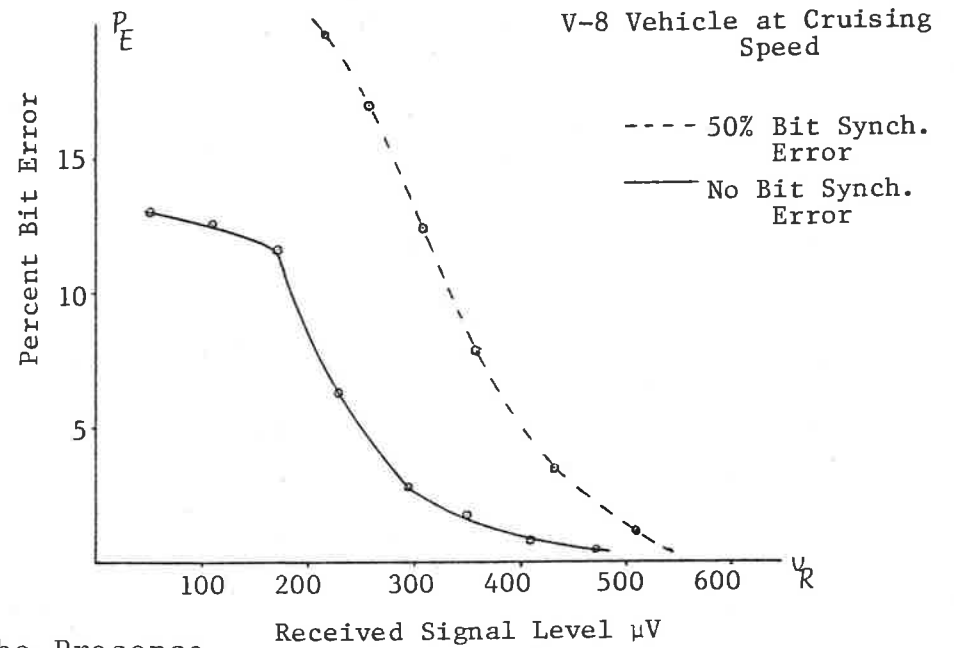
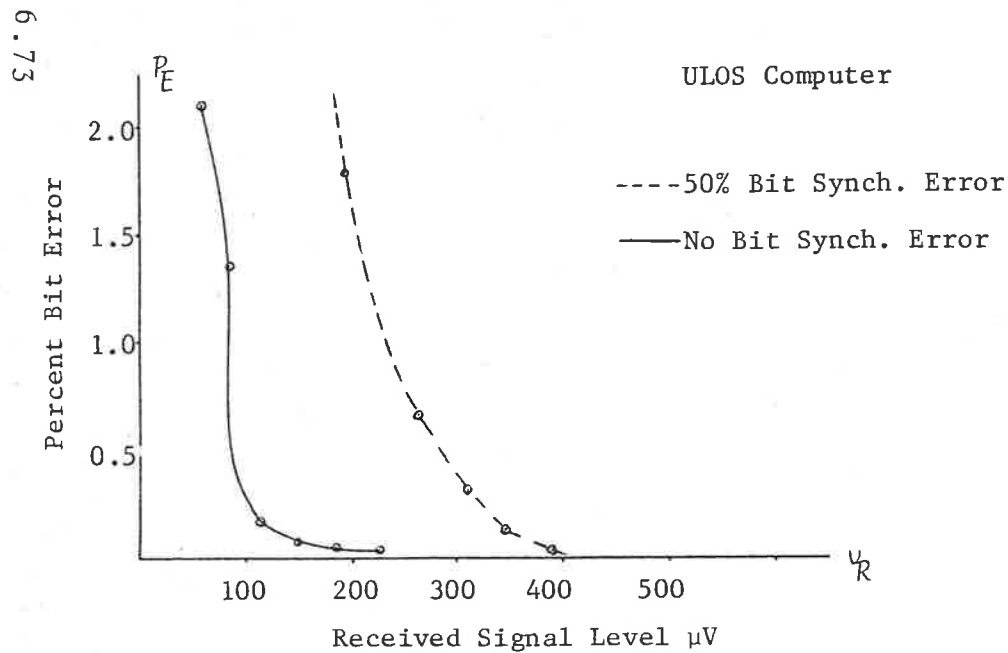
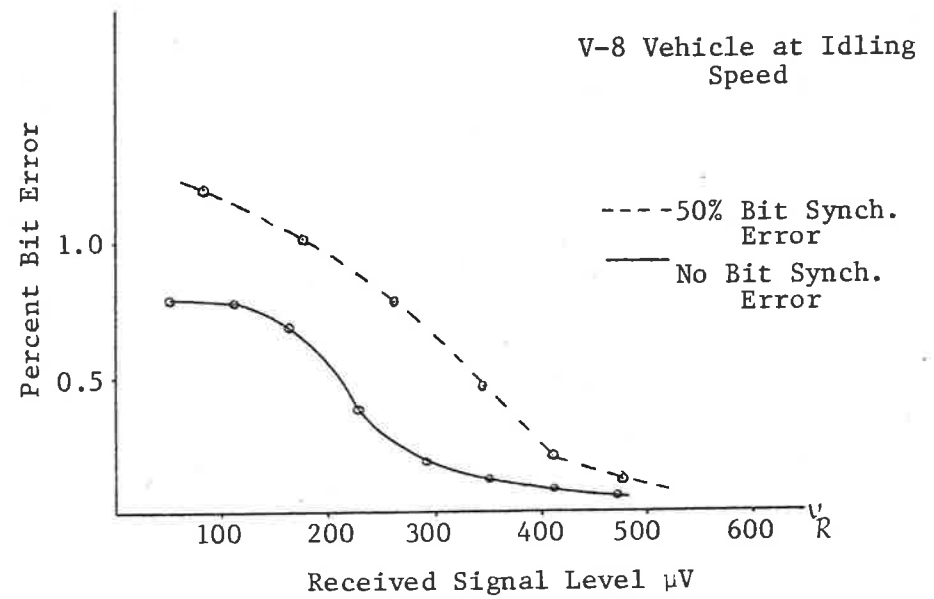
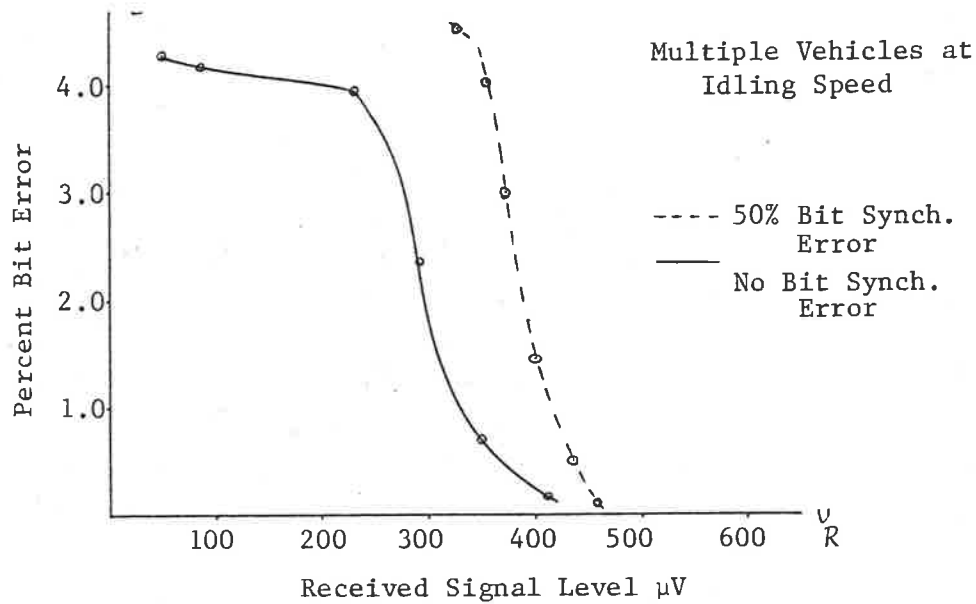
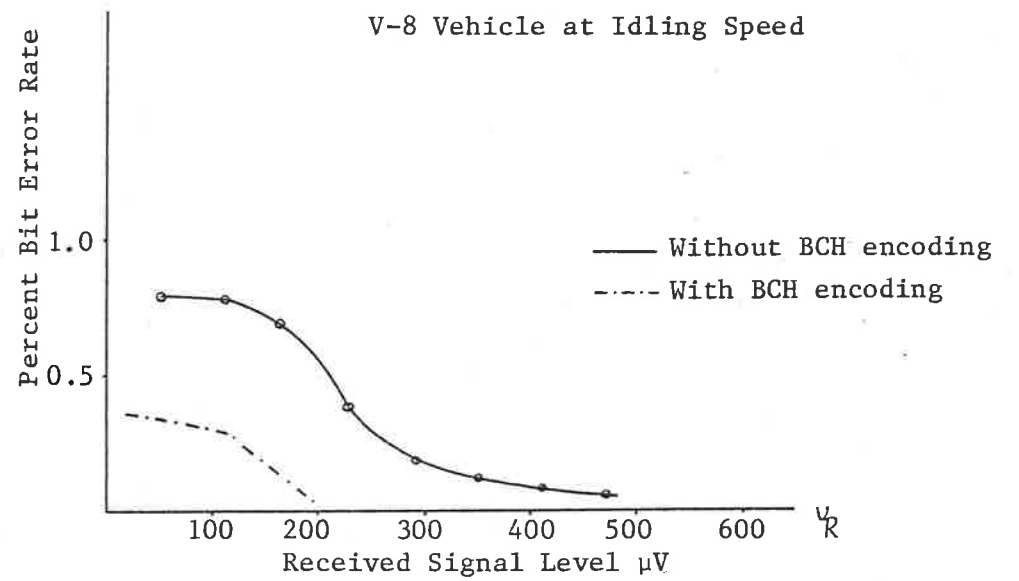
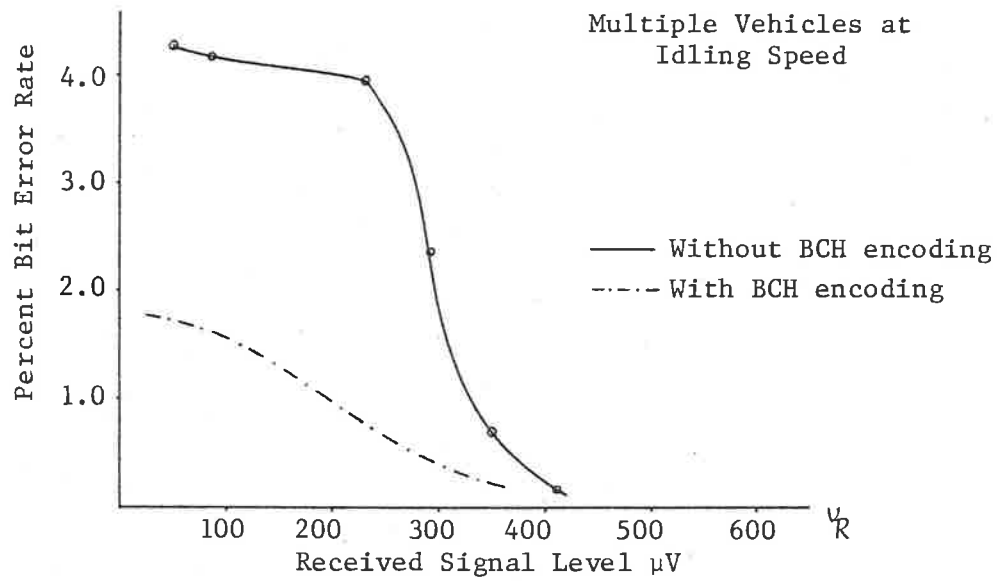


Figure 6.29 Noise Performance of a DPSK System in the Presence of Impulsive Noise



6.74

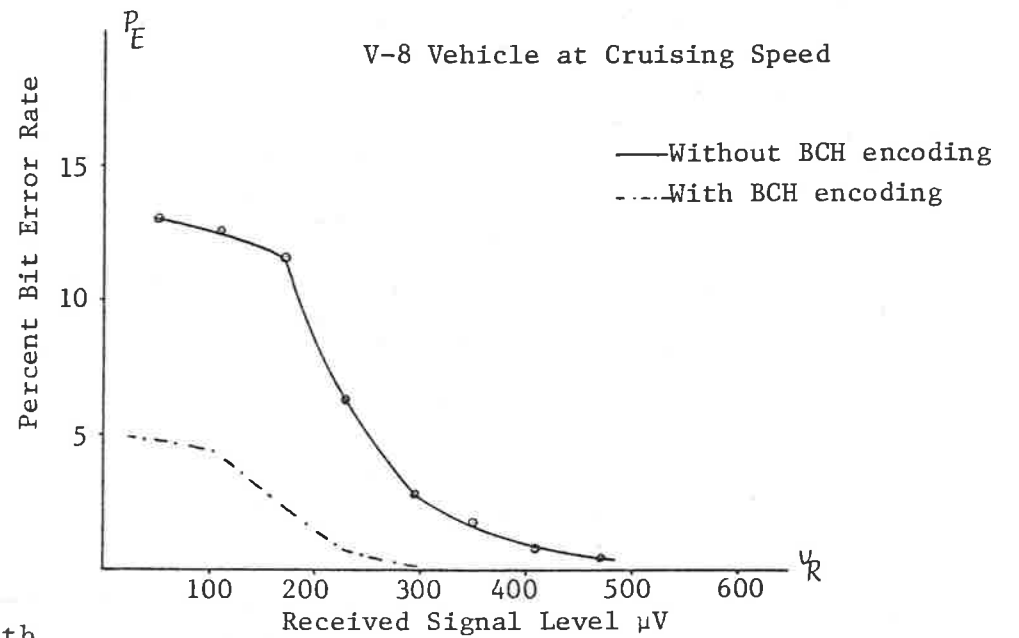
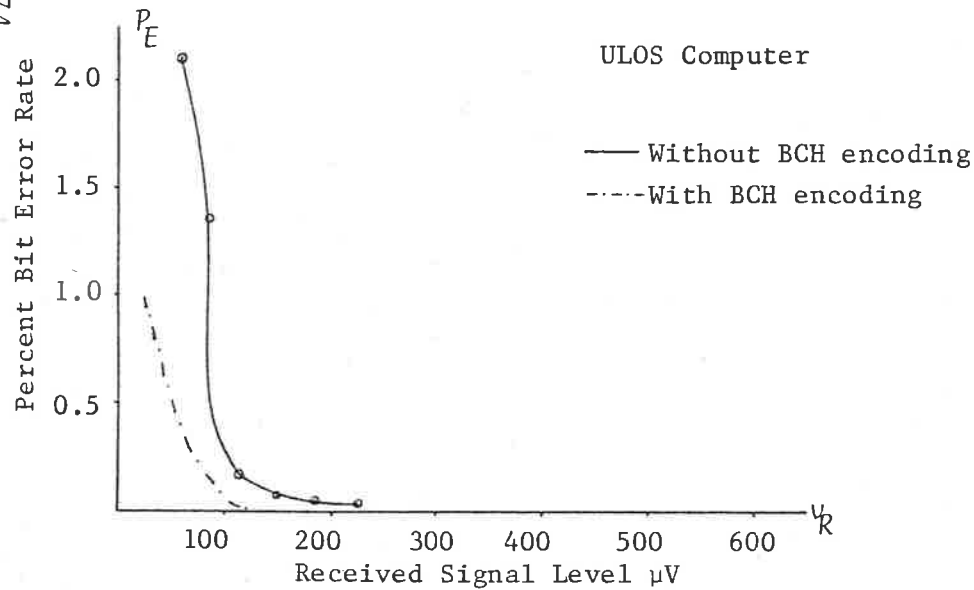


Figure 6.30 Improvement in Noise Performance with BCH Encoding

errors, and correcting single errors achieves a bit error rate acceptable in vehicle identification situations. The simulation was performed before the receiver hardware was built, to ensure that the proposed receiver scheme is suitable for PST applications.

CHAPTER VII

TECHNOLOGIES FOR TRANSPONDER REALIZATION

7.1 Introduction

In most object identification applications the transponders are required to be small. Although the need to integrate the transponder electronics is important from the point of view of overall size, the major advantages of integration result from the low cost, high reliability, reduced power consumption, and lower parts count. There are several technologies available for integration of these devices. Most of the existing technologies, such as NMOS, CMOS, I²L etc., offer the advantages mentioned above in varying degree. The suitability of any of these technologies for the integration of low-frequency passive transponding devices depends upon the requirements of these transponders, and on the characteristics of the available technologies.

7.2 Choice of Technology

Some of the factors that must be considered in choosing a technology for the passive low-frequency object identification applications include circuit density, availability of various circuit functions, circuit performance versus power consumption characteristics, topological properties of circuit layout, compatibility with the remaining parts of the system, and availability of design tools and fabrication facilities. With the rapid advance of semiconductor technology, it is expected that the function and performance of all currently available technologies will continue to improve without a corresponding increase in cost.

The cost per circuit is a function of the yield, and the packing density. The yield is related to the complexity of the technology, and the active chip area. The factors which determine the complexity of an IC technology include the design rules, number of masking steps, and mask alignment accuracy. Quantitatively, performance is expressed in terms of the fault density d_F per unit area. The higher the complexity of the fabrication process, the higher the corresponding fault density. The yield, Y , can be expressed as a function of the fault density, d_F , and the active area of the chip, A_C . The simplest model for the yield, or the frac-

tion of the chips fabricated that do not contain any fatal flaws, is based on the assumption that the defects are randomly distributed over the total area of the slice and one or more defects within a chip cause it to be non-functional. For this simple model of the flaw characteristic, the probability that a chip has n defects is given by the Poisson distribution, $P_n(d_f A_c)$ (Mead and Conway, 1980)

$$P_n(d_f A_c) = \frac{(d_f A_c)^n}{n!} \cdot \exp(-d_f A_c) \quad (7.1)$$

The probability of a good chip, therefore, is given by the Equation (7.2) below:

$$P_0(d_f A_c) = \exp(-d_f A_c) \quad (7.2)$$

While this equation is not the most accurate representation of the behaviour of real-life fabrication processes, it is a widely used model for initial yield estimates. In practice, the performance of most of the modern-day IC processes is better than that predicted by Equation (7.2), and the yield drops off at a rate considerably less than exponential function of the active chip area (Warner, 1974). Circuit designs employing fault tolerant schemes such as functional redundancy also help in improving the yield for the larger circuits.

The packing density, d_p , is determined by two factors, namely, the component density, which is inversely proportional to the area of an elementary device, and the interconnection density. The packing density, which takes into consideration both the component and interconnection densities, represents the ability of the technology to produce a highly dense chip and, hence, a complex circuit at a low cost.

Consideration should also be given to the performance of the technologies as characterized by their digital and analogue capabilities. The most important performance criterion for digital circuits is the power-delay product per gate, or logic function. For the analogue circuits, on the other hand, availability of matched active devices with high gain and low noise, matched resistors and capacitors are important performance parameters.

The technologies which have emerged to date, capable of providing high functional density, low power consumption, excellent reliability and capable of being scaled to submicron geometries without an exponential increase in the power consumption per unit chip area, are the n -channel metal oxide semiconductor silicon gate (NMOS) process, complementary metal oxide semiconductor silicon gate (CMOS) process,

and the integrated injection logic or the merged transistor logic (I²L) process. Several variations of MOS and I²L technologies have been proposed by various authors (Henning et al, 1977), (Rodgers and Meindl, 1974). Another planar technology which shows a great deal of potential in very high speed logic circuits is the gallium arsenide (GaAs) process (Eden et al, 1978). Although the planar GaAs technology is in experimental stages, it provides a serious challenge to the existing silicon technologies. Experimental GaAs devices offering nearly two orders of magnitude advantage in power-delay product have been produced. The three technologies mentioned above which are in wide-spread use at a commercial level are considered for the object identification application. Various experimental processes, although capable of better performance, are not considered in this text. A review of the NMOS, CMOS and I²L technologies is presented, and the characteristics of these technologies are considered, in the light of their possible use in passive transponder applications.

7.2.1 NMOS Technology

The cross-section of a typical high performance NMOS structure is shown in Figure 7.1 (Dennard et al, 1974). The device has a scaled-down geometry.

The characteristics of the device include high substrate resistivity to minimize junction capacitance, short channel length, thin gate oxide, self-aligned polysilicon gate, channel and field implants, shallow source and drain junctions, double polysilicon interconnect layers, depletion or polysilicon loads, low voltage operation, and on-chip substrate voltage generation. These characteristic features of the NMOS structures are given in Table 7.1. Small device geometries allow a reasonably high device density. The interconnect density is improved by using thin metallization for moderate current levels.

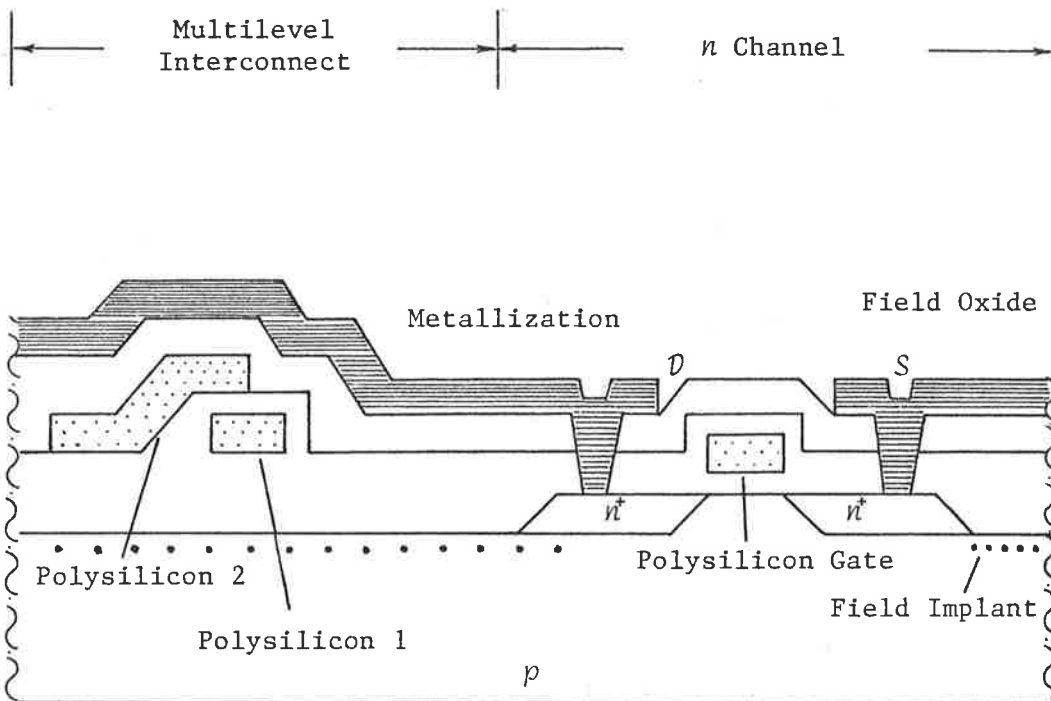


Figure 7.1 Cross-section of an NMOS Structure

Metal migration, which causes the metal atoms to move slowly in the direction of the current flow when the current density exceeds a certain limit, can be reduced by the addition of small percentages of copper and silicon to the aluminium. This technique also prevents diffusion of the silicon in the aluminium and, hence, the resulting shorted junctions. Metal migration poses less of a problem in switching circuits. Short current pulses contribute much less to metal migration than steady direct current (Mead and Conway, 1980).

Table 7.1 Typical Characteristics of NMOS Digital Circuits

CHARACTERISTIC	SYMBOL	VALUE	UNIT
Channel Length	ℓ_c	3	μm
Gate Oxide Thickness	t_{ox}	500	\AA
Channel Width	w_c	1.2	μm
Voltage of Operation	V_{dc}	5	V
Packing Density	d_p	250	Gates/ mm^2
Gate Delay	τ	2	ns
Power-Delay Product	$E_{\Delta w}$	1	pJ

The NMOS technology, therefore, is medium in complexity, offering reasonably high backing density. Fast switching at reasonably low switching energies is possible using NMOS devices. These features make NMOS the most widely used commercial process for digital integrated circuits at the present time.

Although the analogue capability of the NMOS technology is not as good as its digital capability, the application of NMOS technology to analogue circuits has, nevertheless, received much attention. The key elements of analogue circuits in NMOS process include precision ratioed capacitor arrays, analogue switches, operational amplifier, voltage references, comparators, and charge transfer devices (Hodges et al, 1978). The major disadvantages of NMOS technology in the realization of analogue circuits are the unavailability of complementary devices, the low transconductance of the MOS transistors, the high offset voltage associated with MOS transistor pairs, and the high equivalent input noise of the device, particularly the $1/f$ component. Some of these shortcomings are circumvented by improved processing techniques to reduce offset and noise, as well as improved design techniques such as sampled data approach for offset cancellation (Hodges et al, 1978). Typical characteristics of NMOS analogue circuits are given in Table 7.2.

Table 7.2 Typical Characteristics of NMOS Analogue Devices

CHARACTERISTIC	SYMBOL	VALUE	UNIT
Single Stage Voltage Gain	A_V	100-1000	1
OPAMP Unity Gain Bandwidth	f_T	1-5	MHz
1/f Corner Frequency	f_N	5-10	kHz
RMS Input Equivalent Noise Voltage	V_n	30-100	μ V
Input Offset Voltage of Matched Pair	V_{os}	5-10	mV
Capacitance Range	C	0.1-100	pF
Capcitanace Ratio Accuracy	R_c	0.02-0.3	%

Although NMOS technology has its limitations in some analogue applications, it offers significant advantages in some others. One of the most useful applications of this technology results from the self-isolating property of the MOS transistor. The high impedance of the MOS transistor in the off state enables NMOS circuits to store charge on a node for periods ranging up to several milliseconds. In addition, it is also possible to sense the voltage at that node continuously and non-destructively due to the essentially infinite input impedance of the

MOS transistor in the active mode of operation. These properties of MOS devices provide a precision sample and hold capability, and have led to the development of new A/D and D/A conversion schemes (Gray et al, 1978), as well as switched capacitor filtering techniques (Bordersen et al, 1979). Another advantage is that MOS transistors make nearly ideal analogue switches with no dc paths between the gate and the source or the drain. In NMOS circuits, the threshold voltage limits the switched voltage range in analogue applications to one threshold below the gate switching voltage.

In digital circuit applications of NMOS technology, downward scaling of device dimensions is a useful technique, as it provides improved performance otherwise unattainable with the same process parameters. However, it is not as attractive for analogue circuit implementations, for several reasons. The $1/f$ noise appears to be inversely proportional to the square root of the total gate oxide capacitance (Ronen, 1973). Therefore, smaller devices have higher equivalent input noise characteristics. The increased transconductance of the scaled device provides only partial compensation for the increased noise. Even the transconductance increase is limited by velocity saturation effects as the channel length is reduced. The accuracy of matching

various components is adversely affected when the devices are scaled down, and the mismatched component values then appear as gain and offset errors. Furthermore, as the devices are scaled down, sub-threshold conductance does not scale, and the use of sample and hold properties of the MOS transistor are impaired (Dennard et al, 1974).

For analogue functions, high packing density cannot be achieved in NMOS technology. This is due to the increased complexity of the analogue circuits to compensate for the shortcomings of the process. Typical performance of an NMOS operation amplifier is shown in Table 7.3 (Hodges et al, 1978).

Table 7.3 Typical Performance of NMOS OPAMP

CHARACTERISTIC	VALUE	UNIT
Area	0.18	mm ²
Open Loop Voltage Gain	1500	1
Input Offset	<15	mV
Power Consumption	18	mW
Unit-Gain Bandwidth	2.1	MHz
Slew Rate, Rise/Fall	2.6/4	V/ μ S
RMS Input Noise	35	μ V
Power Supply	\pm 5	V

7.2.2 CMOS Technology

The cross-section of a high performance p-well CMOS structure is shown in Figure 7.2. The devices have scaled-down geometries, and are characterized by features including high resistivity substrate to minimize parasitic capacitances, short channel lengths, thin gate oxide, shallow junctions, self-aligned p-well and polysilicon gates, channel and field implants, double layer polysilicon interconnects, single polarity (n^+) polysilicon gates and low voltage operation.

The p-well CMOS technology is slightly more complex than NMOS technology, and circuit yields are, therefore, somewhat lower. The packing density is also lower than NMOS process. Typical characteristics of CMOS technology are given in Table 7.4.

Table 7.4 Typical Characteristics of CMOS Technologies

CHARACTERISTICS	VALUES	UNIT
Channel Length	<3	μm
Gate Oxide Thickness	<500	\AA
Voltage of Operation	3-5	V
Packing Density	100-150	g/mm^2
Gate Delay	<0.5	ns
Power-Delay Product	0.2	pJ

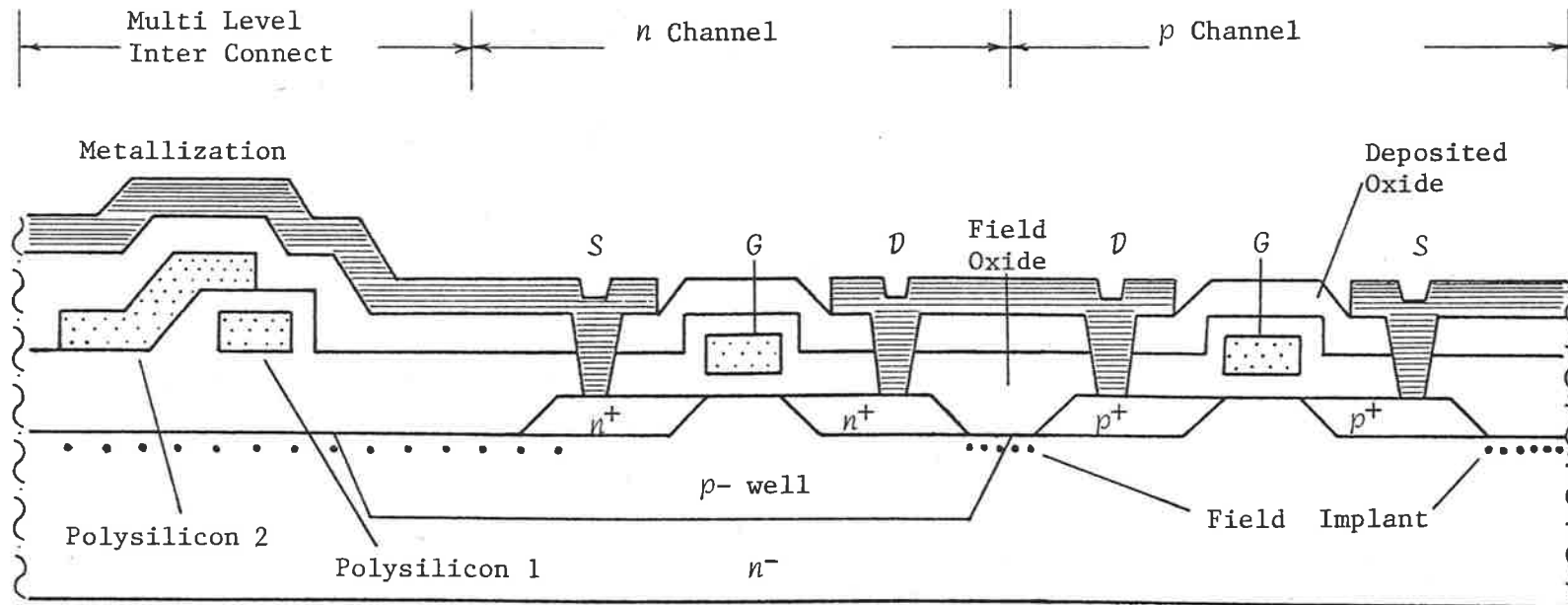


Figure 7.2 Cross-section of a p-well CMOS Structure

One of the most important advantages of CMOS over NMOS is that standby power dissipated in CMOS is negligible. This property is significant when the heat generated by the dense circuitry on the chip could become excessive. Another advantage of CMOS results from the use of active pull-up transistors, and it thereby offers higher resistance to soft errors in memory applications than NMOS devices using polysilicon pull-up loads.

One of the major drawbacks of this technology arises due to the tendency of CMOS devices to latch-up when components are replaced with power applied. This tendency of CMOS is a potential reliability problem. A simple solution to this problem involves the use of an n or n^+ epitaxial starting material to reduce substrate resistance, and prevent sustained latch-up. In addition, in all critical input-output circuits, n -channel devices may be surrounded by p^+ guard rings grounded to p -well, and p -channel devices may be surrounded by n^+ contacts tied to the substrate to reduce lateral potential build up. This technique is believed to eliminate the latch-up problem completely (Payne et al, 1980).

The analogue capabilities of the technology are considerably better than those of NMOS due to the availability of complementary devices, the fact

CMOS analogue switches do not have the threshold drop limitation inherent in NMOS technology, and the availability of the *npn* bipolar device inherent in CMOS structure. Although this *npn* transistor can only be used as an emitter follower with its collector tied to the substrate, it is extremely useful in a variety of applications, such as an output stage pull-up transistor or band-gap voltage reference (Tzanateas et al, 1979). The reduction of power consumption by the use of switched CMOS amplifiers is a very interesting application (Hosticka, 1980).

The CMOS technology loses its power dissipation advantage for analogue functions because all the transistors conduct current all the time. Typically, a low gain OPAMP with characteristics similar to those of the NMOS OPAMP mentioned previously occupies approximately 25% less area. The power dissipation of the amplifier is also correspondingly lower.

Several alternatives to the *p*-well CMOS technology are being investigated for VLSI systems. One of these which may be suitable for telecommunications applications, such as the passive transponder technology, is the *n*-well CMOS process. The cross-section of an *n*-well CMOS structure is shown in

Figure 7.3 (Ohzone et al, 1980). This technology combines the low power of CMOS with the high speed of NMOS. The n -well CMOS gets its speed by putting the faster n -channel device in a high resistivity p -type substrate and the p -channel device into diffused n -well which gives the process its name. By using a high resistivity substrate, the parasitic diffusion capacitances and the substrate body effect of the n -channel device are minimized, resulting in higher switching speed capabilities. The process achieves its higher packing density than the p -well process, thereby making itself competitive with NMOS, due to the fact that, in most circuits, p -channel devices are in the minority, which results in

As far as analogue functions are concerned, the use of n -channel devices, with their inherent high gain and low body effect as the main gain elements, provides improved analogue performance over the p -well technology, where, due to the high body effect, n -channel devices are in the minority.

Another recent development in CMOS technology involves the diffusion of both well types in a very lightly doped or intrinsic substrate in order to optimize both n - and p -channel devices independently (Parrillo et al, 1980).

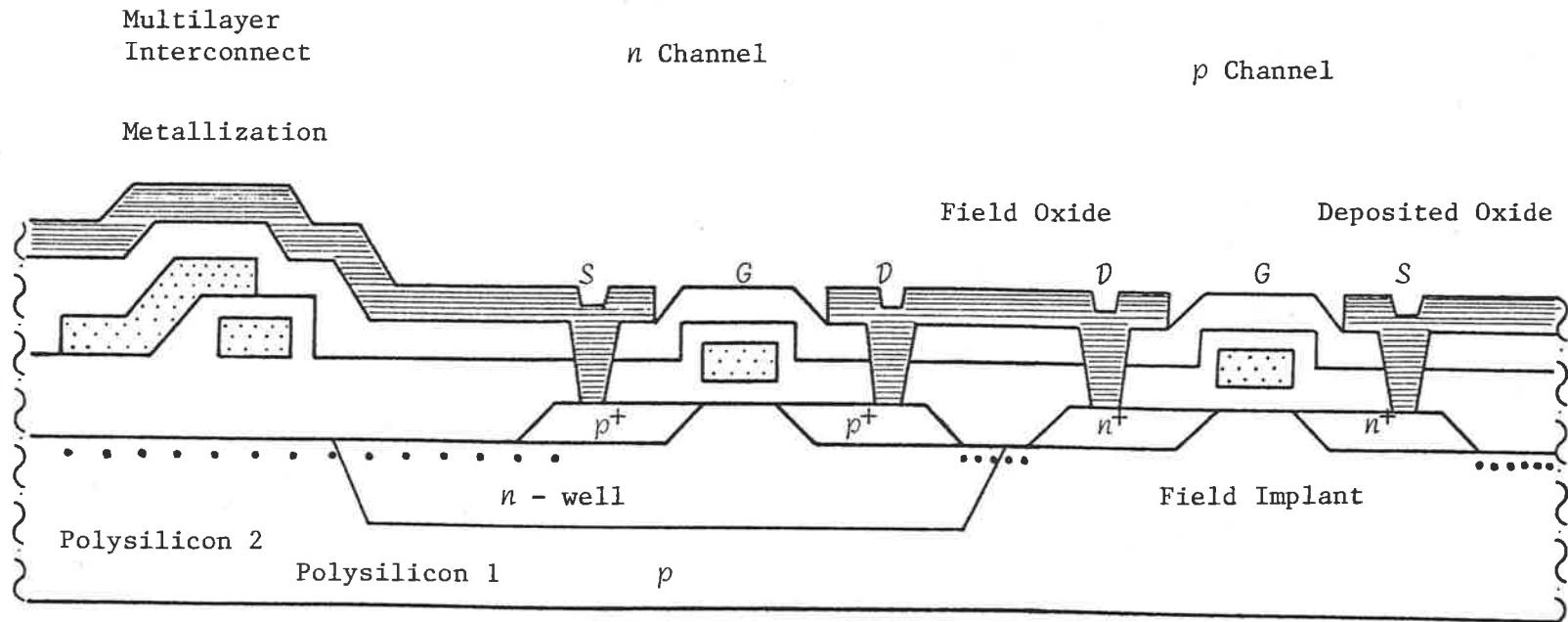


Figure 7.3 Cross-section of an n -well CMOS Structure

7.2.3 I²L Technology

The third technology which has a lot of potential in digital as well as analogue large-scale integrated circuit implementation is the I²L technology, and is an extension of the well-known bipolar technology (Hart and Slob, 1972; Berger and Wiedmann, 1972). The cross-section of an I²L structure is shown in Figure 7.4.

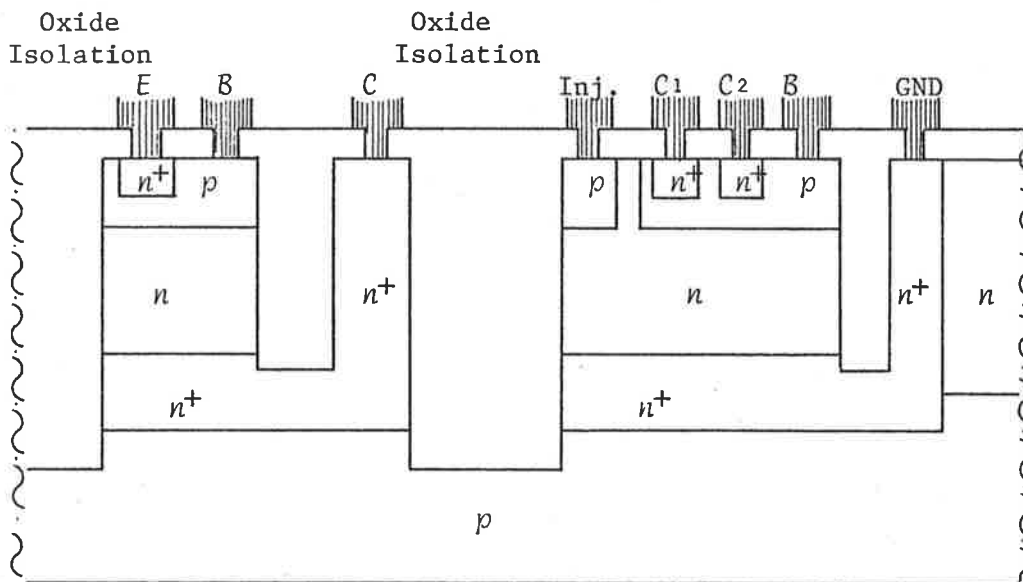


Figure 7.4 Cross-section of an I²L Structure

The I²L technology is capable of providing logic function at high density. It is possible to provide analogue circuit functions on the same chip by the addition of some more processing steps which,

of course, increases the complexity of the resulting technology. The I²L structure shown in Figure 7.4 is characterized by features which include scaled devices, ion-implanted self-aligned base, laser-trimmed resistors, and shallow junctions. The technology can be optimized for high speed digital applications by using Schottky diodes instead of the base-collector junction diodes to eliminate minority carrier charge storage in the I²L gate. Polysilicon interconnects, and base contacts can also be used to achieve high functional density (Tang et al, 1979).

It may seem, at the first glance, that I²L digital circuitry combined with bipolar linear circuitry offers the best of both worlds: high density digital circuit and high voltage, high current linear. But the bipolar process that yields high performance, high density I²L does not necessarily yield high voltage linear circuits. A certain amount of compromise and optimization is required to achieve a compatible process which is commercially viable. Technological improvements in I²L processing have resulted in an order of magnitude improvement in the current gain of the I²L transistors. The analogue *npn* transistor exhibits current gain in the range 100 to 200, with the collector-emitter breakdown voltage in the range 40 to 70 volts and a cutoff frequency greater than 1 GHz (Ozawa et al, 1979).

The technology to achieve the above overall performance is fairly complex. The present-day technology is capable of high packing densities comparable with the NMOS process. The power-delay products achievable are usually lower than the NMOS at similar gate delays (Henning et al, 1977).

The analogue capabilities of the technology are much better than the existing MOS technologies due to the availability of high gain complementary bipolar transistors, low offset voltages for device pairs, low noise, and devices capable of handling high voltage and high current. The functional density for analogue implementations is seldom high due to the isolation requirements within the circuit. However, modern developments in the isoplanar process, such as FAST and I³L, where components are isolated by a selectively grown thick oxide, allow higher functional densities than those achievable by normal planar processes.

7.3 Comparison of the IC Technologies

The objective of integration in any system application is the fabrication of as much of the system as possible on single chips of silicon. With advances in technology, it is possible to put

more and more system modules on single chips without a corresponding increase in the chip area or power consumption. There are fundamental limitations in each of the existing technologies which must be considered to achieve the ultimate goal of complete integration. For this goal to be attained, any signal that is required in the system, other than inputs, outputs and power supplies, must not be generated by means of the technology in the chip. In other words, the subsystems must not require a different technology for the generation of internal signals.

The three silicon technologies described previously have some common features which make these technologies economically viable for analogue and/or digital LSI/VLSI circuit implementations. One of the important common features is the availability of two types of device in each technology. In order to provide some kind of nonlinear threshold operation, there must be a device which is normally off when the lowest available system voltage is applied to its control input. This objective is achieved in the bipolar technologies by means of *npn* transistors, whereas in the MOS process *n*-channel enhancement mode devices are used. A separate type of device is required, in addition, to allow the output of a driver device to reach the highest voltage in the circuit. In the bipolar technology,

this is provided by a lateral *pnp* transistor, in the NMOS technology a depletion mode device is used, and in the CMOS process a *p*-channel enhancement mode device is used. Therefore, in each of these technologies threshold operation is possible, and the output can reach that maximum supply voltage. Other common features of these technologies include high packing density, and a capability to scale to sub-micron dimensions without a corresponding increase in power dissipation per unit area of the chip.

Although present forms of the commercially available I²L process lack the additional level of interconnect available in the MOS technologies, the studies indicate that there is no inherent reason such a level could not be provided (Tang et al, 1979).

From the above discussion, it appears that no single technology provides the optimal solution to the requirements in various applications. In applications where both analogue and digital functions are required, and the minimization of power consumption is the most important objective, the CMOS technology has a clear advantage over NMOS, but the cost of the chip goes up as well, due to the increased processing complexity and chip area. The problem is further complicated, and the resulting choice of the appropriate technology made more difficult, by

the availability in NMOS of dynamic techniques that reduce the power consumption a great deal, at the expense of more area. Furthermore, I²L has the advantage over the MOS technologies that the power per unit area and, hence, the effective delay of elementary logic functions, can be controlled by an off-chip voltage. The decision concerning at what point on the speed versus power curve to operate may thus be postponed until the time of application or even changed dynamically in I²L technology. The choice of technology is further complicated by the rapid advances in the technologies discussed above, as well as the ones in the experimental stages. At the present levels of complexities, the expected applications of various technologies are shown in Table 7.5 (Salama, 1981). In the passive transponder applications, where both analogue and digital circuitry is required to be integrated on the same chip, operating at the minimum power level, both CMOS and I²L technologies offer low power operation, but I²L, with higher packing density, requires smaller silicon area and, therefore, has a cost advantage over CMOS.

Table 7.5 Possible Applications of Various Technologies

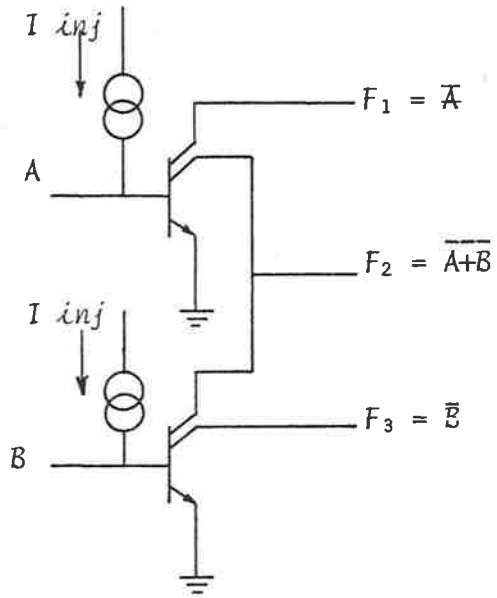
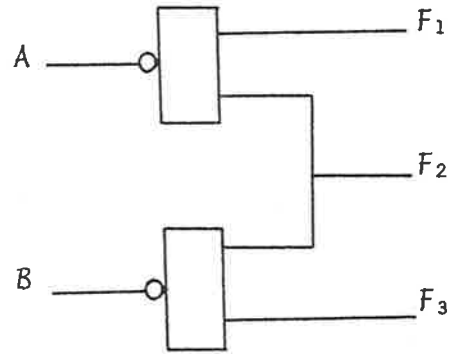
NMOS	CMOS	I ² L/BIPOLAR
Memory	Static RAMs	OPAMPs
Microcomputer	Low Power Microcomputers	PCM Repeaters
Standard MSI & LSI	Standard MSI & LSI	Programmable Logic Arrays
Programmable Logic Arrays	Digital Signal Processors	SLICs
Digital Signal Processors	Multiplexers	PST Labels
	Codecs/Filters	
	Line Powered Telephone Circuits	
	Cross-point Arrays	

7.4 Integration of the Two-port PST Label

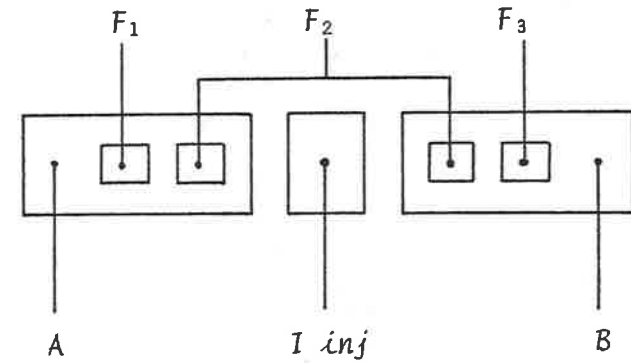
In order to apply the passive subharmonic transponder to a practical object identification application, it is necessary to integrate the transponding label in a suitable technology. It appears, from the previous discussion, that I²L technology is the most suitable one currently available. An attempt has been made to produce a design of the two-port passive subharmonic transponder shown in Figure 2.5. Conversion of this design to I²L layout has not been attempted due to inadequate design facilities of the time.*

The basic structure available in I²L technology is the inverter with multiple collectors. The basic I²L building block is a NOR gate, as shown in Figure 7.5. The functional blocks of the two-port transponder described earlier are implemented in the I²L technology, and are shown in Figure 7.6. The use of large numbers of collectors is avoided to achieve uniformity in gain, and speed of all the collectors. It is, however, possible to

* This work was undertaken during the period when the idea of the one-port device had just evolved from the discussions of the author with Dr. P. H. Cole and Dr. K. Eshraghian. An IC layout software has been developed since that time.

a. An I²L NOR Gate

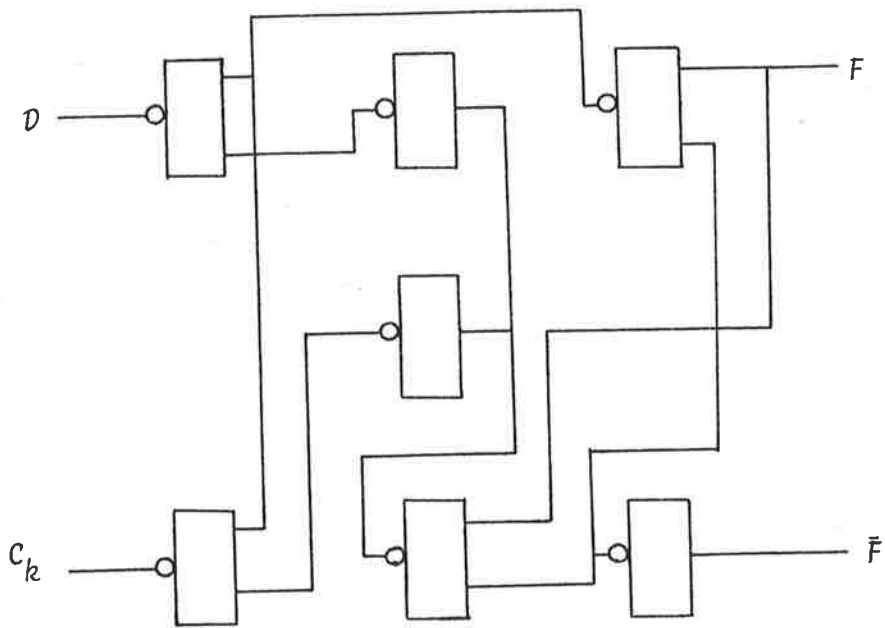
b. NOR Gate Schematic



c. Layout of NOR Gate

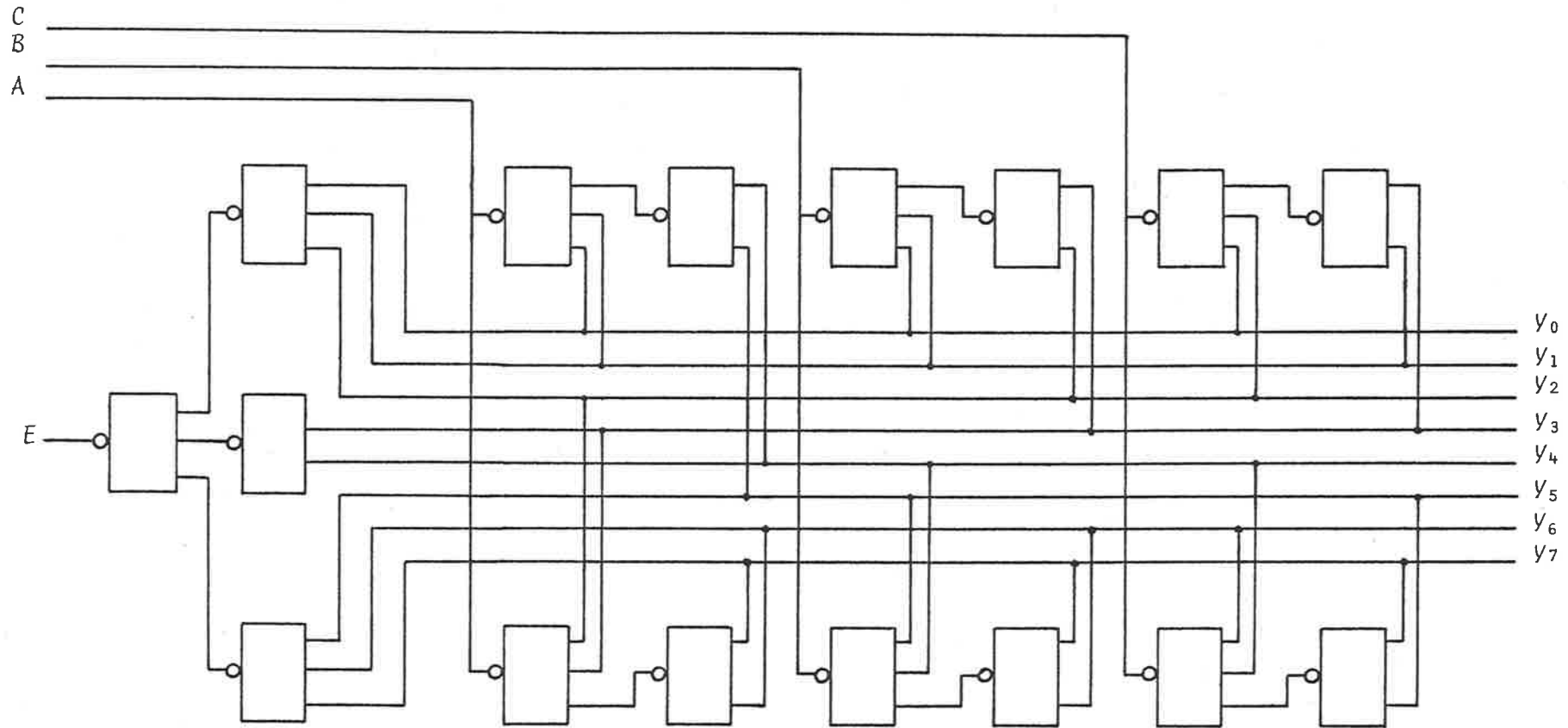
Note : A and B are logic inputs, $F_1, F_2,$ and F_3 are logic outputs and I_{inj} is the injector current.

Figure 7.5 Basic I²L Gate



Note : D is the data input, C_k is the clock input and F is the PSK output.

Figure 7.6b I²L PSK Modulator



Note : A, B and C are the select inputs and E is the enable input.
 $Y_0, Y_1 \dots Y_7$ are the data outputs.

Figure 7.6c An I²L 8-line Address Decoder with Enable

7.30

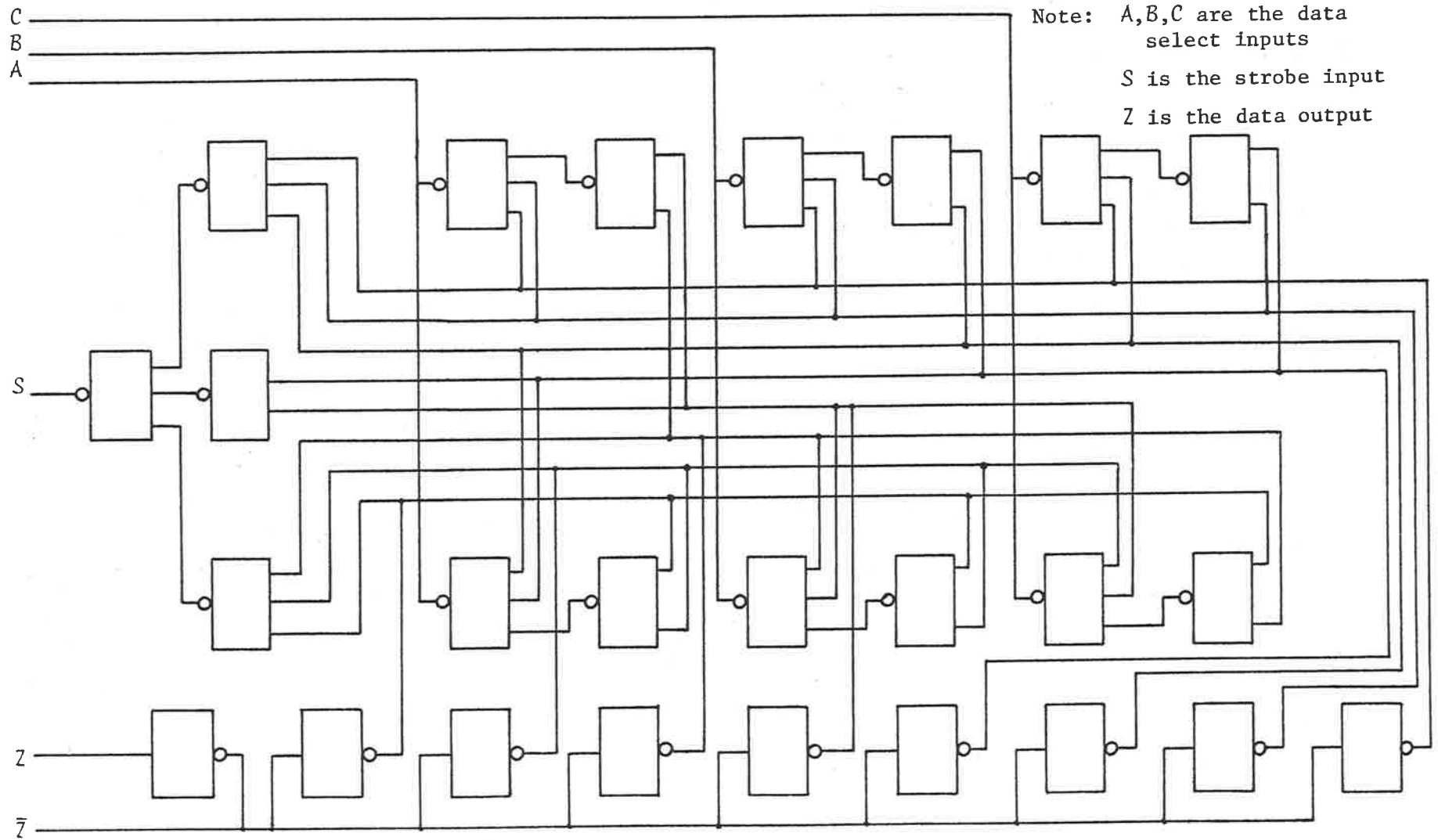


Figure 7.6d An I²L Multiplexer Schematic

increase the fan-out capability of the I²L gates, either by means of circuit design techniques or special layout techniques. These ideas are not employed in the design of two-port transponder building blocks.

The D-type flip-flop is used as a divide-by-two block. The output of the divider chain is used as the select input for the memory address decoder, the data select input for the multiplexer, and the PSK modulator. The two-port transponder using the building blocks described above is shown in Figure 7.7.

7.5 Future Considerations

The recent advances in the VLSI technologies is expected to have a large impact on analogue and digital communication ICs. Although the passive subharmonic transponder may be classified only as an LSI device, rather than VLSI, there are a lot of benefits to be accrued from these developments. In the past few years, higher levels of integration in IC technologies have been achieved by increasing the component density by scaling down the devices, increasing the interconnect density by reduction in design clearances and using a second interconnect layer, increasing the chip area without a corres-

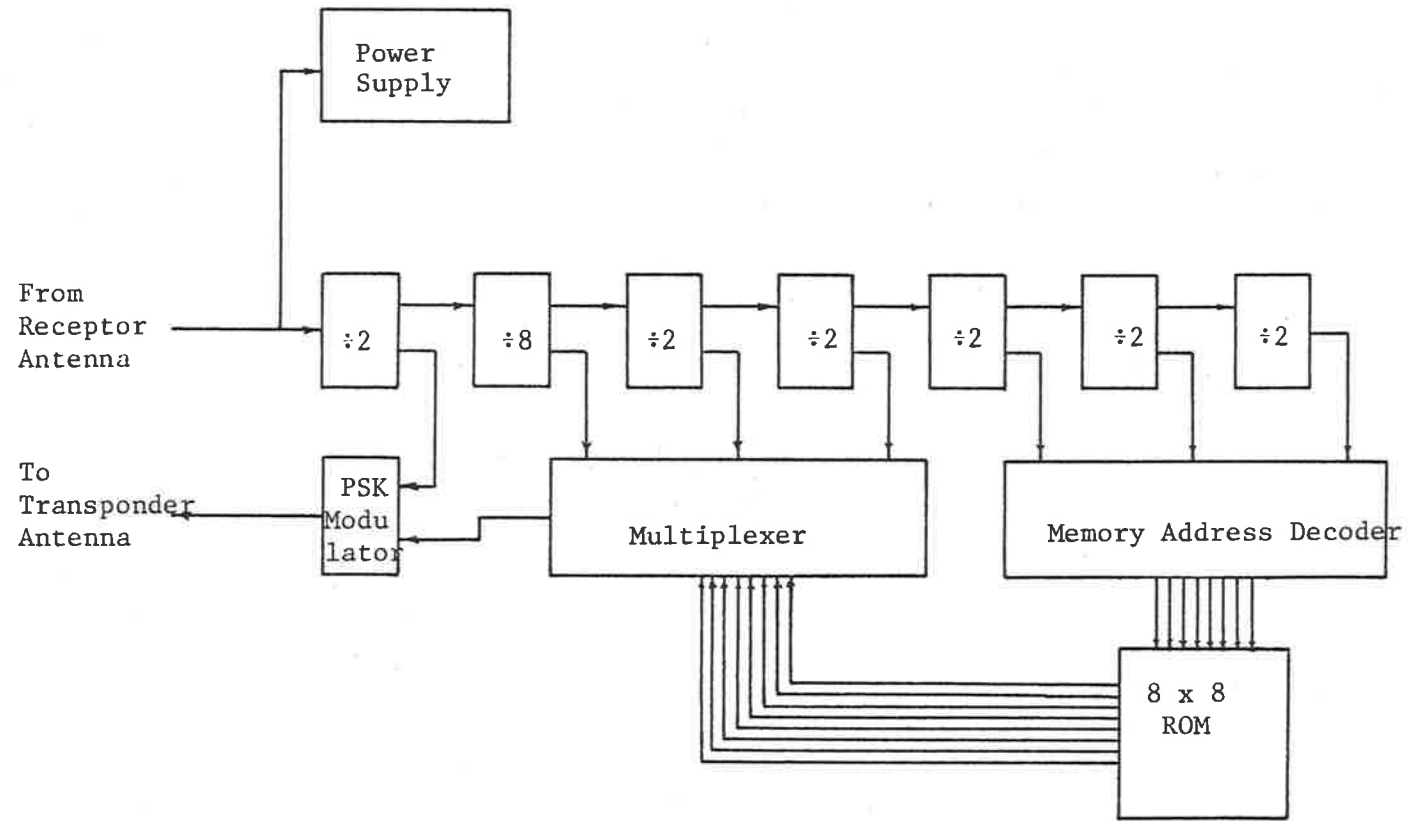


Figure 7.7 Implementation of I²L Two-port Passive Subharmonic Transponder

ponding increase in power dissipation, and by the use of innovative IC design techniques to achieve increased packing densities. This trend is likely to continue in the near future. The benefits to be gained from these advances, from the point of view of passive transponder applications, are reductions in the chip size and power consumption.

The scaling of voltages is not applicable to the I²L technology, since the supply voltage used in I²L circuits is already at the minimum level possible. When the devices are scaled down by a factor s in I²L technology, employing the constant voltage scaling model, the power-delay product scales down by a factor s^2 as compared with a factor s^3 in the MOS technologies (Klaassen, 1980). The metallization paths and the contact windows, in any of the IC technologies, do not scale as advantageously as the actual devices. The resistance and the time constant of the metallization paths increase as the lines are made narrower. The current density and the contact resistance also increase. The scaling limits in the I²L/Bipolar process are reached as the base width reaches the submicron region. This limitation occurs mainly due to punch-through effects experience at very small base widths.

Another limitation on scaling which is likely to be encountered in future VLSI logic circuits is related to the thermal limit which governs the trade-off between the scaling down of the device geometries and the propagation time delay. The scaling limit is reached when the power dissipated per unit area of the chip starts to exceed the practically achievable rate of heat removal per unit area from the chip. The I²L technology has the advantage over the MOS technology that it is possible to dynamically control the operation of the device such that it may be operated in any part of the speed-power curve by changing the supply voltage. The thermal limitation may be controlled, to a certain extent, by means of providing adequate cooling to the circuits. Providing external cooling other than that provided by natural air-flow is generally unacceptable due to the size and cost considerations. Another limitation which occurs when the devices are scaled down without a corresponding scaling in metallization is that the distributed capacitance of the metallization paths becomes a major contributor to signal delays and, in the limiting case, there is no advantage of the scaling to be gained in terms of signal delays.

Apart from the limitation in the devices themselves, there are technological drawbacks in the

processing facilities currently available which must be overcome to continue the progress in VLSI technology. One of the major limitations is in the area of photolithography used for mask making. The problems in the photolithographic process include accuracy over large chip dimensions, pattern alignment on wafers, photo-resist sensitivity, and the etching process. The present ultraviolet techniques of pattern definition and alignment on slice have resolution of the order of 2 microns. It is possible to improve this resolution by about 3dB if pattern generators are used to produce the wafer directly. The resulting elimination of the mask making step has the additional advantage of reduced fault density. The electron beam pattern generation and x-ray lithography techniques which are expected to be used in future have an order of magnitude better resolution than the photolithography techniques currently in use. Other improvements in photolithography are being proposed to extend the life of the photolithography process until the electron beam pattern generation and x-ray lithography become more economical. These include the use of multilayer photo-resist to compensate for wafer surface variations which make fine line lithography otherwise unsuitable, and the use of dry etching methods such as plasma etching, reactive ion etching, and ion milling, instead of the wet etching,

to improve resolution, as well as to reduce undercutting (Deckert and Ross, 1980).

Among other technological limitations, metallization is the most serious one. There exists a need to develop interconnection techniques resulting in very low sheet resistance, low contact resistance, minimum electromigration, and capable of multilevel interconnections. Alternatives to doped aluminium and doped polysilicon, which are the current techniques, such as metal silicides, are being proposed. The metal silicides offer lower sheet resistance than any of the currently available processes (Murarka, 1979).

More systematic CAD/CAM methods are being developed to reduce design times by at least an order of magnitude. Without the use of these advanced techniques, design costs may prove to be a limiting factor in the development of VLSI circuits (Mead and Conway, 1980). Testing complex circuits in a cost effective manner is another limitation faced in the IC industry. The increasing complexity and size of the circuits make the complete functional testing of such devices an almost impossible task. There are various testing schemes currently under investigation by various workers (Frank and Sproull, 1981). Designing fault tolerant systems with limit-

ed self-testing capabilities may overcome the limitations in testing techniques until better approaches to this problem are formulated (Mead, 1982).

The present trends in miniaturization in IC technologies is likely to continue. The impact of these advances, from the point of view of passive transponder applications, is that more efficient transponders can be designed economically. The success of the passive subharmonic transponders as a viable object identification technique lies in the availability of design and manufacturing facilities where the system may be tested during the development stages in an economic way (Clarke and Mudge, 1982).

A major part of the cost in a low frequency object identification system is the cost of the receiver signal processor. The signal processing in the receiver can be implemented very economically by employing the techniques developed for VLSI technology. This leads to major cost reduction in the passive subharmonic transponder system, hence making it a commercially viable object identification technique.

CHAPTER VIII

CONCLUSIONS

The principal objective of the work presented in this thesis has been to undertake a theoretical and experimental study of those aspects of an object identification system employing low-frequency passive subharmonic transponders which are relevant to producing optimum and practically realizable designs.

Attention was first focussed on the two-port Passive Subharmonic Transponder (PST) system. One of the findings of the experimental study was that it is feasible to design control circuits with such low loss that it is possible, in practice, to achieve transfer, in the transponder, of the majority of the energy supplied to the transponder at the interrogation frequency to the reply signal frequency so that the transponder may be regarded as a reasonably efficient converter of energy. A consequence of this fact is that the overall system performance is determined by the loss characteristics of the electromagnetic propagation links.

Another significant result of the work has been the development of the coupling volume approach used in the study of the electromagnetic propagation path. This approach allows the separation of propagation path parameters in a manner suitable for local optimization.

Yet another result arising from the study of the two-port PST systems is that an optimum design of the transponder coil is achieved, for an optimum operating voltage of the transponder circuitry, by the iterative procedure outlined in Chapter 2. In this result, the damping in the transponder coil required to accommodate the information transfer rate is provided solely by the transponder output impedance at the operating voltage.

The thesis has also considered a comparison between a two-port PST and the one-port design, in which a single transponder antenna is used for both reception of interrogation energy at one frequency and transmission of a code modulated reply signal at a subharmonic.

The factors which lead to the greater efficiency of the one-port design, which have been confirmed by experiment, have been identified.

The importance of environmental noise is recognized as being critically important for the operation of a practical passive transponder system. Although measurement and analysis of environmental noise has received considerable attention in the literature, there is little information available on man-made noise in the low-frequency band. The experimental study of the man-made impulsive noise, and the characterization of such noise by means of the statistical parameters presented in Chapter 4, provides a further understanding of the low-frequency impulsive noise.

The results of the noise study are also applicable to the prediction of performance of low-frequency PST systems. For this purpose, computer simulation has been found to be a powerful tool.

Another major finding of the noise study is that the signal-to-noise ratio in a practical passive subharmonic transponder system is expected to be low and, for an optimum operation of such a system, it is necessary to incorporate an error control strategy in the system. This investigation has, therefore, involved a survey of error control strategies applicable to PST systems. An exposition of basic theory, sufficient for its application to the design of a variety of possible error control

schemes, is found in Chapter 5. An optimum practical approach is identified by comparing several encoding strategies. Modifications to the BCH codes to improve the code performance in a practical application are discussed.

A significant part of the thesis has been the prediction, in Chapter 6, of expected error performance of a PST system using computer simulation techniques. These predictions are based on noise measurements discussed earlier, and theoretically derived and experimentally confirmed signal levels.

The final aspect of the work has been to survey present-day technologies suitable for the integration of low-frequency passive transponders.

The NMOS, CMOS and I²L technologies have been considered for this purpose. The advantages and disadvantages associated with each technology have been studied, and a transponder design to prelayout stage, using I²L technology, has been presented.

APPENDIX A

RF INTERFERENCE REGULATIONS

The specification of the tolerable limits to stray radiation is attempted only loosely in Australia. The Federal Department of Communication is responsible for managing and regulating all aspects of radio communications, including frequency allocation and emission control (DOC, 1980). Pending the promulgation of adequate rf interference standards, the U.S. and European emission regulations may be used as a guideline for the design of communication equipment (Head, 1982).

In the U.S.A. the Federal Communication Commission (FCC) regulates rf interference generated by all users of the radio frequency spectrum (FCC, 1959). The FCC has classified equipment into three categories. Transmitters and receivers used for communication purposes only are placed in one class. The focus of interest for the object identification application lies in the other two classes, which are

governed by entirely different rf emission regulations than the ones for communication equipment. One such class of equipment includes the devices that generate rf energy as an intermediate step in performing an end function other than communication, e.g. switching power supplies, and the equipment that uses rf energy directly for an output function other than communication, e.g. medical diathermy, is placed in the other class. The devices in each of these categories are further divided into two classes. Class A includes devices and equipment intended for use in business, commercial or industrial environments, and Class B includes equipment intended for use in residential environments.

In Europe, the rf interference standards are set by bodies such as Verband Deutscher Elektrotechniker (VDE). The VDE interference regulations are based on the recommendations of the International Electrotechnical Commission (IEC). VDE also has two major interference specifications but, unlike the FCC specifications, which are based on the type of the equipment, the VDE specifications are based on the type of interference source. VDE 0875 interference regulation covers equipment not intended to generate rf emission above 10kHz and VDE 0871 covers the equipment that contains sources above 10kHz.

In the PST object identification applications, the transmitter is required to be designed and operated at power levels well within the FCC or VDE regulations. In the low-frequency band, the FCC regulation requires that the permissible electric field, at a standard distance S_{D1} of 305 metres (1000 ft) from the source, should not exceed a value E_{M1} given by the equation below

$$E_{M1} = \frac{2.4}{f} \text{ V/m} \quad (\text{A.1})$$

where f is the frequency of the rf source in Hz

For the proposed object identification system where the transmitter frequency is selected to be 100kHz, the maximum permissible field strength calculated from Equation (A.1) is

$$E_{M1} = 24 \text{ } \mu\text{V/m} \quad (\text{A.2})$$

The VDE 0871 radiated interference limit is defined as a maximum permissible electric field E_{M2} at a standard distance S_{D2} of 100 metres. For the sources in the low-frequency band this is

$$E_{M2} = 50.12 \text{ } \mu\text{V/m} \quad (\text{A.3})$$

A comparison of the FCC specification at 100kHz given by Equation (A.2) can be made with the VDE specification given by Equation (A.3). Using the inverse proportionality of the radiated electric field with distance as shown in Equation (A.5), the VDE specification can be extrapolated to a distance of 305 metres, i.e. the standard distance for the FCC specification. The value of the maximum allowable field at a distance of 305 meters using the VDE specification is E_{M_3}

$$E_{M_3} = 16.44 \mu\text{V/m} \quad (\text{A.4})$$

A comparison of Equations (A.2) and (A.4) reveals that, at a transmitter frequency of 100kHz, the VDE 0871 specification is more stringent than the FCC regulation.

For the calculation of the radiated field, the transmitting antenna is modelled as a small loop of N_1 turns each of area S_1^2 and carrying a current I_1 . The radiated electric field at a distance S_D is then given by

$$E_D^2 = \frac{\eta^2 \pi^2 I_1^2 S_1^4 N_1^2}{\lambda_1^4 S_D^2} \quad (\text{A.5})$$

where

η is the characteristic impedance of
free space

λ_1 is the wavelength at the transmitter
frequency

The relationships between self-inductance and
power have the forms

$$P_1 Q_1 = \omega_1 L_1 I_1^2 \quad (\text{A.6})$$

$$L_1 = \mu_0 N_1^2 F_1 S_1 \quad (\text{A.7})$$

Using Equations (A.6) and (A.7), the radiated
electric field may be expressed as

$$E_D^2 = \frac{\eta^2 \pi^2 P_1 Q_1 S_1^3}{\lambda_1^4 \omega_1 \mu_0 F_1 S_D^2} \quad (\text{A.8})$$

where

P_1 is the power dissipated in the trans-
mitter coil

Q_1 is the quality factor of the trans-
mitter coil

F_1 is the size parameter of the trans-
mitter antenna

The electric field expected from the proposed
identification system using the transmitter antenna
parameters given in Chapter 3 for a transmitter
power of 90 watts at measurement distances of 305
metres and 100 metres are E_{D_1} and E_{D_2} , respectively.

$$E_{D_1} = 23.56 \text{ } \mu\text{V/m} \quad (\text{A.9})$$

$$E_{D_2} = 71.8 \text{ } \mu\text{V/m} \quad (\text{A.10})$$

It therefore appears that, if the transmitter for the proposed system is operated at 90 watts, it satisfies the FCC regulation, but fails to satisfy the VDE requirement. However, it is expected that, in most applications, the object identification system would be operated at much lower power levels.

The expression for radiated electric field given in Equation (A.8), although suitable for assessment of the radiation characteristics of the proposed system, does not lend itself readily to various optimization problems. For the PST applications, it is convenient to compare directly magnetic field energy density $\mu_0 H^2$ at the transponder position with the electric field energy density $\epsilon_0 E_{D_1}^2$ at the standard distance S_{D_1} at which the FCC radiation constraint is applied. From Equation (A.8), and using the expression for dispersal volume

$$V_1 = \frac{P_1 Q_1}{\mu_0 H^2 \omega_T} \quad (\text{A.11})$$

we obtain for the energy density ratio

$$\frac{\mu_0 H^2}{\epsilon_0 E_{D_1}^2} = F_1 \frac{\lambda_1^4 S_{D_1}^2}{V_1 S_1^3} \quad (\text{A.12})$$

Using the expression given below for V_1 , which was developed in Chapter 3

$$V_1 = \frac{L_1 I_1^2}{\mu_0 H^2} = \frac{4}{\pi} F_1 S_1^3 \left[1 + \pi \left(\frac{S_T}{S_1} \right)^2 \right]^3 \quad (\text{A.13})$$

we get for the density ratio

$$\frac{\mu_0 H^2}{\epsilon_0 E_{D_1}^2} = \frac{\pi}{4} \frac{\lambda_1^4 S_{D_1}^2}{(S_1^2 + \pi S_T^2)^3} \quad (\text{A.14})$$

where S_T is the distance between the transmitter and the transponder (sensing distance)

The importance of the equations developed above lies in the fact that they not only provide the justification for the approach used for the analysis in Chapter 3, but they also provide two different points of view on the question of how to select an optimum size of transmitter antenna.

In the first place, if we have no regard for the radiated field constraint and wish to maximise the dispersal volume V_1 with respect to the transmitter antenna characteristic length S_1 for a given

value of sensing distance S_T , we find, from Equation (A.13), that the optimum value is

$$S_{1opt} = S_T \sqrt{\pi} \quad (\text{A.15})$$

For a circular transmitter antenna, it may be shown that the optimum diameter is given by

$$D_{1opt} = 2S_T \quad (\text{A.16})$$

The optimum value of the transmitter characteristic length derived above makes optimum use of a given amount of transmitter power.

If, on the other hand, we wish to maximize the energy density ratio given in Equation (A.14), subject to the condition $E_{D_1} = E_{M_1}$, but without regard for the transmitter power level required, we find that the optimum occurs as $S_1 \rightarrow 0$. Such a vanishingly small transmitter antenna cannot be used in practice, because the power required to provide interrogation energy at the sensing position then tends to infinity.

REFERENCES

- 1944 RICE, S. O., "Mathematical Analysis of Random Noise",
Bell Systems Tech. Journal, Vol. 23, pp. 282-
332.
- 1947 TERMAN, F. E., "Electronic and Radio Engineering",
McGraw-Hill.
- 1949 GOLAY, M. J. E., "Notes on Digital Coding", Proc. of
I.R.E., Vol. 37, p. 657.
- 1950 HAMMING, R. W., "Error Detecting and Correcting
Codes", Bell Systems Tech. Journal, Vol. 29,
pp. 147-160.
- 1952 GILBERT, E. N., "A Comparison of Signalling Alphabets",
Bell Systems Tech. Journal, Vol. 31, pp. 504-
522.
- WILKINS, O. L., "Recognition System", U. S. Patent
No. 2602160.
- 1954 ELIAS, P., "Error Free Coding", I.R.E. Trans. on
Information Theory, Vol. IT-4, pp. 29-37.

- REED, I. S., "A Class of Multiple-Error-Correcting Codes and Decoding Schemes", I.R.E. Trans. on Information Theory, Vol. IT-4, pp. 38-49.
- 1956 COSTAS, J. P., "Synchronous Communications", Proc. of I.R.E., Vol. 44, pp. 1713-1718.
- SLEPIAN, D., "A Class of Binary Signalling Alphabets", Bell Systems Tech. Journal, Vol. 35, pp. 203-234.
- 1957 MAXWELL, E. L. & WATT, A. D., "Measured Statistical Characteristics of VLF Atmospheric Noise", Proc. I.R.E., Vol. 45, pp. 55-62.
- MAXWELL, E. L. & WATT, A. D., "Characteristics of Atmospheric Noise from 1 to 100 KC", Proc. of I.R.E., Vol. 45, pp. 787-794.
- STORER, J. E., "Passive Network Synthesis", McGraw-Hill.
- 1958 GREEN, J. H. & SAN-SOUCIE, R. L., "An Error-Correcting Encoder and Decoder of High Efficiency", Proc. of I.R.E., Vol. 46, pp. 1741-1744.
- SACKS, G. E., "Multiple Error Correction by Means of Parity Checks", I.R.E. Trans. on Information Theory, Vol. IT-4, pp. 145-147.
- 1959 FCC (Federal Communication Commission), "Rules and Regulations: Industrial, Scientific and Medical Services", Vol. 2, Government Press, Washington, D.C.

HOCQUENGHEM, A., "Codes Correcteurs d'Erreurs",
Chiffres 2, pp. 147-156.

1960 BARTEE, T. C., "Digital Computer Fundamentals",
McGraw-Hill, New York.

BOSE, R. C. & RAY-CHAUDHURI, D. K., "On a Class of
Error Correcting Binary Group Codes", Informa-
tion & Control, Vol. 3, pp. 68-79.

BOSE, R. C., & RAY-CHAUDHURI, D. K., "Further
Results on Error Correcting Binary Group Codes",
Information & Control, Vol. 3, pp. 279-290.

GILBERT, E. N., "Synchronization of Binary Messages",
I.R.E. Trans. on Information Theory, Vol. IT-6,
No. 4, pp. 470-477.

GILBERT, E. N., "Capacity of a Burst-Noise Channel",
B.S.T.J., Vol. 39.

MEGGITT, J. E., "Error Correcting Codes for Correct-
ing Bursts of Errors", IBM Journal of Research
and Dev., Vol. 4, pp. 329-334.

MELAS, C. M., "A Cyclic Code for Double Error
Correction", IBM Journal of Research and Dev.,
Vol. 4, pp. 364-366.

PETERSON, W. W., "Encoding and Error-Correction
Procedures for the Bose-Chaudhuri Codes", I.R.E.
Trans. on Information Theory, Vol. IT-6,
pp. 459-470.

- 1961 MEGGITT, J. E., "Error Correcting Codes and their Implementation for Data Transmission Systems", I.R.E. Trans. on Information Theory, pp. 234-244.
- PETERSON, W. W. & BROWN, D. T., "Cyclic Codes for Error Detection", Proc. of I.R.E., Vol. 49, No. 1, pp. 228-235.
- PETERSON, W. W., "Error Correcting Codes", John Wiley, New York.
- PLONSEY, R. & COLIN, R. E., "Principles and Applications of Electromagnetic Fields", McGraw-Hill.
- 1962 BARTEE, T. C. & SCHNEIDER, D. I., "An Electronic Decoder for Bose-Chaudhuri-Hocquenghem Error-Correcting Codes", I.R.E. Trans. on Information Theory, Vol. IT-8, pp. 17-24.
- KREYSZIG, E., "Advanced Engineering Mathematics", John Wiley, New York.
- PERRY, K. E. & WOZENCRAFT, J. M., "SECO: a Self-Regulating Error Correcting Coder-Decoder", I.R.E. Trans. on Information Theory, Vol. IT-8, pp. 128-135.
- STIFFLER, J. J., "Synchronization Methods for Block Codes", I.R.E. Trans. on Information Theory, Vol. IT-8, pp. 25-34.
- TILSTON, W. V., "Simultaneous Transmission and Reception with a Common Antenna", I.R.E. Trans. on Vehicular Communication, pp. 56-64.

- 1963 MYERS, H. A., "Industrial Equipment Spectrum Signature", I.E.E.E. Trans. on Radio Freq. Interference, Vol. RFI-5, pp. 30-42.
- 1964 CHIEN, R. T., "Cyclic Decoding Procedures for Bose-Chaudhuri Hocquenghem Codes", I.E.E.E. Trans. on Information Theory, Vol. IT-10, pp. 357-363.
- 1965 BENNETT, W. R. & DAVEY, J. R., "Data Transmission", McGraw-Hill.
- BERLEKAMP, E. R., "On Decoding Binary Bose-Chaudhuri-Hocquenghem Codes", I.E.E.E. Trans. on Information Theory, Vol. IT-11, pp. 577-580.
- PAPOULIS, A., "Probability, Random Variables and Stochastic Processes", McGraw-Hill, New York.
- RAMO, S., WHINNERY, J. R. & VAN DUZER, T., "Fields and Waves in Communication Electronics", John Wiley.
- 1966 ALLAN, D. W., "Statistics of Atomic Frequency Standards", Proc. of I.E.E.E., Vol. 54, pp. 221-230.
- BUEHLER, W. E. & LUNDEN, C. D., "Signature of Man-Made High-Frequency Radio Noise", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-8, pp. 143-152.
- VAN DER WAERDEN, B. L., "Algebra", Zurich.
- VITERBI, A. J., "Principles of Coherent Communication", McGraw-Hill, New York.

- 1967 BOSE, R. C. & CALDWELL, J. G., "Synchronizable Error-Correcting Codes", Information & Control, Vol. 10, pp. 616-630.
- MAXWELL, E. L., "Atmospheric Noise from 20 Hz to 30 kHz", Radio Science, Vol. 2, pp. 637-644.
- SKOMAL, E. N., "Comparative Radio Noise Levels of Transmission Lines, Automotive Traffic, and RF Stabilized Arc Welders", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-9, No. 2, pp. 73-77.
- VINDING, J. P., "Interrogator-Responder Identification System", U.S. Patent No. 3299424.
- ZVEREV, A. I., "Handbook of Filter Synthesis", John Wiley.
- 1968 AUSTIN, A. V., "Standards of Measurements", Scientific American, Vol. 218, No. 6, p. 50.
- BUEHLER, W. E. et al, "VHF City Noise", I.E.E.E. Electromagnetic Compatibility Symposium Rec., pp. 113-118.
- GALLAGER, R. G., "Information Theory and Reliable Communication", John Wiley.
- RALSTON, E. L., "15 Ways to Control Moving Materials", Control Engineering.
- 1969 CHEN, C. L., PETERSON, W. W. & WELDON, E. J., "Some Results on Quasicyclic Codes", Information & Control, Vol. 15, pp. 407-423.

- CHEN, C.L. & LIN, S., "Further Results on Polynomial Codes", Information & Control, Vol. 15, pp. 38-60.
- FONTAINE, A. B. & PETERSON, W. W., "Group Code Equivalence and Optimum Codes", I.R.E. Trans. on Information Theory, Vol. IT-5, pp. 60-70.
- OH, L. L., "Measured and Calculated Spectral Amplitude Distribution of Lightning Spherics", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-11, No. 4, pp. 125-130.
- SKOMAL, E. N., "Distribution and Frequency Dependence of Incidental Man-Made HF/VHF Noise in Metropolitan Areas", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-11, No. 2, pp. 66-75.
- SKOMAL, E. N., "Analysis of Airborne VHF/UHF Incidental Noise Over Metropolitan Areas", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-11, No. 2, pp. 76-83.
- TAYLOR, W. L., BURDICK, H. M. & EICHACKER, L. W., "Noise Reduction in Wide-Band Atmospheric Receiving System", I.E.E.E. International Conf. on Communication, Conf. Rec., Session 5, pp. 31-34.
- 1970 ANDERSON, R. L., "Electromagnetic Loop Vehicle Detectors", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-19, pp. 23-30.

- BARKER, J. L., "Radar, Acoustic, and Magnetic Vehicle Detectors", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-19, pp. 30-43.
- HURD, W. J. & ANDERSON, T. O., "Digital Transition Tracking Symbol Synchronizer for Low SNR Coded Systems", I.E.E.E. Trans. on Communication Technology, Vol. COM-18, No. 2, pp. 141-146.
- MATHESON, R. J., "Instrumentation Problems Encountered Making Man-Made Electromagnetic Noise Measurements for Predicting Communication System Performance", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-12, pp. 151-158.
- MILLS, A. H., "Instrumentation System for Measurement of Noise Characteristics", I.E.E.E. Regional Electromagnetic Compatibility Symposium Rec., San Antonio, Session II-B, p. 5.
- MILLS, M. K., "Future Vehicle Detection Concepts", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-19, pp. 43-49.
- RYLEY, J. E., "Improvements in or Relating to Vehicle Identification Systems", British Patent No. 1187130.

SKOMAL, E. N., "Conversion of Area Distributed Incidental Radio Noise Envelope Distribution Function by Radio Propagation Process", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-12, No. 3, pp. 83-88.

ZAMILES, C. J. & HURLBUT, K. H., "Measurements of Interference Levels in the UHF Band from Aircraft Altitudes", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-12, No. 3, pp. 88-96.

1971 MARHA, K., "Microwave Radiation Safety Standards in Eastern Europe", I.E.E.E. Trans. on Microwave Theory and Techniques, Vol. MTT-19, No. 2, pp. 165-168

STIFFLER, J. J., "Theory of Synchronous Communication", Prentice-Hall, Englewood Cliffs, New Jersey.

1972 ARTOM, A., "Choice of Prefix in Self-Synchronizing Codes", I.E.E.E. Trans. on Communications, pp. 253,254.

BERGER, H. H. & WIEDMANN, S. K., "Merged-Transistor Logic (MTL) - a Low-Cost Bipolar Logic Concept", I.E.E.E. Journal of Solid State Circuits, Vol. SC-7, No. 5, pp. 340-346.

GUPTA, S. N., "Time Characteristics of Atmospheric Radio Noise", Proc. of I.E.E.E. Conf. on Electromagnetic Compatibility.

HART, K. & SLOB, A., "Integrated Injection Logic: a New Approach to LSI", I.E.E.E. Journal of Solid State Circuits, Vol. SC-7, No. 5, pp. 346-351.

HOENEISEN, B. & MEAD, C. A., "Fundamental Limits in Micro-electronics - II. Bipolar Technology", Solid-State Electronics, Vol. 15, pp. 891-897.

LINDSEY, W. C., "Synchronization Systems in Communication and Control", Prentice-Hall, Englewood Cliffs, New Jersey.

PENNEY, W. M. & LAU, L., eds., "MOS Integrated Circuits", Van Nostrand Reinhold Co.

PETERSON, W. W. & WELDON, E. J., "Error Correcting Codes", MIT Press.

REMIZOV, L. T., "Spectrum of the Fluctuation Component of Atmospheric Radio Interference in the VLF Range", Radio Eng. Electron Physics, Vol. 17, No. 2, pp. 224-226.

UNISEARCH LTD., "Passive Label for Use in Electronic Surveillance Systems", British Patent No. 1298381.

1973 CLORFEINE, A. S., "Driving under the Influence of Electronics", I.E.E.E. Spectrum, Vol. 10, pp. 32-37.

FREEDMAN, N., "RAYTAG, an Electronic Remote Data Readout System", Carnahan Conference on Electronic Crime Countermeasures Proc., pp. 104-107.

- GUPTA, S. N., "Distribution of Peaks in Atmospheric Radio Noise", I.E.E.E. Trans. on Electro-magnetic Compatibility, Vol. EMC-15, No. 3, pp. 100-103.
- KLENSCH, R. J., ROSEN, J. & STARAS, H., "A Microwave Automatic Vehicle Identification System", RCA Review, Vol. 34, pp. 566-579.
- MIDDLETON, D., "Man-Made Noise in Urban Environments and Transportation Systems: Models and Measurements", I.E.E.E. Trans. on Communication, Vol. COM-21, pp. 1232-1241.
- RITER, S., JONES, W. B. & DOZIER, H., "Speeding the Deployment of Emergency Vehicles", I.E.E.E. Spectrum, Vol. 10, pp. 56-62.
- RONEN, R. S., "Low Frequency 1/f noise in MOSFETs", RCA Review, Vol. 34, pp. 280-307.
- SHEFER, J. & KLENSCH, R. J. "Harmonic Radar Helps Autos Avoid Collisions", I.E.E.E. Spectrum, Vol. 10, pp. 38-45.
- SKOMAL, E. N., "Analysis of Metropolitan Incidental Radio Noise Data", I.E.E.E. Trans. on Electro-magnetic Compatibility, Vol. EMC-15, No. 2, pp. 45-57.
- 1974 AUGENBLICK, H. A. & ENGLE, W. J., "Recognition System", U.S. Patent No. 3798642.

- BURGETT, R. R. et al, "Relationship between Spark Plugs and Engine Radiated Electromagnetic Interference", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-16, No. 3, pp. 160-172.
- CONSTANT, J. N., "Microwave Automatic Vehicle Identification (MAVI) System", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-23, No. 2, pp. 44-54.
- GINSBERG, L. H., "Extremely Low Frequency (ELF) Propagation Measurements along a 4900 Km Path", I.E.E.E. Trans. on Communication, Vol. COM-22, pp. 452-457.
- GINSBERG, L. H., "Extremely Low Frequency (ELF) Atmospheric Noise Level Statistics for Project Sanguine", I.E.E.E. Trans. on Communication, Vol. COM-22, pp. 555-561.
- HSU, H. P. et al, "Measured Amplitude Distribution of Automotive Ignition Noise", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-16, No. 2, pp. 57-63.
- PAGE, C. H., "Definition of Electromagnetic Field Quantities", American Journal of Physics, 42, p. 490.
- PAGE, C. H. & VIGOUREUX, P., eds., "The International System of Units (SI)", NBS Special Publication 330.

- SCHULZ, R. B. & SOUTHWICK, R. A., "APD Measurements of V-8 Ignition Emanations", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-16, No. 2, pp. 63-70.
- SHEPHERD, R. A., "Measurement of Amplitude Probability Distributions and Power of Automobile Ignition Noise at HF", I.E.E.E. Trans. on Vehicular Technology, VT-23, pp. 72-88.
- SHINDE, M. P. & GUPTA S. N., "A Model of HF Impulsive Atmospheric Noise", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-16, No. 2, pp. 71-75.
- STERZER, F., "An Electronic License Plate for Motor Vehicles", RCA Review, Vol. 35, pp. 167-175.
- THOMPSON, R. & CLOUTING, D. R., "Digital Angle Modulation for Data Transmission", Systems Technology, No. 19, pp. 14-18.
- WARNER, R. M., "Applying a Composite Model of the IC Yield Problem", I.E.E.E. Journal of Solid-State Circuits, Vol. SC-9, pp. 86-95.
- 1975 DAVIES, D. E. N. et al, "Passive Coded Transponder Using an Acoustic-Surface-Wave Delay Line", Electronics Letters, Vol. 11, No. 8, pp. 163, 164.
- DOW, I. M., "Improvements in or Relating to the Identification of Vehicles", British Patent No. 84867.

- KANDA, M., "Time and Amplitude Statistics for Electromagnetic Noise in Mines", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-17, No. 3, pp. 122-129
- KLAASSEN, F. M., "Device Physics of Integrated Injection Logic", I.E.E.E. Trans. on Electron. Devices, pp. 145-152.
- KOELLE, A. R., DEPP, S. W. & FREYMAN, R. W., "Short Range Radio Telemetry for Electronic Identification Using Modulated RF Backscatter", Proc. of I.E.E.E., Vol. 63, pp. 1260,1261.
- LEAVER, R. F., "Track-to-Train Communication", Systems Technology, No. 20, pp. 30-32.
- McEWEN, C. D., "Identity Transponder System Using C.W. Interrogation", Electronics Letters, Vol. 11, Nos. 25/26, pp. 642,643.
- NAGY, L. L. & LYON, J. A. M., "An Ultrashort Pulse Radar Sensor for Vehicular Precollision Obstacle Detection", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-24, No. 4, pp. 41-45.
- ORANC, H. S., "Ignition Noise Measurements in the VHF/UHF Bands", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-17, No. 2, pp. 54-64.
- REDFERN, T. P., "54c/74c Family Characteristics", National CMOS Data Book, pp. 220-225.

THOMPSON, R. & CLOUTING, D. R., "Digital Angle Modulation for Data Transmission, Part 2: Performance Characteristics", Systems Technology, No. 20, pp. 33-37.

- 1976 ALLARD, R. et al, "An Automatic Headway Control System Based on the Use of a Microwave Telemetry Link", I.E.E. Conference Publication No. 41, Automobile Electronics, pp. 120-123.
- BOLTON, E. C., "Man-Made Noise Study at 76 and 200 kHz", I.E.E.E. Trans. on Electromagnetic Compatibility, Vol. EMC-18, No. 3, pp. 93-96.
- DAVIES, D. E. N., MAKRIDIS, H. & McEWEN, C.D., "Harmonic and Two-Frequency Radars for Vehicle Headway Applications", I.E.E. Conference Publication No. 41, Automobile Electronics.
- DENNARD, R. H. et al, "Design of Ion Implanted MOSFETs with Very Small Physical Dimensions", I.E.E.E. Journal of Solid-State Circuits, Vol. SC-9, pp. 256-268.
- DESLANDES, P. A. & LAST, J. D., "A System for Automatic Vehicle Location", I.E.E. Conference Publication No. 41, Automobile Electronics, pp. 112-115.
- DOMANN, H., "AFF - an Automatic Guidance System for Road Vehicles", I.E.E. Conference Publication No. 41, Automobile Electronics, pp. 140-143.

- HAHLGAUSS, G. & HAHN, L., "Headway Radar Using Pulse Techniques", I.E.E. Conference Publication No. 41, Automobile Electronics, pp. 132-135.
- HULTON, T. J. & KRAMER, J. W., "Transponder for an Automatic Vehicle Identification System", U.S. Patent No. 3964024.
- MAXAM, G. L. et al, "Radiated Ignition Noise Due to the Individual Cylinders of an Automobile Engine", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-25, No. 2, pp. 33-38.
- SHEPHERD, R.A. et al, "New Techniques for Suppression of Automobile Ignition Noise", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-25, No. 1, pp. 2-12.
- 1977 BIRKHOFF, G. & MacLANE, S., "A Survey of Modern Algebra", MacMillan, New York.
- COLE, P. H. & ROY, A. K., "Analysis and Optimization of the Performance of a Passive Subharmonic Transponder Labelling and Interrogation System". Research Report No. 4/77, University of Adelaide.
- HENNING, F., HINGARH, H. K., O'BRIEN, D. & VERHOFSTADT, P. W. J., "Isoplanar Integrated Injection Logic: a High Performance Bipolar Technology", I.E.E.E. Journal of Solid-State Circuits, Vol. SC-12, No. 2, pp. 101-109.

- HERMANN, J. M., EVANS, S. A. & SLOAN, B. J., "Second Generation I²L/MTL: a 20 ns Process/Structure", I.E.E.E. Journal of Solid State Circuits, Vol. SC-12, No. 2, pp. 93-101.
- LAWRENCE, W. T., "A Magnetic Signpost AVM System with Limited Dead Reckoning", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-26, No. 1, pp. 23-29.
- OLSSON, K. O., "Transponder System", U.S. Patent No. 4019181.
- STREET, A. P. & WALLIS, W. D., "Combinatorial Theory: an Introduction", CBRC, Winnipeg.
- 1978 BRENIG, T., "Data Transmission for Mobile Radio", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-27, No. 3, pp. 77-85.
- EDEN, R. C., WELCH, B. M. & ZUCCA, R., "Planar GaAs IC Technology: Applications for Digital LSI", I.E.E.E. Journal of Solid-State Circuits, Vol. SC-13, No. 4, pp. 419-426.
- FIELD, E. C. & LEWINSTEIN, M., "Amplitude Probability Distribution Model for VLF/ELF Atmospheric Noise", I.E.E.E. Trans. on Communications, Vol. COM-26, No. 1, pp. 83-87.
- FRENCH, R. C., "Error Rate Predictions and Measurements in the Mobile Radio Data Channel", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-27, No. 3, pp. 110-116.

- GRAY, P. R. & HODGES, D. A., "All MOS Analog-Digital Conversion Techniques", I.E.E.E. Trans. on Circuits and Systems, Vol. CAS-25, pp. 482-489.
- GUIBAS, L. J. & ODLYZKO, A. M., "Maximal Prefix-Synchronized Codes", SIAM Journal of Appl. Math., Vol. 35, No. 2, pp. 401-418.
- HODGES, D. A., GRAY, P. R. & BORDERSEN, R. W., "Potential of MOS Technologies for Analog Integrated Circuits", I.E.E.E. Journal of Solid-State Circuits, Vol. SC-13, No. 3, pp. 285-294.
- JOHNSON, W. J., "Technique for Correlating Radiated Noise with Individual Spark Events in an Automotive Ignition System", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-27, No. 3, pp. 138-141.
- PORTIS, A. M., "Electromagnetic Fields: Sources and Media", John Wiley.
- WEBER, W. J. "Differential Encoding for Multiple Amplitude and Phase Keying Systems", I.E.E.E. Trans. on Communications, Vol. COM-26, No. 3, pp. 385-391.
- WILLIARD, M. W., "Introduction to Redundancy Coding", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-27, No. 3, pp. 86-98.

YUE, O. C. et al, "Series Approximations for the Amplitude Distribution and Density of Shot Processes, I.E.E.E. Trans. on Communication, Vol. COM-26, No. 1, pp. 45-54.

1979 BERGMANN, G., "A One-Chip I²L Controller for Appliances", I.E.E.E. Journal of Solid-State Circuits, Vol. SC-14, No. 3, pp. 569-572.

BORDERSEN, R. W., GRAY, P. R. & HODGES, D. A., "MOS Switched-Capacitor Filters", Proc. of I.E.E.E., Vol. 67, No. 1, pp. 61-75.

COLE, P. H., ESHRAGHIAN, K. & ROY, A. K., "Measurement and Characteristics of Environmental Noise Affecting Low-Frequency Transponder Operations", I.R.E.E. International Conv. Proc., pp. 457-459.

COLE, P. H., ESHRAGHIAN, K. & ROY, A. K., "Theory and Operation of Passive Subharmonic Transponder", I.R.E.E. International Conv. Proc., pp. 51-54.

GARDNER, F. M., "Phaselock Techniques", John Wiley, New York.

MURARKA, S. P., "Refractory Silicides for Low Resistivity Gates and Interconnections", I.E.D.M. Tech. Digest, Washington, DC, pp. 454-457.

- OETTING, J. D., "A Comparison of Modulation Techniques for Digital Radio", I.E.E.E. Trans. on Communication, Vol. COM-27, No. 12, pp. 1752-1762.
- OZAWA, O. et al, "A High Speed I²L Compatible with High Voltage Analog Devices", I.E.D.M. Tech. Digest, Washington, D.C., pp. 188-191.
- PASUPATHY, S., "Minimum Shift Keying: a Spectrally Efficient Modulation", I.E.E.E. Communications Magazine.
- TZANATEAS, G., SALAMA, C. A. T. & TSIVIDIS, Y. P., "A CMOS Band Gap Voltage Reference", I.E.E.E. Journal of Solid-State Circuits, Vol. SC-14, pp. 655-657.
- 1980 CAHN, C. R. & LEIMER, D. K., "Digital Phase Sampling for Microcomputer Implementation of Carrier Acquisition and Coherent Tracking", I.E.E.E. Trans. on Communication, Vol. COM-28, No. 8, pp. 1190-1196.
- COLE, P. H., ESHRAGHIAN, K. & ROY, A. K., "Object Identification System", European Patent No. 79301035.6.
- DECKERT, C.A. & ROSS, D. L., "Microlithography - Key to Solid State Fabrication", Journal of Electrochem. Society, Vol. 127, pp. 450-455.
- DOC (Department of Communications), "Australian Table of Frequency Allocations 9kHz-400GHz".

- ENGELMANN, G., "SICARID: Identification System for Process Automation", Siemens Publications, No. F573/113.101.
- ESHRAGHIAN, K., "Electromagnetic Traffic Sensing and Surveillance", Ph.D.Dissertation, Dept. of Elec. Eng., University of Adelaide.
- FOOTE, R. S., "Automatic Vehicle Identification Tests and Applications in the Late 1970's", I.E.E.E. Trans. on Vehicular Technology, Vol. VT-29, No. 2., pp. 226-229.
- FRANKS, L. E., "Carrier and Bit Synchronization in Data Communication - a Tutorial Review", I.E.E.E. Trans. on Communication, Vol. COM-28, No. 8, pp. 1107-1120.
- HOSTICKA, B. J., "Dynamic CMOS Amplifiers", I.E.E.E. Journal of Solid-State Circuits, Vol. SC-15, pp. 887-893.
- JACKSON, C. L., "The Allocation of the Radio Spectrum", Scientific American, Vol. 242, No. 2, pp. 30-35.
- MEAD, C. A. & CONWAY, L. A., "Introduction to VLSI Systems", Addison-Wesley Publishing Co.
- OHZONE, T. et al, "Silicon Gate n-well CMOS Process by Full Ion Implantation Technology", I.E.E.E. Trans. on Electron. Devices, Vol. ED-27, pp. 1789-1795.
- PARRILLO, L. C. et al, "Twin Tub CMOS - a Technology for VLSI Circuits", I.E.D.M. Tech. Digest, pp. 752-755.

PAYNE, R. S., GRANT, W. N. & BERTRAM, W. J.,

"Elimination of Latch-up in Bulk CMOS",

I.E.D.M. Tech. Digest, pp. 248-251.

SCHOLTZ, R. A., "Frame Synchronization Techniques",

I.E.E.E. Trans. on Communication, Vol. COM-28,

No. 8, pp. 1204-1212.

TAKASAKI, Y., "Optimizing Pulse Shaping for Baseband

Digital Transmission with Self-Bit Synchroniza-

tion", I.E.E.E. Trans. on Communication,

Vol. COM-28, No. 8, pp. 1164-1171.

1981 ESHRAGHIAN, K. & COLE, P. H., "Modulation Principles

in One-Port Passive Transponders", I.R.E.E.

International Conv. Proc., pp. 375-377.

FRANK, E. H. & SPROULL, R. F., "Testing and Debug-

ging Custom Integrated Circuits", ACM Journal

of Computing Surveys, Vol. 13, No. 4, pp. 425-

451.

ROSENBAUM, S. D., "Special IC's in Digital Switching

- an Overview", I.E.E.E. Journal of Solid-State

Circuits, Vol. SC-16, No. 4, pp. 247-252.

ROY, A. K. & COLE, P. H., "Error Control in a

Vehicle Identification System", I.R.E.E.

International Conv. Proc., pp. 381-383.

ROY, A. K., "User Manual - RADDAT", Internal Report,

University of Adelaide.

SALAMA, C. A. T., "VLSI Technology for Telecommuni-

cation IC's", I.E.E.E. Journal of Solid-State

Circuits, Vol. SC-16, No. 4, pp. 253-260.

SLINGER, E., "Design Guide to Data Acquisition",
Digital Design, pp. 72-75.

1982 CLARKE, R. J. & MUDGE, J. C., "Organization of the
Australian Multi-Project Chip", Micro-
electronics 82, Adelaide, Conv. Digest.

ESHRAGHIAN, K., COLE, P. H. & ROY, A. K., "Electro-
magnetic Coupling in Subharmonic Transponders",
Journal of Electrical and Electronics Engineer-
ing, Australia, Vol. 2, No. 1, pp. 28-35.

HASKARD, M. et al, "Radio Noise Measurements in an
Urban Environment", Journal of Electrical and
Electronics Engineering, Australia, Vol. 2,
No. 2, pp. 94-102.

HEAD, M., "U.S., European RF-Emission Regulations",
Australian Electronics Engineering, pp. 56-62.

MEAD, C. A., "VLSI - the Next Revolution", Micro-
electronics 82, Adelaide, Conv. Digest.