THE DEVELOPMENT OF A RULE BASED EXPERT SYSTEM TO AUTOMATE THE DIGITAL ANALYSIS OF CONDITION MONITORING PARAMETERS CAPTURED ON ROLLING ELEMENT BEARINGS SUBJECTED TO SIMULATED FAILURE

BRAD LEGGAT
DEPT OF MECHANICAL ENGINEERING
UNIVERSITY OF CAPE TOWN (RSA)

SUBMITTED IN FULFILMENT OF THE DEGREE OF
MASTER OF SCIENCE IN ENGINEERING

SUPERVISORS :
R B NOYES (MECH. ENG)
A W D JONGENS (ELEC. ENG)

SYNOPSIS

THE DEVELOPMENT OF A RULE BASED EXPERT SYSTEM TO AUTOMATE THE DIGITAL ANALYSIS OF CONDITION MONITORING PARAMETERS CAPTURED ON ROLLING ELEMENT BEARINGS SUBJECTED TO SIMULATED FAILURE

Author : Brad Leggat

DESCRIPTIVE ABSTRACT

This paper deals with the use of rule based expert systems to automate the detection and diagnosis of rolling element bearing fatigue failure under variable operating conditions. Statistical parameters are used to analyse vibration patterns and measured and derived operating conditions are used to assist the diagnosis. A simple prognosis is provided by using vibration patterns to determine the percentage of bearing rated life used.

1. INTRODUCTION

This synopsis provides a brief summary of the development of a rule based expert system to diagnose bearing failure. Firstly it covers the proposal of a generic, expert system based industrial condition monitoring system. It then discusses in more detail the development of a specific aspect of the system, viz. the analysis of rolling element bearing condition.

The bearing test rig and data capture system are described, followed by primary research to define the bearing analysis solution space. This includes the use of vibration parameters, measured and derived operating conditions and the bearing running condition. It then explains the development of rulebases for the three analysis tasks of detection, diagnosis and prognosis. Included is a discussion on techniques used to normalise and adjust the vibration parameters to allow analysis under any operating conditions.

Finally the synopsis is concluded with a discussion on the performance of the system and contributions made to the developing field of condition monitoring using expert systems.

2. AN INDUSTRIAL CONDITION MONITORING SYSTEM

The first stage of the project involved development of the framework for a portable, computerised, condition monitoring system. This would be designed around typical industrial machinery and perform diagnosis using expert systems.

The importance of defining this framework was that the bearing expert system would be developed as one of the components. In addition recommendations for further research were given in the light of such a system. An overview of the system is illustrated in figure 1.

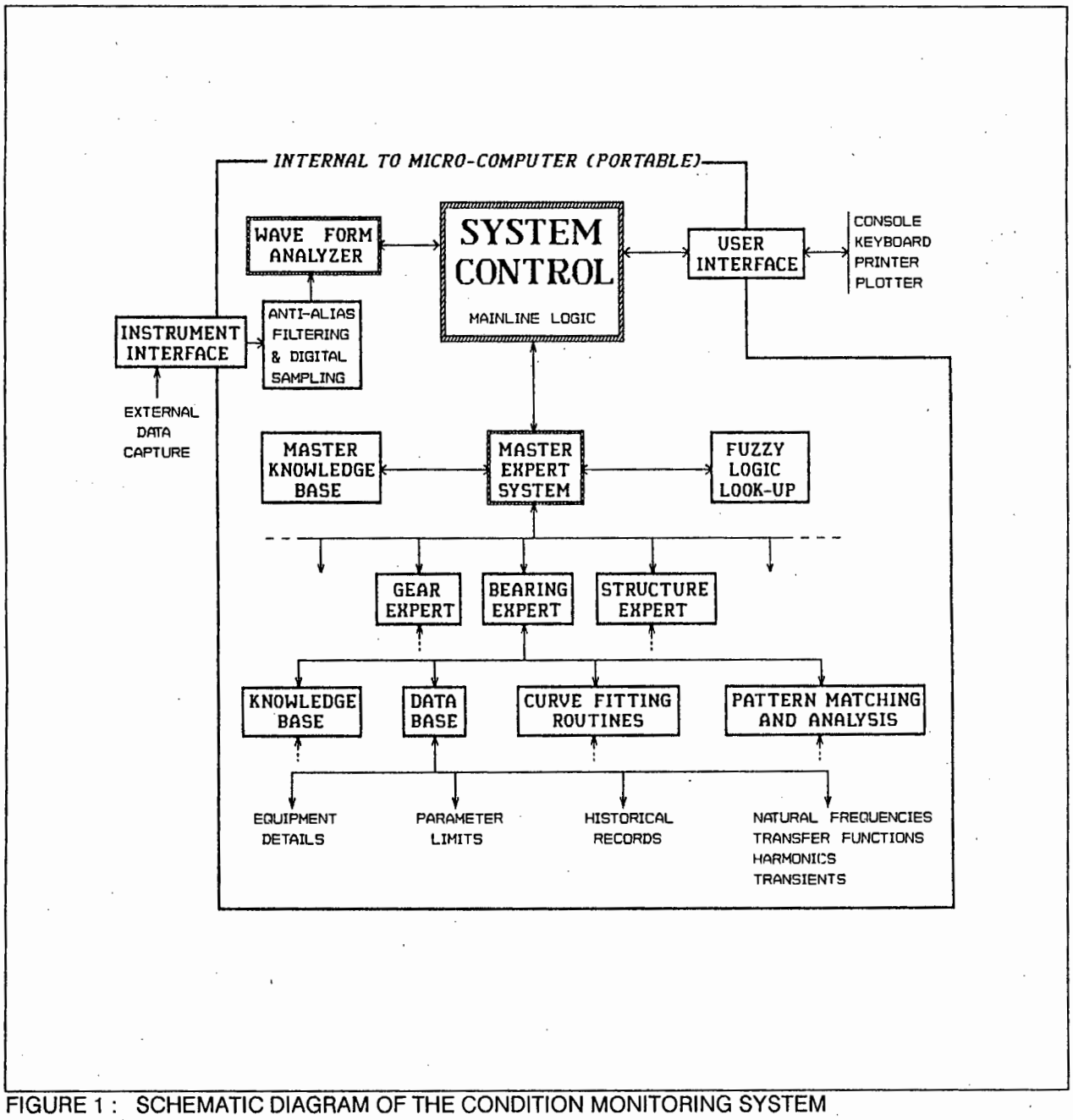

FIGURE 1 : SCHEMATIC DIAGRAM OF THE CONDITION MONITORING SYSTEM

3. THE TEST RIG AND DATA CAPTURE SYSTEM

The test rig was designed for the systematic destruction of bearings under simulated industrial conditions [figure 2]. The two outer bearings provided support for the central bearing which could be hydraulically loaded to 1.6 times its rated load. Power was provided by a variable speed DC drive with a maximum speed of 2000 rpm.

Attatched to the rig was a computer based data capture and processing system. This system measured the operating conditions and vibrations directly from the bearing using various transducers and sensors. Data collection was performed by digital sampling up to a maximum frequency of 27.6 kHz.

The measured operating conditions included the LOAD, SPEED and TEMPERATURE of the bearing, while the vibrations were in the form of TIME vs ACCELERATION data. Various parameters were then extracted from the raw digital data in preperation for analysis and interpretation. This processing included averaging and smoothing techniques.

The system was flexible and enabled easy modification for the various experiments performed. Data could either be interpreted directly after capture and processing or stored for later analysis. Typically the whole process of data capture, processing and storage could be performed in under two seconds. Thus the failure patterns of the bearings could be studied in detail.

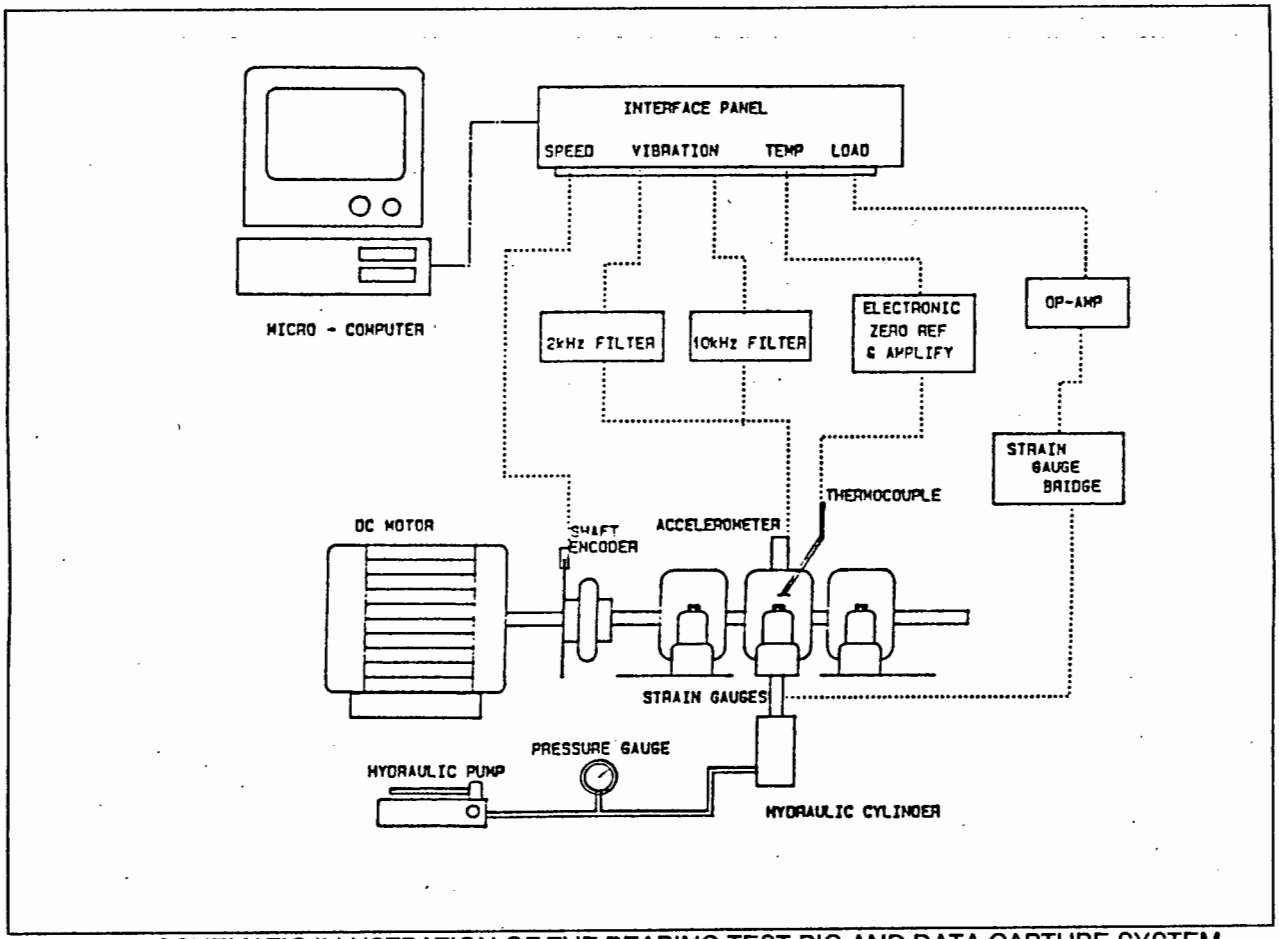

FIGURE 2 : SCHEMATIC ILLUSTRATION OF THE BEARING TEST RIG AND DATA CAPTURE SYSTEM

4. EXPERIMENTATION TO DEFINE THE BEARING ANALYSIS SOLUTION SPACE

Various experiments were performed using the test rig and data capture system as described above. These experiments were designed to develop and quantify the bearing analysis solution space. This then enabled the development of the bearing analysis expert system.

The solution space was the multi-dimensional space within which the expert system would operate to analyze bearing condition. Any vector within the space represented the bearing condition under the particular operating conditions and vibration patterns. The boundaries of this space were defined by the limits of all analysis parameters.

The parameters were grouped into 3 related subsets as shown in figure 3. These were the actual bearing condition, the operating conditions and vibration measurements. Primary research involved defining the relationships between these three subsets. This was achieved by the parameter dependency and bearing destruction experiments and a restructuring of the bearing rated life equation.

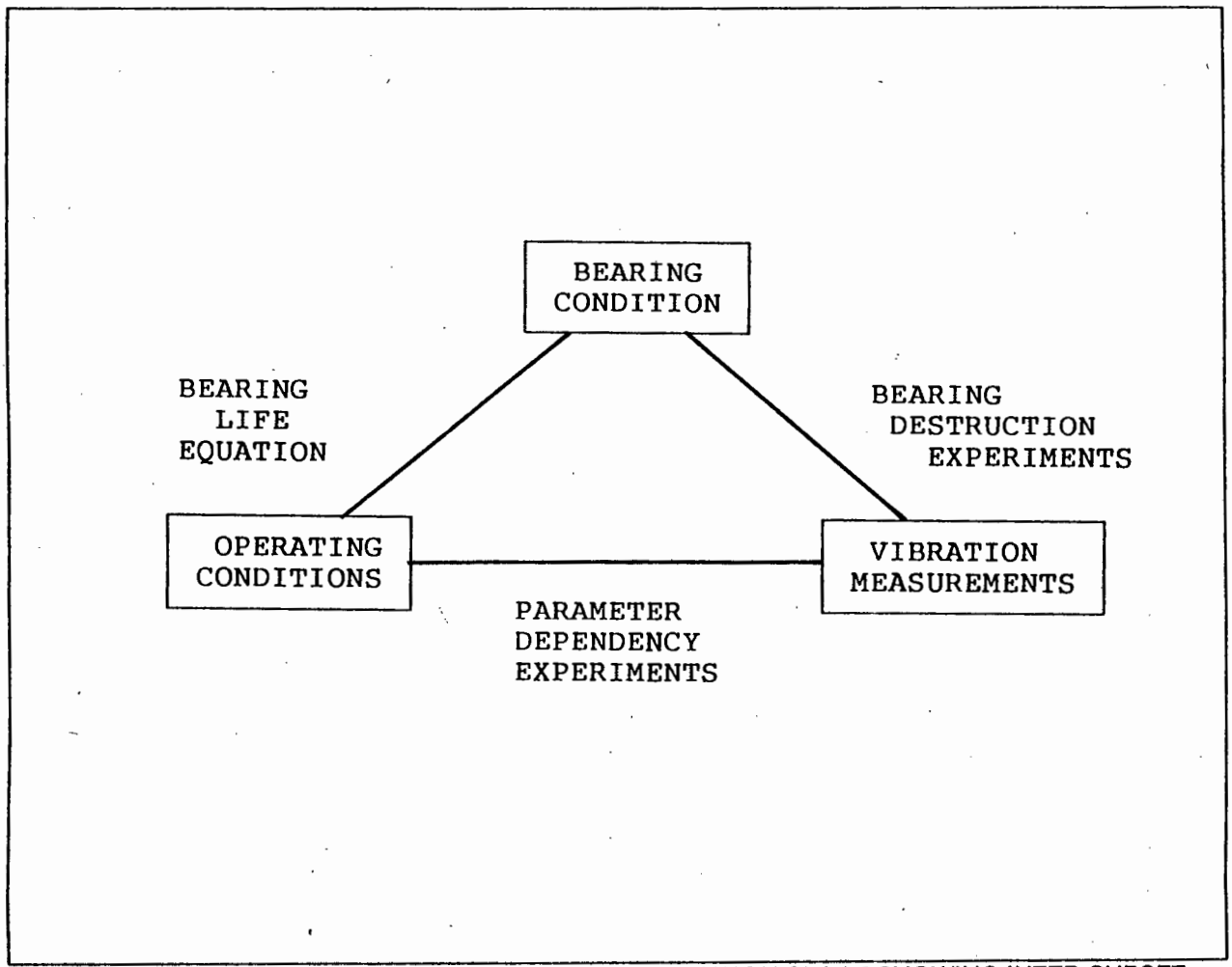

FIGURE 3 : SCHEMATIC DIAGRAM OF THE ANALYSIS SOLUTION SPACE SHOWING INTER-SUBSET EXPERIMENTS

THE PARAMETER DEPENDENCY EXPERIMENTS

Firstly the parameter dependency experiments were performed by varying one operating condition while keeping the others constant. The vibration parameters were then measured over this variation. During these tests the bearing condition was known to be clear of any fatigue spalling or wear damage. The operating conditions included Bearing Load, Speed and Temperature while the vibration parameters were the RMS and Peak Acceleration, Kurtosis and Crest Factor.

The method was to vary slowly the given operating condition while constantly capturing all measured parameters. These values were then stored whenever the varying operating condition changed by a predetermined amount. The stored data was then analysed to determine the relationship between the operating condition and the vibration parameters.

For example the dependence of the Peak Acceleration on bearing speed is shown in figure 4. Best fit curves (as shown) were obtained by regression analysis. Thus these experiments provided empirically defined relationships between each of the operating conditions and vibration parameters. Equation 1 shows the relationship between Peak Acceleration and Speed.

$$P = C\,S^{M} \qquad \text{..... 1}$$

where : P = Peak acceleration [G's]
S = Shaft speed [rpm]
$C = 1.647 \times 10^{-5}$
M = 1.554

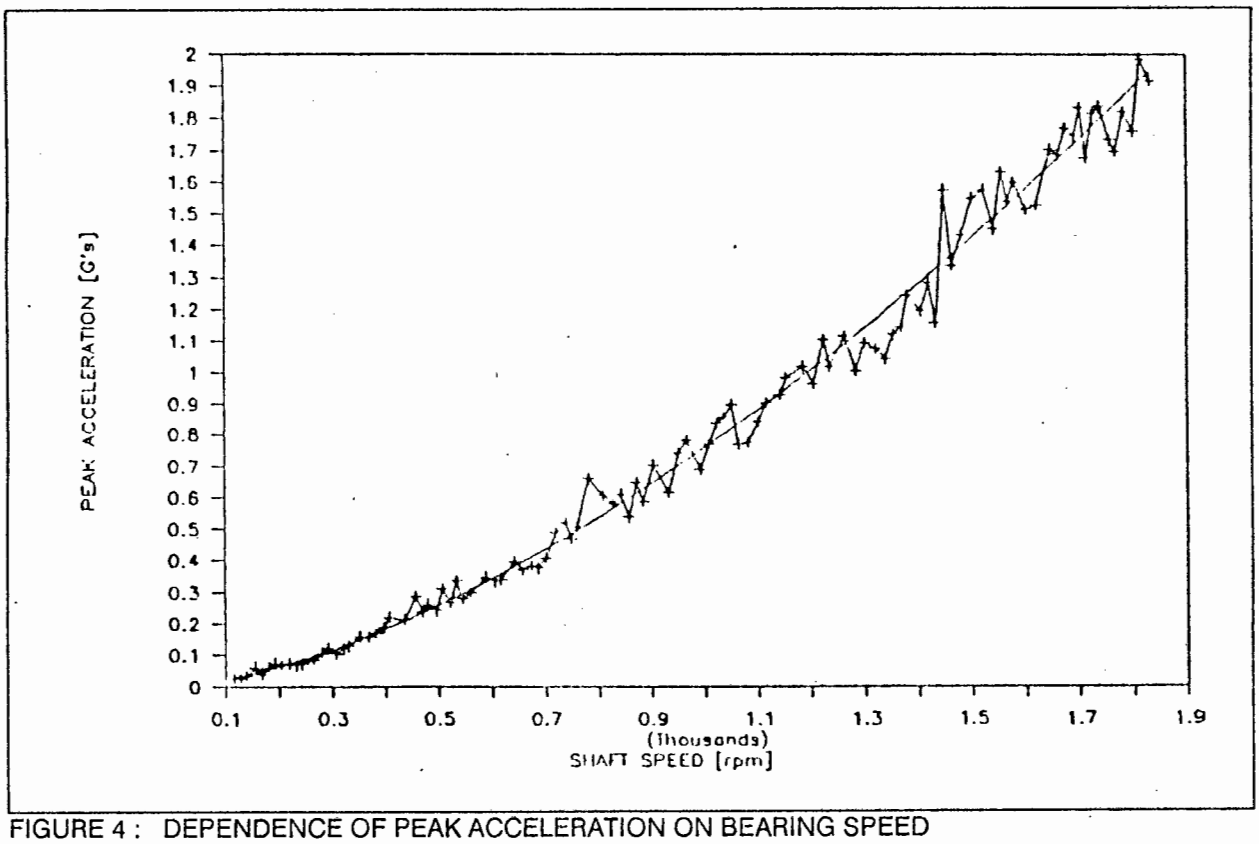

FIGURE 4 : DEPENDENCE OF PEAK ACCELERATION ON BEARING SPEED

It was found that the vibration parameters Peak and RMS acceleration increased rapidly with increasing speed, while the parameters Kurtosis and Crest Factor were largely independent of speed. In addition the Peak and RMS decreased approximately linearly with increasing load, while Kurtosis and Crest Factor again were largely independent of load.

THE BEARING DESTRUCTION EXPERIMENTS

The second set of experiments involved the destruction of bearings under constant operating conditions. From this the relationships between bearing condition and vibration parameters were established.

The method was to run the bearings under constant (reference) operating conditions. This was continued until fatigue failure rendered the bearings unfit for further use. During this time all measured parameters were continuously stored approximately every two seconds. This enabled a detailed analysis of the vibration patterns indicated by the various parameters.

The vibration parameters were then related to various stages of bearing failure. For example figure 5 shows the variation of Kurtosis for the last hour of the experiment. This particular bearing lasted for approximately 11 hours. It can be seen that the graph has been divided into 5 separate time zones. These zones were selected by combined analysis of the vibration parameters.

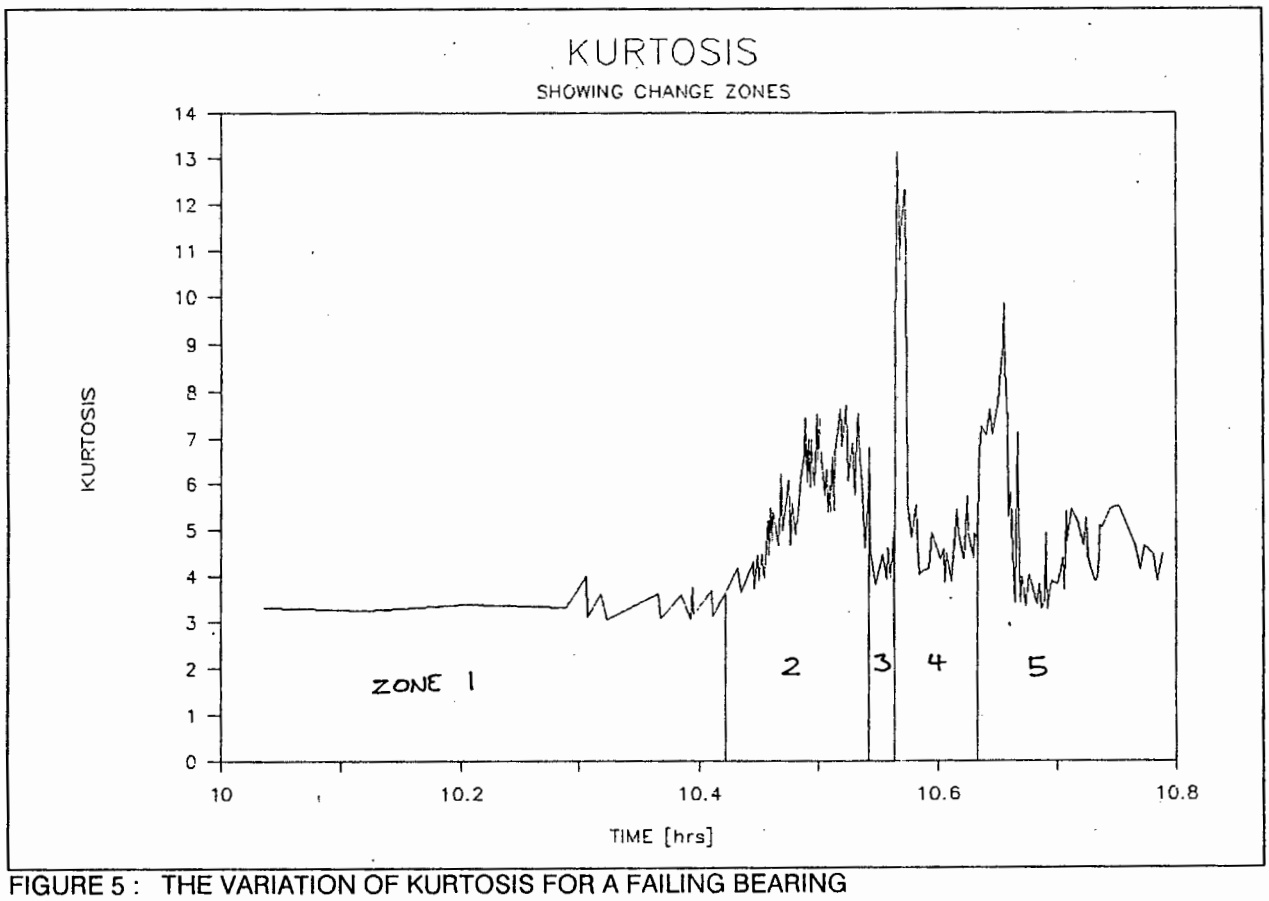

FIGURE 5 : THE VARIATION OF KURTOSIS FOR A FAILING BEARING

For example the Kurtosis introduced each new stage of fatigue damage by suddenly rising as a new spalling pit appeared. In addition Kurtosis gave the earliest warning that something was occurring in the bearing. This can be seen in Zone 2.

The bearing condition was developed as a 'linguistic variable' for expert system diagnosis. This involved assigning characteristic descriptions to the 5 time zones. These descriptions are shown in table 1.

TABLE 1 : DESCRIPTION OF BEARING CONDITION IN THE TIME ZONES

ZONE	CHARACTERISTIC DESCRIPTION OF BEARING CONDITION
1	Satisfactory and running normally.
2	Showing incipient fatigue failure in the form of surface scoring and light flaking.
3	Showing advanced surface flaking with spalling and pitting imminent.
4	Showing failure in the form of early pitting and spalling
5	Showing advanced fatigue failure with well defined pitting and spalling. Total failure is imminent.

RESTRUCTURING THE BEARING LIFE RATING EQUATION

The final set of relationships needing definition were those between the bearing condition and the operating conditions. For this purpose, use was made of the extensive research by bearing manufacturers to define these relationships. The results were embodied in the well known and commonly used 'Bearing Life Rating Equation' shown in reduced form as follows.

$$L_{10h} = a_{23} (C/P)^3 \, 10^6/(60\ N) \quad \text{[hours]} \qquad \ldots\ldots 2$$

where :
- L_{10h} = Rated life in hours for a failure probability of 10% (i.e. $a_1 = 1$)
- a_{23} = Combined life adjustment factor for materials and operating conditions
- N = Bearing speed [rpm]
- C = Dynamic load rating [kN]
- P = Equivalent applied dynamic load [kN]

This equation was restructured to enable the direct application of the operating conditions Load Speed and Temperature. These operating conditions were inherent in the equation and restructuring involved exposing them. From the restructured equation the remaining life of the bearing could be predicted by inserting the operating conditions directly.

The load was clearly evident in equation 2 as the 'equivalent applied dynamic load'. The speed and temperature were less evident, being contained in the factor a_{23}. This adjustment factor was empirically derived from the lubricant viscosity index 'K'. 'K' in turn was defined as the ratio of the operating viscosity to the required lubricant viscosity.

Firstly the required lubricant viscosity was derived from the speed of the balls relative to the bearing raceways. Secondly the operating viscosity was derived from the viscosity temperature chart for the particular lubricant being used. Thus the operating and required viscosities could be replaced by equations having speed and temperature respectively as the variables for a given bearing and lubricant.

In addition the lubricant viscosity index 'K' provided significant diagnostic information on the adequacy of the lubricant. A value of one indicated that the operating viscosity was equivalent to the required viscosity and the lubricant was therefore satisfactory for the purpose. Higher values represented better lubrication while lower values poor lubrication conditions.

The final equation for the given bearing and lubricant was as follows.

$$L_{10h} = C\, T^{Mt}\, N^{Mn}\, P^{Mp} \qquad \ldots\ldots 3$$

where :

L_{10h} = Bearing life [hrs.] for 10% probability of failure
N = Speed in rpm
T = Temperature in °C
P = Applied load in kN
C = Combined constant
Mi = Exponents for the operating parameters, chosen according to the value of 'K'

For a given bearing and lubricant, the only variables contained in this equation were the operating conditions. This was as required by the objectives of restructuring the bearing life rating equation. In addition the restructured equation also provided useful insight into the differing effects of the operating conditions on bearing life.

This concluded the defining of the bearing analysis solution space. From this position it was possible to proceed with the development of the expert system.

5. DEVELOPMENT OF THE EXPERT SYSTEM

The expert system was developed as three separate rulebases. These were divided according to the three levels of machinery condition analysis, viz Detection, Diagnosis and Prognosis.

THE DETECTION RULEBASE

Detection, being the simplest level, merely provided a yes/no answer to the possibility of something being wrong with the bearing. There was only one detection rule as illustrated by figure 6. This rule was designed to execute as rapidly as possible, acting only as a primary filter to reduce the amount of data needing diagnosis. The critical values for the four parameters were statistically derived from the bearing destruction experiments.

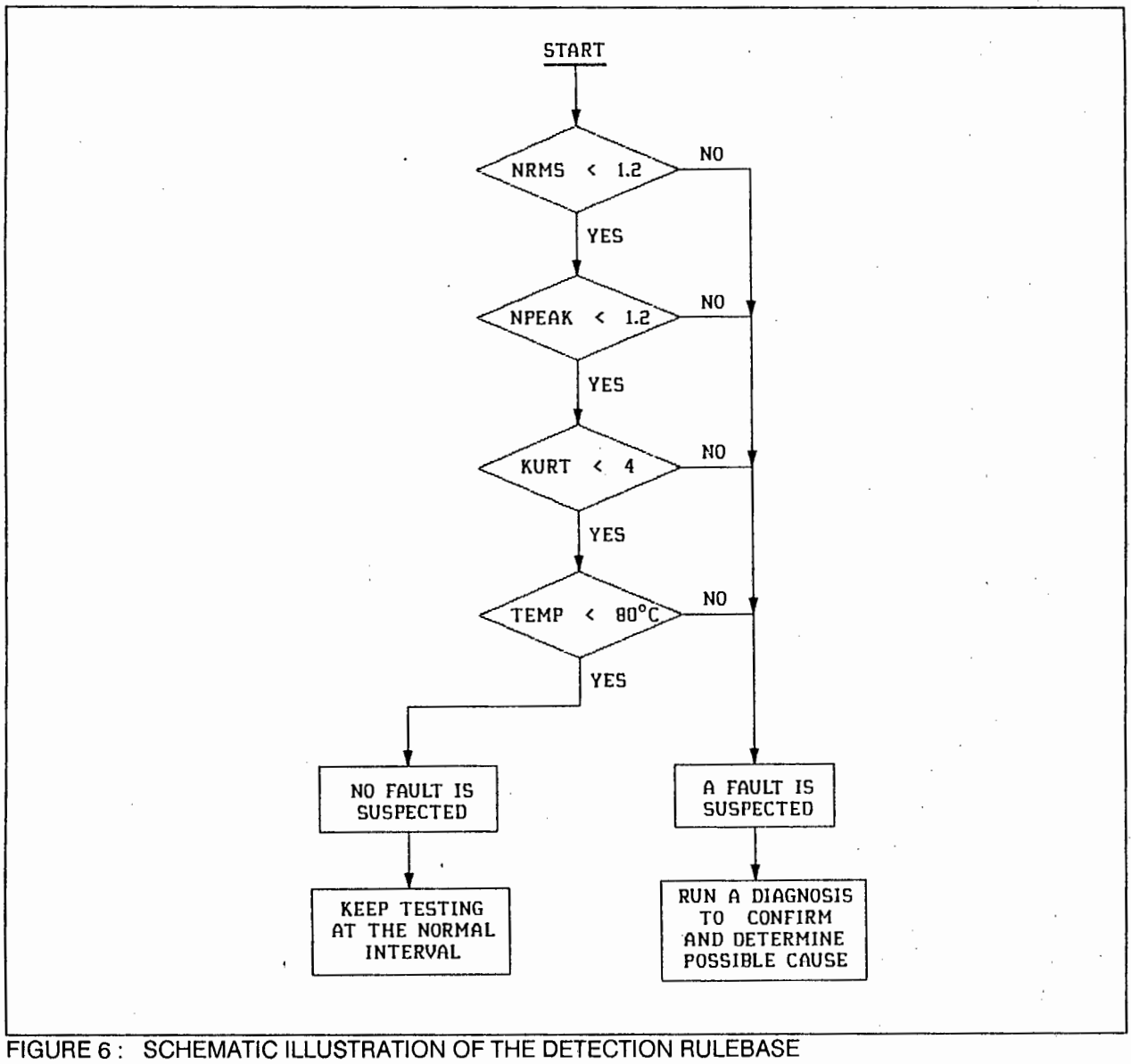

FIGURE 6 : SCHEMATIC ILLUSTRATION OF THE DETECTION RULEBASE

THE DIAGNOSIS RULEBASE

The diagnosis rulebase was designed to determine the cause of failure once possible damage was detected. The diagnosis rules were developed in two groups, viz. the vibration parameter rules and the operating condition rules.

Before the vibration parameters were diagnosed, they were adjusted for the operating conditions and then normalised by the average value for a bearing in good condition. The adjusting was done so as to enable diagnosis over a range of loads and speeds, while the normalisation was to enable the development of a more generic set of rules.

Adjustment factors were obtained by applying the current and reference operating conditions to the empirical equations derived in the parameter dependency experiments. These factors were then used to adjust the measured vibration parameters up or down to enable diagnosis at the reference operating conditions.

The vibration parameter rules were developed using using the five zones of bearing condition. These zones were established by the bearing destruction experiments [figure 5] and included statistical selection of zone limits. In addition the probabilities for each zone were selected using the binomial distribution.

These concepts are illustrated in the following rule for diagnosing operating zone using NRMS (adjusted and normalised RMS reading).

IF:

NRMS >= RMS_LIMIT_3
and NRMS < RMS_LIMIT_4

THEN:

BEARING CONDITION : Zone 1 Probability = 19/100
BEARING CONDITION : Zone 2 Probability = 47/100
BEARING CONDITION : Zone 3 Probability = 63/100
BEARING CONDITION : Zone 4 Probability = 47/100
BEARING CONDITION : Zone 5 Probability = 19/100

The descriptions for each zone are given in table 1. Similar rules were developed for each vibration parameter. Thus it could be established with some measure of confidence what the internal condition of the bearing was. It was then the task of the operating condition rules to decide which operating condition was most probably causing the failure.

The operating condition rules were developed by categorizing each of the measured and derived operating conditions into 'fuzzy logic' lookup tables. As an example the categories of the lubricant viscosity index are shown in table 2.

TABLE 2 : THE LUBRICANT VISCOSITY INDEX CATEGORIES

K RANGE	CHARACTERISTIC DESCRIPTION. The Lubricant Viscosity is ...
< 0.4	Totally inadequate
0.4 - 1.0	Barely sufficient
1.0 - 4.0	Satisfactory
> 4	Very good

Two separate sets of rules were developed for the operating conditions. These were for the 'bearing rated life' and the 'lubricant viscosity index'. This enabled a divide and conquer approach to rule base development. A portion of the flow chart for the bearing rated life rules is shown in figure 7.

The decision as to whether the bearing was failing or not was provided by a separate rule stating that if the bearing condition was in zone one then it was not failing, else it was failing. This simplified the diagnostic rulebase extensively because one set of operating condition rules applied to all the vibration rules.

In particular figure 7 represented the following rule.

IF : The bearing is failing
AND : The rated life is very short
AND : The load is fairly low
AND : The lubricant viscosity is barely sufficient

THEN : The cause of failure is : Incorrect Lubrication Probability = 60%
OR : The cause of failure is : Overloading Probability = 30%
OR : The cause of failure is : Some other cause Probability = 10%
RECOMMEND : Use a lubricant with higher viscosity

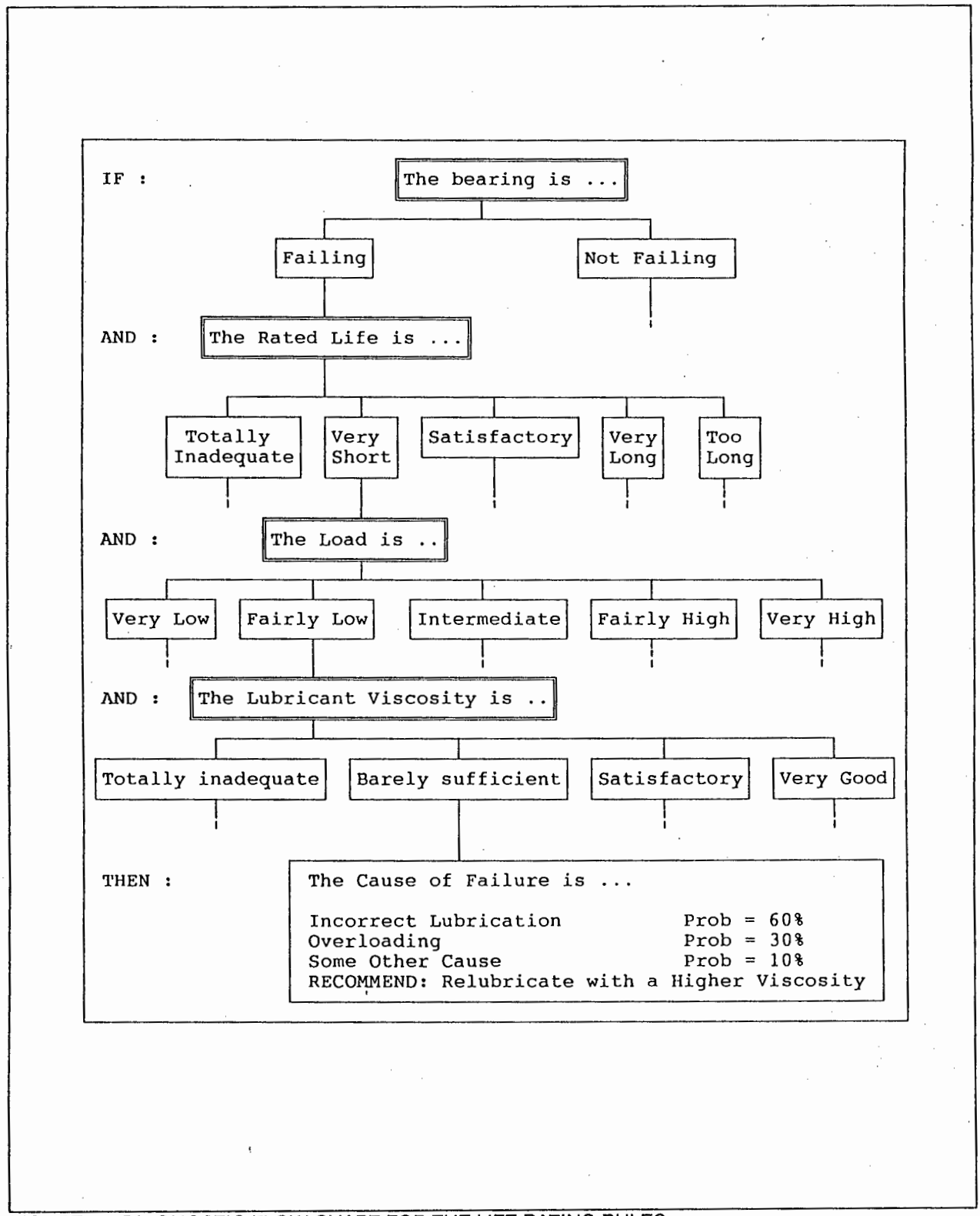

FIGURE 7 : DIAGNOSTIC FLOW CHART FOR THE LIFE RATING RULES

THE PROGNOSIS RULEBASE

Finally the task of the prognosis rulebase was to provide an estimate of the remaining bearing life and the progression of failure. This was accomplished by first of all establishing the bearing condition zone. A simple prognosis was then given using the sum of the percentage durations of the remaining time zones and the rated bearing life under the current operating conditions. In addition the prognosis rulebase provided a suitable testing interval by the same process.

Thus the three tasks of detection, diagnosis and prognosis were sucessfully accomplished by the various rulebases.

6. CONCLUSION

In conclusion, the system performed well and provided good insight into the possibility of developing the proposed industrial condition monitoring system. However more work was needed to develop the bearing diagnosis into a generic system which could be adapted for any rolling element bearing.

One of the most interesting aspects of the project was the combination of different types of measurable information to form a more complete analysis of bearing condition. In particular the operating conditions provided a significant contribution to the tasks of diagnosis and prognosis.

Another important contribution was the detailed study of the performance of the vibration parameters over the life of a bearing failing by typical fatigue. More specifically the interplay of the parameters provided insight into bearing failure patterns and how they may be detected using a combination of parameters.

The work also revealed that it was a complex task to achieve real expert-like analysis with an expert system. The expert provided no unexpected conclusions and the performance was predictable. However, to a person less skilled in diagnostics the system could provide new insight and relieve an otherwise difficult task. To the more experienced user it could relieve the routine and free up time for the more interesting diagnostic problems.

The future of this type of system looks bright with a large amount of development being done internationally. It is hoped that work in this country will keep abreast of this trend and that the recommended research and development will come to fruition.

CONTENTS

CONTENTS

CONTENTS (cont)

LIST OF FIGURES

LIST OF FIGURES (Cont)

LIST OF TABLES

LIST OF APPENDICES

LIST OF APPENDICES (CONT)

SECTION 1

INTRODUCTION

1. INTRODUCTION

This thesis is associated with the development of expert systems to automate the analysis of industrial machinery running condition. In particular it deals with the use of vibrations to analyze the running condition of rolling element bearings. Also included is a proposal for the framework of a generic industrial condition monitoring package.

After the introduction the next two sections deal with the initial motivation for the project and a statement of the problem to be addressed by research. This leads to the research objectives and the scope and limitations related to achieving them.

Section six deals with a brief literature survey, developed according to the flow of information through a typical condition monitoring system. Included in the survey are the physical aspects of bearings related to condition monitoring, analog and digital signal processing techniques for vibrations and various aspects on the theory and use of expert systems. This includes a discussion on the particular expert system shell used for rulebase development.

Section seven deals with the directing of research work towards the condition monitoring needs of industry. This is followed by section eight containing the proposal of a generic, expert system based condition monitoring package for industrial machinery monitoring.

Sections nine and ten discuss the development of the bearing test rig and associated data capture system. This includes the design, building and calibration of all mechanical, electrical, electronic and computerized systems for the controlled destruction and monitoring of test bearings.

Section eleven develops the concept of a solution space for bearing condition analysis. This involves the three subsets of the solution space, namely the operating conditions, the vibration measuring parameters and the variables describing bearing running condition. Section twelve then goes on to define and develop the relationships between these subsets using experimental analysis.

In section 13 a rule based expert system is developed to analyze the bearing running condition. This system emphasizes the automated collection and analysis of data with a minimum of consultation with the user. It is structured around the three stages of analysis, viz. detection, diagnosis and prognosis. The relationships developed in sections 11 and 12 are used extensively in the construction of the rules. The system is developed as generically as possible to allow for extension to other types and sizes of bearings.

Sections 14, 15 discuss the conclusions reached and recommendations made for further research. Finally section 16 concludes the report.

SECTION 2

MOTIVATION FOR THE PROJECT

2. MOTIVATION FOR THE PROJECT

The merits of condition based or predictive maintenance had been demonstrated on a range of industrial plants [1] and were becoming more widely recognized. Typical advantages cited were reduced down time, prevention of catastrophic failure and increased plant reliability. These factors alone were enough to motivate interest and research on the application of condition monitoring techniques.

However many of the condition monitoring techniques avaiable were sophisticated and complex, being implemented by highly skilled personnel. Thus the tendency on many industrial programs was to rely on simple diagnostic techniques. While these were relatively easy to implement, their simplicity provided inherently unreliable diagnosis.

Thus the need for easier implementation of the more sophisticated techniques provided further motivation for the project. What was needed was a system which would assist relatively unskilled personnel to use better and more reliable diagnostic techniques without necessarily having in-depth knowledge of the techniques themselves.

However this would mean that the system would have to perform its own analysis of the results. In other words some form of artificial intelligence was required to extract diagnosis from the analysis techniques used. This could also assist the maintenance decision making process by providing recommendations in simple english as opposed to relatively complex analysis data.

Such a system would have to be capable of analysing most types of machinery components commonly occurring in industry. For example a damaged gear could affect the analysis of a bearing and the system would have to be able to distinguish between the two. Thus it became apparent that a system would have to be conceived in the context of the wider needs of condition monitoring in industry.

SECTION 3

STATEMENT OF THE PROBLEM

3. STATEMENT OF THE PROBLEM

The problem that the research was to address was developed from a growing understanding of the problems faced by industry in attempting to implement vibration monitoring methods [2].

The following factors were considered to define the problem :

- The lack of reliability and consistency of commonly used vibration monitoring methods.
- The shortage of first hand experience in vibration monitoring.
- Insufficient specific knowledge of machinery vibration failure patterns, leading to conflicting reports on monitoring methods.
- The need for earlier warning of incipient machinery failure.
- The need for automated conversion of large amounts of condition monitoring data into useful information such as recommended maintenance action.

The problem as defined above led to the following 'working hypothesis' :

THE WORKING HYPOTHESIS

That the interpretation of rotating machinery running condition using multiple condition assessment techniques can be simplified in implementation and increased in reliability by utilizing the data capture and processing power of computers incorporating the diagnostic capabilities of expert systems.

SECTION 4

OBJECTIVES OF THE RESEARCH

4. OBJECTIVES OF THE RESEARCH

To address the problem as stated above, the following research objectives were established :

i) To investigate the current state of condition monitoring in industry including the techniques being used and the type of machinery being monitored.

ii) To develop the framework of a portable computer based condition monitoring system for analyzing typical industrial machinery.

iii) To develop a specific aspect of the framework in detail, namely the analysis of rolling element bearing running condition.

The last objective was to constitute the bulk of the research and was therefore further broken down into the following steps :

iv) To design and build a mechanical test rig for investigating the failure of rolling element bearings under varied operating conditions in a simulated industrial environment.

v) To design and build a digital data capture and signal processing system for the investigation of bearing failure patterns as provided by the test rig in i) above.

vi) To investigate the most suitable and reliable digital signal analysis techniques for detecting incipient bearing failure, diagnosing the possible cause of failure and prognosing the remaining useful operating life.

vii) To develop an expert system to perform bearing analysis using the techniques as determined in iii) above, including the task of making recommendations for maintenance action.

The scope of the research objectives given above was subject to various limitations. This is discussed in the following section.

SECTION 5

SCOPE AND LIMITATIONS OF THE RESEARCH

5. SCOPE AND LIMITATIONS OF THE RESEARCH

The scope of research was limited by a number of factors including available time, research funding and available equipment, the data capture and processing system and departmental experience in condition monitoring and related fields. These are discussed in more detail as follows :

AVAILABLE TIME

The magnitude of the research objectives was too large to be covered in detail in the given two year period. However the objectives also formed part of a proposed long term project [3] and were therefore given in detail. The time constraints dictated that objective ii) was to receive the minimum attention as it would evolve with proposed future research.

As regards bearing testing there was only time for one type of bearing to be tested, namely 'self aligning ball bearings'. Thus a limitation of the research was that it would have to be assumed that other rolling element bearings produced similar failure patterns.

RESEARCH FUNDING AND AVAILABLE EQUIPMENT

Firstly there was a critical shortage of the type of electronic equipment needed for data capture and analysis. For example there was only one spectrum analyser avaiable for the whole of Mechanical, Acoustics, Electrical and Electronics departments. This instrument was already being used to capacity.

The problem was being addressed by making recommendations to the Heads of the concerned departments [4]. However if these were considered, the lead time to purchasing such equipment would be too long to affect the first two years of the project.

In addition such equipment as analogue filters and amplifiers were not readily available and would have to be designed and built as part of the project.

Secondly, departmental funding for the purchase of the required equipment was limited. An attempt was made to obtain research funding from industry for this purpose [5, 6]. For various reasons this was unsucessful. However, funds were available to purchase a basic IBM compatible computer for general laboratory use.

THE DATA CAPTURE AND PROCESSING SYSTEM

The data capture and processing system placed a limit on the scope of research for two reasons. Firstly such a system was nonexistent in the department and therefore had to be designed and built from the ground up. This included all analog signal processing equipment and digital signal processing routines. In addition all software to drive the analog to digital conversion card had to be developed from first principles.

Secondly the available computer provided limitations because of processing speed. The large volumes of data required to be processed placed a limitation on the number and type of data analysis parameters that could be used. In addition the digital storage space was found to be insufficient for the storage of 'raw' time data records for later analysis.

DEPARTMENTAL EXPERIENCE

The lack of in-depth experience within the department provided a further limitation on research. In a sense the undergraduate research of the few previous years was pioneering and only provided a basic framework on which to proceed.

Thus choices on such issues as frequency ranges, signal processing techniques, data capture and processing systems etc. had to be made on information gleaned from the literature survey and the experience gained from previous work [7, 8]. This inevitably led to various choices having to be remade, with the consequent narrowing of research scope.

The literature survey and discussion of relevant theory are covered in the following section.

SECTION 6

LITERATURE REVIEW AND DISCUSSION OF RELEVANT THEORY

6. LITERATURE REVIEW AND DISCUSSION OF RELEVANT THEORY

The literature review and discussion of the relevant theory are developed by considering the flow of information through a typical condition monitoring system. This is illustrated in figure 6.1.

Firstly, machinery being considered for condition monitoring will produce measurable parameters giving an indication of its running condition. Typically the machinery will consist of rotating and reciprocating components while the measurable parameters could be speed, temperature, vibration, etc.

Some form of transducer can be used to convert any of these parameters to a proportional analog voltage signal. This signal will usually be processed in an analog manner by amplification and filtering to reduce noise and other effects. The signal is then prepared for digital signal processing.

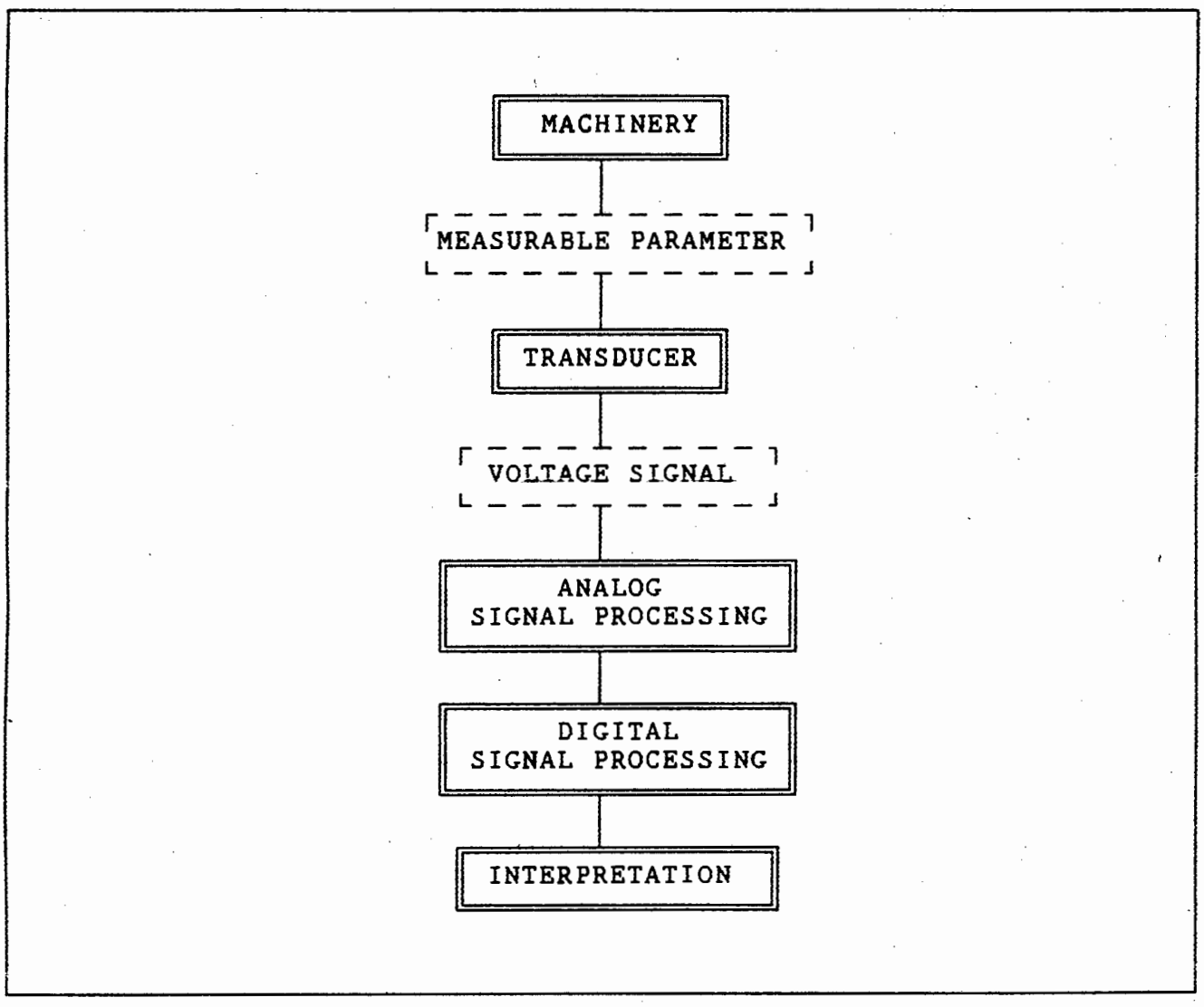

FIGURE 6.1 : THE FLOW OF INFORMATION THROUGH A TYPICAL CONDITION MONITORING SYSTEM

The reason for using digital signal processing lies in the flexibility afforded by computers. Many otherwise impossible signal analysis techniques may be applied to a digitized signal [9].

The first stage involves the analog to digital conversion. This is followed by various techniques such as digital filtering and conversion to domains other than that in which the signal was measured. Digital signal processing usually involves some form of data reduction [10] and preparation for interpretation.

Finally the signal is interpreted as an indication of the running condition of the machinery. The interpretation may be performed by a person skilled in such a task, or automatically by a computer using some form of artificial intelligence.

The machinery that this review is primarily concerned with is the rolling element bearing. It covers the measurable parameters applicable to bearings, namely load, speed, temperature and most importantly vibrations. In addition the digital signal processing techniques are mainly concerned with extracting information from the vibration signals.

The first stage of the review discusses the important physical aspects of bearing operation.

6.1. PHYSICAL ASPECTS OF ROLLING ELEMENT BEARINGS

This section deals with various physical aspects of bearings relating to monitoring their running condition. It starts with common failure modes such as wear, fatigue and others. Then bearing service life rating techniques are discussed, followed by bearing vibrations. This includes a section on general vibration theory and terminology.

6.1.1. BEARING FAILURE MODES

Before attempting to analyze bearing condition it is important to define the common types of failure occurring. Essentially there are two modes of failure, namely fatigue and wear. In the final analysis most causes of bearing damage lead to either or both of these failure modes.

The first of these to be discussed is fatigue. This includes the fatigue mechanism and its causes.

FATIGUE

Fatigue can probably be cited as the most common mode of failure for rolling element bearings [11]. This can also be deduced from the so called 'Fatigue life rating' equation as proposed by [12] and commonly used by bearing manufacturers [13].

Most bearing applications involve cycled stress fatigue in their normal operation. This is due to the fact that the load applied to the bearing is transferred by means of the rolling elements. These continuously cycle the load over any given point on the loaded zone of the raceways and are cyclically loaded themselves due to rolling. This stress reversal usually occurs many times during the life of a bearing [append. 1].

In addition it is not unusual to encounter high stresses in the bearing elements. This can be demonstrated using Hertzian contact stress theory for point contact [15]. Even though Hertz considered dry conditions it can be shown that the pressure distribution under EHD (elastohydrodynamic) lubricant conditions is very similar to the hertzian contact pressure as shown in figure 6.2. [15]. Thus the combined conditions of high cycling and high stresses lead to fatigue failure being common in bearings.

The typical progression of fatigue failure is as follows : The cyclic loading and alternating stresses cause cracks to grow from subsurface defects. These defects consist of microscopic inhomogineities in the bearing material such as unevenly distributed alloying elements and slag inclusions.

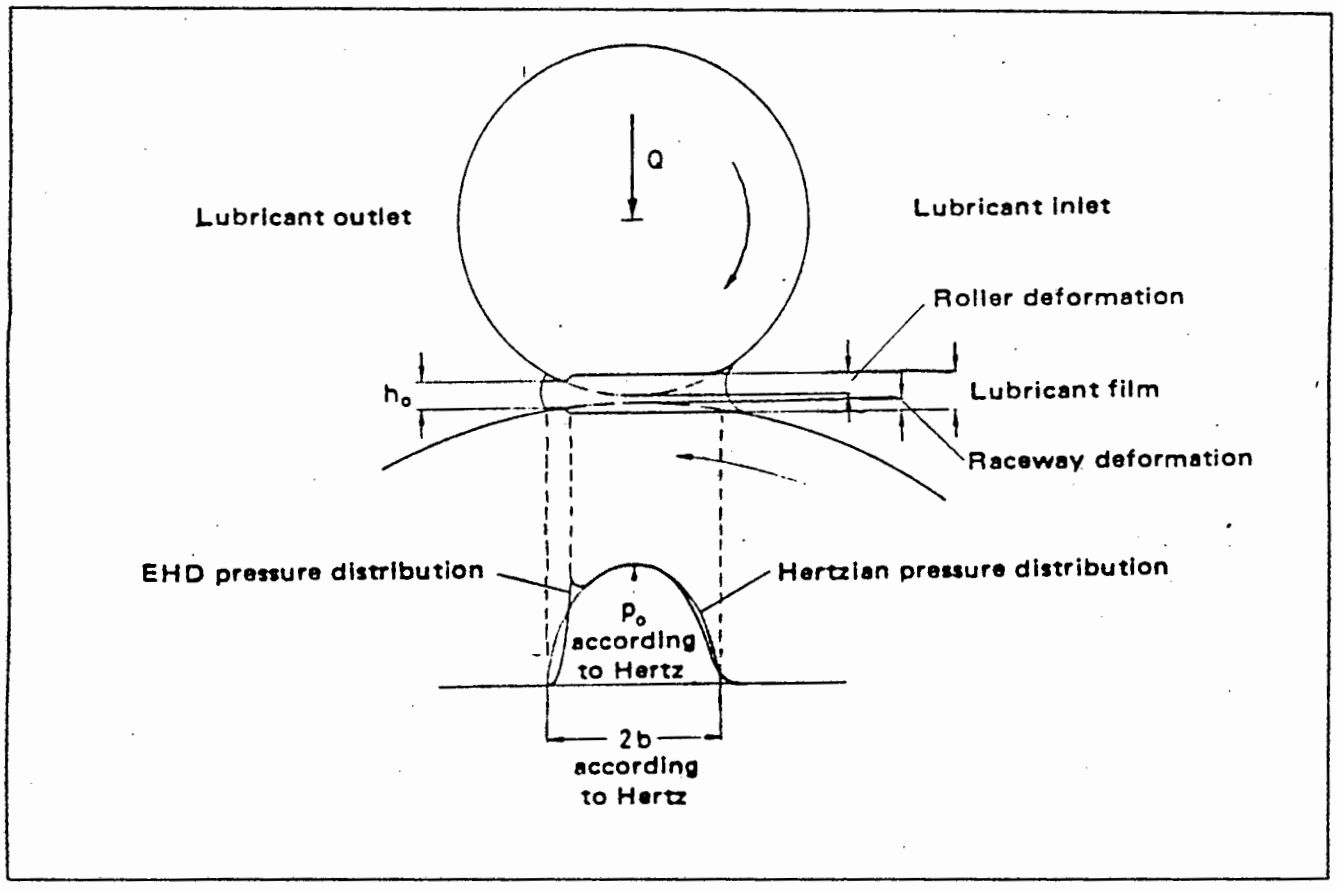

FIGURE 6.2 : ELASTOHYDRODYNAMIC PRESSURE DISTRIBUTION IN A ROLLING ELEMENT BEARING

The cracks grow and spread until they reach the bearing surface, causing a flake of material to break off. This results in pitting or spalling on the surface and flakes in the lubricant. Both of these cause further stress concentrations and accelerate the fatigue process. Thus the bearing fails more rapidly once fatigue spalling has occurred.

It has been demonstrated by experiment [15] that the fatigue cracks form at a depth corresponding to the maximum shear stress. This corresponds to a depth of approximately half the width of the contact area.

WEAR

Wear is a condition which occurs in all bearings which are not ideally lubricated. Ideal lubrication can be defined as follows:

- The lubricant has the correct viscosity at the given operating conditions. i.e.. a lubricant film of sufficient thickness is formed (> 0.5 microns).
- The lubricant must be sufficiently clean, meaning that included solid particles are of an order of magnitude smaller than the minimum lubricant film thickness.
- There are no chemically aggressive constituents or moisture in the lubricant.

It has been demonstrated in a laboratory situation that effectively infinite life can be achieved under ideal conditions [15]. However lubricants degrade with time and bearing seals wear out, ensuring that lubricant conditions in the industrial environment are seldom ideal. Hence wear is a commonly encountered mode of failure for bearings.

The mechanism of wear usually takes the form of microscopic pitting of the bearing surfaces. Severe wear in bearings can be detected by a large increase in the radial clearance and is characterized by a dull appearance on the raceways and rolling elements [7, 16].

A bearing will only be considered to have failed by wear if it ceases to perform its required function. For example large bearing clearances may cause excessive vibrations, misalignment and wear in related machinery such as gears, shafts, impellers etc.

OTHER FAILURE MODES

There are other types of failure which are less commonly encountered. These failure modes usually result in fatigue or wear and the diagnosis will therefore be similar. From a vibration measurement point of view it is usually not possible to tell the difference between say normal fatigue and fatigue caused by another failure mode [7].

These mechanisms are manifold and include the following :

- False brinelling
- Fretting corrosion
- Corrosion and erosion by chemical and electrical means such as welding current passing through the bearing
- Faulty mounting practise such as hammering, incorrect clearance fits, misshaped or wrongly sized housings and shafts
- Misalignment
- Vibrations from excessive clearances or external sources such as unbalanced loading.
- Foreign material ingression, such as in commonly lubricated gearboxes.
- Incorrect or faulty lubrication such as low levels, wrong viscosity, ingression of moisture, etc.
- Overloading, including excessive preloading.
- Cage failure
- Overheating due to faulty mounting practise or running conditions.
- Overspeeding
- Manufacturing errors such as improper hardening of bearing surfaces, material inhomogineities, improper surface finish, etc.

The above mechanisms may be interrelated, such as overheating due to incorrect lubrication, or overloading due to misalignment, etc. The failure modes have a direct effect on the bearing life. This is discussed in the section that follows.

6.1.2. BEARING LIFE

The ultimate goal of monitoring the condition of a rolling element bearing is to detect incipient failure, to diagnose possible causes of the damage and to prognose how the damage will progress. Because these are closely related to the bearing life it is important to know what factors determine the life and how they are used to predict it.

In general the factors determining bearing life are related to the mode of failure of the bearing as discussed in the previous section. The most important of these are the materials used to construct bearing components, the applied operating conditions and the performance of the lubricant. These factors are closely related to and dependent on each other.

THE LIFE RATING EQUATION

The most common method of calculating bearing life is that proposed in ISO 281/I-1977 [12]. This has been further developed by bearing manufacturers to cater for their design criteria such as materials used, etc.

The fatigue life equation starts with the standard definition of bearing fatigue life used all over the world [15] and given in ISO recommendation R281 and DIN 622 as follows :

> "The rating life of a sufficiently large number of dimensionally identical bearings is expressed by the number of revolutions (or hours) at constant speed reached by 90% of this bearing group before the first signs of fatigue appear."

Using this definition the rating life is then determined experimentally for a specific bearing type running under various loads. The results are plotted on Weibull paper for the failure probability 0.10 (i.e. 90% reliability) and typically produce a straight line [14, 15].

Then for convenience of calculation, the dynamic load carrying capacity 'C' is defined as the load under which the bearings will run for 10^6 revolutions before fatigue appears. This value is derived from an extrapolation of the Weibull plot.

10^6 revolutions only represents a few tens of hours of bearing life at typical speeds [append. 1]. However 'C' is purely used as a reference point for calculation of the rated life at other loads.

This assumes that the Weibull plot will be a straight line for all operating conditions. However this is not always the case and manufacturers have devised various methods of accounting for the operating conditions and other factors.

The basic equation for the fatigue life derived form the Weibull analysis is as follows.

$$L_{10} = (C/P)^p \qquad .....6.1$$

where

L_{10} = Life rating in millions of revolutions
C = Dynamic load rating
P = Actual service load applied
p = 3 for ball bearings (empirically derived)

This basic equation is commonly recommended by bearing manufacturers and usually modified by various adjustment factors as follows.

$$L_{nA} = A_1 A_2 A_3 L_{10} \qquad .....6.2$$

where

A_1 = Adjustment factor for reliability
A_2 = Adjustment factor for materials
A_3 = Adjustment factor for operating conditions

Usually A_2 and A_3 are combined into one factor because they are related. Tables are given for A_1 and A_{23} in bearing manufacturers handbooks [13].

Two things are worthy of note in this equation. Firstly it does not account for the fatigue life limit experienced by metals. There will be a load (or related stress) below which a bearing will 'never' fail by fatigue [14]. Secondly it assumes cleanliness of the lubricant which is rarely the case in industrial situations.

The SKF researchers have recently attempted to include these effects into the life rating equation [47]. However at present not enough detail is known of their results to include a discussion.

Also of note is that the above life rating calculations depend heavily on accurately knowing the dynamic loads applied to the bearing in service. This information is often not clearly definable, with corresponding loss of life prediction.

Advances in the materials used to construct bearings have had a significant effect on their fatigue life. In particular purity of the metal and more even dispersion of alloying elements have been achieved by new smelting techniques. This has led to gains in hardness without sacrificing toughness and through-hardening [15].

However the higher quality of materials has led to a greater scatter in the fatigue life of bearings. This is because fatigue is initiated at subsurface defects in the bearing. If a defect happens to lie in the loaded zone then the bearing will fail more rapidly.

Previously with many defects this was likely to occur whereas with the new materials it is less likely. This also contributes to the difficulty in predicting bearing fatigue life.

Equation 6.2 relates to a failure probability of 10%. However the median life of a specific bearing type (i.e.. 50% probability of failure) can be as much as an order of magnitude larger. This is even more so in the light of the above discussion.

The following section discusses the vibrations produced by bearings.

6.1.3. BEARING VIBRATIONS

Before the vibration of bearings can be discussed it is necessary to consider vibration theory and terminology in general. This will then be applied to bearings.

GENERAL VIBRATION THEORY AND TERMINOLOGY

In general the vibration produced by a complex dynamic mechanical system can be characterized by the amplitude of the vibration of a particular point in the system over time. This is known as a time series.

The amplitude may be defined as the displacement ($x(t)$), velocity ($dx(t)/dt$) or acceleration ($d^2x(t)/dt^2$) of that point relative to another fixed point in the system or relative to an inertial reference frame. It also has a related phase relative to some reference point in time.

The vibration signal measured at this point represents the sum of the vibration produced by each moving component in the system. If some part of the signal carries no useful information then it is regarded as 'noise'. In addition each components vibration will be modified in some way by the path through which it travels. This modification is dependent on the physical properties of the path and can have a significant effect on the content of the signal.

There are various levels of classification for these signals. At the first level a signal may be either deterministic or random. A deterministic signal is one which can be predicted exactly at any point in time using some specific law. More often the signal will be random in nature meaning that the amplitude can not be predicted exactly over time. In this case it is necessary to revert to a statistical description of the waveform.

Signals can be further classified as continuous or transient. The classification only has meaning over a specified time interval. Usually the interval will be larger than the duration of some transient phenomenon associated with the machinery being analyzed.

Signals will usually consist of some combination of the two. For example in a bearing the periodic cycling of the rolling elements will cause a continuous signal while some lubricant contaminant passing between two bearing surfaces would cause a transient impact. This transient effect would normally be short in comparison with the signal being analyzed.

Random signals may be further categorized as being sationary or non-stationary. Again this only has meaning within a specified time interval. This classification is particularly important from a condition monitoring point of view. The signal will be regarded as being stationary for a specified time interval being used to analyze running condition at that particular point in time. However the signal will not be stationary over the life of some machinery as it deteriorates.

More specifically a signal can be defined as being stationary if specified statistical properties do not vary with time. A process can be defined to have degrees of sationarity by specifying increasing orders of statistical moments (defined below) to be time-invariant.

The time series x(t) representing the random signal can be described by various probability concepts. The first of these is the probability density function (p.d.f). This function is defined as the percentage of time for which the signal lies within a particular amplitude range. Thus for x(t) we get the probability density function for values of x given by p(x) as follows.

$$p(x) = \lim_{\Delta x \to 0} \frac{\text{Prob } \{x < x(t) \leq x + \Delta x\}}{\Delta x} \qquad \ldots\ldots 6.3$$

Where the probability will always be greater than or equal to zero and the sum of all probabilities will be 1.

The most commonly encountered p.d.f. is the Gaussian (normal) distribution function, the definition of which may be found in most texts on statistical methods [e.g. 10]. A Gaussian distribution is typical for the signal produced by a bearing in good running order [17].

The next step in describing the signal involves the calculation of various statistical moments. Using the p.d.f of the series the moment M of order k can be defined about zero as follows.

$$M_k = \int_{-\infty}^{\infty} x^k p(x) dx \qquad \ldots\ldots 6.4$$

It is assumed that the signal is stochastic. That is to say the first moment (k = 1) known as the mean, will be zero while the second moment (k = 2) known as the variance will be finite. This is typical for the time series of a vibration signal.

Higher order moments are usually derived by removing the mean if it is non zero and dividing by the standard deviation raised to the power of the moment. Thus they are normalized and become known as the 'central' moments [17].

In general the odd moments (k = 1, 3, 5 ...) indicate the position of maximum density relative to the mean while even moments (k = 2, 4, 6 ...) indicate the spread or flatness of the distribution.

The second moment (k = 2) is called the variance, the root of which is the standard deviation (σ) while the third moment (k = 3) is known as the skewness. The square root of M_2 in this case will be recognized as the RMS (root mean square) value, an important parameter in vibration analysis.

Also of particular interest to vibration analysts is the fourth moment (k = 4). This is normalized by the fourth power of the standard deviation (σ^4) as described above, thereby becoming a dimensionless parameter. This parameter is known as the 'Kurtosis'. i.e.

$$\text{Kurtosis} = M_4/\sigma^4 \qquad \text{.....6.5}$$

and was first proposed for use by Dyer and Stewart in 1978 [17].

Intuitively, Kurtosis can be interpreted as measuring the impulsiveness of the signal. An impulse of short duration would cause the fourth moment to rise sharply while having a relatively small effect on the standard deviation. Thus Kurtosis would indicate this condition.

Calculating Kurtosis for an integer number of periods of a sine wave yields a value of 1.5. For gaussian noise it can be demonstrated that the Kurtosis will be 3 [10]. Thus a value of 3 is considered to indicate machinery in good condition.

Two other parameters commonly derived from vibration signals are the Peak value and the Crest Factor (C.F.). The True Peak value will be the maximum absolute magnitude of the signal. The Crest Factor is then obtained by this True Peak value divided by the RMS (defined above). i.e.

$$\text{C.F.} = \text{Peak} / \text{RMS} \qquad \text{.....6.6}$$

The Crest Factor is not regarded as a robust parameter [10]. That is to say it is overly sensitive to outliers, i.e. large spurious signals of short duration that can be regarded as noise. This was also found by Matthew and Alfredsen [50].

The signal as discussed above is said to be in the time domain. In general this raw time domain signal usually contains too much information or is too complex for visual analysis. Therefore it is often necessary to transform or map to another domain so as to extract certain information about the signal.

If the transformation involves a mapping of less than one-to-one there will be an associated data reduction [10]. This is usually a desirable effect as various aspects of the signal can be more easily recognized.

One of the domains for which it is particularly instructive to transform to is the amplitude domain. The calculation of the statistical moments as described above can be regarded as a mapping to the amplitude domain.

As can be seen this involves a severe data reduction from the original time signal. In other words one particular statistical moment is being used to characterize the signal at that point in time. It is more common to use a number of moments together to describe the signal. For example a pure Gaussian signal may be fully characterized by the first and second moments alone [10].

One of the most commonly used domains is the frequency domain. Mapping from the time domain to the frequency domain is a one-to-one process. However the data in the frequency domain is represented by magnitude and phase information. Thus data reduction may be achieved by considering only one of these at a time (usually the magnitude). Transformation to the frequency domain (Fourier Transform) is discussed in more detail in appendices 10, 11 and 12.

Related to the frequency domain is Cepstral (derived from spectral) analysis. This involves repeating the fourier transform on the frequency spectrum to detect periodicity in the frequency domain. The units revert back to time and are known as quefrency (derived from frequency). The technique is usually used for gears but has interesting possibilities for the detection of bearing harmonics.

It is often the case that the transformations may be more conveniently performed on a digitized version of the signal. This is discussed in more detail in section 6.3.

The following section discusses the vibration of bearings in particular.

BEARING VIBRATIONS

A study on the techniques being used for rolling element bearing analysis [18] revealed a considerable variety. Among the more commonly used were overall (unfiltered) vibration detection, low frequency (0-500Hz) narrow band analysis, high frequency (1-40kHz) spectral analysis (ring vibrations), envelope-spectrum analysis, statistical analysis, shock pulse monitoring (SPM) and acoustic emission.

Each technique requires its own specific set of signal processing equipment, some of which was not available for this thesis. Therefore only those techniques which are applicable will be discussed in more detail. Some of the techniques have already been covered in the previous section on general vibration theory.

Vibrations produced by a bearing are probably the most important source of diagnostic information for assessing its running condition without physically opening the bearing up. However a bearing is a complex dynamic system where the vibrations are dependent not only on the physical design of the bearing but also the applied operating conditions such as load, speed, lubricant etc. In addition the effect that structural paths have on the vibrations can be difficult to quantify.

Therefore a common method used on bearings and in fact most machinery, is to establish a 'baseline'. This is a set of values (for selected parameters) which characterize the vibration of the bearing in good running order. This then takes into account any structural effects and also any effect which the processing and measuring system might have on the signal. Of course this will be specific to the particular machine and is often called the machines 'signature' because of its individuality.

It is necessary to have some a priori knowledge of the bearings vibrations before meaningful techniques can be selected for testing. Firstly it is necessary to know the expected frequency content of the signal so that frequency ranges can be optimized.

The first step is to develop a simple model of the bearings kinematic behaviour. This is based on the rotational frequency of the inner or outer ring of the bearing, depending on which is rotating. From this it is possible to estimate the rotational frequencies of each element provided that there is no slipping between the contact surfaces. These are known as the element passage frequencies. The equations for this model are given in appendix 22 [48].

Should a defect occur in one of the bearing components then this should be indicated by a growth in the magnitude of the signal at the passage frequency of that particular element. In fact the impacting of the elements in the defect causes harmonics of the element passage frequencies to occur. This coupled with structural modification of the signal often causes higher harmonics from about 5 to 10 times the fundamental frequency, to predominate [7].

Within the usual operating speeds of industrial bearings ($n \leq 3000$ rpm) these element passage frequencies can be covered by frequency ranges below about 1kHz. This can be regarded as low frequency analysis.

However the frequency content of the signal is known to extend beyond these harmonics of the element passage frequencies. The so called ring frequencies have been detected as high as 120kHz [19]. These frequencies are supposedly the resonant response of the bearing and housing structure to transients (ring down). These are measured in acceleration which becomes more sensitive with increasing frequency. Each system has its own characteristic ring frequencies ranging from about 5 kHz to grater than 100 kHz.

It is evident that the excitation source is the impacting of asperities through the lubricant film [7]. Bearings running under poorly lubricated conditions were shown to have a higher ring down response evidenced as two broad peaks in the 5 to 10 kHz range. This also indicated that the excitation was similar to band-limited random noise in this frequency range.

A similar concept is that of shock pulse monitoring. This is said to be based on elastic shock waves emanating from the impacting of asperities on the bearing surfaces. There are a number of commercially available bearing detectors based on the shock pulse method. These systems are designed to respond to particular high frequency ranges corresponding to electrical and mechanical resonances. For example the Shock pulse Meter centered on 32 kHz and the Endevco Incipient Failure Detector centered on about 100 kHz. [18, 49, 50].

As far as the parameters used to define bearing condition are concerned, the most commonly used are the RMS and Peak values of the vibration signal. Less common are the Crest Factor and Kurtosis. The Kurtosis is sometimes used in a number of frequency ranges, usually in 5kHz steps from a few Hertz to 20 kHz. In addition many other parameters have been proposed and tested with varying success [17, 50]. However these are very seldom if ever used in industry.

There have been some attempts to define acceptable levels of RMS and Peak vibration, using velocity. See for example ISO 2372, 2373 and 3945. However these are only used as a guideline until experience has been built up on a particular machine. In addition they are usually only defined over the lower frequency ranges (10Hz-1kHz).

6.2. ANALOG SIGNAL PROCESSING TECHNIQUES

Analog signal processing encompasses all operations performed on a measured signal before it can be digitally processed. Typically this involves the initial measuring by the transducer, some form of amplification followed by low, high or bandpass filtering, further amplification and then analog detection if digital analysis is not to be performed.

This section briefly discusses these processes with particular emphasis on the techniques used in this thesis.

The transducers most commonly used for industrial vibration measurements are accelerometers. That is to say transducers which provide an output voltage proportional to the acceleration of the surface on which they are mounted. Reasons for this are their relative robustness, wide frequency and amplitude dynamic range, accuracy and reliability.

Accelerometers have high impedance low voltage outputs and therefore require some form of charge amplification [51]. The charge amplifier therefore serves the dual purpose of impedance matching with low impedance circuits and amplification to reduce triboelectric and other forms of noise. Usually the charge amplifier will have a rounded output such as 100mV/G where G is the acceleration due to gravity.

Most accelerometers used in industrial applications have built in charge amplifiers. This serves the purpose of amplification close to the voltage source to minimize external electrical interference and obviates the need to select and match amplifiers with transducers. It also adds an important dimension of simplification and robustness to the system.

An important aspect of accelerometers is that they have a non-linear response in certain frequency ranges. More specifically they have a natural resonance at some critical frequency, owing to their construction [51]. This is much the same as a simple mass/spring/damper 1 degree of freedom system.

The useful frequency range of the accelerometer is therefore usually quoted for a 5% or 10% deviation from linearity. Typically this range is about 30% of the resonance frequency. The accelerometer will also be rated at some low cutoff frequency below which the response will be negligible. This is usually of the order of less than 1Hz. Note that the higher the designed resonance frequency the smaller the voltage output of the accelerometer.

In addition to the natural resonance of the accelerometer it will also have a mounted resonance frequency. This will be heavily dependent on the mounting method used and the cleanliness of the mounting surfaces. Under ideal conditions the mounted resonance frequencies shown in table 6.1 can be expected. These values are approximate [51].

TABLE 6.1 MOUNTING METHOD RESONANCE FREQUENCIES

MOUNTING METHOD	RESONANT FREQ.	LINEAR RANG.
Threaded Stud	32 kHz	9 kHz
Beeswax	30 kHz	7 kHz
Epoxy Cement	28 kHz	7 kHz
Magnet	7 kHz	1.5 kHz
Hand Held	2 kHz	400 Hz

The 'raw' signal coming from the accelerometer/charge amp contains information in all frequency ranges, including the effects of resonance as described above. Therefore it is common practise to filter the signal and retain the 'linear' portion. This is usually a bandpass filtering with a low and high cutoff frequency. In some instances the resonance is the required portion and is bandpass filtered and demodulated as in envelope detection.

The important features of a bandpass filters design are the low and high pass frequencies, the roll-off of the skirts and the passband ripple. These concepts are discussed further in appendices 29 to 31.

If the signal is to be transformed to the frequency domain by a discrete fourier transform then filtering is particularly important to prevent aliasing. The concepts of aliasing and the fourier transform are discussed in the following section along with digital signal processing techniques.

6.3. DIGITAL SIGNAL PROCESSING TECHNIQUES

The statistical methods described in section 6.1 are given for a continuous waveform. However with digital analysis it is only possible to work with a discrete version of the waveform. Thus the first stage of digital analysis is the conversion of the analog signal to its digital equivalent.

In the conversion process x(t) will be sampled (quantized along time) to obtain x(nT), where T is the time interval between consecutive samples and n is an integer. Thus x(nT) is a discrete approximation of the original analog signal x(t).

There are a number of precautions to be taken in the process of digitization. Most notable of these is to avoid aliasing. This effect is similar to that of a stroboscope on a rotating shaft. The shaft can be perceived to be rotating at different speeds depending on the strobe flashing frequency. See appendix 49.

To avoid aliasing the sampling must conform to the Nyquist sampling criterion. In essence this states that the original waveform of a sampled wave can only be fully recovered if it was sampled at greater than twice the maximum frequency content of the original wave [append. 49].

The major portion of digital signal analysis in this thesis involves the derivation of statistical moments. For this purpose equation 6.4 is more commonly estimated by a time integral which obviates the need for deriving the p.d.f. Thus the equation for the moment of order k becomes :

$$M_k = \frac{1}{\Gamma} \int_0^{\Gamma} x^k(t)\, dt \qquad \text{.....6.7}$$

Where Γ is the time length of the sample being used to represent the signal.

A discrete equivalent of equation 6.7 is required for approximating the statistical moments using the digital representation x(nT) of the waveform. For the moment of order k, this is seen to be :

$$M_k = \frac{1}{N} \sum_{n=1}^{N} x^k(nT) \qquad \text{.....6.8}$$

Where N is the number of samples taken. Clearly the larger N is the better the approximation of the moment will be. Note that equation 6.8 assumes a stochastic signal.

The other parameters such as Peak and Crest Factor still hold the same meaning. They are also estimated from the digitized waveform. The Peak will be the maximum absolute value of x(nT) which will necessarily be ≤ the Peak of x(t).

Thus the Crest Factor may be estimated by :

$$C.F. = \frac{\text{Max } |x(nT)|}{\sqrt{M_2}} \qquad \text{.....6.9}$$

Interpretation of vibration parameters requires experience from a vibration analyst. Experience in the form of an understanding of machinery dynamics in general and also the behaviour of specific machines being monitored. In addition the meaning and significance of parameters being used to monitor condition must be clearly understood.

Interpretation is usually simplified by trending parameters over time. It is assumed that the initial or baseline values represent a machine in good running order. Where such historical trend information is not available, reliance can be placed on values provided by standards organizations, or the dimensionless parameters such as kurtosis can be used on a once off basis.

Recently much interest has been shown in computer aided diagnosis using artificial intelligence. This is discussed in the next section.

6.4. EXPERT SYSTEMS

The task of analyzing machinery condition can be a repetitive if not difficult task. The machinery condition analyst is presented with large amounts of data which need to be interpreted and converted into useful maintenance information.

It is often tempting to ease the workload by relying heavily on specific parameters. This is compounded by the difficulty of perceiving trends in a group of parameters at once, especially if a simple increase does not necessarily indicate a problem arising.

Therefore it makes sense to attempt some form of computer aided diagnosis. Using a system to capture and apply the experience gained by 'experts' on particular items or components of machinery, a divide and conquer approach could be pursued.

The use of computers to perform reasoning or at least mimic human intelligence is known generically as Artificial Intelligence (A.I.). The particular branch of A.I. where systems are developed to capture human experience and apply it to reasoning is known as Expert Systems (E.S.).

The field of condition monitoring is particularly suited to the application of expert systems. This is because it can be divided into small domains of reasonably clearly defined knowledge. For example a select number of analysis parameters used to diagnose certain types of bearing failure could constitute such a domain. This is an essential requirement for developing an expert system [20].

This section discusses the use of expert systems for condition monitoring. Included is a generalized functional description and specifics on the particular system used for this thesis. Some examples of other successful implementations are also provided.

6.4.1. THE BASIC STRUCTURE OF AN EXPERT SYSTEM

Basically an expert system consists of two essential components, the 'Knowledge base' and the 'Inference engine' [20, 21].

The knowledge base consists of knowledge and experience about a specific and well defined domain of expertise. It is usually obtained from a so called 'domain expert' who is a person experienced in his particular field. The knowledge may be represented in a number of ways, dependent on type and structure.

The Inference engine is a software mechanism designed to reason and infer conclusions about some specific instance within the domain of expertise. It interacts with a 'user' and uses the static knowledge in the Knowledge base to reason and provide conclusions. This interaction is illustrated in figure 6.3.

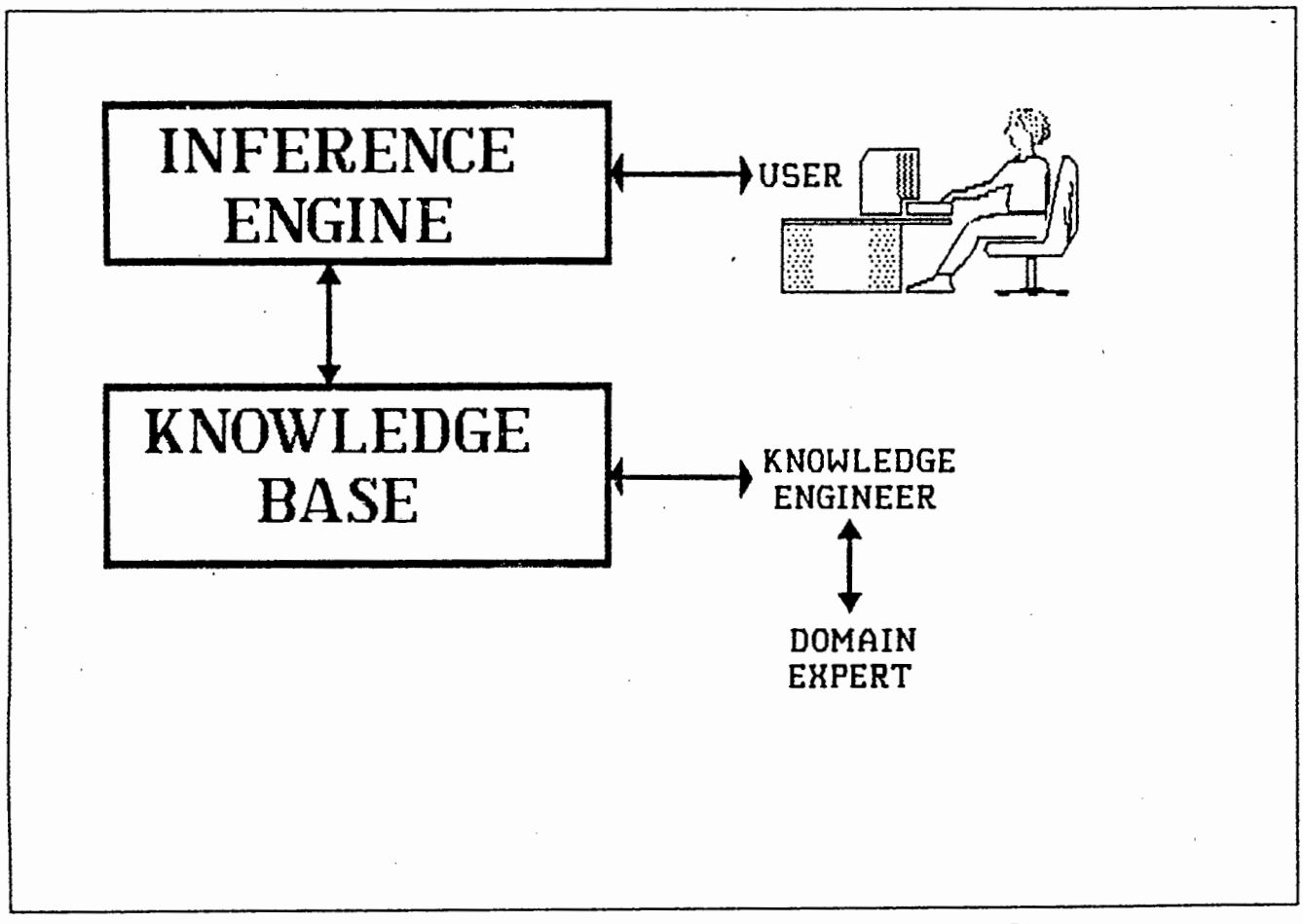

FIGURE 6.3 : SCHEMATIC DIAGRAM OF AN EXPERT SYSTEM

The knowledge base is built up by a 'Knowledge engineer'. This is a person adept at interviewing the Domain Expert and formulating his experience and expertise into a format acceptable to the inference engine.

In particular the methods of knowledge representation and the basic functioning of the Inference mechanism are important. These are discussed in the following sections.

6.4.2. KNOWLEDGE REPRESENTATION

There are a number of different ways of representing knowledge, each depending on the application and type of information at hand. However only those that are specifically applicable to this thesis are discussed.

Firstly a structure is used to contain the knowledge. The structure consists of a combination of selected 'Attributes' describing the domain [20, 21]. Each Attribute has associated with it a specific 'Value' which is a variable, not necessarily numeric. An 'Object' is then defined as a particular instance of the structure in which the Value of each Attribute is specifically instantiated.

The Knowledge is further represented by 'Rules'. A rule is a statement about conclusions that can be drawn from a particular combination of Attributes and their instantiated values. The rules consist of clauses where the Attribute is the subject and the predicate is essentially the Value of the Attribute.

The rules are usually presented in the 'If : Then' format. That is if a certain combination of clauses is found to be true then the related conclusions are also 'true'. Often associated with the rule are certain probabilities that the conclusions are in fact true. The following rule illustrates these concepts.

```
IF:      Clause 1
   and   Clause 2
   and   .....

THEN:    Conclusion 1   (Probability = 56%)
   and   Conclusion 2   (Probability = 84%)
```

There are many possible variations to this rule format. For instance the conclusions may themselves be instantiations of other Attribute Values. In addition the conclusions and even their associated probabilities can also be used to build clauses in the 'IF' part of the rules.

Hence it is possible to build rules about rules, the so called 'Meta-Rules'. These are often used to 'prune' the solution space by applying rules selectively. The Meta-Rules are used to select the rules applicable to the case at hand. Thus layers of knowledge about knowledge can be built up to high levels of abstraction.

The knowledge contained in the rules is usually heuristic in nature. This means that it is derived from the experience, educated guesses and 'rules of thumb' often used by domain experts. In many cases it cannot be proven as hard fact and relies on the art of the expert as opposed to rigorous scientific analysis. In other cases it can be supported by some statistical evidence.

One of the single most powerful aspects of knowledge based systems is that they are designed to be easily modified. Rules can be added, deleted and modified at will, using simple English type constructs and interactive editors. The process of fine tuning a rulebase is continuous, akin to the learning process of a person.

However the Knowledge base is static in that it does not perform reasoning on the information it contains. This is the task of the Inference engine as discussed in the following section.

6.4.3. THE INFERENCE ENGINE

As the name implies, the task of the Inference Engine is to infer conclusions from known information. It is the reasoning centre of the Expert System.

A fundamental task of the Inference Engine is to keep track of what is currently known to be true. In other words the conclusions that have already been drawn and the attributes with their values that have been instantiated.

The Inference Engine then applies what is known to the rules so as to infer new knowledge. This is then added to what is known and the process continues. Of course at the start nothing is known to be currently true and so the Inference Engine may start by consulting the user.

For example it may ask the user to provide values for all the attributes which cannot be inferred from other rules. These values are then used in the rules to instantiate further attribute values until some conclusions are reached. This process is known as 'Forward Chaining' because it starts with the basic data and proceeds forward to the conclusions.

Proceeding in the reverse direction is known as 'Backward Chaining'. In this case the Inference engine has to start by assuming one of the conclusions to be true. It will then attempt to prove this by applying the rules in which the assumed conclusion is inferred.

Thus the rules are reversed through until the attribute values cannot be instantiated by further rules. At this stage the Inference Engine will consult the user to provide the required values.

The process of Backward Chaining is very similar to the way a human expert might perform diagnosis. For example a medical practitioner may assume a patient to have a particular illness. He will then set about trying to prove this by testing the values of various attributes of the patient such as his body temperature, to support his assumption.

If any attribute denies the truth of the assumption then a new assumption will have to be made and tested. This process will continue until a conclusion is reached. Failing all else the final conclusion may be that the illness can not be diagnosed.

However there is clearly a disparity between a human's reasoning and the application of his knowledge to computer reasoning. One attempt to overcome this is the concept of 'fuzzy logic'. This is discussed in the following section.

6.4.4. FUZZY LOGIC AND LINGUISTIC VARIABLES

The process of building a knowledge based system involves an attempt to apply the subjective concepts of a human expert to the exacting demands of computer processing. This seeming incompatibility has spawned the application of so called 'Fuzzy logic'.

As an example the term 'very hot' is a fuzzy and subjective concept. 'Very hot' is also sensitive to the context in which it is being applied. The computer cannot differentiate between 'very hot' weather and 'very hot' fire unless given specific ranges of values.

As discussed earlier this type of knowledge is represented in the Knowledge base as an Attribute-Value combination. The value is defined as a 'variable' but it becomes necessary to distinguish the type of variable. It can be numeric but this does not lend itself well to abstraction from programming.

For this reason 'Linguistic Variables' are used to replace the numeric variables. This is facilitated by means of a fuzzy logic lookup table. A possible lookup table is illustrated in table 6.2.

TABLE 6.2 : AN EXAMPLE OF A FUZZY LOGIC LOOKUP TABLE

TEMPERATURE	LINGUISTIC VALUE
< 25°C	Not Hot
25-30°C	Mildly Hot
30-35°C	Hot
35-40°C	Very Hot
> 40°C	Extremely Hot

Thus the table forms a bridge between the human understanding of a concept and the computer understanding of the same concept.

This type of knowledge and the processes of reasoning about it has to be applied within some framework in the computer. This is discussed in the next section.

6.4.5. EXPERT SYSTEM SHELLS AND PROGRAMMING LANGUAGES

The supporting framework for the expert system is developed using some programming language which the computer can process. There are different techniques used, each constituting some degree of abstraction from computer processing and corresponding ease of implementation.

The commonly used AI (Artificial Intelligence) languages are Prolog and Lisp, while the framework may also be developed using one of the conventional high level compiled or interpreted languages such as 'C', Pascal, Fortran or Basic. However there is an essential difference between programming using the AI 'Heuristic' languages and the conventional languages.

The heuristic languages have been specifically developed for programming using logic (Pro - log) which simulates human knowledge retention and reasoning. As such the rules and facts can be entered in free format without particular cognisance to the order in which they will be processed. The processing is the task of the system itself.

In contrast conventional programming involves the development of algorithms which have a clearly defined and structured flow of operation. The implementation is rigid and the programmer must give full cognisance to the order of processing.

The ease of development and high level of abstraction and flexibility of heuristic programming is played off against the more rapid run time processing of the conventional languages.

However, every computer program, even if it is heuristic, distills down to an algorithm at the machine code level. Therefore it is very tempting to consider a heuristically developed expert system and conclude that it would be far more elegantly implemented if it were developed with a conventional language.

This is not true when it is considered that a Knowledge base is a collection of knowledge needing constant modification, adaptation and fine-tuning. In general the cost and time to modify conventional software is proportional to the overall program size, whereas that for heuristic programs is theoretically proportional to the amount of modification [22].

Providing an even further stage of abstraction are the Expert System Shells. These originated from the development of application-specific expert systems. It was found that if the rules and knowledge were removed then the remaining shell could be adapted for other applications by adding new rules and information [23].

Thus the ES Shell consists of an inference engine and some form of interactive editor for constructing rules and objects of knowledge. Typically they are written in one of the compiled languages such as 'C' and are characterized by ease of use and modification to the Knowledge base.

Due to their popularity, new ES shells regularly appear on the market, each supporting new and varied features. For reasons of availability the shell chosen to be used for this thesis was EXSYS. Some of it's important features are discussed briefly in the next section.

6.4.6. FEATURES OF THE EXSYS EXPERT SYSTEM SHELL

Exsys is an ES shell in which knowledge can be encoded as rules in the typical 'IF-THEN' format described above [24]. It uses backward chaining as the inference mechanism and incorporates the use of calculated probabilities for the conclusions.
In particular there are three different probability modes which can be selected. These are based on three probability ranges ranges as follows :

0 to 1	:	0 = False 1 = True
0 to 10	:	0 = Definitely False (Locked) 10 = Definitely True (Locked) 1-9 = Degrees of Certainty
-100 to 100	:	-100 = Most Probably False 0 = Uncertain +100 = Most Probably True

Exsys has a number of variations on the 'IF-THEN' theme which allow for the building of flexible expert systems.

For example instructions which cause external programs to run, may be included in the 'IF' or 'THEN' part of the rules. This allows external variables to be gathered or new parameters to be calculated using external programs.

In addition a program may be caused to run at the beginning of a session. This is typically used to instantiate the necessary attributes directly without consulting the user. The external program can be used to obtain readings from an instrument and then pass these back to the expert system for interpretation.

Various commands may also be imbedded in the rules to call up other rule bases. This makes it possible to develop rule bases for specific tasks, including the control and application of other rule bases with 'meta-rules'. This divide and conquer approach is powerful for encoding large domains of knowledge into sub domains, thereby keeping the logic more simple.

Essential to the application of the above features is the communication of information between rule bases, external programs and other rule bases. Also needed is some form of overall system control. In the case of Exsys these are provided by two applications as described in the following section.

6.4.7. EXPERT SYSTEM CONTROL AND EXTERNAL COMMUNICATIONS

The process of building and running expert systems which divide knowledge domains into specific sub-domains is known as 'blackboarding' [22, 52]. The rulebases are then said to be in 'object-network' format.

The name 'blackboarding' is derived from the way in which sub-domain expert systems communicate with each other. In a run-time session, a sub-domain system can write its conclusions to a so-called blackboard. These conclusions are then available for any other expert system to consider. Thus the expert systems communicate with each other in a simulated consulting session.

To achieve this communication, Exsys requires two specific items of application software. These are the 'Generic Control Blackboard^{R}' and the 'Generic Data Blackboard^{R}'. They are generic because they are designed to be invoked and used by software in general.

The first of these, the Generic Control Blackboard serves the purpose of interfacing diverse applications. Basically it is a batch processing system acting as an extension to DOS and performing conditional job queueing.

Powerful features include modular development of batch processes, optional queue loop detection and suppression and procedural attachment to expert system rulebases. In other words capture and preprocessing of sensor data becomes a background function.

The second of these, the Generic Data Blackboard serves as a structured interface between expert systems [53]. A specific communication protocol is defined through which small volumes of data may be written to, updated, retrieved by query or deleted from communication files (blackboards). Typical data would be attributes and their values, either instantiated by other rule bases or by sensor data processing software.

Thus a sophisticated diagnostic system could be built using any combination of data collection and preprocessing software, sub-domain knowledge bases developed in object-network format and communication via blackboards. All processing could be transparent to the user with results and conclusions reported as recommendations for maintenance action.

The next section discusses examples of expert systems developed for machinery diagnostics.

6.4.8. EXAMPLES OF EXPERT SYSTEMS DEVELOPED FOR MACHINERY DIAGNOSTICS

Two examples of expert systems developed specifically for machinery diagnostics will be discussed. These are a blackboard expert system developed by Stewart Hughes to monitor helicopter rotor head systems [22] and COMOS, an on-line condition monitoring system for main coolant water pumps and passive primary components in nuclear reactors [54].

The Stewart Hughes rotor monitoring system is a good example of the use of blackboarding techniques. The system was well modularized, with seven separate expert systems each performing their own task.

These included a track fault expert, a pattern matching expert for analysis of vibration harmonics, coupled to a track rod flight expert, a transient test expert, a forward flight expert and a historical expert. Finally the master expert system combined and processed the results of the others and incorporating a fuzzy logic lookup table, provided the final diagnosis.

Each expert system 'fed' off an associated data base holding historical or current captured data. Also included in the structure were various routines for curve fitting and pattern matching. Different types of inference engines were developed for the various experts using Prolog while the data processing routines were developed in either Fortran or Pascal.

An interesting concept in this work was that of a fault symptom matrix. This was a matrix relating typical faults in a system to the kind of symptoms observed in frequency spectra of the related vibrations. This type of matrix was used to build up the expert systems rules.

The COMOS system is mentioned because of some interesting aspects. This system was designed for on-line monitoring in two different modes. The first was for detecting shaft fatigue failure on main cooling water pumps and was continuously on-line. The second monitored reactor pendular vibrations and various associated process parameters leading to the detection of various reactor faults.

Of particular interest in the COMOS system was spectral feature extraction using various forms of pattern recognition and signal classification. In particular various statistical discriminants [54] were used to quantify spectral changes in narrow spectral windows containing resonance peaks or other important spectral information.

These discriminants quantified various aspects of spectral peak changes such as amplitude and frequency, but more interestingly the width and shape of the peaks. If spectral analysis is available then these techniques would provide valuable diagnostic information.

This concludes the literature survey and discussion of relevant theory. The following section deals with directing the project toward the needs of industry.

SECTION 7

DIRECTING THE PROJECT TOWARD THE NEEDS OF INDUSTRY

7. DIRECTING THE PROJECT TOWARD THE NEEDS OF INDUSTRY

To direct the project towards industrial needs, a total of four months were spent working in industry. This consisted of two months on the SANS (South African Nylon Spinners) plant and two months on the Koeberg Nuclear Power Station.

Following is a discussion of the work done during these two periods and how this related to the thesis.

7.1. CONDITION MONITORING ON THE SANS POLYMER PLANT

To start the masters program a two month project was completed at the SANS Polymer Plant in Belville. The project was initiated due to the condition monitoring needs of SANS. This coincided with the author seeking to gain experience on condition monitoring in industry.

The task set by SANS was to investigate the use of condition monitoring on the Polymer and Bottle Polymer plants. The aim was to reduce maintenance costs and lost production. This involved specifying which machinery could benefit from condition monitoring. Also required was a plan of action for implementing condition monitoring on the plant

To achieve these objectives the work covered the following aspects :

- Investigating general preventive and diagnostic maintenance methods applicable to the plant.
- Defining critical machinery on the polymer plant by using flow diagrams, maintenance history, a criticality rating study and scenario planning.
- Assessment of the vibration equipment being used by SANS.
- Successful use of spectral analysis to diagnose various machinery faults, leading to maintenance action.
- Research on monitoring equipment available on the market.
- A financial feasibility study to justify the use of condition monitoring on the plant.
- Recommendations for setting up a vibration monitoring program.

The SANS project provided an initial assessment of industrial condition monitoring needs. This included the type of machinery needing monitoring and the specific condition monitoring methods applicable to different machines and components.

During this time a systematic method of monitoring machinery vibrations was developed [append. 2]. This helped to understand some of the difficulties involved in implementing condition monitoring on an industrial plant.

Finally the SANS library was used to access the CSIR's CSTI (Centre for Scientific and Technical Information). This provided many relevant articles on condition monitoring for the literature survey.

A full report was written on the work completed during this period [ref. 25].

7.2. CONDITION MONITORING ON THE KOEBERG NUCLEAR POWER PLANT

The Koeberg situation was entirely different to that of SANS. Koeberg had a proficient condition monitoring team well versed in vibration measurements and the use of spectrum analyzers. In particular the Palomar Micrologⁿ data collector was used extensively for routine vibration monitoring on the plant.

The main objective of this period was to observe a well set up condition monitoring system in operation and to find the state-of-the-art in data logging systems and off-line condition monitoring. It was a time to critically analyze and assess the confidence of diagnosis being provided by such a system.

The method of vibration monitoring was as follows : A route was set up on the computer including all the measuring points on every monitored machine within a certain section of plant. The route was then downloaded to the data logger and used to collect an RMS velocity level for each point.

Over a period of time an `acceptable' level of vibration was established as a baseline for each point. This was then used as a guideline to set two alarm levels, one representing `caution' and the other `take action'. Essentially these alarm levels and the time interval between successive measurements were subjectively set by experience.

In addition it was possible to automatically collect a 400 line spectrum if either of the alarms was exceeded. The frequency range of the spectrum would be predetermined. Because of the data logger memory space consumed by the spectra this feature was usually only used if overall RMS levels were exceeded.

After collection the RMS values would be downloaded to the computer which would then generate an `exception report' of all alarm levels exceeded. In addition the history of the RMS levels could be displayed as a `Time vs RMS Amplitude' plot. In this plot it was apparent if there were any upward trends in the overall levels. Thus paramount importance was placed on the RMS level of vibration to detect any potential failures.

On occasion when the RMS levels seem to indicate an upward trend, a dedicated spectrum analyzer was used to do more in-depth analysis of the vibration patterns. However these spectra usually did not have a baseline for comparison and tended to be used with an element of conjecture.

If spectra were analyzed then these were usually in the lower frequency ranges to a maximum of about 2 kHz, but more usually to about 500 Hz. In addition hand held accelerometers were usually used. These would provide a useful frequency range of about 500 Hz with some variability due to pressure applied and point of application.

Thus the RMS levels would be collected over the same frequency range. This would ignore the valuable information contained in the higher frequency ranges [7], with corresponding loss of detection.

A personal evaluation of the system was as follows : In general the system had had a reasonable amount of success with failure **detection**. This was often the case where damage was already significantly advanced. However as far as **diagnosis** and **prognosis** was concerned there was relatively little confidence expressed.

These inadequacies were ascribed to the following factors : Firstly the parameter RMS Velocity was being measured in a frequency range which would show relatively little increase for the early stages of common failures such as bearing damage. In addition too much reliance was placed on this parameter, thereby ignoring other patterns in the vibrations such as `impulsiveness'.

Secondly, attempts at diagnosis and prognosis were affected by a lack of in-depth knowledge of the failure patterns of the machinery concerned. This included the lack of use of combined techniques to confirm diagnosis.

The time spent at Koeberg was useful to confirm the direction of research of the project, namely that expert systems could be used to assist machiney condition analysis by implementing combined analysis techniques and assisting in their interpretation.

The next section deals with the proposed framework of such an expert system.

SECTION 8

DEVELOPING THE FRAMEWORK OF THE CONDITION MONITORING SYSTEM

8. DEVELOPING THE FRAMEWORK OF THE CONDITION MONITORING SYSTEM

The first stage of the project was to sketch the basic framework for a computerized, expert system based condition monitoring system. The emphasis was placed on industry in general, including the most commonly found types of machinery. The bulk of the work for this dissertation was then to develop a specific aspect of the system, namely that for bearings.

The framework was to be used as the basis for a long term project involving the Mechanical and Electrical departments and possibly also the Computer Science department at the University. The long term goal was a saleable product with the possibility of production for industry.

A two-way relationship between the project and industry was proposed. Interested parties in industry would provide research funds in exchange for the expertise to implement condition monitoring and partake of the developing technology. This would include condition monitoring seminars and assistance with the setting up of condition monitoring systems on related plant.

This section discusses the proposed framework of the system including the required computer hardware and software. Also included is a summary of machinery needing further research so as to be included in the condition monitoring system.

8.1. DESCRIPTION OF THE SYSTEM

It was proposed that the diagnostic system be based on a lightweight ruggedized portable computer. This would enable a flexible system to be built, incorporating the advantages of digital signal processing and expert systems.

An overview of the system is illustrated in figure 8.1

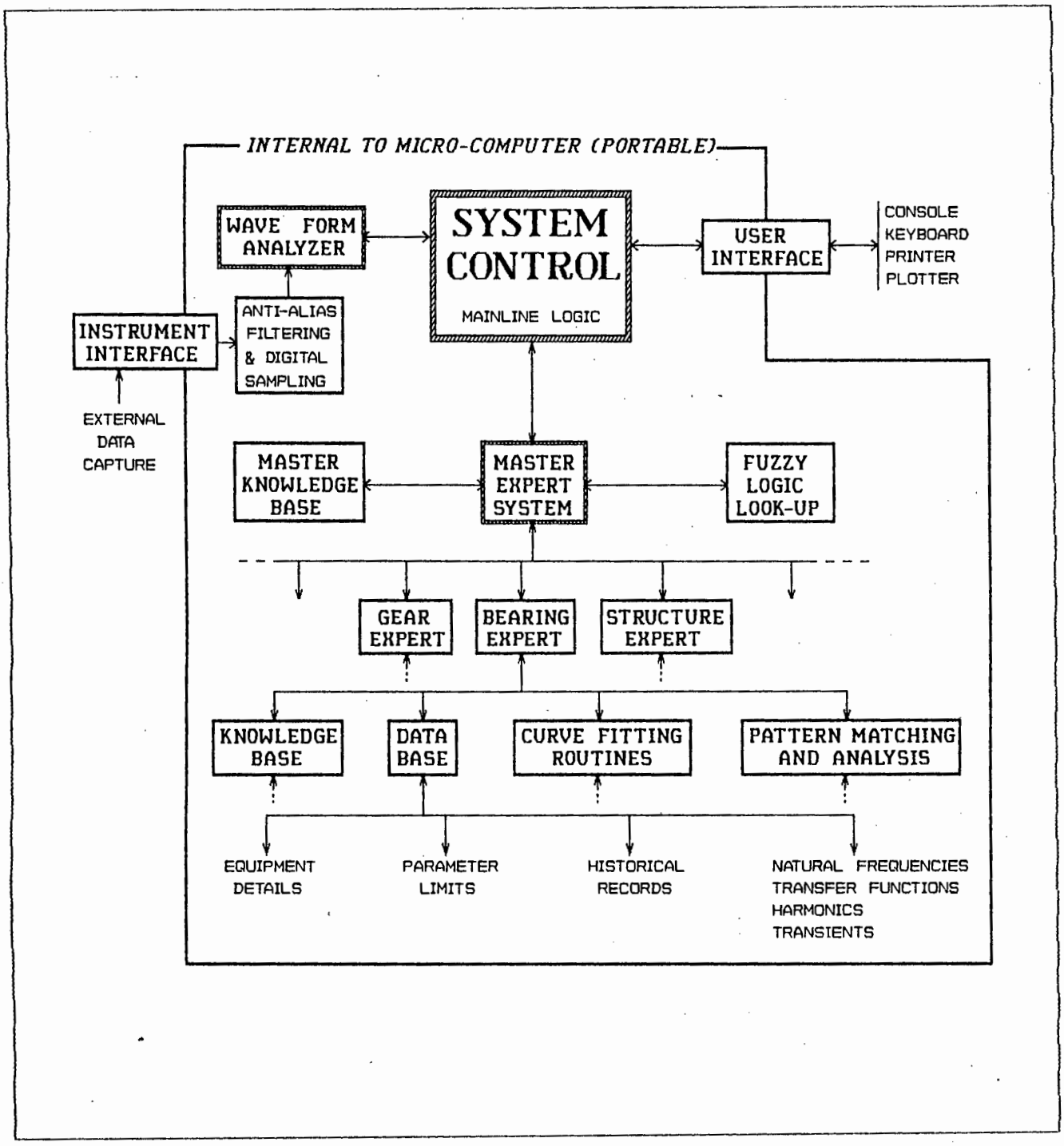

FIGURE 8.1: SCHEMATIC OF THE CONDITION MONITORING SYSTEM

The functioning of the system would be as follows. Firstly it would have an external interface for attaching transducers, shaft encoders and any other necessary instrumentation. Typically the interface would be through universal connectors such as BNC plugs and would serve to measure analog voltages.

These analog signals would then be filtered to prevent aliasing, amplified if necessary and then digitized. This would happen internally to the computer. The digital signals would then be processed by a dedicated on-board waveform analyzer. Such processes as digital filtering, discrete Fourier transforms, integration and differentiation and other transforms would would be carried out by this system. This processing would proceed as a background process.

The mainline logic would be maintained by the system control center. This would be the software that decided which processes were currently active. Further tasks would be the handling of requests and interfacing between all active systems, including the user interface.

A typical session could be visualized as follows. The user would instruct the system to diagnose a particular machine. The system would then call up a database containing information on this particular machine. Such information would include the components in the machine along with their details and the monitoring points and transducers used. The user would then be instructed what transducers to use and where to place them.

Control would then be passed to the master expert system which would in turn control the individual experts. Each relevant expert would then be given a chance to perform its first level analysis. The active expert would consult its data base for the types of analysis to be performed on this component. Through this a request would be generated for particular parameters to be collected for analysis.

These parameters could consist of any combination of process variables, operating conditions, vibrations and other condition monitoring techniques. They would be collected through the instrument interface, processed by the waveform analyzer and prepared by the control system to be passed to the expert.

Typically this first level analysis would constitute a detection phase of relatively simple parameters by each expert. Each would then provide the results of its analysis to the master expert. The master would then decide if a fault was suspected and if the second level of analysis, namely detection, should be applied.

Each expert would then receive its turn to request more parameters and perform a diagnosis, asserting a probability or confidence level that the suspected fault was being caused by its component. The master expert would then decide which component(s) were failing and allow the expert of that component to perform a prognosis. Thus recommendations could be made to the user on what maintenance action to take.

The individual experts would perform their analyses using the knowledge base (rules etc.), curve fitting routines, pattern matching and the data base with its parameter limits, historical records and other information. The master expert would rely on its knowledge base and fuzzy logic lookup tables to provide the diagnosis, prognosis and recommendations in a format easily understood by the user.

In addition the various levels of analysis could be suppressed for any session. For example it might be required to perform a first level detection on the whole machine and then come back for a diagnosis and prognosis if required. Thus a large quantity of equipment could be scanned and only analyzed in detail if required.

There would be a number of advantages to having the diagnosis and prognosis performed at the machine. Firstly, if a fault was suspected, this could be confirmed by more measurements at that time. Secondly the expert could also enquire of the user as to visual problems such as leaking seals, oil levels or the proper functioning of some support system such as coolant circulation. In addition the system would be self diagnosing of transducer faults, requesting them to be corrected by the user.

A notable aspect of the system would be a minimum of consultation with the user. The system would first attempt to resolve any problems by direct measurements on the machine. Only relatively simple questions would be asked if and when required.

Clearly the system would require a sophisticated platform of hardware and software to perform the type of analysis required. Therefore it was necessary to do an initial investigation on the type of equipment available for the task. Research on hardware and software are discussed in the next section.

8.2. RESEARCH ON HARDWARE AND SOFTWARE

The hardware envisaged for the system consisted of a portable or 'luggable' microcomputer with expansion slots to accept digital signal processing and waveform analysis cards. It was evident that this kind of hardware was developing at a rapid pace.

For example in the Stewart Hughes helicopter monitoring system [sect 6 & ref 22] the use of 32 bit processors was discussed as a future possibility. Requirements cited were a large amount of memory (> 1MByte) and rapid processing. These had now become freely available on the market with the possibility of expanded memory up to 16 Mbytes and 32 bit processing [55].

As far as the waveform analysis and digital signal processing cards were concerned, there were some available for the kind of processing required. One such system demonstrated was the CSI Wavepak® [56]. This consisted of two expansion cards compatible with IBM® microcomputers. They were capable of a wide range of signal processing tasks including anti alias filtering and Fast Fourier Transforms.

Included with the Wavepak expansion cards was a set of software to drive them. This included all aspects of data capture, processing and display, data base maintenance and also for building up routes of monitoring points. Using such a system would obviate the need for developing equivalent software.

However the Wavepak system did not have all of the required analysis tools. Such aspects as Cepstral analysis and statistical parameters were not features offered. In addition the system was based on an 8 Bit bussing architecture which could not make full use of the speed of a 32 bit bus. This would probably be a future development.

Other future developments to be aware of would be the use of transputers and parallel processing. This would enable the simultaneous processing of multiple tasks, thereby greatly enhancing the speed of the system.

Other software required included a high level programming language for the overall control and a heuristic programming language or expert system shell. These were discussed in section 6.

The following section covers the type of machinery needing research.

8.3. MACHINERY NEEDING RESEARCH

In the industrial work completed [sect. 7], it was found that the larger portion of industrial machinery needing monitoring could be represented by all or part of a generic machine. This is shown schematically in figure 8.2.

The machine consisted of a motor coupled with a flexible or rigid coupling to some form of power transmission and finally to some form of impeller doing mechanical work. The transmission would be achieved by gears, belts and other systems while the impellers could be pumps and fans or even winders, etc. In addition the whole system was supported by various types of bearings.

For example a water pump could consist of the motor and belt driven impeller without the gears, while a mill might have all of these components.

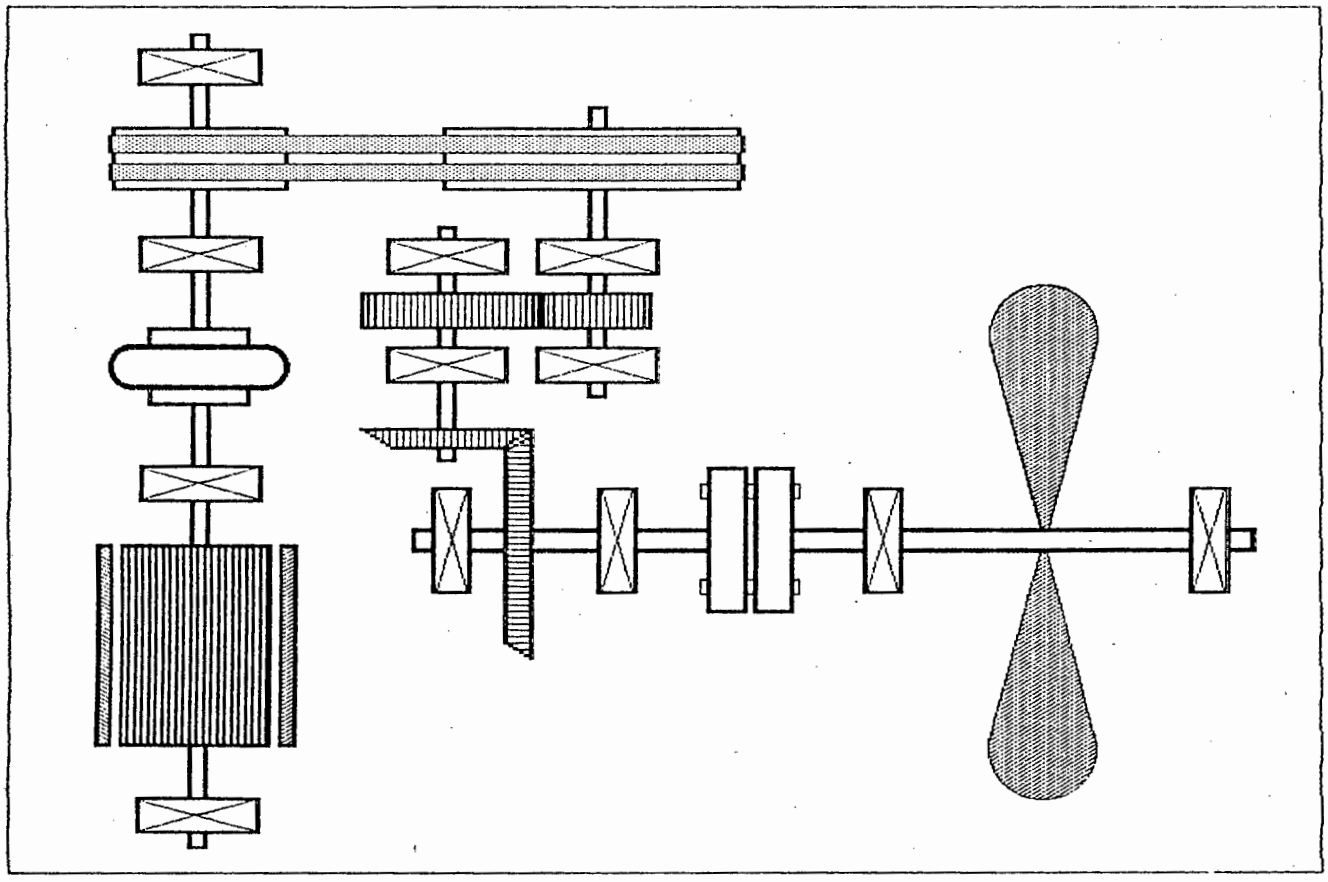

FIGURE 8.2: SCHEMATIC ILLUSTRATION OF A TYPICAL INDUSTRIAL MACHINE

To develop a system for monitoring this type of equipment would mean research on each component. This would involve isolating the component in a laboratory environment so that the vibration patterns and their causes could be more clearly defined. For most of the equipment this would include simulating typical industrial failure modes while monitoring all relevant condition monitoring parameters.

The concept of isolating machinery and simulating faults was carried through by proposing a number of research projects for undergraduates in the department of mechanical engineering. These involved the study of cavitation in pumps [57], further study on bearings [8] and gear vibrations [58].

Figure 8.3 Shows a hierarchy of the typical categories of machinery to be included in research.

The following section discusses the development of a test rig to monitor bearing condition.

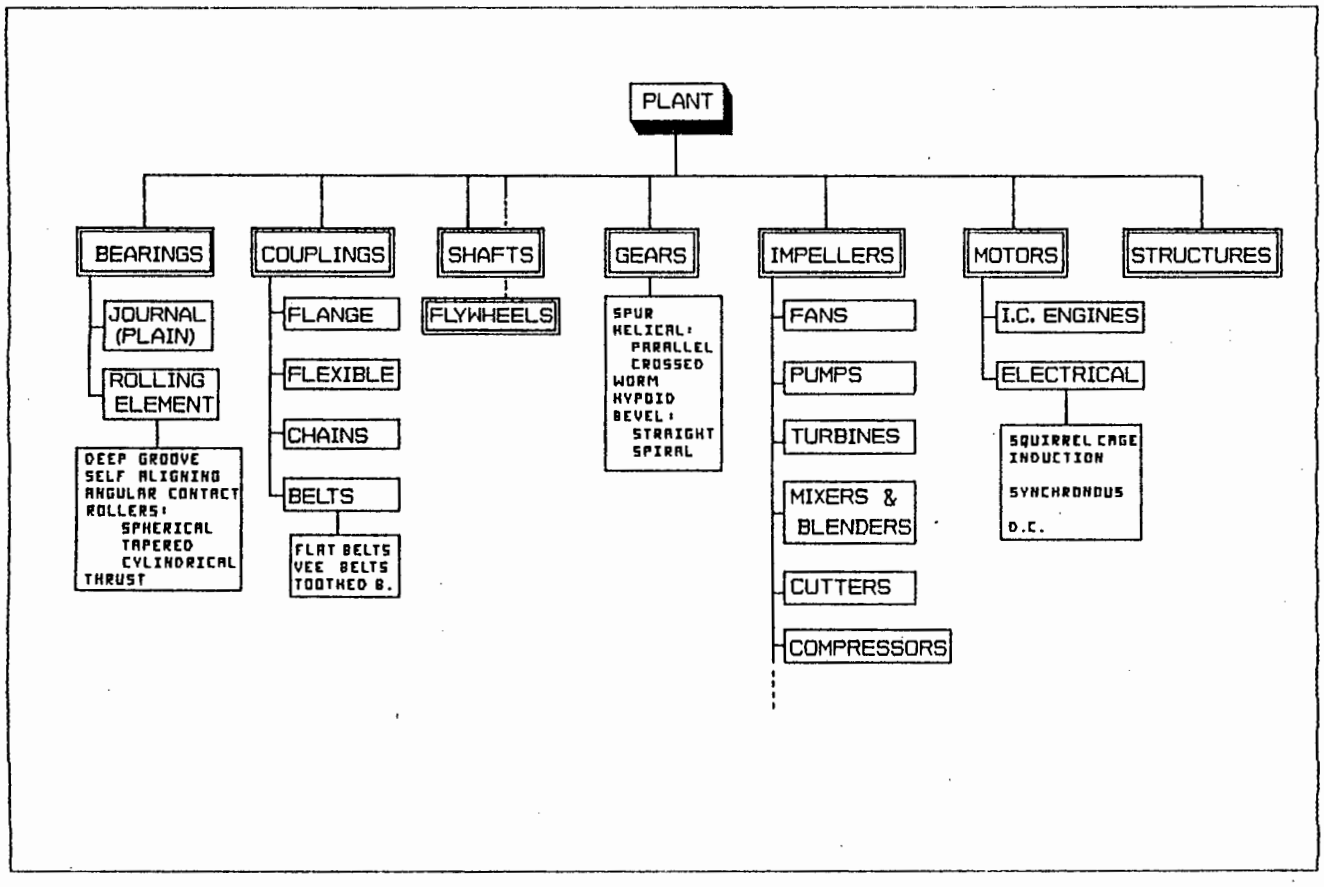

FIGURE 8.3: HIERARCHY OF MACHINERY NEEDING RESEARCH

SECTION 9

DESIGN OF THE BEARING TEST RIG

9. DESIGN OF THE BEARING TEST RIG

The bearing test rig was designed to enable the systematic destruction of a rolling element bearing by applying an excessive load. This presented a wide variety of physical configurations and thus specific choices had to be made.

Each aspect of the test rig design is discussed separately in the sections that follow, with justification given for the decisions made.

9.1. GENERAL DESCRIPTION OF THE TEST RIG

The test rig consisted of a channel iron frame with a dc variable speed motor, three bearings on a shaft aligned axially with the motor and a hydraulic loading mechanism. This arrangement is shown in figure 9.1

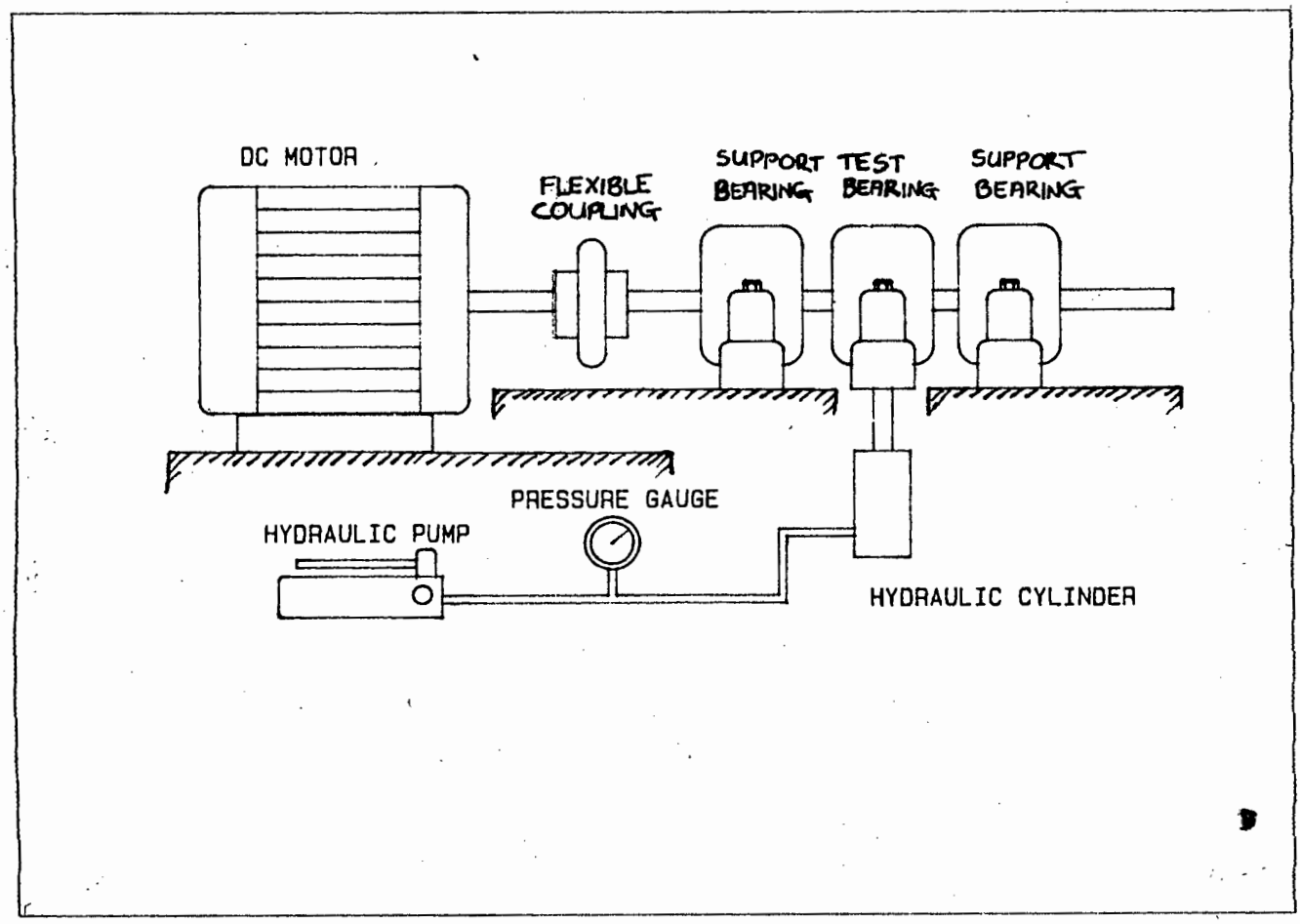

FIGURE 9.1 : SCHEMATIC DIAGRAM OF THE BEARING TEST RIG

The central test bearing could be radially and unidirectionally loaded to a maximum of approximately 1.6 times its dynamic load rating. This was achieved by a hydraulic cylinder acting on a hinged lever arm. The test bearing was axially located while the two outer support bearings were axially floating to allow for thermal expansion.

9.2. THE BEARING CONFIGURATION

On selecting bearings to be tested and used for support, consideration was given to the following aspects:

- Dynamic load rating
- Mounting methods for shaft and housing
- Ease of inspection
- Misalignment capabilities
- Widespread use in industry

For the test bearing a 'self-aligning' bearing was chosen for its misalignment capabilities, ease of mounting using tapered adapter sleeves, ease of inspection without destruction and relatively low dynamic load rating as compared to say spherical roller bearings.

For the outer support bearings, spherical roller bearings were chosen for their relatively high dynamic load rating, ease of mounting using tapered adapter sleeves and misalignment capabilities.

The test bearing was centrally mounted between the two support bearings so that it would carry twice the load of the support bearings when radially loaded [figure 9.2]. In addition the dynamic load rating of the support bearings was approximately four times that of the test bearing.

This meant that if the test bearing was carrying its full load rating the support bearings would only be running at 12% of full load. Therefore the support bearings would easily outlast the test bearings. They would also provide relatively low vibration levels and thus prevent undue contamination of the test bearing vibrations.

All of the bearings were mounted in plummer blocks and sealed with 'V-ring' seals at the shaft. This system allowed easy acess to the bearings for inspection and replacement and also the application of vibration and temperature sensors.

Appendix 3 gives the full specifications of the bearings and related equipment.

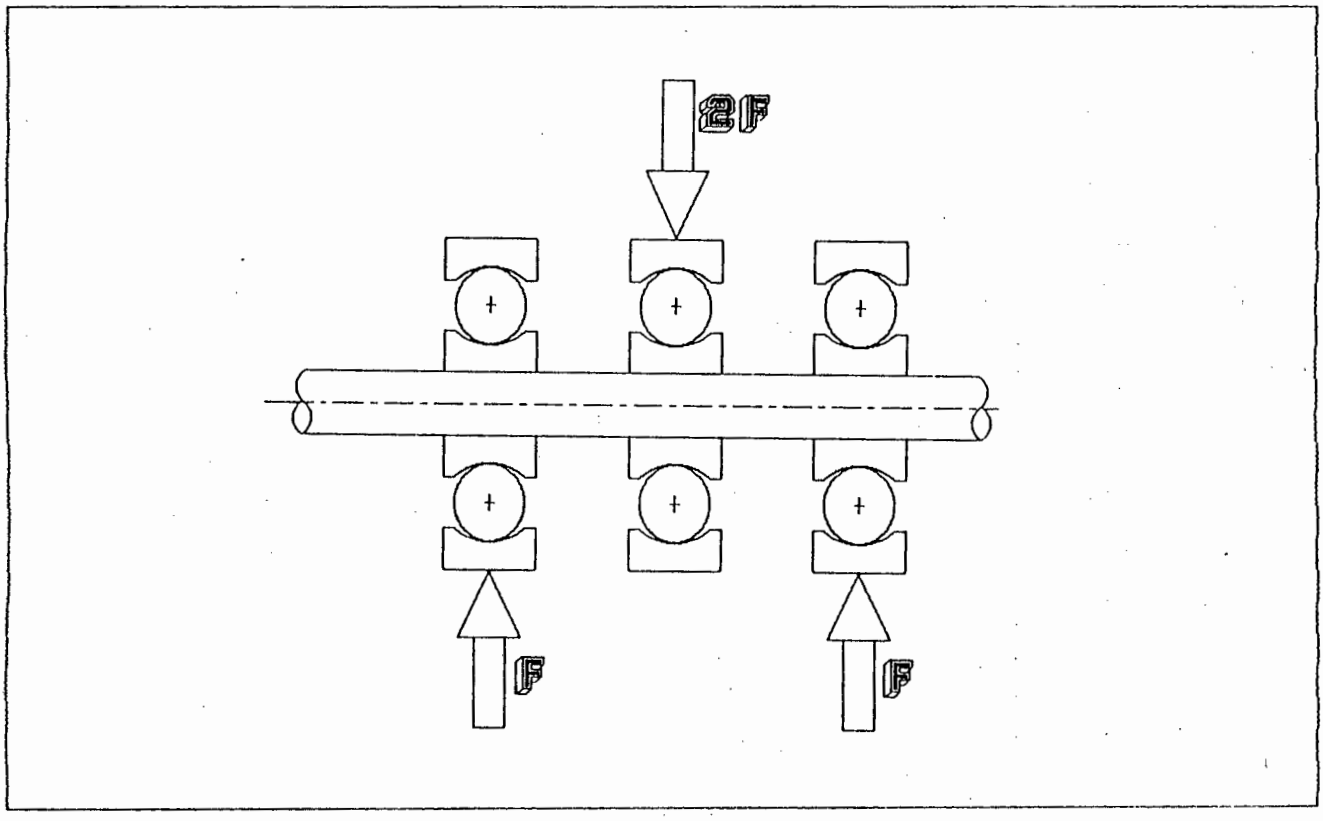

FIGURE 9.2 : THE BEARING LOADING ARRANGEMENT

9.3. THE LOADING SYSTEM

The loading system was required to provide a radial load on the central test bearing well in excess of its dynamic load rating. The selected test bearing was rated at 15·9 kN [13] which was regarded as a conservative estimate [26].

The maximum applicable load was selected as 25 kN or approximately 60% greater than the load rating. From previous experience [7] this maximum load would fail the bearing within a few hours.

In addition the load would have to be easily applied and easily varied according to the requirements of testing. This would also have to incorporate a limit on overloading to prevent damage to the shaft.

A further requirement of the loading system was that it should not misalign the bearing by more than the rated angle [append. 3]. Bad misalignment could lead to a different type of failure such as balls being ejected from the cage.

All of these requirements were fulfilled by the design shown in figure 9.3.

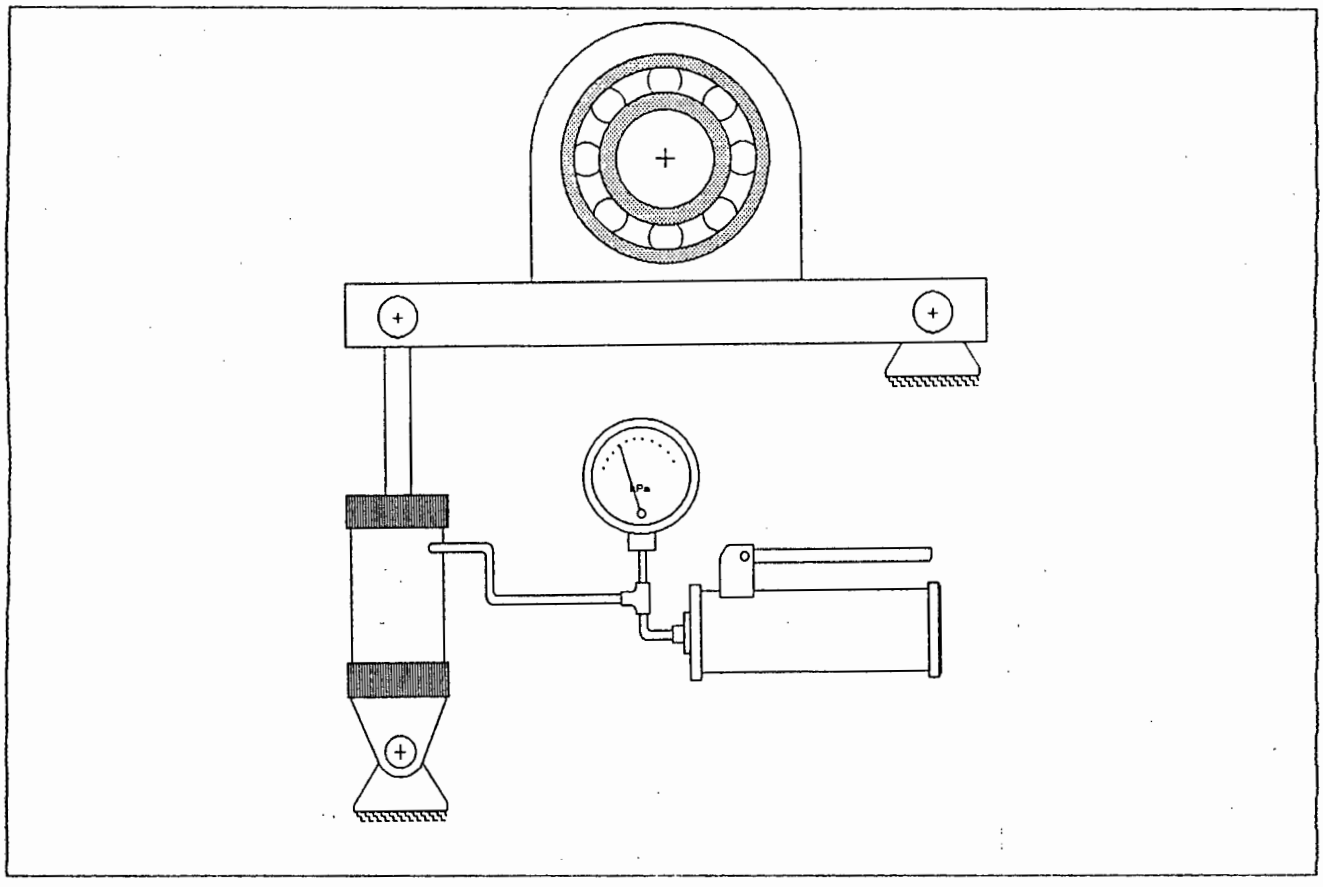

FIGURE 9.3 : SCHEMATIC DIAGRAM OF THE BEARING LOADING MECHANISM

The loading system was operated by applying oil pressure into the hydraulic cylinder which then loaded the bearing through the lever arm. The load on the test bearing could be infinitely varied between 0 and 25kN.

The load was increased by pumping or reduced by opening the return valve. The calibrated pressure gauge served as an overload protector by giving an indication of the magnitude of the applied load.

9.4. THE DRIVE MOTOR

It was considered desirable to test vibration monitoring parameters at varying speeds. This required the use of a variable speed motor. A 4kW dc drive was available to be used for the task. This incorporated a Saftronics single phase regenerative controller using a thyristor bridge to provide variable power.

The system featured powering in all four quadrants as well as a closed loop feedback system for speed control. The speed could be varied by adjusting a variable reference resistor in the control loop. The minimum speed was set for 116 rpm and the maximum for 1935 rpm to cover the greater part of the 100 to 2000 rpm range.

The power demand on the motor was considerably less than 4 kW. All that was required was to overcome the friction in a failing bearing under full load. Thus the motor was running lightly loaded and the closed loop feedback system tended to be unstable. This produced vibration noise as the control system was `hunting' to find the correct speed.

The Saftronics controller featured adjustable torque settings as well as an adjustable response time for the feedback system. Thus the system could be tuned for the specific application. Once properly set, the motor ran very smoothly with vibration levels almost unmeasurable at the bearing pedestal.

9.5. THE SHAFT AND COUPLING

The shaft was recognized as a critical component in the design of the test rig. This had been confirmed by previous testing [7]. Firstly the diameter was restricted to 30mm by the inner race of the bearing and secondly the applied moment could not be reduced due to the bearing spacing being determined by the plummer blocks.

It was calculated that the shaft would be subject to high cycle fatigue conditions ($> 10^6$ cycles) with full load stresses in excess of 200 MPa [append. 4]. Therefore a shaft with an ultimate tensile stress in excess of 2000 MPa on the surface was required [append. 4]. This was achieved by case hardening an M120 steel to a hardness of 57 Rc [27].

In addition the shaft and bearing arrangement was designed so as to avoid stress-raisers as far as possible. For example there were no keyways or steps on the shaft and the bearings were mounted using tapered adapter sleeves.

For the coupling between the shaft and motor a Fenaflex® rubber coupling with Taper Lock® adapters was used. This allowed for easy removal of the shaft for bearing replacement. In addition the rubber coupling ensured that a minimum of vibration was transmitted from the motor to the shaft.

The motor was mounted on a separate baseplate attached to the frame by threaded fasteners. These allowed for easy alignment of the bearing and motor shafts and only transmitted low levels of vibration to the rest of the rig.

9.6. THE TEST RIG FRAME

The test rig frame was built using various sections of channel iron butt welded into a rectangular frame. The frame was mounted on rubber mats to lessen external vibrations. All components were bolted to the frame.

The test rig was set up with all equipment except for the central test bearing. The motor was then run at various speeds while measuring the vibrations on the test bearing pedestal. It was found that the vibrations at this point were negligible even when other machinery in the laboratory was being used. Thus the vibrations measured on the test bearing would not be significantly contaminated by external sources.

Attached to the test rig was a system to capture and process various condition monitoring parameters. The design and calibration of this system is discussed in the following section.

SECTION 10

DESIGN AND CALIBRATION OF THE DATA CAPTURE AND PROCESSING SYSTEM

10. DESIGN AND CALIBRATION OF THE DATA CAPTURE AND PROCESSING SYSTEM

To achieve a fully automated expert system to diagnose machinery vibrations required a data capture and processing system based on digital analysis of various mechanical parameters. This involved the interfacing of analog and digital systems.

In particular the parameters shaft speed, bearing load, bearing temperature and bearing vibrations had to be accurately measured and calibrated. These were then digitized and processed in a form that the expert system could interpret. This chapter details the components and processes used to measure and calibrate these parameters.

Figure 10.1 schematically illustrates the test setup showing the relationship between the various systems.

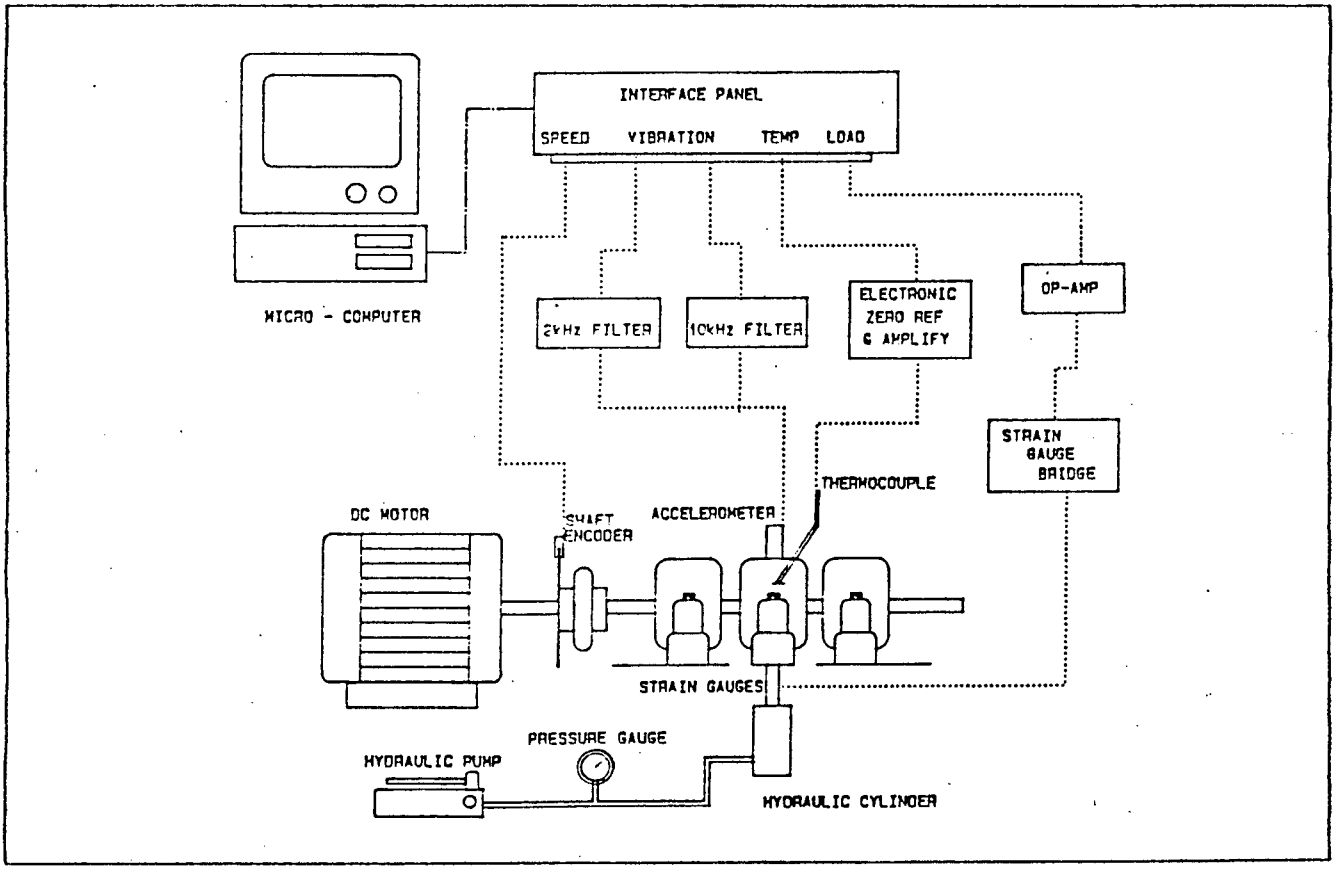

FIGURE 10.1 : SCHEMATIC ILLUSTRATION OF THE TEST RIG SETUP

10.1. GENERAL DESCRIPTION OF THE TEST RIG SETUP

The system was designed to enable the capture and digital analysis of all parameters relating to the condition of the test bearing. This was achieved by using analog transducers to convert the parameters to a voltage.

These voltages were then filtered, amplified, digitized and bussed into the computer for further digital processing and interpretation. The digitization was achieved by means of an analog to digital conversion card mounted in the microcomputer.

The sections that follow describe the individual components of the test rig setup as shown in figure 10.1.

10.2. THE MICROCOMPUTER

The microcomputer used was a Bondwell XT which was IBM® PC XT compatible. In other words it used the equivalent of an Intel® 8088 CPU (Central Processing Unit). Thus the whole computer system centered around the 8-bit architecture of this chip.

Very importantly the computer was fitted with the optional 8087 maths coprocessor chip. This chip handled the floating point operations separately from the CPU. The large amounts of floating point calculations needed for the real time processing made this chip essential.

The additional peripherals on the Bondwell were a 20 MByte hard drive, one 320kByte floppy drive, a Hercules graphics card and an Epson MX 80 dot matrix printer. This hard drive proved to be of insufficient size to store enough time or frequency records for a full bearing test.

The Bondwell also had 5 internal expansion slots, each with 8-bit wide busses. One of these was used to mount the analog to digital conversion card.

10.3. THE ANALOG TO DIGITAL CONVERSION CARD AND INTERFACE PANEL

The analog to digital conversion was achieved by an A to D (Analog to Digital) card. This was a key component in the system because it enabled the digitization of the voltages produced by the various transducers. The card used was the DT2801-A made by DATA TRANSLATION and was mounted in one of the internal expansion slots of the Bondwell microcomputer.

The card was connected to a DT707 interface panel by means of a 20 strand ribbon cable. The interface panel had standard BNC connections for 8 differential analog inputs. In addition there was an input for an external trigger and an external clock pulse.

The card was calibrated for the various input ranges according to the procedure laid out in the manual [28]. It was therefore operating within specifications.

The whole data capture system had to be designed around the specifications of the card. These are listed fully in appendix 5. The more important aspects of these specifications and their implications on the test setup are discussed as follows.

Firstly, a digitized version of a signal was only a discrete representation of the analog version. Hence the analog to digital conversion limited the resolution with which the signal could be characterized digitally. In particular this card performed a 12-bit A to D conversion.

A 12-bit conversion meant that only $2^{12} = 4096$ discrete values of signal amplitude could be represented. This meant that the card had a maximum dynamic range of 72 dB [append. 6]. The digitized values from the card were presented as integers between 0 and 4095. These integers could be converted to a voltage using FSR (Full scale Range) in the following equation [append. 7].

$$\text{Output Voltage} = \frac{\text{FSR} * (\text{A/D Integer} - 2048)}{4096}$$

In addition the board had to be pre-configured to either Unipolar (0 to 10V) or Bipolar (-10 to +10V) ranges. In other words if any of the signals were to be represented by a voltage varying from negative to positive then a global Bipolar range would have to be selected.

This was indeed the case with the vibration amplitude. Selecting this global bipolar range meant that the unipolar signals such as temperature could only be represented with a maximum of $4096 \div 2 = 2048$ discrete values, or a dynamic range of 66 dB.

Therefore to use the dynamic range to its fullest extent the incoming signals would have to be matched with voltages accepted by the card. To assist in this the the card had four software adjustable gains to achieve the following voltage ranges: ±10 Volts, ±5 Volts, ±2.5 Volts and ±1.25 Volts. In some instances the signals had to be pre-amplified to adjust them to these ranges.

Another very important specification was that of sampling frequency. The card had a software configurable sampling interval which could be increased in multiples of 1·25 microseconds. The manufacturers advised that at least 29 * 1·25 microsecond intervals be used because of the clock pulses needed to perform the A to D conversion . This corresponded to a maximum sampling frequency of 27·586 kHz [append. 8].

An advanced feature of the card was that it could make use of DMA (Direct Memory Access). This meant that the digitized values could be loaded directly from the card buffer into the DRAM (Direct Random Access Memory) under the control of the card.

In other words the capture of data bypassed and therefore was not limited by asynchronous and relatively slow communications between the card and the CPU (Central Processing Unit) of the microcomputer. The card could be given control of DRAM by taking over the DMA controller chip. This DMA feature was essential for achieving the highest digitization rates.

However the DMA feature and indeed all communications between the card and the microcomputer had to be programmed using advanced techniques. This required the selection of a powerful computing language which could be used for this type of programming. This is discussed in the following section.

10.4. THE TEST SYSTEM CONTROL SOFTWARE

Developing an automated bearing condition assessing system required the development of software using advanced computer programming techniques. Therefore it was necessary to select a powerful high-level language.

This section describes how a suitable language was selected and also gives a brief overview of how the test software functioned.

10.4.1. SELECTING A SUITABLE PROGRAMMING LANGUAGE

A "real time" data capture and processing system was needed for the rapid processing of large amounts of data. For example calculating the bearing related parameters typically involved in the region of 15 000 data points [append. 9]

To capture and process this amount of data in real time was an impossibility using for example the interpreted BASIC programming language. Previous experience [7] showed that processing relatively small amounts of data took up to 5 minutes with BASIC, a far cry from the few seconds needed for real time processing.

In addition the involved process of DMA to achieve high sampling frequencies and similar "machine level" functions compounded the need for an advanced compiled programming language.

The compiled languages available for the IBM compatible computer consisted of TRUE BASIC, QUICK BASIC, FORTRAN, PASCAL and "C". The BASIC languages were not considered because of their inherent lack of advanced features.

As a postgraduate course in FORTRAN had been attended [29] it was decided to use this language. However it was subsequently discovered that FORTRAN was totally unsuited for this type of programming because it had no commands for machine level communication.

The choice was then between PASCAL and C. A comparison of the two languages revealed that C was eminently suited for this type of programming and included advanced memory management features and the ability to handle large data arrays. This was not a standard feature of PASCAL and could only be achieved with nonstandard techniques. In addition PASCAL was found to be far more rigid in structure than C [30]. Thus C was chosen as the language to be used.

To become proficient in a new language would involve a long learning curve. However it was decided that the gain in power and flexibility of programming would outweigh this incubation period. In retrospect this proved to be the case.

In addition the C language was gaining in popularity at the university and was becoming the language of choice for instruction of computer science undergraduates [31]. Thus the long term project to develop a full expert system diagnostics package could more easily include the computer science department.

10.4.2. AN OVERVIEW OF THE TEST SYSTEM CONTROL SOFTWARE

The objective of the test system control software was to manage the systematic digital capture and processing of research data from the test rig. This was then presented to the expert system for diagnosis. The functions included communication and control of the various computer hardware components, memory management, data file handling and data processing.

The test system control software was coded in Borland's "Turbo C®" computer language. The modularity of C enabled the software to be written in groups of reusable functions (modules). This modularity was considered good programming practice [29].

These function groups were classified as the main control program, high level functions, intermediate level functions and machine level functions. The levels and functions are shown diagrammatically in figure 10.2.

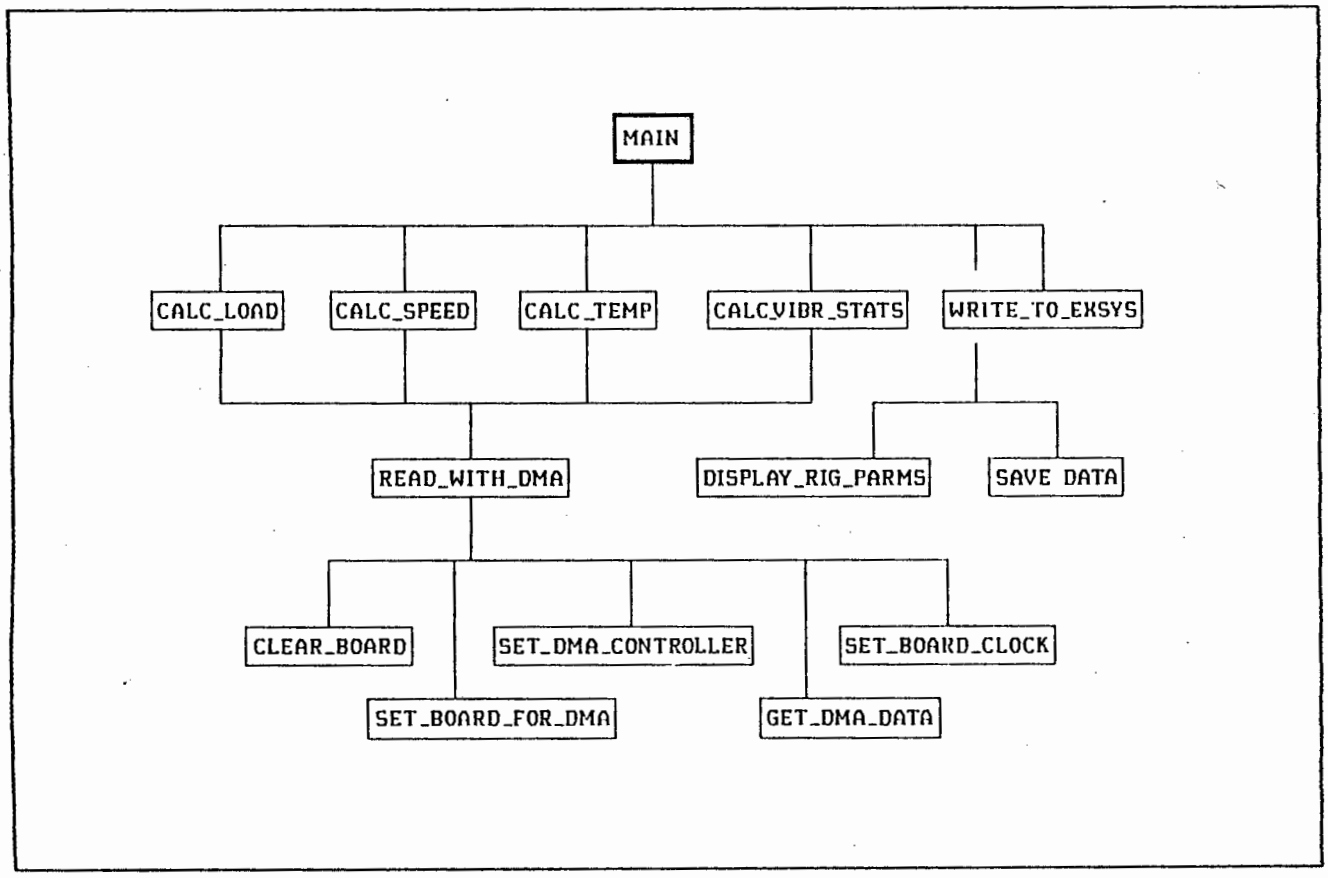

FIGURE 10.2 : HIERARCHY OF THE CONTROL SOFTWARE FUNCTIONS AND LEVELS

Each of these groups were compiled separately to make libraries of reusable functions which could in effect be seen as new commands. Thus the increasing levels of function groups provided further abstraction from the "machine level".

In this way programming was at increasingly "high" levels becoming easier to read and understand. This greatly simplified the software development process and also provided "self documenting" of the software.

The functions were grouped into levels according to how they were accessed by the main control program and the other functions. In other words the machine level functions were accessed by the intermediate level functions and so on.

At the machine level the functions all performed operations which involved the access of hardware registers and communication ports. These included such functions as bit-wise and byte-wise handling of variables and commands and the interpretation of status flags.

The machine level functions were required because no control software was provided with the A to D card. However the communication protocols, command, status and DMA register addresses as well as the necessary commands were all provided in the card's manual [28].

The DMA process was greatly enhanced by C's inherent ability to provide the DRAM location addresses of array variables. In other words the digitized values could be loaded directly from the card to the relevant variable array. These could then be accessed without having to be relocated in memory.

Incidentally this DMA process highlighted the limitations of the 20-bit address word with 8-bit channel architecture used by DOS and the INTEL 8088 CPU. It was therefore decided to recommend the INTEL 80386 32-bit or similar technology when making recommendations for future work. This would also allow for larger arrays and very much faster processing.

The intermediate level functions provided controlled access of the machine level functions by the high level functions. This made the program more readable and also defined a communication protocol to ensure that the machine level functions were accessed correctly.

The high level functions in turn also provided a communication protocol for the main control program. This meant that new control software could be written by anyone without specific knowledge of the machine level functions, as long as the communication protocols were observed. Hence the abstraction from the machine level functions.

The main control software consisted of a number of interchangeable programs. Each of these were designed to do various tasks in the collecting, calibrating and interpretation of data. Only the specific control program for the task at hand was compiled with the rest of the library of functions in each case.

An important aspect of the computer controlled test system was that the measured signals could be converted to any units and scales. This was done by the use of programmable calibration factors. In addition any parameters could be normalized by for example baseline parameters. This greatly simplified the bearing diagnostic process.

The reliability with which parameters could be measured was increased by the use of programmable averaging and smoothing techniques. This was particularly so with the relatively static parameters such as load and temperature.

Other relevant software developed but not included in the above functions were the FFT (Fast Fourier Transform) routines and data windowing routines. In particular the Cooley-Tukey radix 2 algorithm [append. 10] performed on a bit reversed array [append. 11] was used to calculate the FFT's. For data windowing a self designed window similar to the "Bingham" window [32] was used [append. 12].

In addition various data file handling, post-processing and graphics display routines were written to extract information from the research data.

10.5. THE BEARING LOAD MEASURING SYSTEM

This section describes the methods used to measure the load applied to the test bearing. It provides a general description of how the system performed the load measurement and discusses how the various stages of the system were calibrated.

10.5.1. GENERAL DESCRIPTION OF THE SYSTEM

Measuring the load applied to the test bearing was achieved in two different ways. Firstly a pressure gauge was inserted on the hydraulic line as shown in figure 9.3. This gauge was not used as a load transducer but rather as a backup to ensure that the shaft was not overloaded at any stage [section 9.5].

The load transducer consisted of a bonded strain gauge arrangement to measure the strain on the hydraulic cylinder shaft. This strain was directly proportional to the load on the central test bearing. The change in resistance of the strain gauges was measured by a Control Instruments P-3500 Strain Indicator.

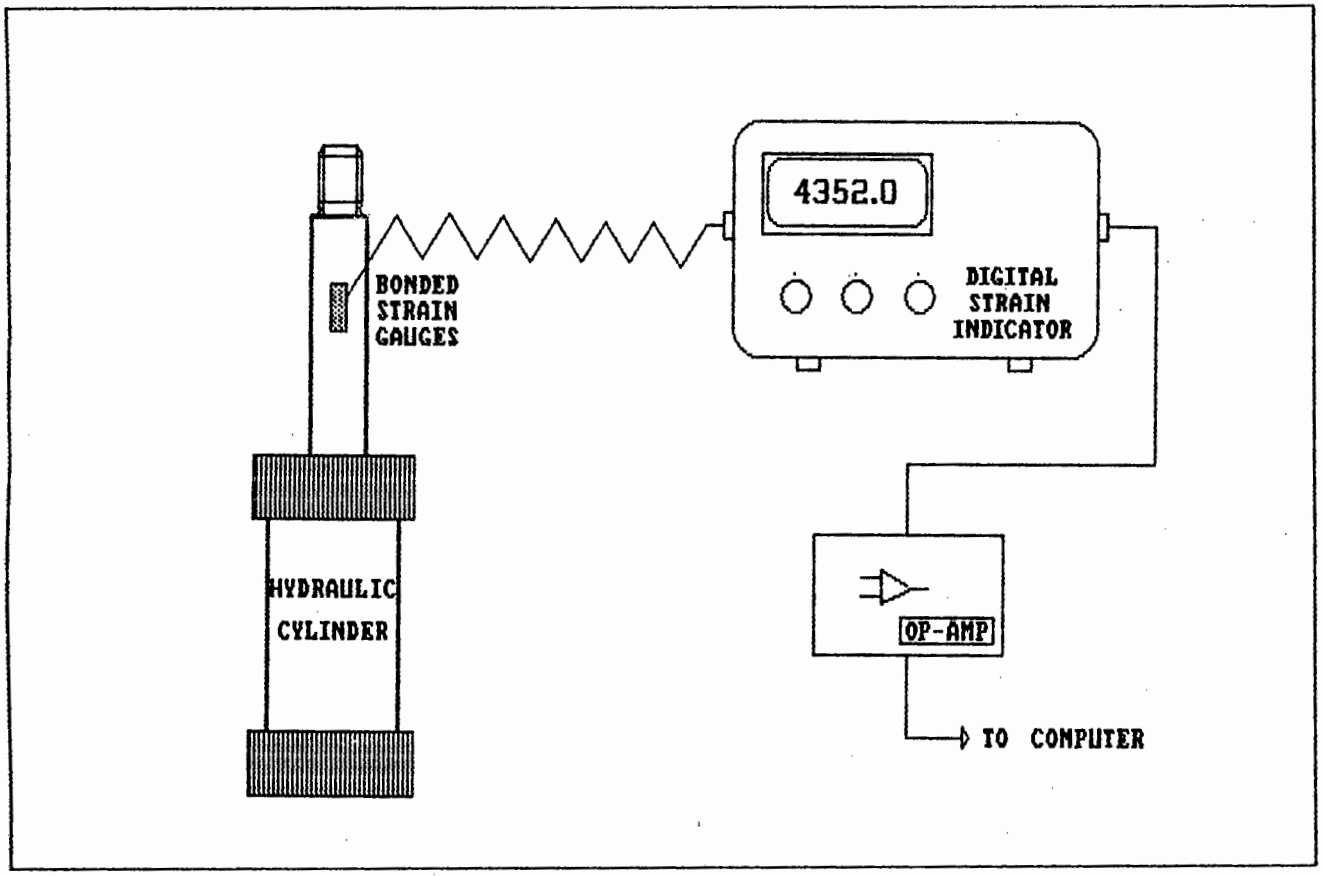

FIGURE 10.3 : SCHEMATIC DIAGRAM OF THE LOAD MEASURING SYSTEM

This instrument provided a digital readout in microstrain when the correct gauge factor was applied. In addition there was an adjustable analog voltage output proportional to the strain on the digital readout. This voltage was used as an input to the A to D card.

The analog output voltage was too small to utilize the full range of the A to D card and so it had to be amplified to an acceptable level. This was achieved by developing a two stage operational amplifier as shown in figure 10.4.

The variable resistance R2 was used to amplify the bridge output voltage range from (0 to 20 mV) to (-0·69 to 2·12 V) or a gain of 46 dB. The negative value was achieved by having a bipolar power supply driving the op-amps. Thus most of the ±2·5 voltage range of the card was utilized.

The strain gauge arrangement was designed to compensate for temperature variations and the effect of bending. This was achieved by mounting four similar strain gauges on the shaft in a "full bridge" configuration as shown in figure 10.5.

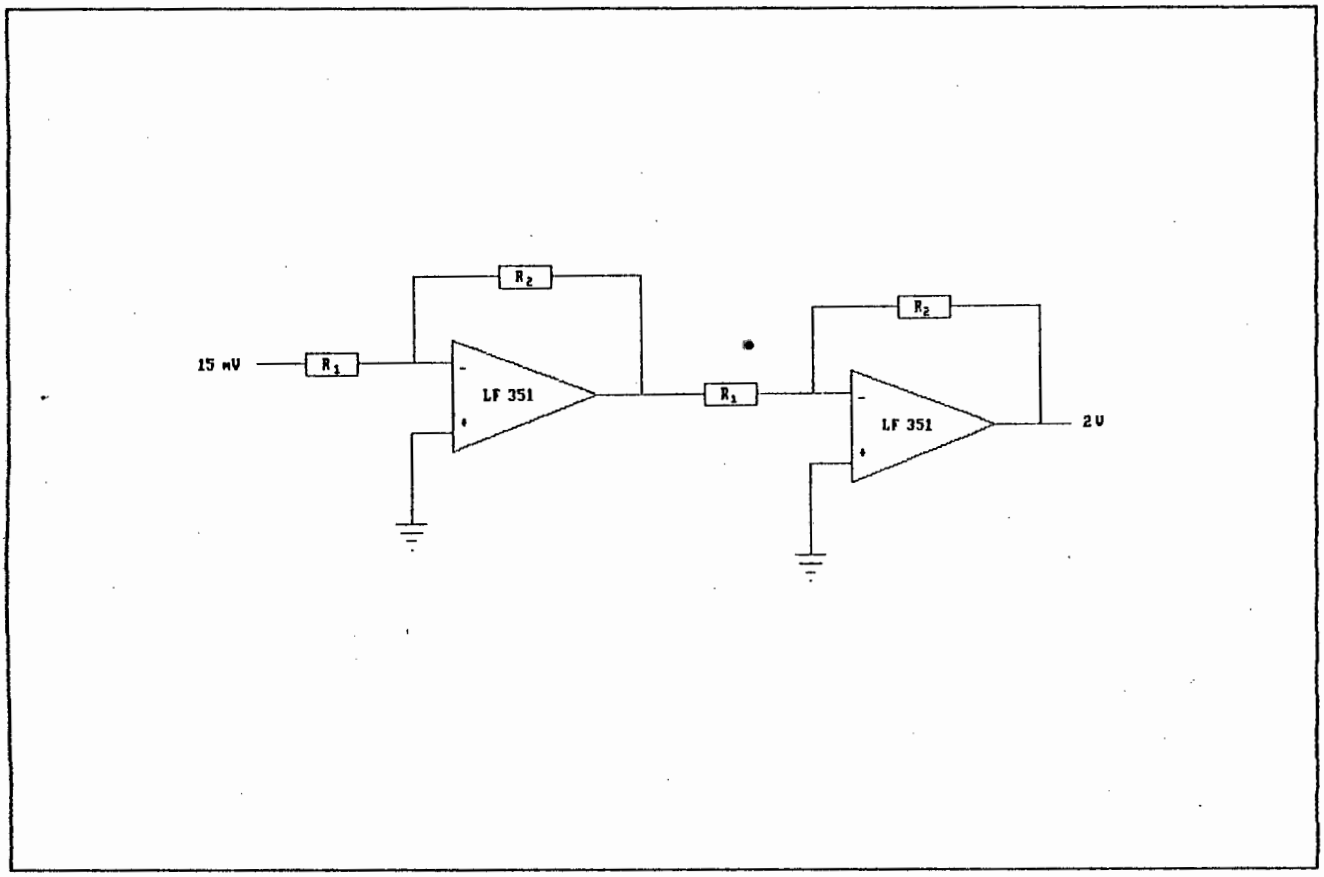

FIGURE 10.4: CIRCUIT DIAGRAM OF THE LOAD SIGNAL AMPLIFIER

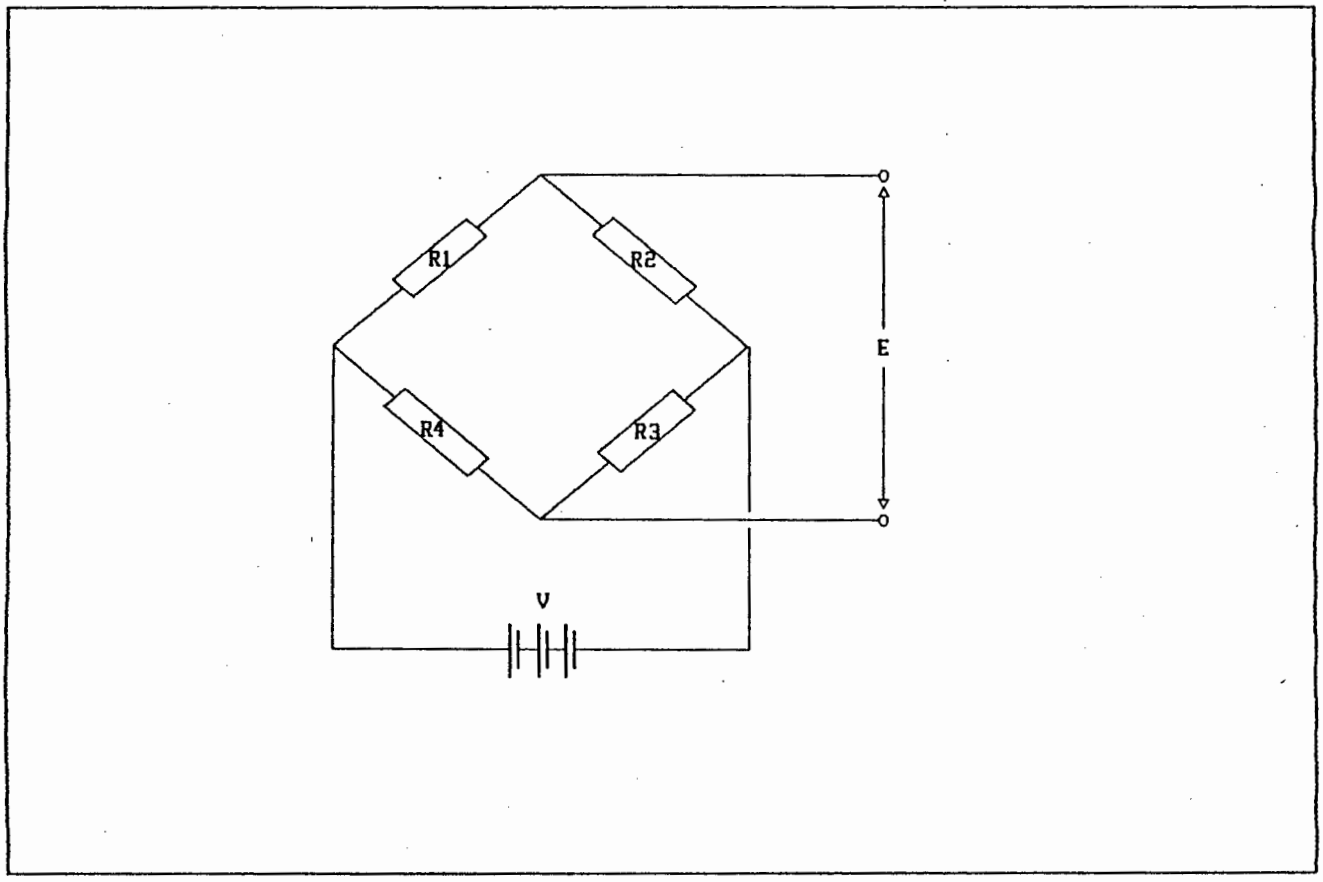

FIGURE 10.5 : THE STRAIN GAUGE BRIDGE CIRCUIT

The strain gauges used were Kyowa^{R} bonded resistance gauges with a gauge factor of 2·08. The specifications of these gauges are given in appendix 13. Gauges 1 and 3 were mounted axially on the shaft and therefore measured the axial strain which in turn was proportional to the bearing load.

The voltage through the full bridge arrangement was governed by the following equation [33]:

$$dE = \frac{R_1 R_2}{(R_1+R_2)^2} * \left(\frac{dR_1}{R_1} - \frac{dR_2}{R_2} + \frac{dR_3}{R_3} - \frac{dR_4}{R_4}\right) \quad \ldots 10.1$$

Where : dE = The change in voltage across the bridge
dR_i = The change in resistance of gauge `i'
R_i = Resistance `i'

Compensation for bending strain was achieved by mounting gauges 1 and 3 on opposite sides of the cylinder shaft. If a bending moment was applied to the shaft the one side would be in compression and the other in tension. Thus the decrase and increase in resistances respectively would cancel out and negate the effect of the bending.

Temperature changes were compensated for by placing gauges 2 and 4 perpendicular to gauges 1 and 3. Because the metal shaft would expand uniformly in all directions with temperature change, all gauges would register the same change in resistance.

Then according to eqn. 11.1, gauges 2 and 4 would subtract from 1 and 3 because of their position in the bridge and change in temperature would have no nett effect. In addition gauges 2 and 4 were unaffected by axial strain because of their transverse mounting.

Thus temperature effects were successfully accounted for. This was important because the the temperature of the cylinder shaft changed significantly during testing.

10.5.2. CALIBRATION OF THE BEARING LOADING SYSTEM

It was needed to have a digital representation of the load on the bearing in kN. This was used to test the dependence of the various vibration parameters on the bearing load and also to calculate the predicted life of the bearing. To have such a reading of load required calibration of three stages of the load measuring system.

Firstly the pressure gauge was calibrated. This was done on an Instrom `Universal Testing Instrument' model TT-D. The hydraulic cylinder was mounted in the tester and loaded in tension. The load was varied either by adjusting the Instrom or using the Enerpack hydraulic pump. Thus the load-pressure curve of figure 10.6 was obtained [append. 14].

The straight line (shown by squares) indicated a least squares regression line through all of the data points. The curves below this line corresponded to pumping up the cylinder (+) and releasing the load on the Instrom (x) while the curves above the line were releasing the pressure (diamonds) and increasing the Instrom load (triangles).

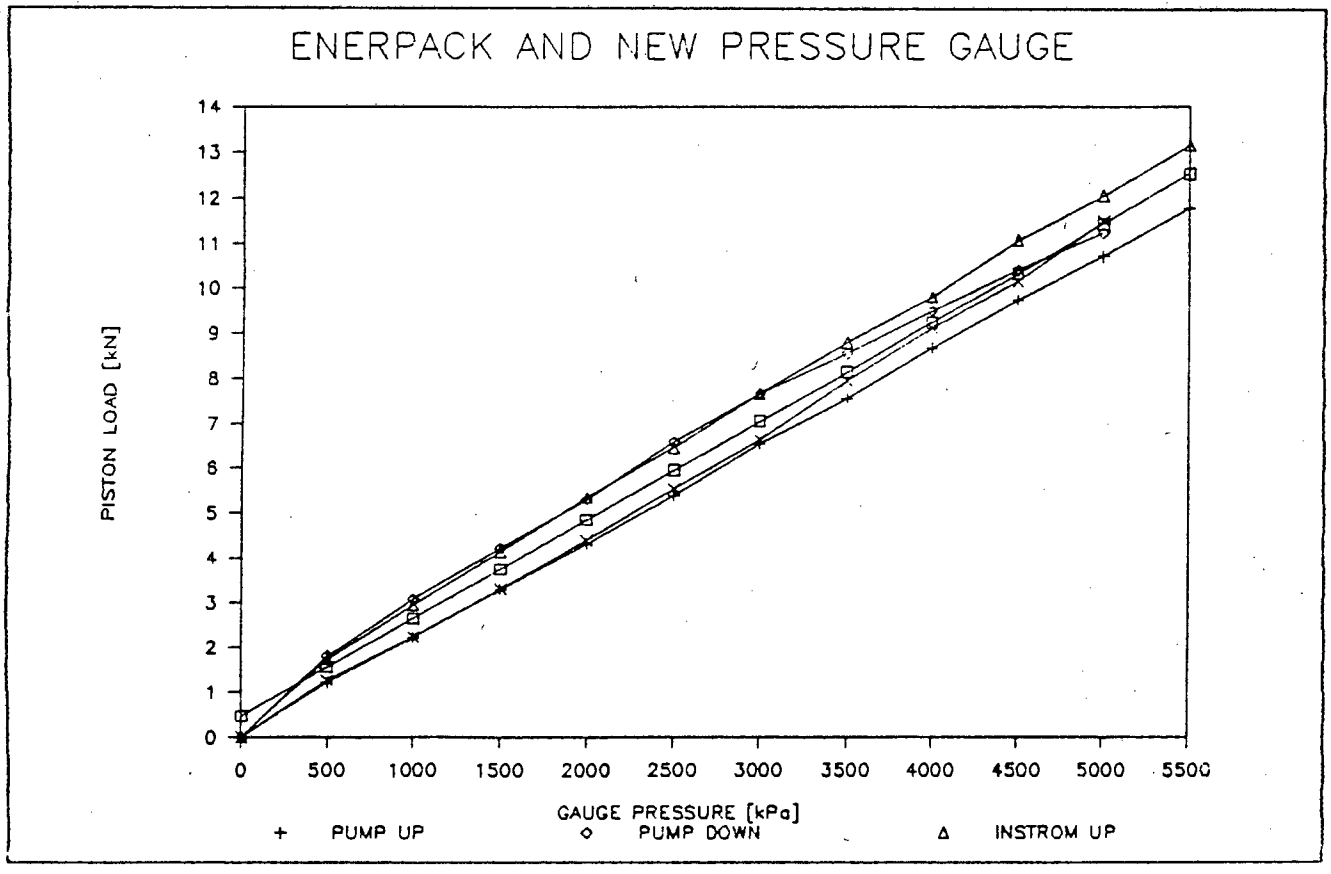

FIGURE 10.6 : CALIBRATION CURVE FOR THE HYDRAULIC PRESSURE GAUGE

This hysteresis effect indicated that there was a considerable amount of friction between the piston seals and the cylinder walls. Hence the pressure gauge could only give a general indication of the applied load. Of course this hysteresis effect due to friction would not affect the strain gauge readings.

The strain gauge bridge was calibrated on the Instrom in a similar manner to the pressure gauge. Thus the load-strain calibration curve of figure 10.7 was obtained [append. 15]. It was unknown whether the non-linearity of the curve was due to the mechanical or electrical components.

However it was decided that the regression line through these points should be taken as the calibration curve for the bridge. This line was accurate enough in the higher load zone which was where the system would be most used. Note that the applied load in figure 10.7 would be doubled on the test bearing due to the lever effect of the load beam.

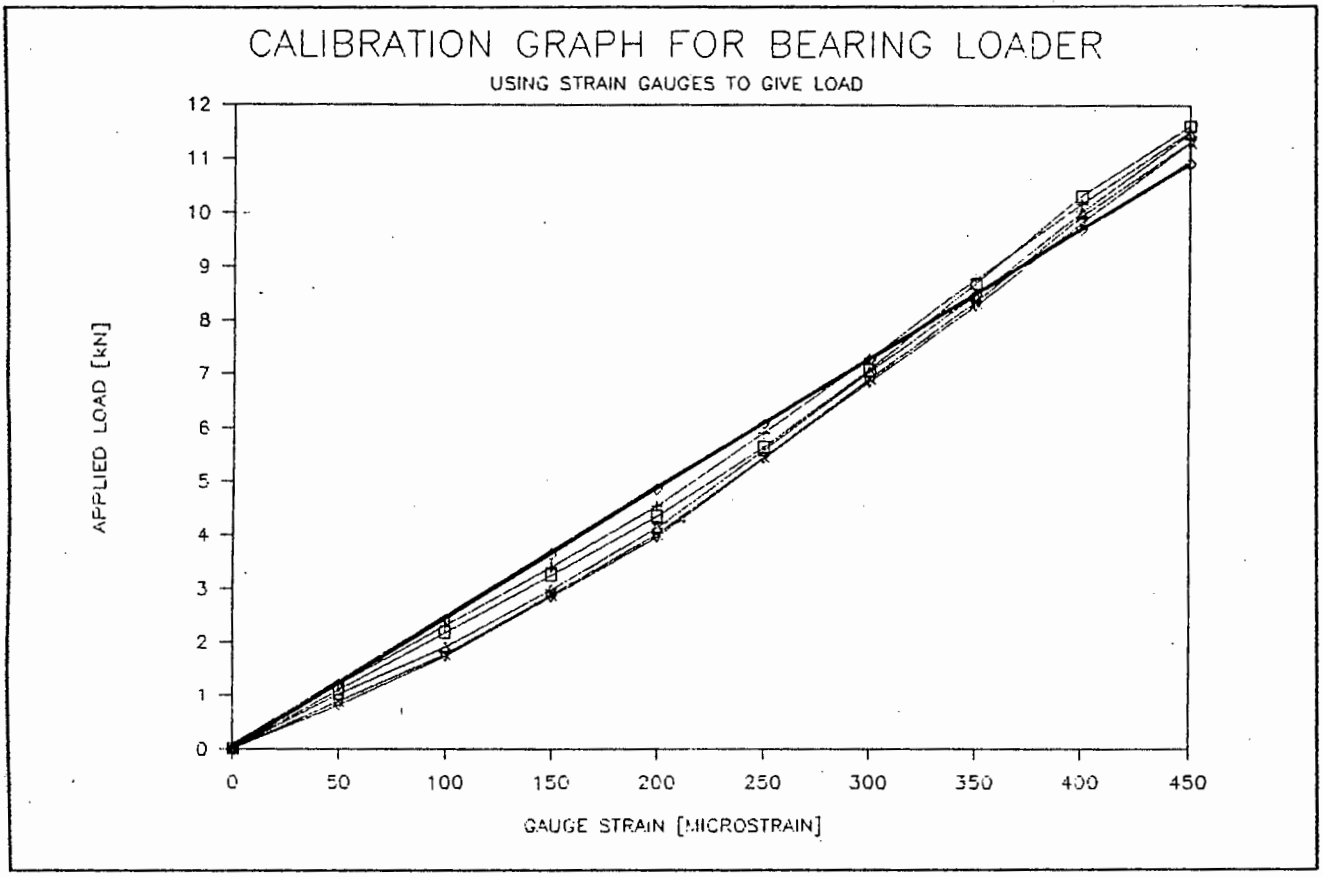

FIGURE 10.7 : CALIBRATION CURVE OF THE STRAIN GAUGE BRIDGE

The analogue output of the Strain Indicator and the operational amplifier were calibrated together. Firstly the operational amplifier was adjusted to make full use of the A to D card's dynamic range. This was done using the variable resistance R_1 [fig. 10.4].

The Strain Indicators microstrain reading was then varied while readings of the corresponding voltages seen by the A to D card were taken [append. 16]. These are presented in figure 10.8.

It can be seen that this stage of the circuit was very linear as was expected. The two calibration curves of figures 10.7 and 10.8 were then combined [append. 17] to provide an overall calibration curve represented by the following linear equation :

Test Bearing Load = 2 * (A-D val * 0.01012 - 16.005) ...10.2

This equation was used to calculate the bearing load from the A-D value read at the input to the A to D card. In theory this load could be calculated with a resolution of 0·01 kN, corresponding to a full scale dynamic range of about 67 dB [append. 18].

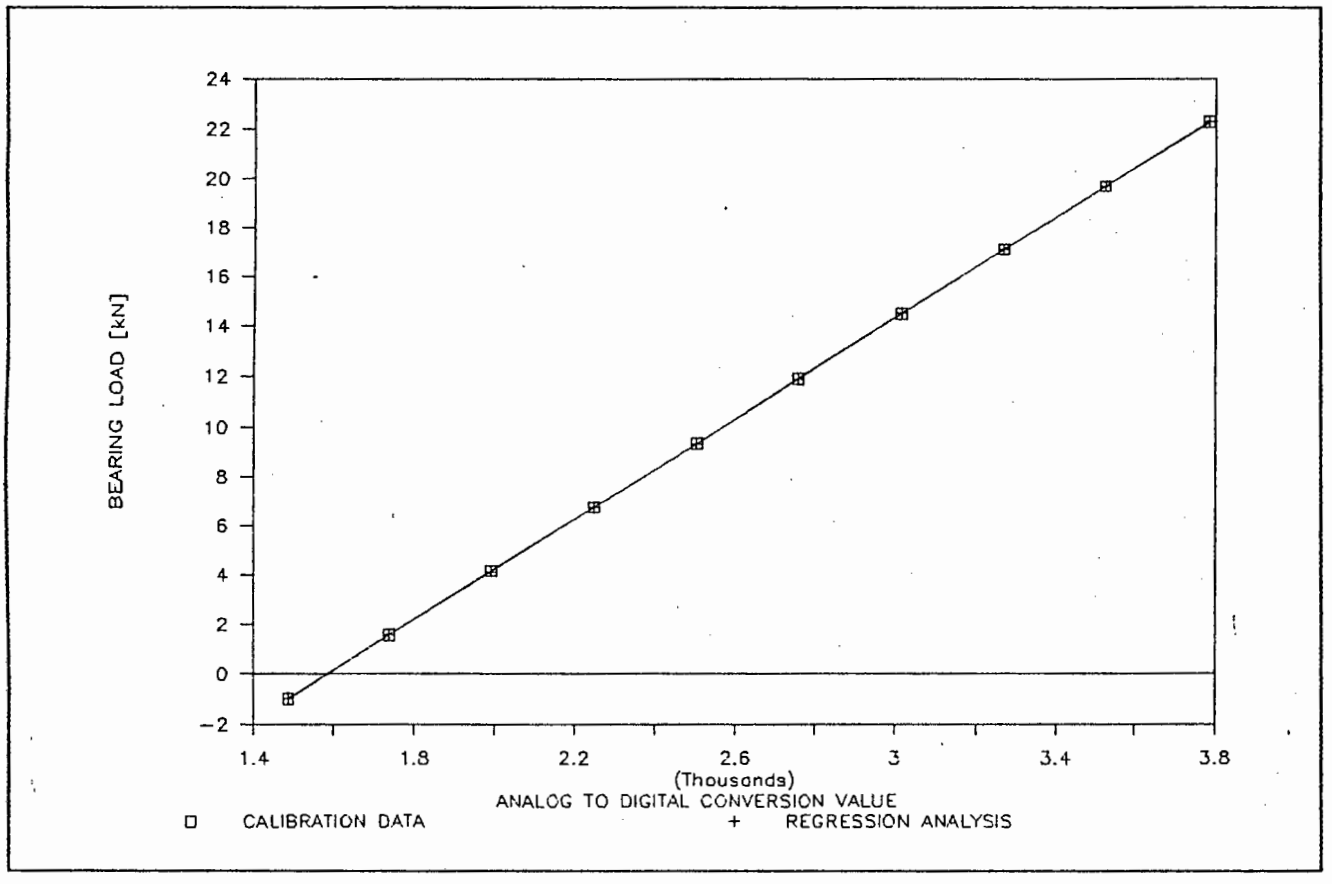

FIGURE 10.8 : CALIBRATION CURVE OF THE STRAIN GAUGE AMPLIFIER

However the load reading on the computer showed a large apparently random fluctuation. This was surprising as a stable dc value was expected. It was suspected that the load signal was contaminated by noise. The fact that this fluctuation was seen on the computer and not the strain gauge display was probably due to differing digital sampling mechanisms of the two instruments.

The rapid speed with which the A to D card obtained each individual sample would mean that the sample was often occurring on the peak or trough of some noise source. However this noise source would have to be strong to provide the fluctuations observed. Thus the signal was analyzed dynamically using an analog oscilloscope (Tektronix T922 15MHz oscilloscope).

It was found that this signal was badly contaminated by a periodic high frequency (>350 kHz) noise spike. It was postulated that this noise was created by electromagnetic interference from the dc motor. Various attempts were made to locate the source and thereby reduce its effect.

Firstly the motor and support frame was properly earthed to the ground. When this had no effect a ground loop was suspected. Thus the strain gauge bridge was isolated and powered by an independent dc source. This still had no effect.

After this all cabling was replaced by shielded cabling and grounded only at one end. This had a significant attenuating effect on the noise level, confirming that the source was electromagnetic interference.

At this stage the signal to noise ratio was calculated to be about 19 dB [append. 19]. However this was well below the 67 dB dynamic range of the A to D conversion and so further attempts were made to smooth the signal.

Because the load signal was required to be a dc value it was decided to dc-couple this voltage. This was done by mounting a capacitive filter across the end of the final analog stage. This also had an attenuating affect on the noise.

In addition the signal was smoothed by averaging a set of 100 consecutively sampled points. Each set of 100 points was collected in 5 milliseconds, corresponding to a sample frequency of 20 kHz.

These combined actions improved the signal to noise ratio to about 30 dB [append. 19]. This meant that the load could be displayed with a realistic accuracy of about 0·7 kN for a full scale load of 23 kN. It was decided that this provided a sufficiently good representation of the bearing load.

Thus the bearing load measuring system was completed and functioning satisfactorily. The next system to be designed and calibrated was that to measure temperature.

10.6. THE BEARING TEMPERATURE MEASURING SYSTEM

The bearing temperature was an important part of diagnosing bearing condition. Temperature gave an indication of the lubricant viscosity which related directly to the lubricant film thickness. This in turn affected the bearing life.

The temperature was measured using a Chromel-Alumel thermocouple. It was mounted on the outside face of the outer race of the test bearing as shown in figure 10.9. This was the closest static position to the heat source, namely the bearing contact surfaces. It was noted that the temperature of this position would be approximately 10°C lower than that of the inner race [15].

The expected maximum temperature of a damaged bearing under full load was about 130°C [7]. For this temperature the thermocouple junction would produce a voltage of 5·33 mV [append. 20]. This would have to be amplified to make full use of the A/D card's dynamic range. In addition some form of ice point compensation would have to be used [33, 34].

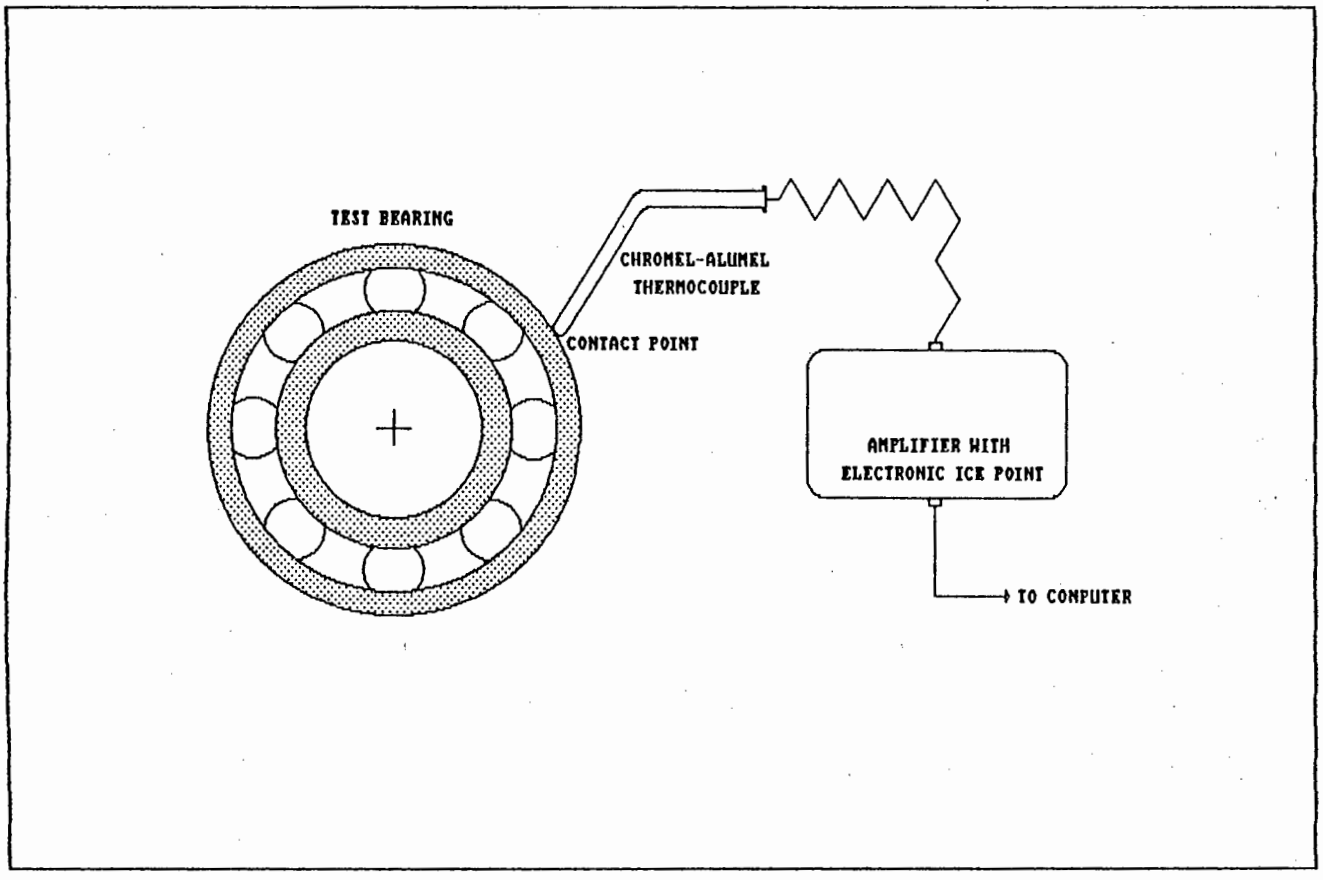

FIGURE 10.9 : SCHEMATIC DIAGRAM OF THE TEMPERATURE MEASURING SYSTEM

For these purposes an integrated circuit thermocouple amplifier was obtained. It was a K-type AD595 made by ANALOG DEVICES. This amplifier was designed to provide an output of 10 mV/°C for a K-type thermocouple such as Chromel-Alumel [35]. In addition the amplifier provided electronic cold junction compensation. Thus the output was proportional to the °C temperature scale.

The system was tested at 0°C using an ice and water mixture and at 100°C using boiling water. It was found that at 100°C the output voltage was 0·038 volts too high while at 0°C it was 0·026 volts too high. These corresponded to 3·8°C and 2·6°C respectively. This error was compensated for by modifying the digital values before display [append. 21].

The system could theoretically display the temperature to a resolution of 0·12°C [append. 21]. However a resolution of 1 or 2°C was regarded as sufficiently accurate to identify any potential bearing faults.

Thus the temperature measuring system was calibrated and functioning correctly. The next system to be developed was the shaft rotational speed measuring device.

10.7. THE SHAFT SPEED MEASURING SYSTEM

It was required to know the shaft speed as accurately as possible. The accuracy was required for calculating bearing element passage frequencies [append. 22] and for normalizing the vibration parameters with respect to speed.

The speed was measured by using an optical shaft encoder. This consisted of a light beam being interrupted by a slotted disk attached to the rotating shaft. The encoder was an integrated unit with circuit as shown in figure 10.10.

The light beam was provided by the light emitting diode and received by the photoconductive transistor. This transistor would then provide a reference voltage V_{out} as shown. If the beam was interrupted, V_{out} would fall to zero.

Thus the encoder would effectively provide a square wave with frequency proportional to the speed of the shaft. Speed could then be calculated from this square wave by counting the number of rising and falling edges.

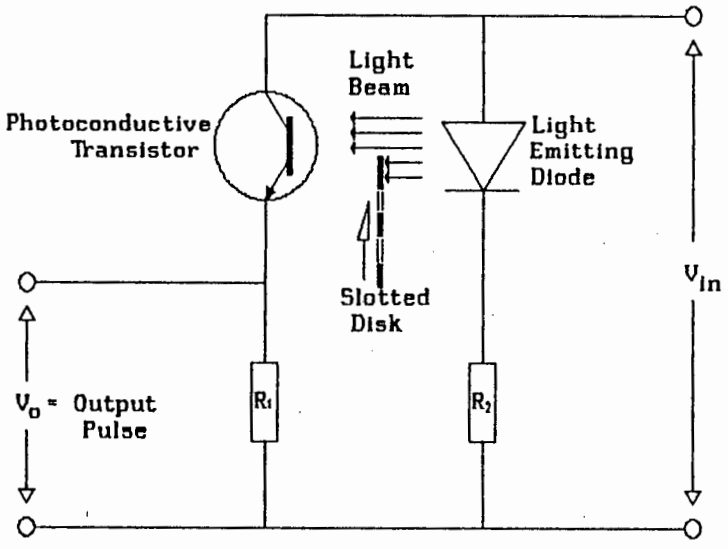

FIGURE 10.10 : THE OPTICAL SHAFT ENCODER CIRCUIT

However the speed of beam interruption was too fast for the encoder to produce a square wave. Instead the wave was triangular in shape. In addition the amplitude of this wave converged on a certain positive voltage level with increasing speed. This level coincided with the mean voltage of the triangular wave.

The triangular wave was digitally sampled by the A/D card and was therefore represented by discrete values. To calculate speed from these values amounted to counting the frequency of the triangular wave. In other words a counter was incremented every time subsequent values crossed over a preset threshold level in either direction.

This threshold level was set so as to ensure that an equal number of samples would fall above and below it. The best level was found to be the convergence voltage of the triangular wave as discussed earlier.

The speed of the shaft (RPM) was calculated using the total number of counts (C) in the following equation [append. 23] :

$$RPM = \frac{C * F_s * 60}{S_T * C_R} \qquad \ldots\ldots 10.3$$

Where : F_s = Sampling frequency
S_T = Total number of samples collected
C_R = Number of counts per revolution

From this equation it was shown that the resolution with which speed could be calculated (delta RPM) was given by the following equation [append. 23] :

$$\text{delta RPM} = K / S_T \qquad \ldots\ldots 10.4$$

Where K was a constant.

Note that a smaller value of delta RPM would represent a better resolution.

Equation 10.4 also demonstrated that a better resolution would be obtained by increasing the number of samples collected. It was calculated using this equation that a resolution of 1 RPM could be achieved by using 4455 samples [append. 24]. This was increased to 6000 samples to ensure a resolution of better than 1 RPM.

For calculating speed in real time these samples would have to be collected as rapidly as possible. In other words the highest sampling frequency of the A-D card and the most number of slots in the encoder disk would have to be used.

It was decided that to improve the statistical reliability of the counting process, at least two consecutive samples on average would have to fall on the same side of the threshold value at maximum rotational speed. In other words it was required to have four samples per cycle of the triangular wave.

However any machining inaccuracies on the width of the encoder disk slots could cause less than 4 samples to fall on a particular cycle. Therefore it was decided to increase the samples per cycle to 4·5 on average.

The number of slots in the disk was limited by the maximum sampling frequency of the A-D card, the maximum shaft speed and the number of samples per cycle. Therefore using the maximum sampling frequency of 27·586 kHz [append. 8], a maximum shaft speed of marginally less than 33 Hz and on average 4·5 samples per cycle, the maximum number of slots allowable on the disk was calculated as 185·76 [append. 25].

However shaft encoder disks were usually designed with the number of slots being some integer multiple of 60. This would enable a frequency counter to give a direct reading of the equivalent multiple of shaft speed in revolutions per minute. The closest multiple of 60 to 185·76 was 180 slots. This was used as the final value for the disk.

The diameter of the disk was determined by the required width of the slots and the spaces between the slots. The spaces required the same dimensions as the slots and had to be accurately machined. This was to ensure that the correct minimum number of samples per half cycle of the triangular wave occurred while the light beam was either complete or interrupted.

It was decided that a 2mm slot and 2mm space would would be sufficiently large to prevent damage and distortion of the slots. Thus the nominal diameter of the encoder disk was calculated as 230·5 mm [append. 26].

The output voltage of the encoder was applied directly to the A-D card without amplification. In addition the cabling had to be coaxially shielded to reduce electrical interference from the motor.

The final digital speed reading was compared with that produced by a "Jaquet's" mechanical speed indicator and a digital frequency counter attached to the output of the encoder. The digitally computed speed was seen to correlate well with these other methods [append. 27] and therefore it was considered accurate enough for the required purpose.

The bearing vibration measurement system was then designed and calibrated. This is discussed in the next section.

10.8. THE BEARING VIBRATION MEASURING SYSTEM

The bearing vibration measurements were the ultimate aim of the whole test rig measuring system. This was because the calculated vibration parameters were considered the most important tools for effective bearing condition analysis.

In addition some of the vibration parameters such as RMS acceleration were not dimensionless. Therefore to allow comparisons with other vibrations work such as laboratory research, international standards organizations and industrial programmes, the system would have to be properly calibrated.

The first section covers the design and general description of the bearing vibration measuring system. This is followed by a discussion of the calibration procedure used for each component.

10.8.1. THE DESIGN OF THE BEARING VIBRATION MEASURING SYSTEM

Vibration measurements were taken using an accelerometer as the transducer. The accelerometer was mounted on top of the test bearing's plummer block with a threaded stud as shown in figure 10.11. The interfacing surface between the accelerometer and the plummer block was machined and polished smooth to ensure transmission of the higher frequencies without significant attenuation.

The bearing loading mechanism had been designed so that the accelerometer could be easily mounted in the center of the loaded zone of the test bearing. This ensured that the length of the vibration path between source and measurement point was minimized, thereby minimizing structural attenuation of the vibrations.

The accelerometer used was a PCB 308M126 [append. 50] and was of the ICP (Integrated Circuit Piezoelectric) type. It provided an output voltage proportional to the absolute acceleration of the accelerometer. This meant it was a so called "seismic" transducer where the acceleration was relative to an inertial reference frame.

The built in integrated circuit had a dual function. Firstly it amplified the voltage produced by the piezoelectric element to a rounded calibration constant of 100 mV/G. This amplification of the signal close to the source meant that the SN (signal to noise) ratio of the system was greatly improved.

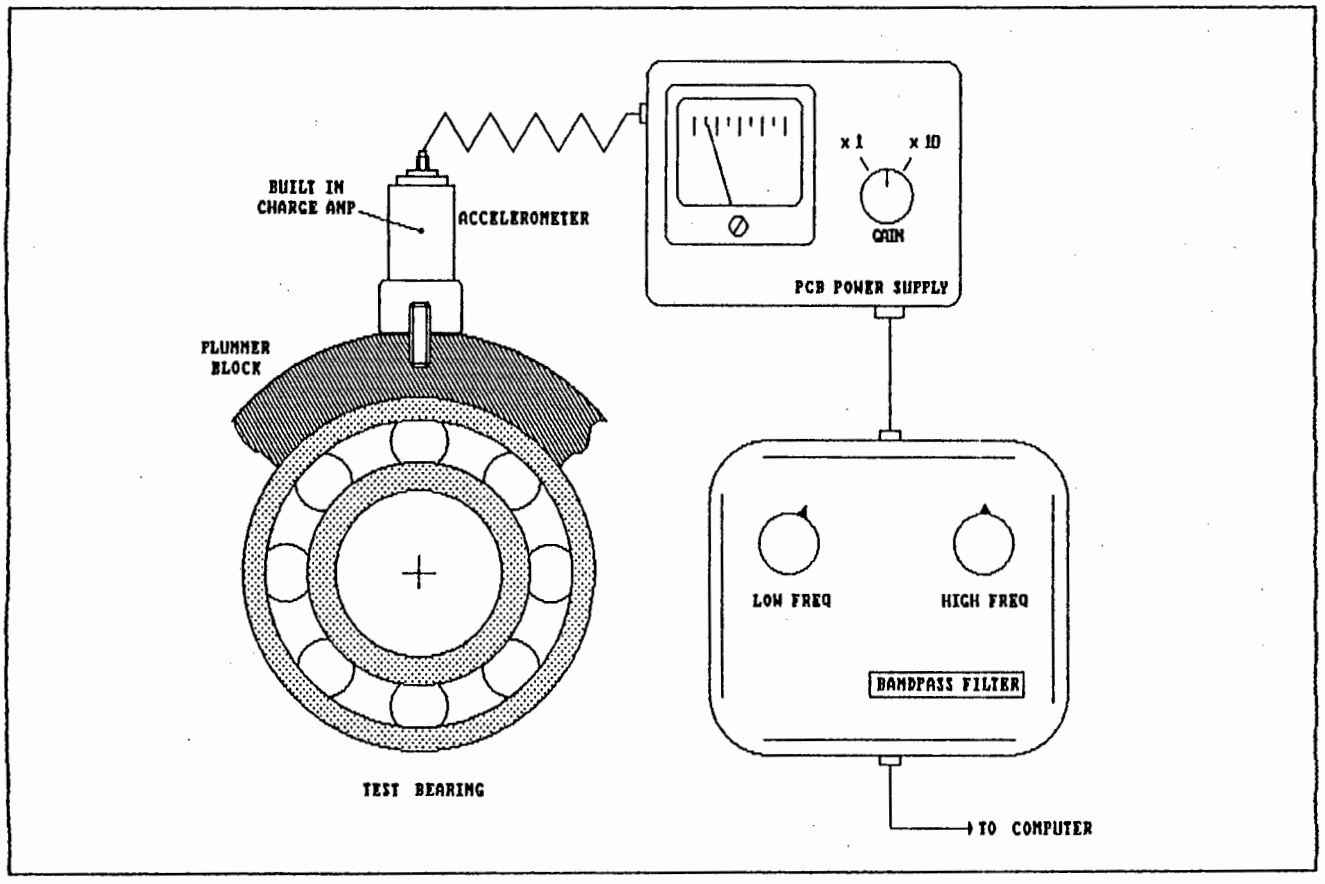

FIGURE 10.11 : SCHEMATIC DIAGRAM OF THE BEARING VIBRATION MEASURING SYSTEM.

Secondly it reduced the electrical impedance of the accelerometer unit. This was to simplify impedance matching with relatively low impedance measuring instruments.

Power for the integrated circuit of the accelerometer was provided by the dedicated PCB 480D06 power unit. This power unit used two 9 Volt batteries to provide an 18 Volt 4 mA DC supply for the integrated circuit. The same coaxial cable was used to transmit the vibration signal and the power supply. The power unit then filtered this DC voltage out of the output signal.

In addition the power unit had selectable amplification gains of 1 and 10. However the output signal was large enough to be used without amplification. This signal would then have to be filtered before being passed on to the interface panel of the A-D card.

The mounted resonance frequency of the accelerometer was given as being > 25kHz. In addition the frequency response was within 10% up to 7 kHz.

Filtering the signal was an important part of the vibration measurement for two reasons. Firstly the signal would have to be low pass filtered to prevent aliasing of the digitized signal. Secondly the amplitude response of the accelerometer was to be regarded as inaccurate outside the specified frequency band. Therefore a low pass filter with a cutoff frequency of approximately 8 kHz was required.

In addition it was desired that the filter should have the following characteristics. Firstly a sharp "knee" at the cutoff frequency was needed. This was to provide a high initial roll-off on the skirt to minimize the effects of aliasing.

Secondly a flat passband with little signal attenuation was desired so that processed signals would show the correct amplitudes. However this could be compromised if necessary and was less important than the steep roll-off of the skirts. The phase shift characteristics were not important because phase was not used for diagnosis.

Research on filters used in digital spectrum analyzers showed that 5th to 8th order filters were typically used [37, 38, 39]. These were usually the 'Chebychev', 'Elliptical' or 'Butterworth' types using active components. Typical values for the initial roll-off were 50 to 120 dB/Octave and about 1 to 3 dB for the passband flatness or ripple.

The only filter available in the faculty was the KROHN-HITE 3500 bandpass filter. This filter could be set by adjusting the low and high cutoff frequencies individually. More specifically both cutoff frequencies could be varied between 18 Hz and 200 kHz to form a passband of any width and in any position between these values.

No documentation was available for this filter so the frequency response curve had to be obtained by experiment. The experiment is discussed in detail in appendix 28. This provided the frequency response curve as shown in figure 10.12.

It could be seen from this curve that the filter characteristics were far from ideal. Firstly the passband attenuated the signal by as much as 12·5 dB in the low frequencies and by 3.9 dB in the higher frequencies. Thus the passband flatness was about 8·5 dB which could not be compensated for when adjusting parameter levels. In addition the high frequency skirt only had an initial roll-off of about 22 dB/octave.

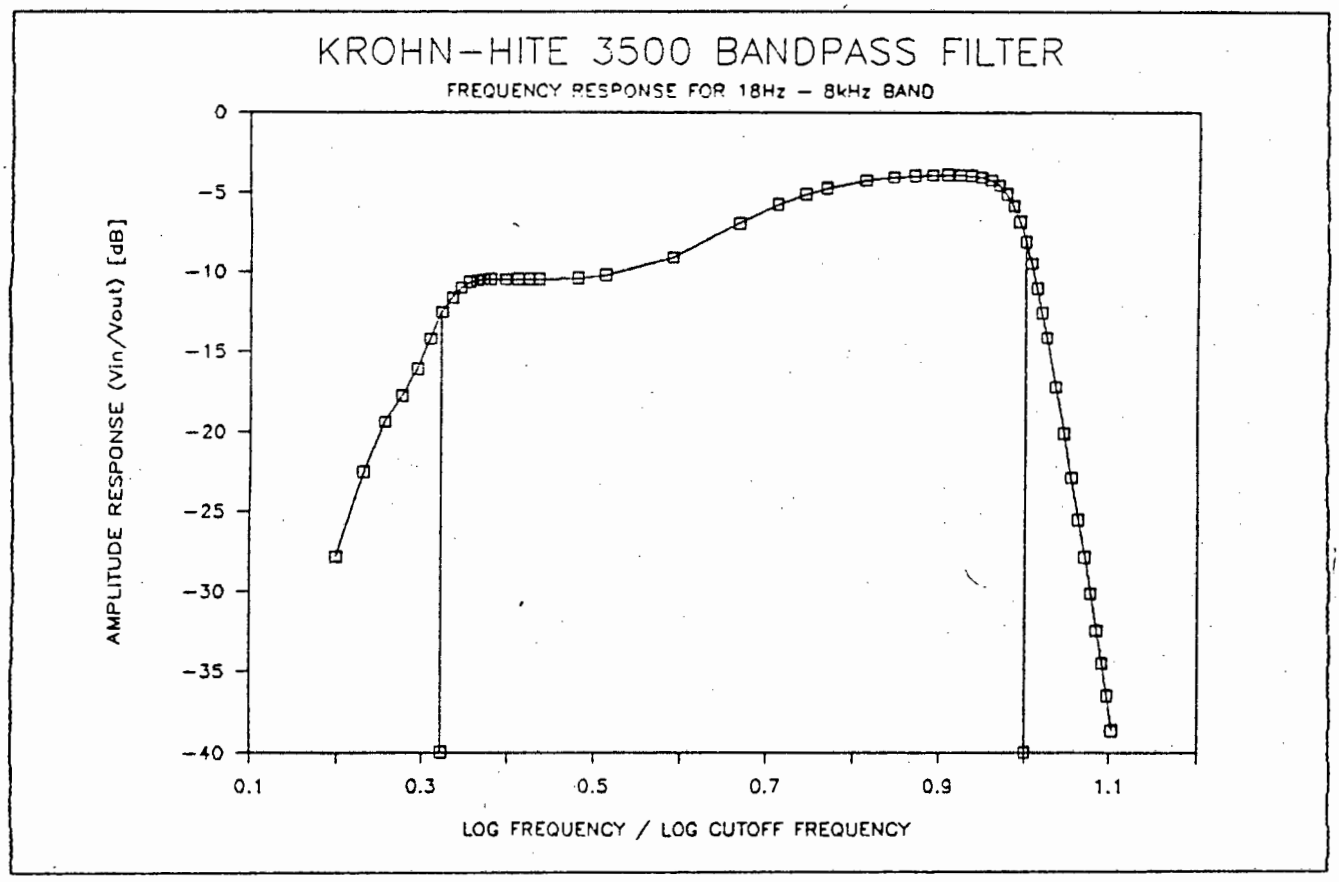

FIGURE 10.12 : THE KROHN-HITE BAND PASS FILTER FREQUENCY RESPONSE FOR AN 18 Hz TO 8 kHz PASSBAND

In addition the low cutoff frequency of 18 Hz was higher than desired. This was because of the lowest expected frequency in the vibration spectrum, namely the shaft rotating frequency, which went as low as 1.9 Hz for the slowest speed (114 rpm). The 18 Hz cutoff corresponded to a shaft rotational speed of 1080 rpm.

The most important low frequency component was that of the BPFO, occurring at 6·285 times the rotating frequency [append. 22]. Thus the lowest speed at which the filter would allow this to be detected was 172 rpm. However it was noted that the higher multiples of BPFO (x5, x6 etc.) were usually more evident [7].

The only potential sub-synchronous frequency was the fundamental train frequency. However it was not commonly found on spectral traces [7].

These filter inadequacies led to experimentation with filter design. It was decided to build two low pass filters with cutoff frequencies of 10 kHz and 2 kHz. The latter would enable more detailed analysis of the lower spectral frequencies. One would have only passive components and the other include active components for comparison.

Seeing that the more popularly used designs were the Butterworth and Chebychev filters these were chosen to be used. The Butterworth design would optimize a flat passband while the Chebychev design would have a steeper skirt roll-off [40, 41].

Firstly a Butterworth 5 Pole passive low pass filter was designed and constructed for a cutoff frequency of 10 kHz [append. 29]. The frequency response curve was obtained using the same technique as for the Krohn-Hite filter [append. 28, 30]. This curve is shown in figure 10.13.

According to theory this filter should have had a 3 dB passband flatness and a 27 dB/octave initial roll-off [append. 31]. However the values recorded were 5.5 dB/octave for the passband flatness and only 12 dB/octave for the initial roll-off [append. 32].

The discrepancy was due to deviations of the used components from the design specifications [append. 33]. This exercise illustrated the need for balancing the stages of the filter by using components close to the specified values.

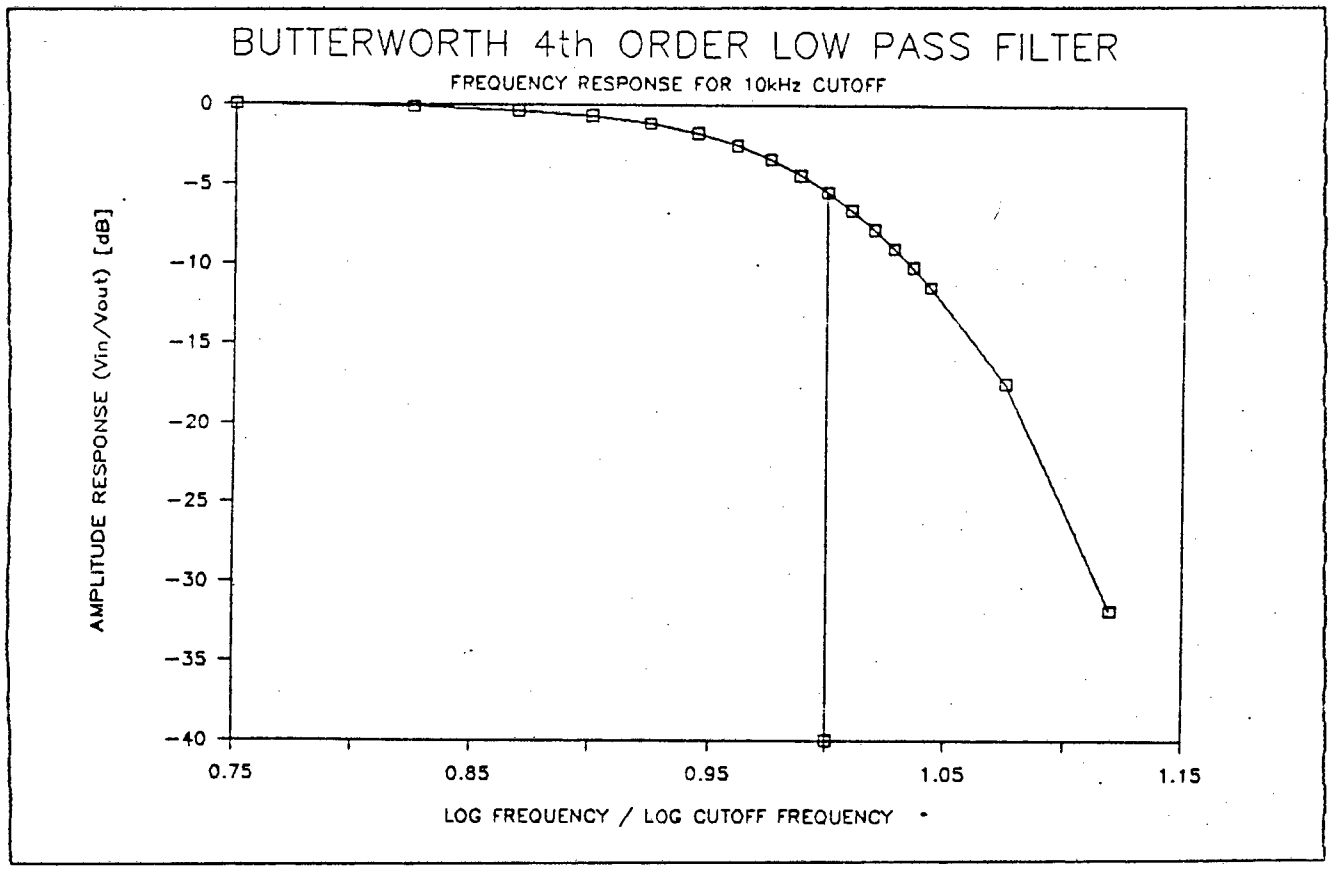

FIGURE 10.13 : THE BUTTERWORTH LOW PASS FILTER FREQUENCY RESPONSE FOR A 10 kHz CUTOFF FREQUENCY

The Chebychev active low pass filter was designed as an 8 Pole filter [append. 34]. However it was difficult to match available components with the design specifications. For example to achieve the required resistance values sometimes required more than 4 resistors. This caused the design to become clumsy and susceptible to component tolerances so this filter was not built.

The poor performance of the Butterworth filter and the design difficulties of the Chebychev filter led to the use of the Krohn-Hite filter for all the filtering work. However the design exercise highlighted the important aspects and problems involved in implementing filters for signal conditioning.

The filter output was then presented directly to the A-D card for digitizing. It was also connected to a Tektronix T922 15MHz analog oscilloscope. This was to provide an immediate display of the vibration pattern to check that the system was functioning correctly. The signal was clarified by using the output of the shaft encoder to trigger the oscilloscope.

SECTION 11

THE BEARING CONDITION ANALYSIS SOLUTION SPACE

11. THE BEARING CONDITION ANALYSIS SOLUTION SPACE

The concept of a solution space was derived from expert system development. A solution space was a specific subset of the domain of knowledge about the particular subject, in this case bearing condition analysis.

In particular the solution space was the clearly defined space within which the expert system would operate to analyze the bearing condition. The solution space was a multidimensional zone with complex interactions and dependencies between the variables of the system. Therefore it was necessary to map the solution space before the expert system could be developed.

The solution space was divided into three subsets of variables which were grouped according to their function. These subsets were the operating conditions, the vibration parameters and the bearing condition. The mapping process consisted of selecting the variables for each subset and then defining the relationships between the subsets by experimentation.

This section describes the selection of variables for the subsets and sets the limits of each variable if these are determined without experimentation. The next section describes the experiments performed to define the inter-set relationships and limits of variables.

The solution space with its three sets of variables is illustrated schematically in figure 11.1

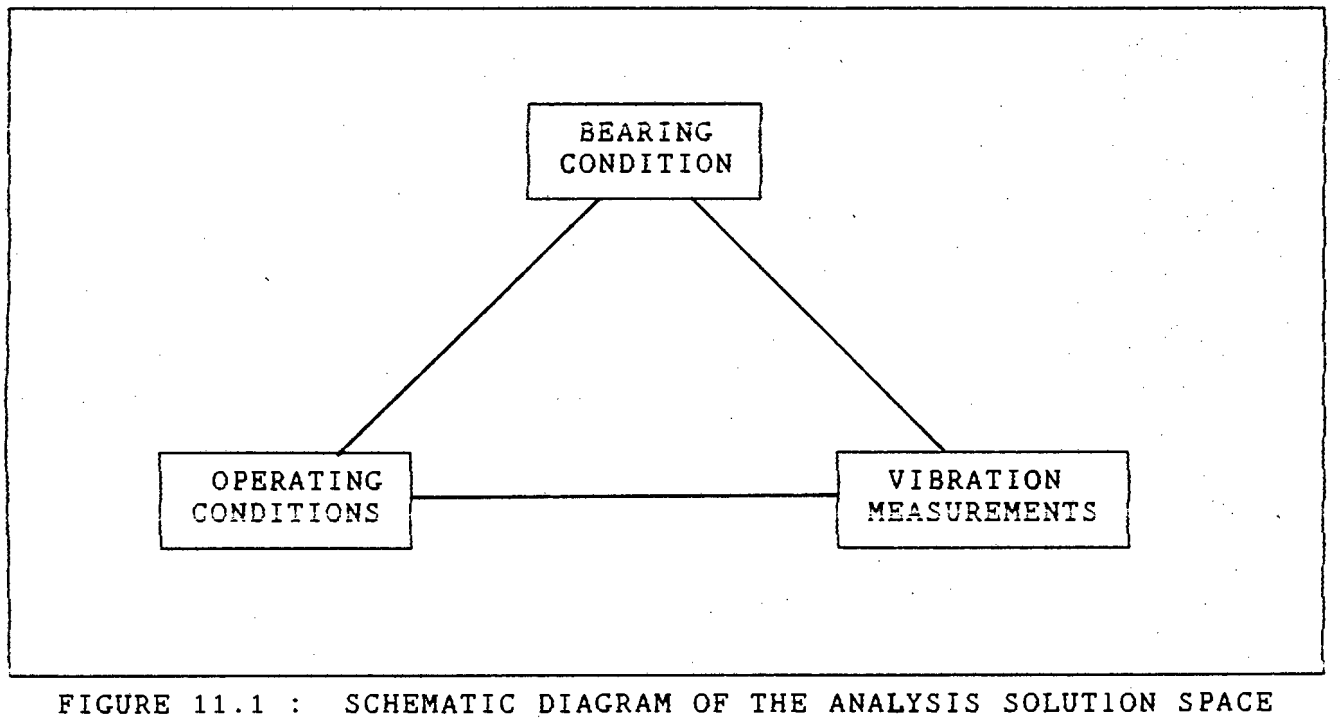

FIGURE 11.1 : SCHEMATIC DIAGRAM OF THE ANALYSIS SOLUTION SPACE

11.1. SELECTION OF VARIABLES FOR THE ANALYSIS SOLUTION SPACE

The first step in defining the solution space was to define the variables involved in each subset of the space. The main criterion for including any variable in one of the sets was its applicability and importance for bearing condition analysis. Thus it was important to bear in mind the three stages of bearing analysis, namely detection, diagnosis and prognosis.

The selection of the variables in the three subsets, namely operating conditions, vibration parameters and bearing condition is discussed as follows. Justification is given for their selection and the limits of each are defined where possible.

11.1.1. THE SELECTION OF BEARING OPERATING CONDITIONS

The operating conditions to be measured were selected according to their ease of measurement and their effect on bearing life. In addition they had to characterize the operating condition of the bearing as fully as possible.

The three most prominent conditions affecting bearing life were given in the extended rating life equation commonly used for bearings [13]. These were the effective load on the bearing, the shaft speed and the bearing temperature. All of these parameters were relatively easily measured on the test rig [section 10].

Firstly the speed was limited at the upper end by the driving motor. The maximum speed was theoretically 2000 rpm. However this was further limited by the speed control system to a maximum of 1933 rpm. At the lower end the speed was purposefully terminated at approximately 100 rpm. This was because slower speeds produced very small vibrations which were difficult to measure. Thus the speed range was from 114 rpm to 1933 rpm.

Secondly load was limited by the cyclic fatigue endurance of the shaft. From calculations [append. 4] the maximum safe load that the shaft could endure for sufficient bearing tests was a maximum of about 23 kN. However a load of up to 25 kN could be applied for short periods without exceeding the elastic limit of the shaft. Therefore the load range varied from 0 kN to 25kN.

The third measurable operating condition was the bearing temperature. However the temperature could not be as directly controlled as the load or speed. In fact it was dependent on a number of factors. Firstly it was dependent on the load-speed combination, secondly on the lubricant specifications namely type, amount and flow rate and thirdly on the amount of bearing damage.

Therefore it was difficult to control the temperature accurately for dependency tests. However it was known that the temperature would range between about 30°C for a lightly loaded slowly moving bearing to as much as 140°C for a badly damaged bearing under the most severe operating conditions.

Thus the boundaries of the solution space relating to operating conditions were clearly defined. It was noted that these conditions encompassed by far the larger portion of industrial operating conditions [5, 6] and thus the diagnosis was applicable to industry.

11.1.2. THE SELECTION OF VIBRATION MEASURING PARAMETERS

In selecting vibration parameters to characterize bearing condition a number of factors had to be taken into account. These factors included their applicability in industry, ease and expedience of calculation and their ability to provide recognition of bearing fault patterns and incipient failure.

Most importantly these parameters were to be selected in the context of diagnosis by an expert system. This meant that each parameter would have to allow for easy comparison with known patterns and levels. Therefore they would almost certainly have to be scalar parameters.

In addition each would have to distinguish its own certain aspect of the waveform with as little redundancy and overlapping with other parameters as possible. This was particularly so with the severe data reduction of scalar parameters.

There were further constraints on the selection of parameters. Firstly it was necessary to know the pattern of each selected parameter accurately for a whole bearing life-span. This would mean an effectively continuous measurement of the parameter level.

Because the vibration parameters had to be obtained digitally it was necessary to select parameters that could be calculated rapidly. This would ensure that they could be captured almost continuously for the life span of a bearing. This posed a limitation on the domains to be selected from.

In particular the frequency domain parameters were limited by relatively long processing times. The processing involved windowing, the FFT calculation, spectral smoothing and usually a numerically intensive calculation from the spectra. In addition it was very difficult to map the spectral variations for the entire solution space.

Hence it was decided to select parameters from the amplitude domain. These parameters characterized enough of the waveform to allow for significant diagnosis. They were also easily and rapidly calculated and therefore often used in industry.

To assist in the selection of vibration parameters there were particular vibration characteristics which gave a good indication of incipient failure. These were the overall level, the magnitude of individual impacts and the impulsiveness of the waveform. The parameters were selected to cover these aspects of the waveform.

The overall level was most easily measured using the RMS acceleration while the individual impacts were characterized by the Peak acceleration and the impulsiveness by Kurtosis. Hence these parameters were selected as the basis for the initial diagnosis.

In addition Crest factor was included for its ease of calculation and its relatively widespread use in industry. However it was noted that it contained information already provided in the diagnosis of Peak and RMS. Therefore it was essentially a redundant parameter.

The vibration parameters were limited in their frequency content by the accelerometer's response range. This transducer had an acceptably linear response up to about 8 kHz. Therefore an 8 kHz low pass filter was applied to the signal conditioning circuit. In fact the filter was a bandpass filter with the high pass frequency set at 18 Hz. This low end frequency was a limitation of the filter design.

This concluded the definition of the vibration parameter subset. The next subset to be defined was that of bearing condition.

11.1.3. THE SELECTION OF BEARING CONDITION VARIABLES

The variables used to define bearing condition were not all true variables in the mathematical sense. In fact they were so called 'linguistic variables' peculiar to expert systems. However they still had definable relationships with the bearing vibration parameters and operating conditions. These relationships were derived according to the concepts of 'fuzzy logic' [section 6.4].

In simple terms they constituted a means of applying a description of particular aspects of bearing condition to the vibration parameters and operating conditions. Therefore they were part of the solution space and needed specific definition. These variables were essentially the interface between machine and human and served to abstract the user from the actual measured and calculated variables.

These linguistic variables were chosen according to their ability to describe the important aspects of the bearing running condition concisely. These important aspects were the lubricant adequacy, the amount of fatigue spalling damage, a description of the correctness of operating conditions and the remaining life of the bearing.

The linguistic values applied to these variables are developed with the discussion on the measurement of the various parameters by experimentation. The experiments are discussed in the next section.

SECTION 12

EXPERIMENTAL ANALYSIS TO DEFINE THE SOLUTION SPACE RELATIONSHIPS

12. EXPERIMENTAL ANALYSIS TO DEFINE THE SOLUTION SPACE RELATIONSHIPS

Once the solution space had been defined in terms of variables it was necessary to define the inter-subset relationships. These relationships were defined by developing empirical equations relating the variables in each subset to the other subsets.

There were three groups of inter-subset relationships needing definition. Firstly it was necessary to know the dependence of bearing condition on the operating conditions. Secondly it was important to know how the vibration parameters depended on the operating conditions. And thirdly it was necessary to know how the bearing condition related to the vibration parameters.

The empirical relationships were determined by performing various experiments. These involved varying the variables in one subset and measuring the effect in another. If necessary the third subset of variables was kept constant for the duration of the experiment.

In effect there were an infinite number of experiments to define the solution space fully. This was because any of the variables could assume any value in their given range. Therefore it was necessary to select specific experiments which could be used to define the whole solution space by extrapolation. The selection and results of the experiments are discussed sections 12.1 to 3. This includes an explanation of the extrapolation process.

The inter-subset experiments are shown schematically in figure 12.1

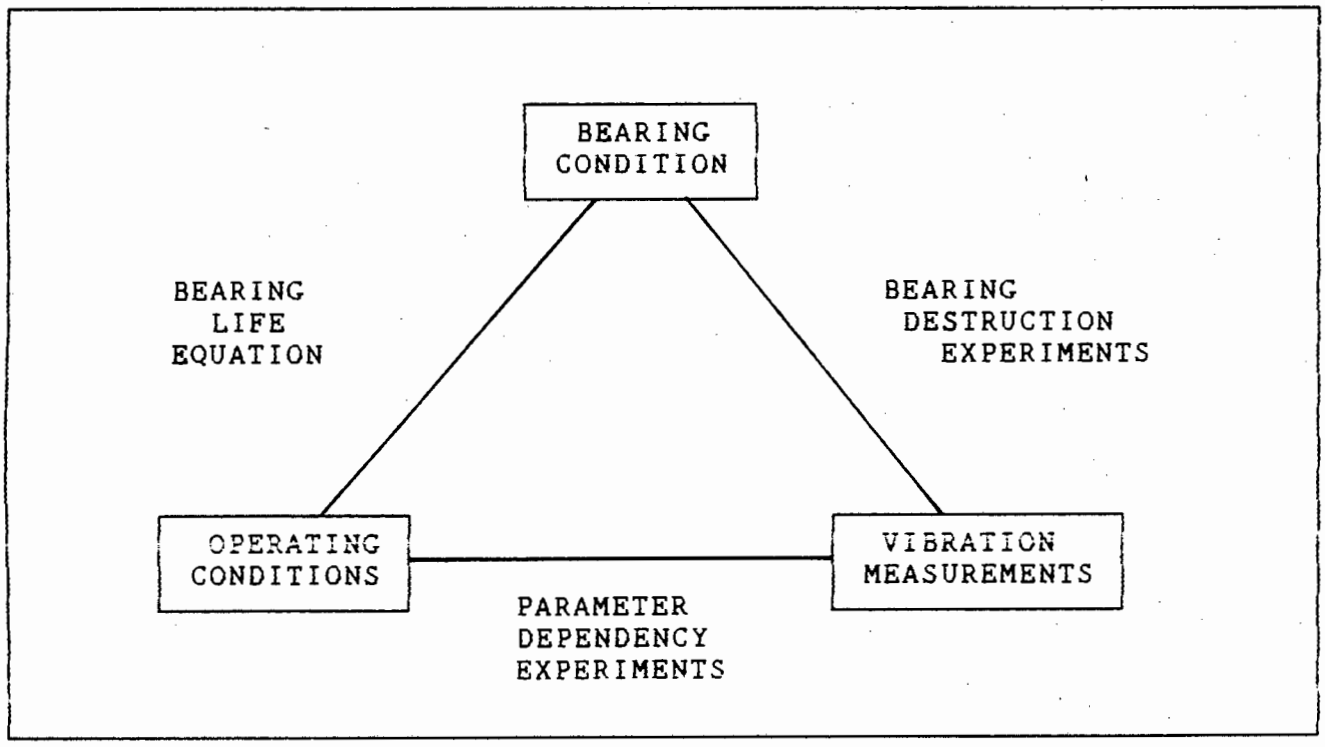

FIGURE 12.1 : SCHEMATIC DIAGRAM OF THE ANALYSIS SOLUTION SPACE SHOWING INTER-SUBSET EXPERIMENTS

12.1. THE SELECTION OF EXPERIMENTS TO DEFINE THE SOLUTION SPACE

It was desired to be able to analyze bearing condition at any point within the solution space. In other words under any combination of operating conditions and for any stage of bearing degradation. This required a knowledge of the vibration patterns over the full solution space.

To define the solution space with full resolution would require a large number of experiments. For example bearings would have to be destroyed under all combinations of operating conditions because the vibration levels would be different for each combination. Considering the time limitations for experimentation it was not possible to define the full solution space in this way.

Therefore it was necessary to select specific experiments which could be carried out within the time limitations. This meant that the more severe operating conditions would have to be applied for the experiments. The results of these experiments would then have to be regarded as representative of the solution space.

At first this posed a severe restriction on the reliability of the diagnosis at operating conditions different from those of the experiments. However this was overcome by devising a new set of experiments which enabled the extrapolation of the test results over the whole solution space.

The experiments consisted of defining the relationships between the vibration parameters and the operating conditions. Once these were known it was possible to adjust the vibration parameters according to the operating conditions under which they were measured.

These adjusted parameters could then be compared with the vibration levels determined by the experiments under the test operating conditions. Thus the experimental operating conditions became the reference point for comparison of all vibration parameters.

Therefore the experiments selected to define the solution space were as follows :

- To destroy a bearing under severe but constant operating conditions (reference operating conditions) while measuring the vibration parameters.

- To measure the dependence of vibration parameters on differing operating conditions for a bearing in good condition.

In addition it was important to know the dependence of bearing condition on operating conditions. This was derived from the life rating equation. These experiments are discussed in the following sections.

12.2. THE DEPENDENCE OF BEARING CONDITION ON OPERATING CONDITIONS

The first relationship between subsets to be analyzed was the dependence of bearing condition on the operating conditions. The specific bearing condition of interest was the remaining life of the bearing. This was partly determined from the life rating of the bearing under the specific operating conditions.

What was needed was a mathematical formulation to predict the expected bearing life using the operating conditions of load, speed and temperature. This formulation could be achieved in three possible ways as follows.

Firstly many experiments could be undertaken to measure how long bearings lasted under specific operating conditions. However this was obviously not a viable alternative due to time constraints. Therefore it was necessary to rely on other experimental work in this area.

There were two possibilities left to pursue, consisting of two empirical equations relating to bearing life. These were namely the well known bearing life rating equation and an equation to calculate the lubricant film thickness.

The lubricant film thickness equation was attractive because it contained all three of the operating condition variables, namely load, speed and temperature [15] and [append. 35]. However it would have to be assumed that bearing life was proportional in some way to the lubricant film thickness. This assumption would probably be fairly accurate given that the lubricant was uncontaminated.

However the lubricant film thickness equation assumed that the rolling elements of the bearing were rollers and not balls. This meant that the reduced lubricant film thickness due to leaking out of the sides of the balls was not accounted for. Therefore this equation was not used to predict the bearing life but it did give an interesting insight into important aspects of lubricant performance.

The third alternative was the extended bearing life rating equation [13]. This equation was the empirical result of extensive research and was regarded as the most reliable method of calculating bearing fatigue life [15]. At first glance this equation was seen to depend only on the bearing load. However on closer analysis the adjusting factors incorporated temperature and speed indirectly.

Therefore if the equation could be reformulated to contain all three operating condition variables then it could be used by the expert system for condition analysis. Following is a discussion on the reformulation of the life rating equation to include all the operating conditions.

The life rating equation in reduced form was as follows [13] :

$$L_{10h} = a_{23} \cdot (C/P)^3 \cdot 10^6/(60 \cdot N) \quad \text{[hours]} \qquad .....12.1$$

where : L_{10h} = Rated life in hours for a failure probability of 10% (i.e.. $a_1 = 1$)

a_{23} = Combined life adjustment factor for materials and operating conditions

N = Bearing speed [rpm]

C = Dynamic load rating [kN]

P = Equivalent applied dynamic load [kN]

In this equation was embodied the effects of all the important operating conditions, namely temperature, load and speed. The effects of temperature and speed were accounted for in the combined life adjustment factor a_{23}.

The factor a_{23} was based on the so called viscosity ratio K. This ratio was an indicator of the adequacy of the lubrication under the particular operating conditions. More specifically K was defined as follows [13].

$$K = V/V_1 \qquad .....12.2$$

where : K = viscosity ratio

V = Operating viscosity of lubricant

V_1 = Required lubricant viscosity

Firstly the operating viscosity of the lubricant was determined from the equation :

$$V = C \cdot T^M \qquad .....12.3$$

where : V = Lubricant operating viscosity [mm^2/s]

T = Temperature [°C]

C = Conversion constant (920667.93)

M = Conversion exponent (-2.3450)

Equation 12.3 was derived from the temperature viscosity chart of the particular lubricant using a log regression analysis [append. 36]. In this case the lubricant was a Caltex multigrade oil. Equation 12.3 is illustrated graphically in figure 12.2.

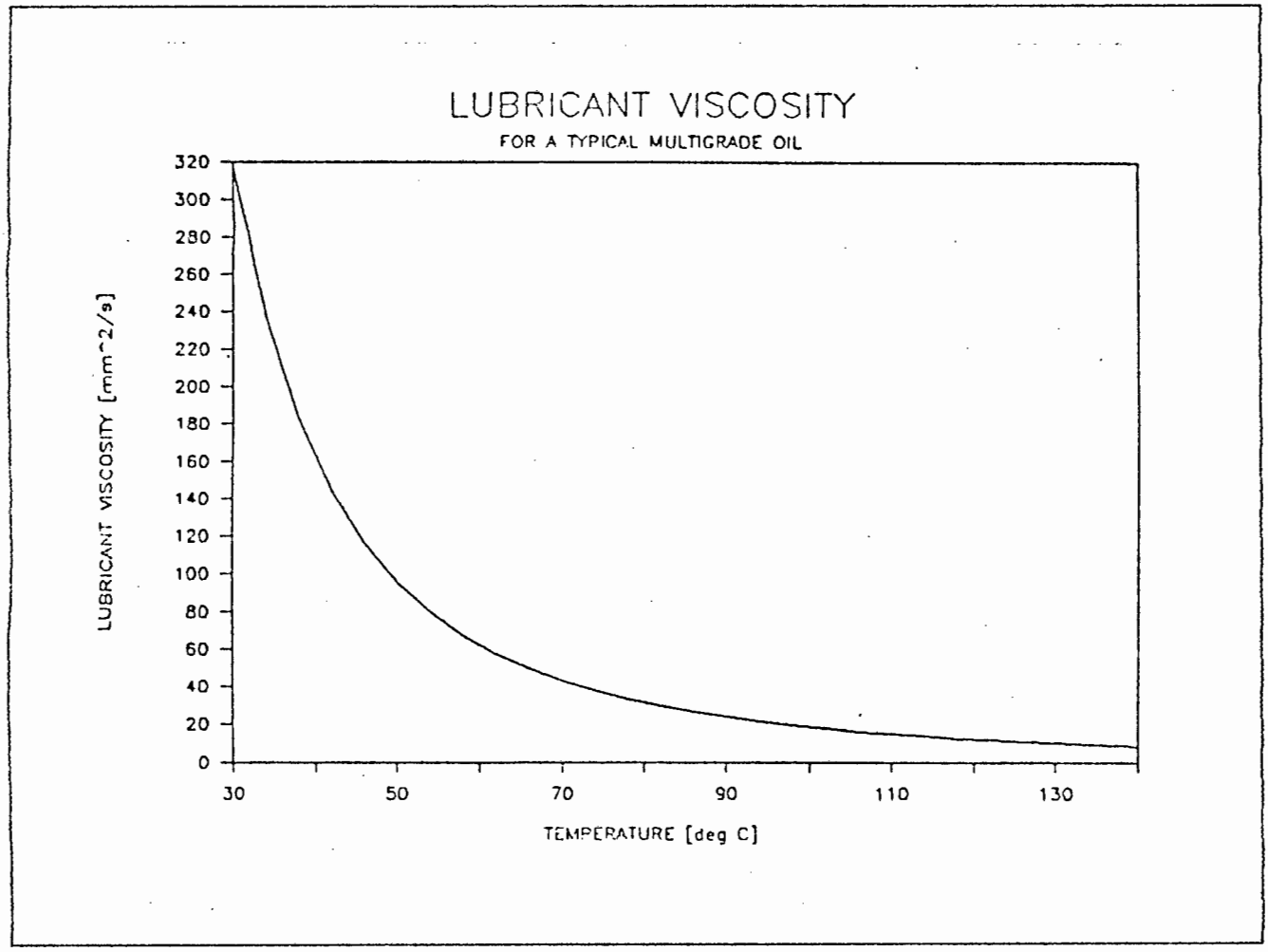

FIGURE 12.2 : LUBRICANT VISCOSITY FOR TEMPERATURE OF A MULTIGRADE OIL

As was expected the lubricant operating viscosity was highly dependent on temperature, becoming particularly thin at high temperature. The limiting temperature for the oil was about 150°C at which stage the additives would start separating from the base oil [42].

The denominator of the viscosity ratio K [eqn. 12.2] was the required lubricant viscosity. In theory this was entirely dependent on the velocity of the balls of the bearing relative to the raceways [15]. In other words it was a function of the rotational speed and the bearing mean diameter [append. 35].

The required lubricant viscosity For the test bearing was obtained from the following equation :

$$V_1 = C \cdot S^M \qquad \text{.....12.4}$$

where : V_1 = Required lubricant viscosity [mm²/s]
S = Bearing rotational speed [rpm]
C = Conversion constant (5164.168)
M = Conversion exponent (-0.8071)

These conversion factors were derived from a generalized speed-required viscosity chart [append. 37] according to the method in appendix 36. Equation 12.4 is graphically presented in figure 12.3.

This curve showed that a high viscosity was required for low speeds and a lower viscosity for high speeds. The reason for this strong dependence on speed related to the elasohydrodynamic lubricant film thickness [section 6, append. 35].

This film separated the bearing contact surfaces by hydrodynamic pressure [15]. The film would cushion the surfaces from impacts and thereby reduce spalling and increase the bearing life. Because the pressure was caused by the rolling action of the bearing balls, the required lubricant viscosity was dependent on bearing speed. Thus the life rating equation was closely related to the lubricant film thickness equation.

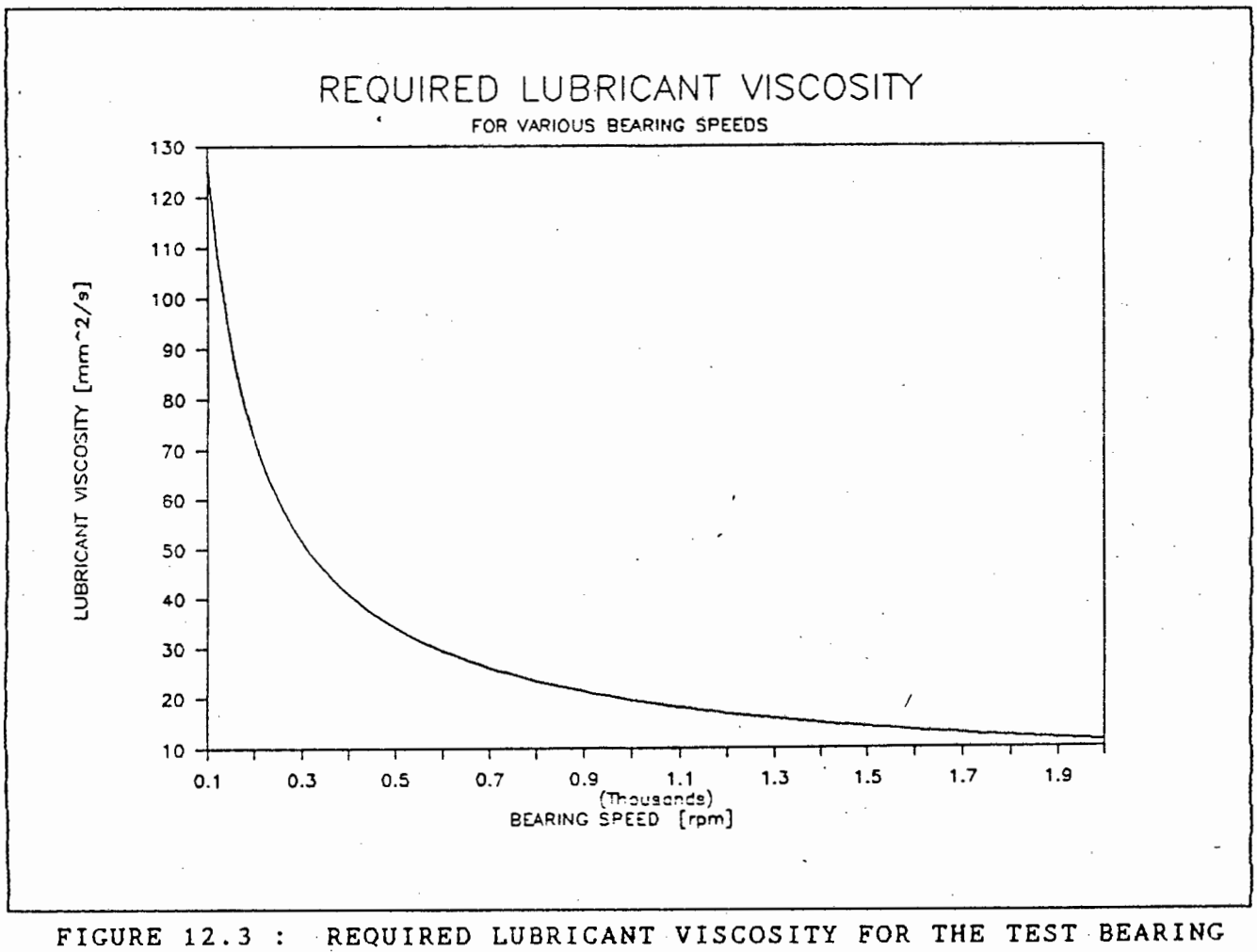

FIGURE 12.3 : REQUIRED LUBRICANT VISCOSITY FOR THE TEST BEARING

Once the operating and required lubricant viscosities had been obtained these could be used to calculate the viscosity ratio K according to eqn. 12.2. This ratio represented the performance of the lubricant for the operating conditions of bearing speed and temperature.

The K ratio was divided into four zones to assist the expert system in diagnosing the lubricant condition. These zones are illustrated in figure 12.4.

Note that figure 12.4 only applied to the specific lubricant being used, in this case a multigrade oil by Caltex. The operating viscosity of this lubricant showed less variation over a wide range of temperatures than a typical monograde oil.

The condition zones were separated by the contours corresponding to K = 0·4, 1 and 4. The critical value was represented by K = 1 where the operating viscosity was equal to the required viscosity. Values of K greater than 1 represented increasingly adequate lubrication while those below showed increasingly inadequate lubrication.

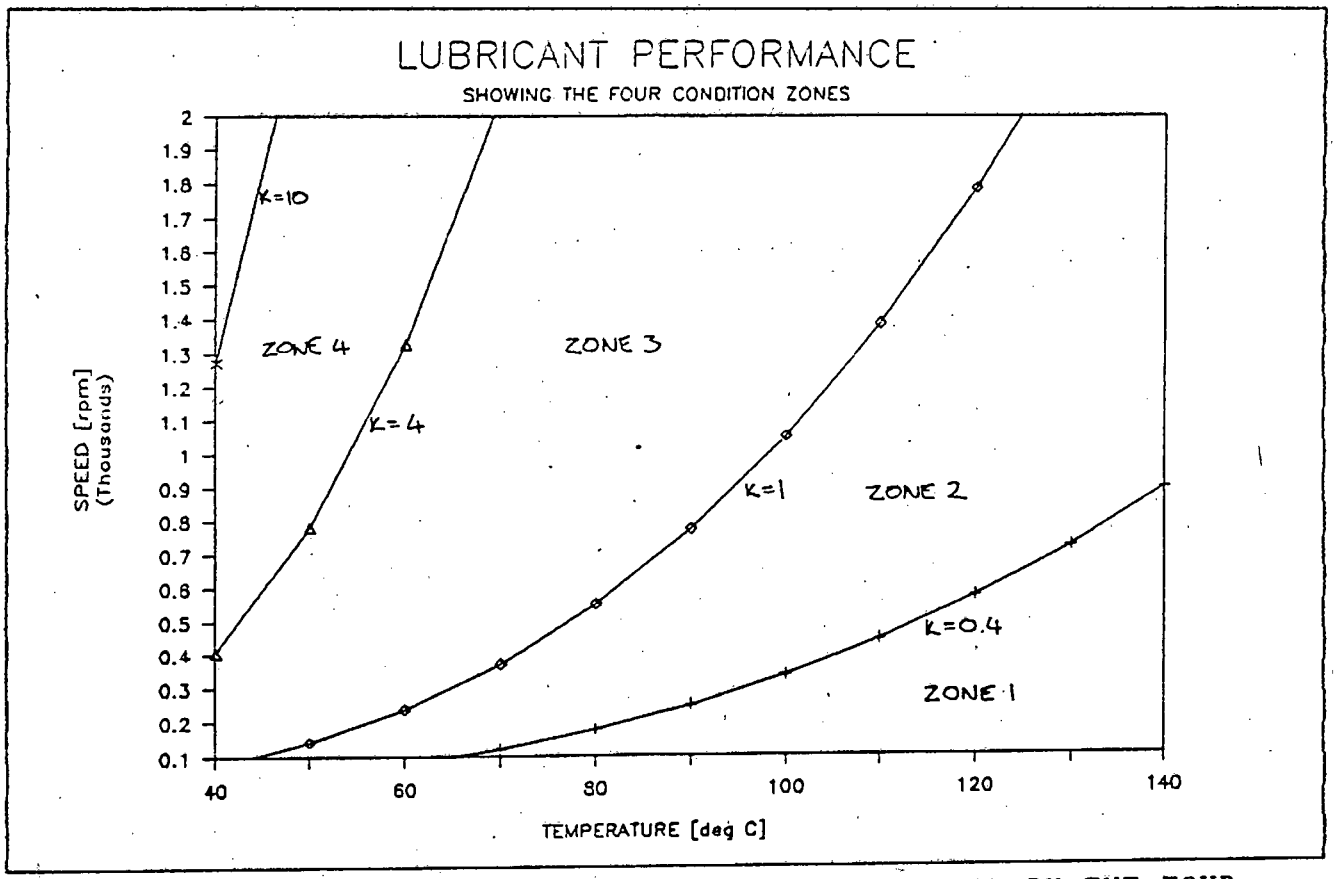

FIGURE 12.4 : LUBRICANT PERFORMANCE AS INDICATED BY THE FOUR CONDITION ZONES OF THE VISCOSITY RATIO K

Above K = 1 (zone 3) the lubricant was classified as having a satisfactory operating viscosity. This would be the zone in which bearings would most commonly operate, i.e. at speeds greater than 100 rpm and usually between 40 to 90°C [15]. At lower speeds and higher temperatures, application-specific lubricants would usually be required.

However applying high loads while operating in zone 3 could still lead to reduction of the elastohydrodynamic film thickness [append. 35]. This would be accompanied by occasional contacting of the bearing surfaces with an increase in wear leading to fatigue and spalling.

The best values for K would be between 4 and 10 (zone 3). In this zone the lubricant would provide practically continuous separation of the bearing contact surfaces at moderate loads. If the lubricant were very clean then wear would be negligible and the bearing life would be effectively infinite.

This had been demonstrated in a laboratory situation under very clean conditions with moderate loads [15]. The load would have to be lower than the fatigue limit, i.e.. that load below which the bearing material would never fail by fatigue [14].

The lubricant film could not prevent fatigue failure under high loads because it would transmit the load similarly to dry contact. This was because the pressure in the lubricant film would be very similar to the Hertzian contact stress distribution for dry contact [section 6].

For values of K above 10 the lubricant would be too viscous. It would therefore heat up due to viscous friction unless specifically cooled. This temperature rise would decrease the viscosity and thereby the viscous friction, leading to stability at a lower viscosity in zone 4 or 3.

Below K = 1 (zone 2) the lubricant could be classified as barely sufficient. Under high loads the lubricant film thickness would be small enough to allow frequent contact between the bearing surfaces. This would generate excess heat, further decreasing the viscosity and compounding the fault. Thus the bearing would have a significantly reduced life.

Finally zone 1 below K = 0·4 represented very poor lubricant viscosity. Contact of the bearing surfaces would be almost continuous and the bearing would not last for long. These extreme operating conditions would require special lubricant additives for low speeds and high temperatures.

It was evident from this discussion that the life of the bearing was heavily dependent on the lubricant performance and therefore on K. Hence the life adjustment factor a_{23} was directly related to K.

The relationship between K and a_{zs} was represented by four equations, each for their own specific K zone. These equations were of the following form :

$$a_{zs} = C \cdot K^{M} \qquad \text{.....12.5}$$

where : a_{zs} = Life adjustment factor for eqn. 12.1.

K = Viscosity ratio

And the conversion factors C and M depended on the K zone of operation as shown in table 12.1.

TABLE 12.1 CONVERSION FACTORS FOR EQUATION 12.5

K ZONE	EXPONENT M	CONSTANT C
1	0.534045	0.32933883
2	1.756313	1.00008692
3	0.662682	1.00000697
4	0.000000	2.50600000

Equations 12.5 were derived from the four curves of the a_{zs} life adjustment factor chart [append. 39]. These equations are shown graphically in figure 12.5.

The last of the four curves was a constant for K greater than 4 (zone 4) as shown in figure 12.4. This showed that under ideal conditions no gain in the bearing life could be achieved by increasing the value of K above 4. This corresponded to the effectively infinite life of bearings as discussed earlier.

Thus the factor a_{zs} could be calculated from the measured values of speed and temperature of the bearing. In turn this meant that the predicted life of the bearing could be calculated using only the measurable operating conditions of bearing load, speed and temperature.

In other words the life rating equation had been reformulated to contain all the operating conditions as required. In addition the the lubricant viscosity ratio K was a useful by-product of the formulation. It was used as a measure of the lubricant adequacy for diagnosis.

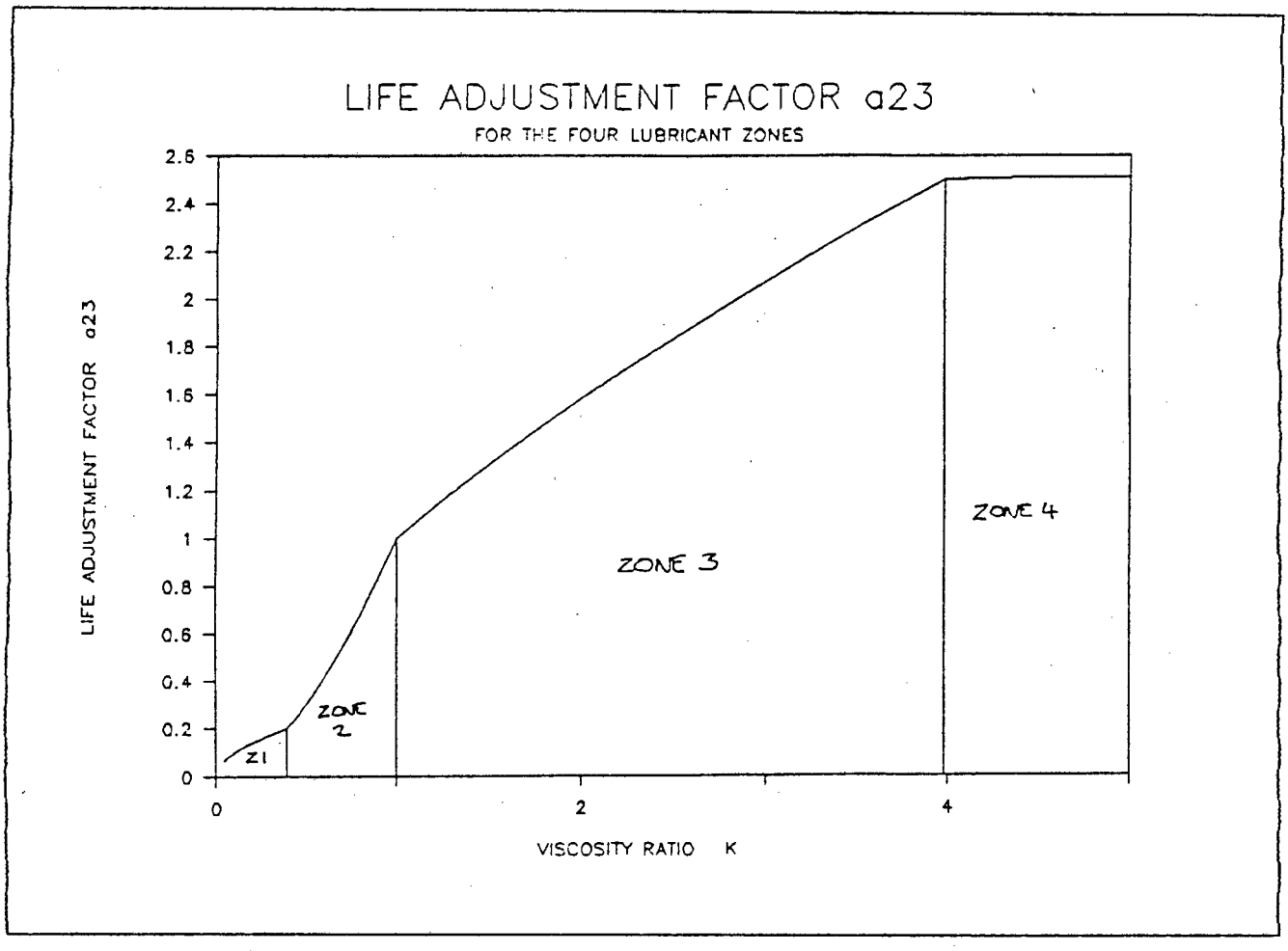

FIGURE 12.5 : LIFE ADJUSTMENT FACTOR a_{23} AS DERIVED FROM THE VISCOSITY RATIO K

The various equations derived in reaching this stage were combined [append. 40] to produce an overall equation as follows:

$$L_{10h} = C \cdot T^{M_t} \cdot N^{M_n} \cdot P^{M_p} \qquad \ldots\ldots 12.6$$

where :

L_{10h} = Bearing life [hrs.] for 10% probability of failure
N = Speed in rpm
T = Temperature in °C
P = Applied load in kN
C = Combined constant
Mi = Exponents for the operating parameters

The combined constant C and exponents Mi had to be selected according to which K zone was being operated in [fig. 12.4]. This meant that the particular value of K would first have to be calculated using eqn. 12.2. The values of C and Mi for the respective K zones are given in table 12.2.

TABLE 12.2 : CONSTANTS AND EXPONENTS FOR EQUATION 12.6

K ZONE	C (x10^6)	M_t	M_n	M_p
1	351.458	-1.25234	-0.56899	-3.0
2	602170.249	-4.11856	0.41746	-3.0
3	2078.784	-1.55399	-0.46517	-3.0
4	167.889	0.00000	-1.00000	-3.0

The importance of equation 12.6 was that the only variables needed to calculate the L_{10} life were speed, load and temperature, all of which could be measured automatically on the test rig. This meant that an estimate for the bearing life could be calculated without consulting the user, which was a desirable aspect of the expert system [section 6.4].

A graphic interpretation of equation 12.6 would be instructive of which operating parameters would have the most significant effect on bearing life. However because this equation had four variables (three independent and one dependent) no single graph would be sufficient for this purpose.

Therefore six families of curves were developed which gave a view of the equation from different angles. This was achieved by varying each of the parameters Load, Speed and Temperature in turn while keeping one of the others constant and drawing contours for the third. These curves are shown in figure 12.6 a) to f).

The interpretation of these curves was aided by considering the exponents in table 12.2. The larger the magnitude of the exponent the more significant the effect of the related parameter on bearing life. However the effect of the differing constants C had to be taken into account simultaneously.

Firstly it was evident that in most cases load was the most significant factor contributing to bearing life. This could be clearly seen on the curves of figure 12.6 a) and b) where the life dropped off rapidly with increasing load.

In addition the load had an exponent of -3 for all K zones. This was the largest magnitude by a factor of between 2 and 6 (except for temperature in zone 2). Therefore load would have to be accurately measured if a reliable extended life rating was to be obtained.

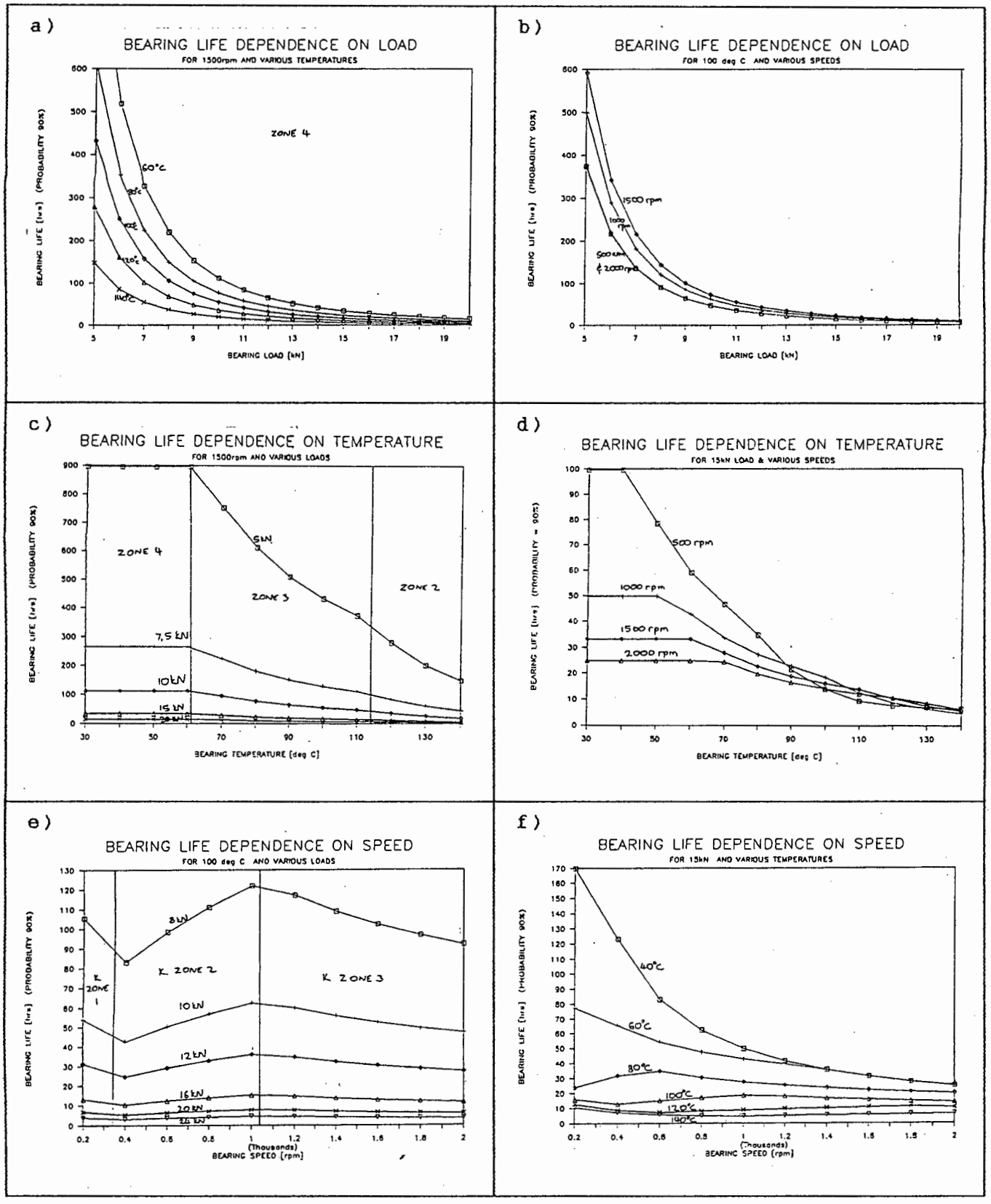

FIGURE 12.6 : BEARING LIFE DEPENDENCE ON OPERATING CONDITIONS

Temperature was the next most influential operating condition on bearing life. According to c) and d) of figure 12.6 the life dropped off constantly for increasing temperature above about 60°C. this as also evident on a) and f). The large exponent M_t for zone 2 [table 12.2] was offset by the fact that the constant C for the same zone was 2 or 3 orders of magnitude larger than for the other zones.

Finally speed had the least effect on bearing life as can be seen by the relatively small exponents M_n for the various zones. However it was interesting to note that it was only in zone 2 that an increase in speed would improve the life of the bearing.

This was because the life rating was calculated in hours and not in cycles. At high speeds the bearing would have a higher life rating in terms of cycles, but these cycles would be used up more rapidly due to the high speed.

This was better understood by considering the exponents M_n. Converting from life cycles to life hours involved adding a factor of 1 to the exponent [append. 41] which changed the normally negative value to positive in the case of zone 2. Hence increasing speed increased the life in hours for this zone.

This explained the apparent paradox of the curves in b) and d). These curves crossed over each other for temperatures above 90°C in d) and coincided for low and high speeds in b). In d) there was still a consistent trend of reduced life at higher temperatures.

The crossing over of the curves in d) was due to a more severe reduction of life (slope of the curve) at the lower speeds, becoming gradually less severe for the higher speeds. However this increase or decrease in life hours due to speed in the various zones was still relatively less significant than the effects of load and temperature.

It was interesting to note the similarities between the L_{10} life rating equation [eqn. 12.6] and the equation for the lubricant film thickness [append. 35]. This confirmed that the life of the bearing was highly dependent on the lubricant film thickness.

The extended rating life equation 12.6 was used as part of the process to calculate the predicted life of the test bearing for use by the expert system in prognosis. This is discussed in more detail in section 13.

The next set of relationships to be derived were those between the vibrations and the operating condition. This is discussed in the following section.

12.3. THE DEPENDENCE OF VIBRATION ON OPERATING CONDITIONS

It was clear from rough experimentation that there was a definite relationship between the vibrations produced by bearings and their related operating conditions. For example the vibration levels produced by a good bearing were higher for higher speeds. Thus acceptable vibration limits would have to be defined for the various operating speeds.

To define these relationships a specific set of experiments was conducted. The aim of these experiments was to provide a mathematical formulation for each specific dependence of vibration parameter on operating condition. Once these relationships were defined mathematically they could be used to adjust the vibration parameters to the reference operating conditions.

The three measurable operating conditions that could possibly influence the vibrations were those defined earlier as load, speed and temperature. The experiments involved measuring the vibration parameters while varying each of these operating conditions in turn. The other operating conditions were kept as constant as possible in each case.

The changes in vibration parameters were then recorded whenever the specific operating condition changed by a specified amount. This process of detecting the change and recording the data was achieved automatically by the computer.

The various experiments and their results are discussed in the sections that follow.

12.3.1. THE DEPENDENCE OF VIBRATION PARAMETERS ON SPEED

The most predominant relationship between vibrations and operating conditions was the dependence of vibration on bearing speed. Therefore the first experiment was to establish the dependence of vibration parameters on speed.

This experiment involved varying the speed over the full range from 114 to 1933 rpm while collecting the vibration parameters. During the test no load was applied and the temperature remained approximately constant at 40°C.

It was decided that about 100 vibration parameter values would sufficiently cover the variation of each parameter over the speed range. Therefore the vibration parameters were calculated and stored every time the speed increased by 20 rpm. The parameters collected were RMS and Peak acceleration, Kurtosis and Crest Factor.

Firstly the results for RMS acceleration are shown in figure 12.7. Clearly there was a very large increase in the overall level of vibration with speed. The value of RMS at the highest speed was about 80 times that of the lowest speed. This was predictable because at higher speeds more energy was entering the system and would show as higher vibration levels.

Superimposed on the experimental points is a best fit curve obtained from a log regression analysis [appendix 36]. This analysis produced a polynomial of the following form :

$$P = C \cdot S^{M} \qquad \ldots\ldots 12.7$$

where :

P = RMS acceleration [G's]
S = Shaft speed [rpm]
C = 3.882×10^{-6}
M = 1.594

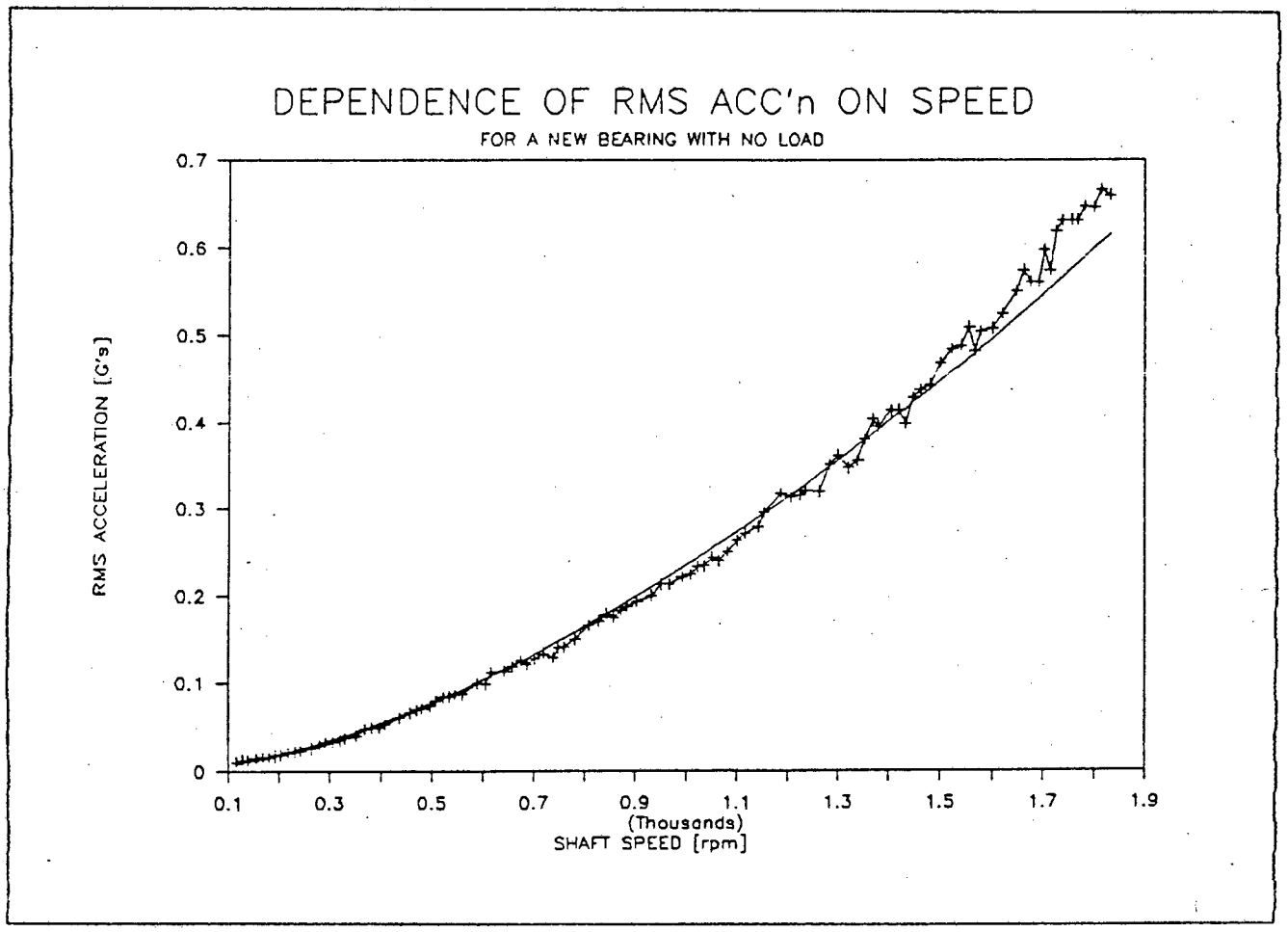

FIGURE 12.7 : THE DEPENDENCE OF RMS ACCELERATION ON SPEED

Equation 12.7 was used to adjust the RMS vibration level to account for speed variation.

Visually the experimental points showed relatively little scatter, especially at the lower speeds. However if the scatter was considered as a percentage of the best fit curve at that particular speed then the scatter at lower speeds was worse. The scatter ranged from an average of about 10% in the low speeds to about 4% at the highest speeds.

At higher speeds there was a slight upward deviation from the regression line. However this deviation was too small to affect the adjustment of the RMS values significantly.

The second parameter measured was Peak acceleration. The results for this parameter are shown in figure 12.8. Peak acceleration was tested for dependence on speed under the same conditions as the RMS acceleration. It was also shown to be heavily dependent on variations in speed. The peak at the highest speed was approximately 72 times larger than at the lowest speed.

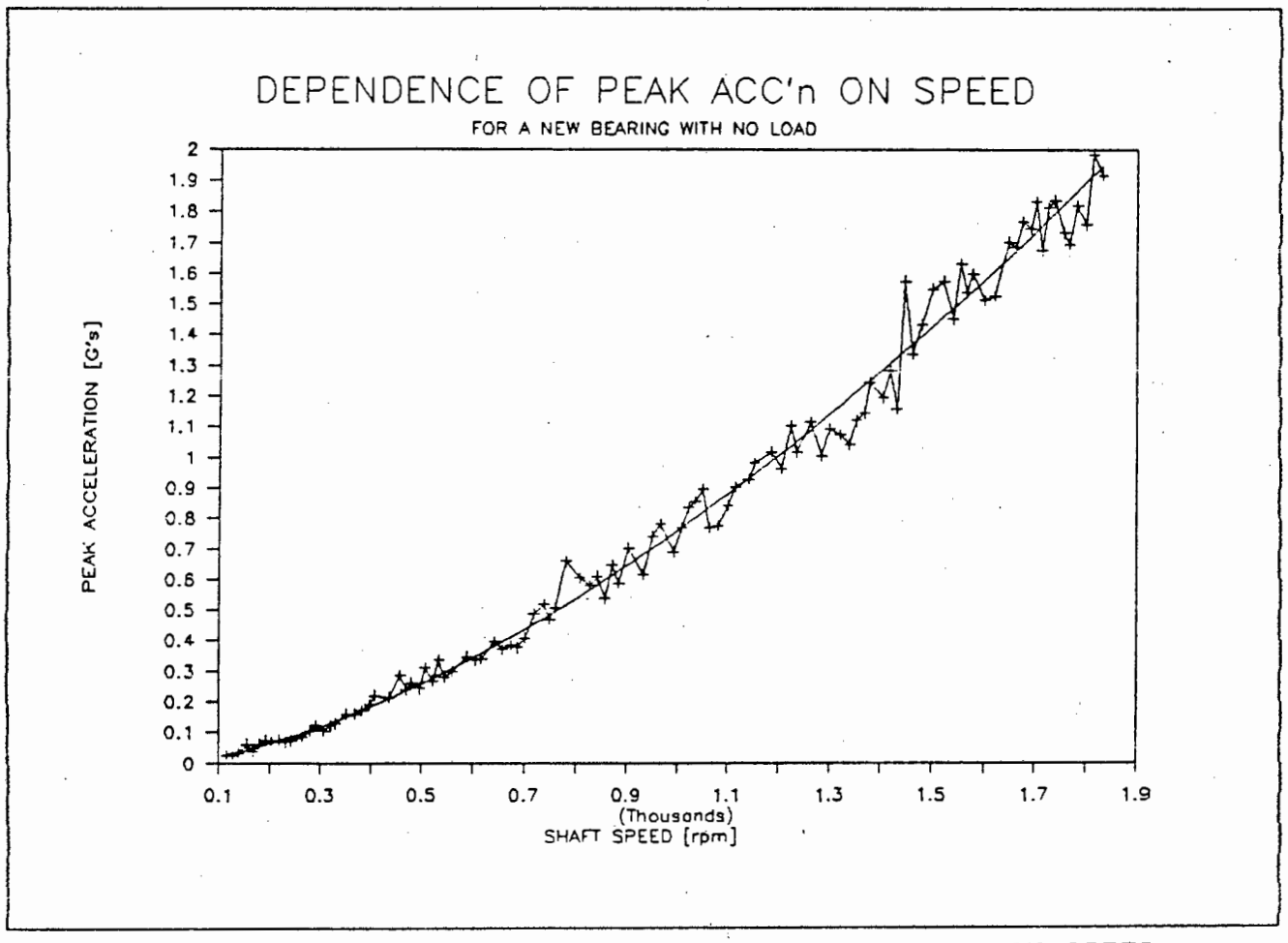

FIGURE 12.8 : THE DEPENDENCE OF PEAK ACCELERATION ON SPEED

The best fit curve through the experimental points showed a similar pattern to RMS acceleration. This was given by the following equation (derived according to the method in [append. 36]) :

$$P = C \cdot S^{M} \quad , \qquad \ldots\ldots 12.8$$

where :

P = Peak acceleration [G's]
S = Shaft speed [rpm]
$C = 1 \cdot 647 \times 10^{-5}$
$M = 1.554$

This equation was used to adjust the Peak acceleration values for speed variations.

The experimental values for Peak showed more scatter than those of RMS. This was predictable because the larger impacts would be less frequently occurring. Therefore it was possible that certain time ensembles would be captured so as to contain larger impacts than others.

It was interesting to note that the Peak levels did not rise with speed as rapidly as the RMS levels. This could also be seen by comparing equations 12.7 and 12.8. The reasons for this were probably that at higher speeds more large peaks were occurring rather than the same number of peaks at higher impact levels. This would cause the overall levels to rise faster than the peak levels.

Thirdly the dependence of Kurtosis is shown in figure 12.9. This showed that there was a barely discernible upward trend with speed. A regression line through the points only showed a 10% increase over the full range of speed. Note the suppressed zero on the ordinate.

This slight increase could be used to adjust the Kurtosis values for speed. However the scatter of Kurtosis values was large enough to make this adjustment invalid. Therefore Kurtosis was considered independent of speed which was very useful. This meant that it could be interpreted directly without adjustment for speed.

It was also noted that Kurtosis did not show more low speed scatter than high speed scatter as did the other parameters. This would make it more reliable than the other parameters for low speed diagnosis.

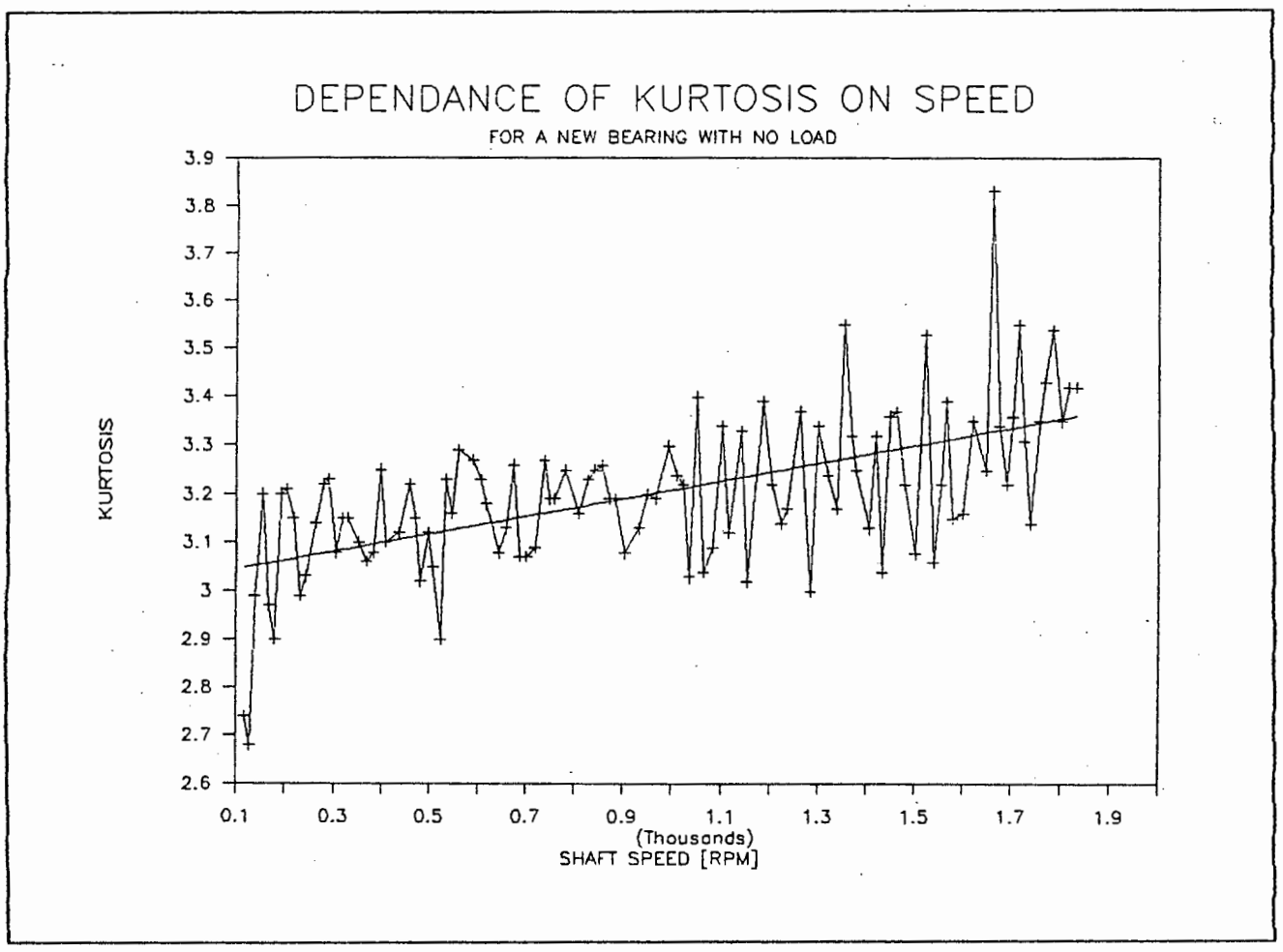

FIGURE 12.9 : THE DEPENDENCE OF KURTOSIS ON SPEED

Finally the dependence of Crest factor on speed is shown in figure 12.10. As was predictable the Crest factor showed a slight downward trend with increasing speed. This was because the Peak acceleration did not grow as rapidly as the RMS level as discussed earlier.

The Crest Factor showed a lot of scatter at the lower speeds compared with the higher speeds. This revealed an interesting aspect of the scatter of the RMS and Peak levels, namely that it was greater at lower speeds than at higher speeds if considered as a percentage of the average at that speed.

This scatter was confirmed by plotting out the percentage of scatter against speed for RMS and Peak. The scatter was particularly high at the very low speeds close to 100 rpm and more so for Peak than for RMS. This high scatter at low speeds was one of the reasons why the monitoring of machinery rotating at low speeds was found to be particularly error prone [43].

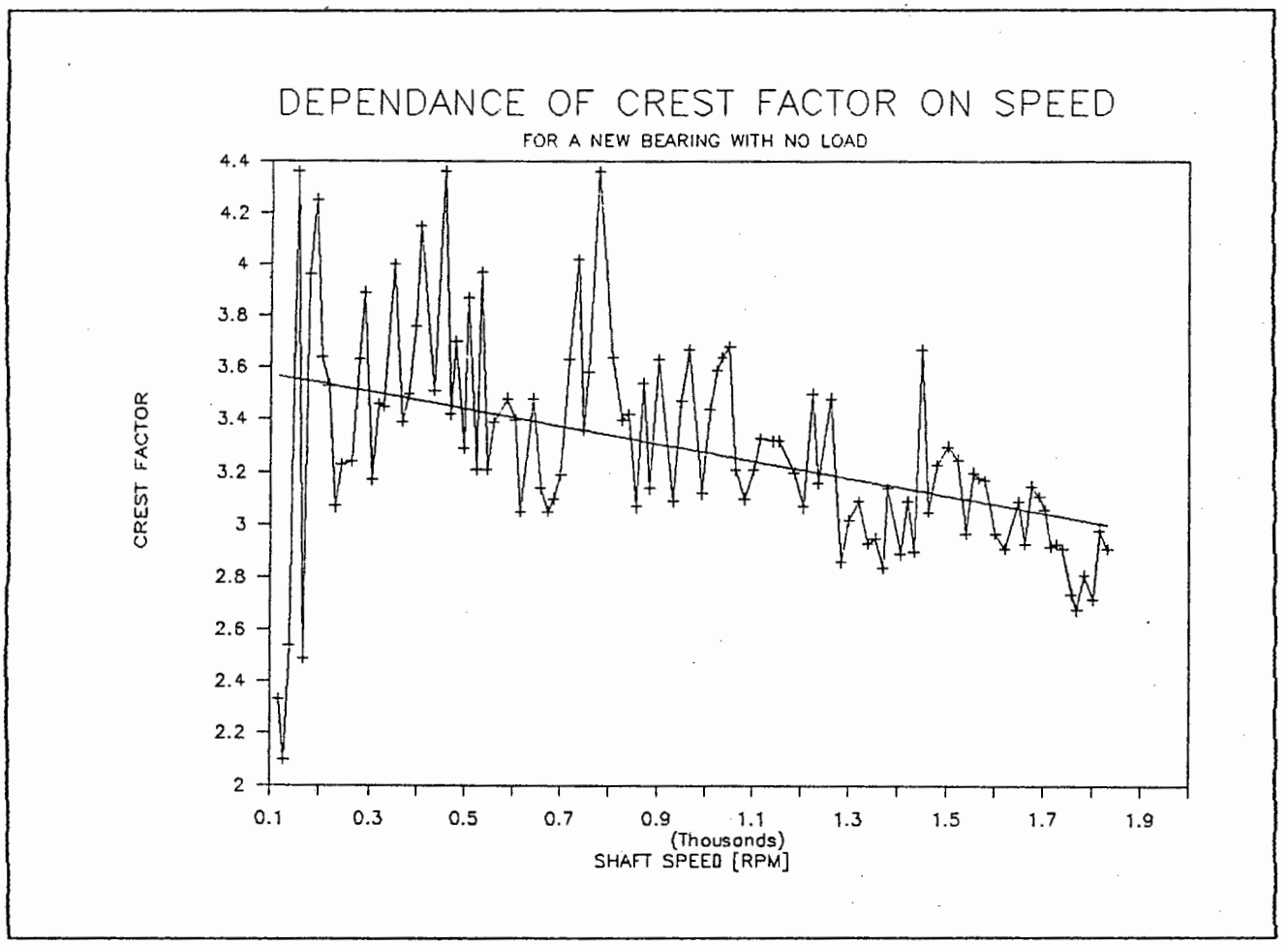

FIGURE 12.10 : THE DEPENDENCE OF CREST FACTOR ON SPEED

Thus the dependence of all the vibration parameters on speed were defined in mathematical terms. The vibrations could therefore be analyzed independently of speed once they had been adjusted for this operating condition.

The next stage was to define the dependence of vibration parameters on load. This is discussed in the following section.

12.3.2. THE DEPENDENCE OF VIBRATION PARAMETERS ON LOAD

The vibration parameters were clearly dependent on the load applied to the test bearing. While the dependence was not as pronounced as with speed, some of the parameters would have to be modified to account for load variations. A similar set of experiments were performed to formulate mathematical relationships for the dependence of the vibration parameters on load.

The method was to set the shaft rotating at full speed (1933 rpm) and apply full load (24 kN). The load was then slowly decreased by letting the hydraulic fluid escape through the partially opened return valve. The flow rate was set so as to provide the computer with sufficient time to collect the samples.

As before it was required to collect approximately 100 samples for each parameter over the whole load range from 0 kN to about 24 kN. Therefore the parameters were collected every time the load decreased by 0·2 kN.

The first parameter to be considered was the RMS acceleration. The results of this test are presented in figure 12.11.

It can be seen that the RMS level showed a fairly significant decrease for increasing load. The highest loaded condition showed approximately 24% lower RMS than the no load conditions. This was probably because once loaded the rolling elements were more constrained in movement.

However it was expected that the slope of the decrease would gradually become shallower for the higher loads. This was because the vibrations could not be decreased indefinitely with increasing load.

There was a perceptible shallowing out of the curve for higher loads but it was found that over this load range a straight line provided a better fit than a curve. This was enhanced by the flatter trend for the first 5 kN.

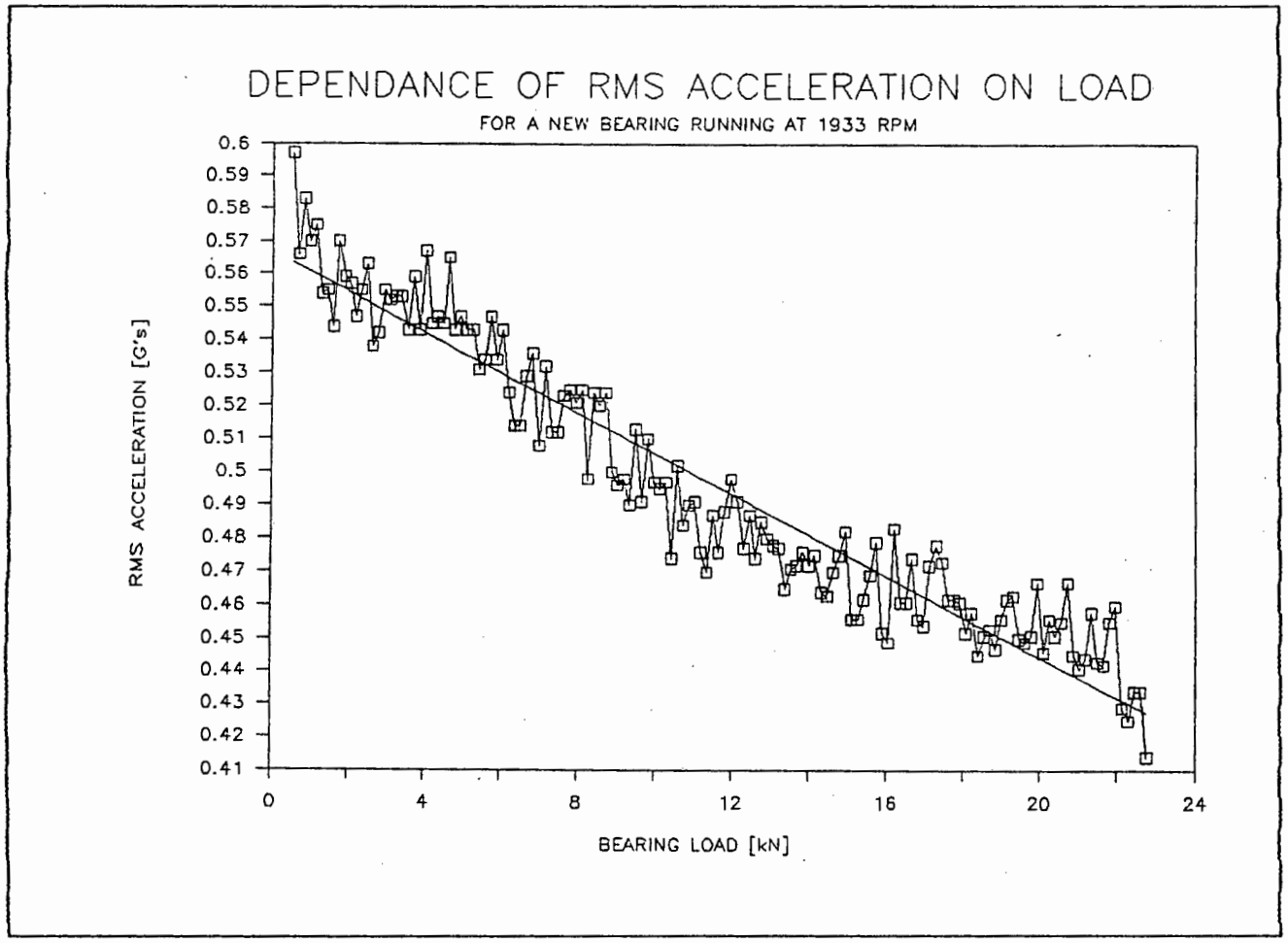

FIGURE 12.11 : THE DEPENDENCE OF RMS ACCELERATION ON LOAD

Considering that the diagnosis would be confined to this load range it was decided to use the best fit straight line. The equation for this line was as follows (calculated using a least squares regression analysis on a LOTUS® spreadsheet) :

$$V = M \cdot P + C \qquad \ldots\ldots 12.9$$

where :

V = RMS acceleration [ms^{-2}]
P = Applied load [kN]
M = $-6 \cdot 11 \times 10^{-3}$
C = 0·567

This equation is shown as a straight line in figure 12.11 and was used to adjust the RMS levels to account for load variations.

The Peak acceleration showed similar trends to RMS but not as pronounced. The dependence of peak acceleration on load is shown in figure 12.12. The level at the highest load was about 17% lower than with no load.

The scatter of peak values was probably large enough to render the adjustment of load ineffective. However it was decided to pursue this because it would make for a uniform processing format for Peak and RMS values. The equation for adjusting Peak values for load was as follows:

$$V = M \cdot P + C \qquad \ldots\ldots 12.10$$

where :

V = Peak acceleration [ms^{-2}]
P = Applied load [kN]
$M = -1 \cdot 13 \times 10^{-2}$
$C = 1 \cdot 528$

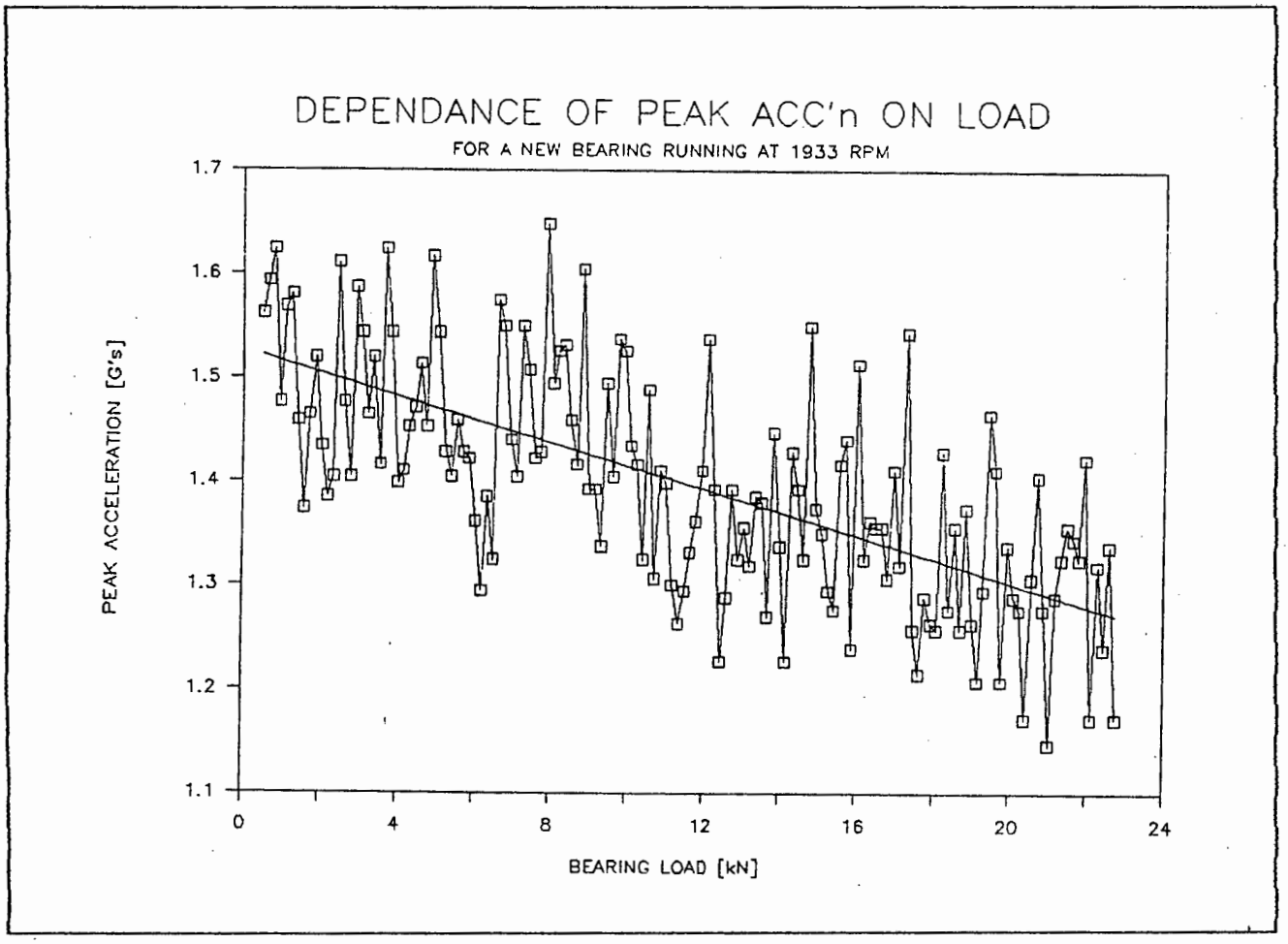

FIGURE 12.12 : THE DEPENDENCE OF PEAK ACCELERATION ON LOAD

Thirdly the Kurtosis showed little variation with increasing load. There was a tendency for the signal to become slightly less impulsive with increasing load, especially above the load rating of the bearing (approx. 16 kN). A regression line through the points only showed a 5% decrease for the whole load range as shown in figure 12.13.

Therefore it was decided that Kurtosis needed no adjusting for varying load. This meant that it could be interpreted directly without adjustment for either load or speed.

The scatter of kurtosis values for load variation was about 24% of the average value. Assuming that the bearing condition remained constant for the test (highly likely) this was rather high. The scatter meant reduced reliability when categorizing the Kurtosis values for condition analysis.

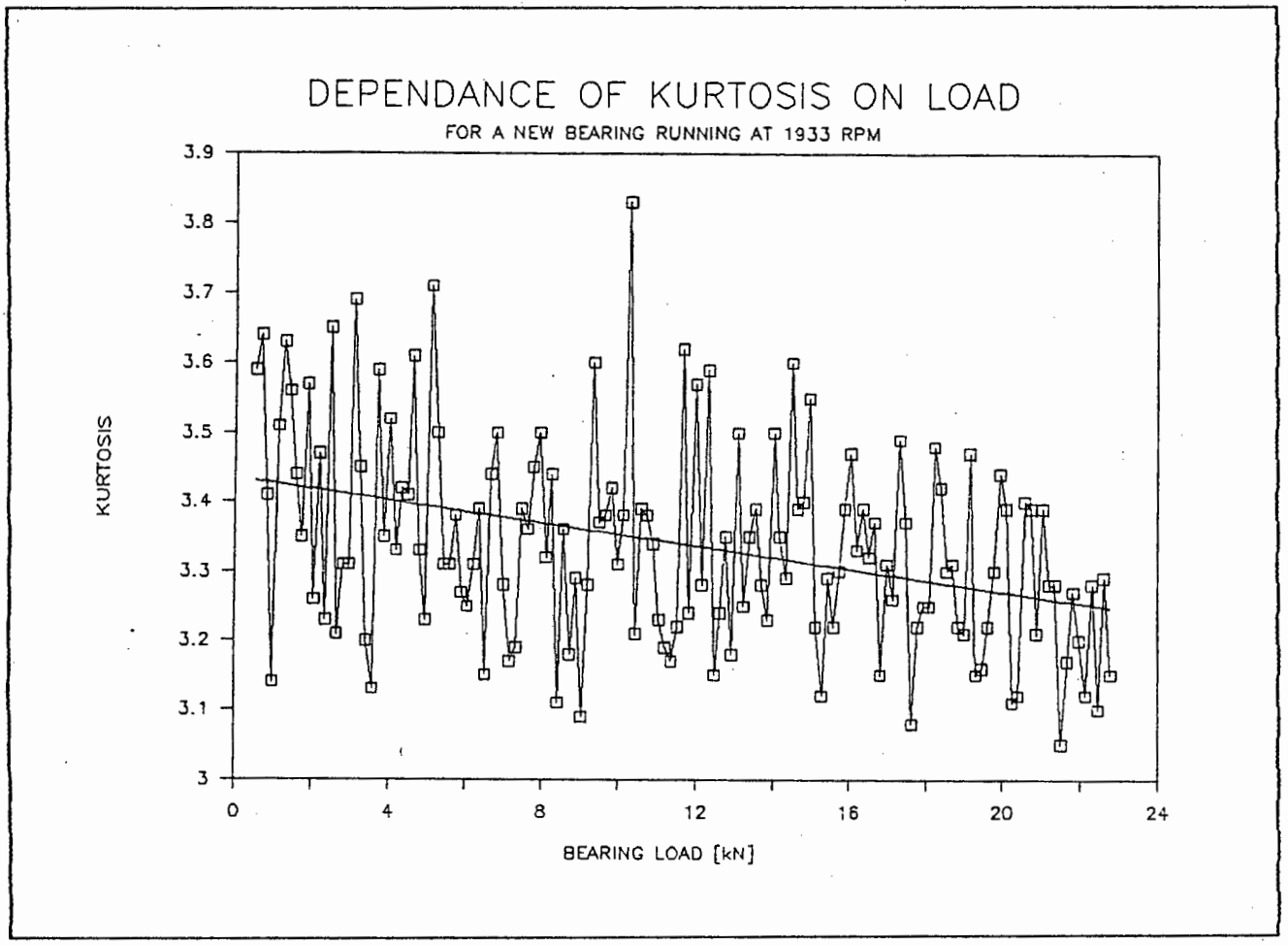

FIGURE 12.13 : THE DEPENDENCE OF KURTOSIS ON LOAD

Finally the dependence of Crest factor on load is shown in figure 12.14. Crest factor showed an upward trend with an increase in about 10%. This was expected from the results of Peak and RMS. The RMS showed a sharper decline with increasing load than Peak and therefore the Crest factor showed the upward trend.

However the Crest factor showed a large scatter about the regression line. Certainly the scatter was larger than the increasing trend. Therefore it would not be useful to adjust the Crest factor for varying load.

The Crest factor showed a scatter of about 34%. This was large and contributed to making the parameter unreliable for condition analysis.

Thus the dependence of vibration parameters on load was fully defined. The last remaining operating condition was temperature. The dependence of vibration parameters on temperature is discussed in the following section.

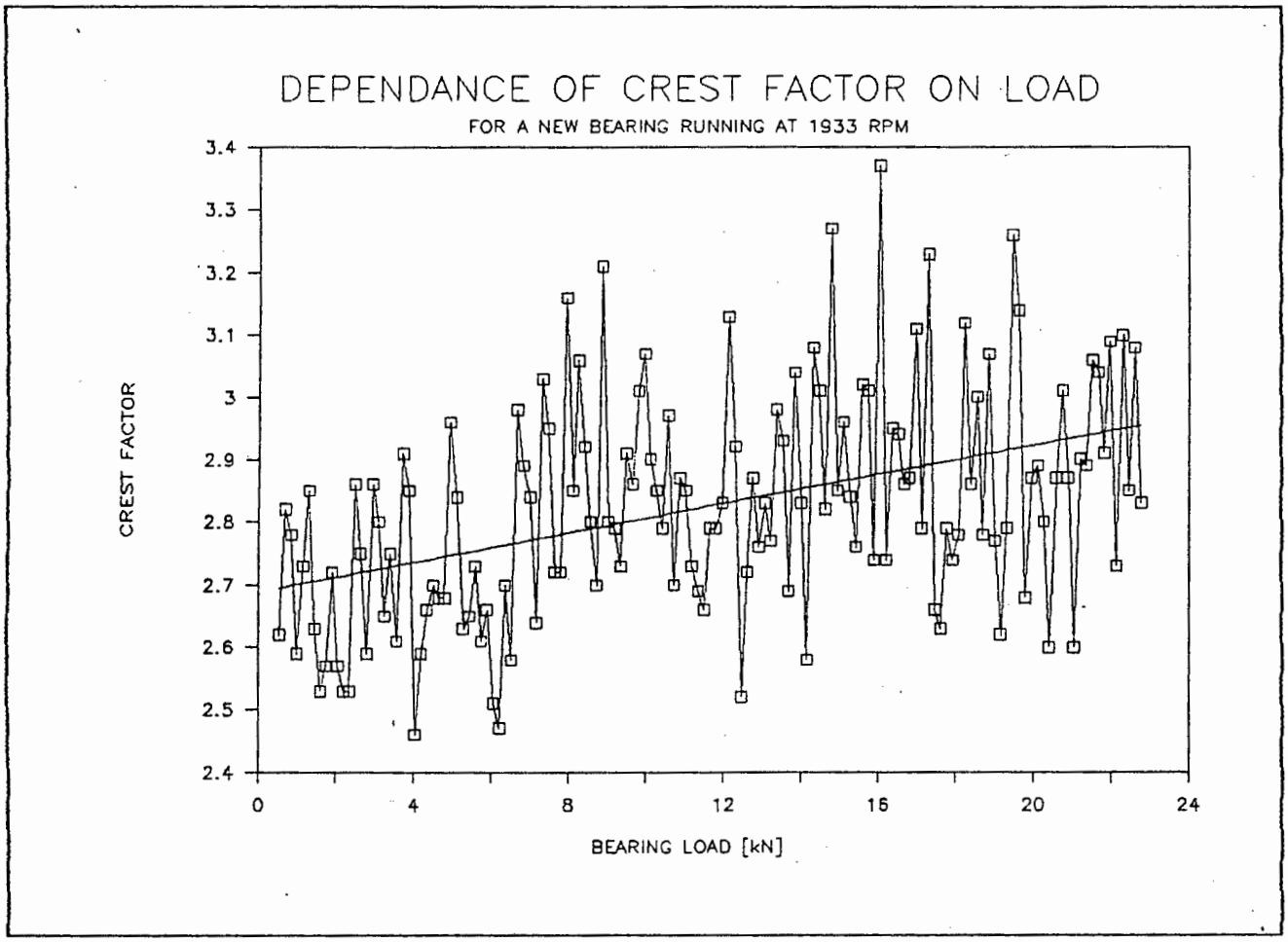

FIGURE 12.14 : THE DEPENDENCE OF CREST FACTOR ON LOAD

12.3.3. THE DEPENDENCE OF VIBRATION PARAMETERS ON TEMPERATURE

The final operating condition on which the vibration parameters were tested for dependence was the bearing temperature. This was achieved by running the test bearing under full load and speed.

At the start the bearing was at room temperature (20°C). The temperature then rose fairly rapidly, levelling off at about 110°C. The heating up process was slow enough to allow for the vibration parameters to be calculated. The vibration parameters were then stored every time the temperature rose by 1°C. This enabled the collection of about 100 values for each parameter.

The results of this experiment are shown in figure 12.15. The lower, middle and upper traces represented the RMS, Peak and Kurtosis values respectively. It was clear from these traces that there was no definite relationship between the bearing temperature and vibration parameters. Therefore no adjusting would be necessary for varying temperature. This concluded the set of experiments to define the relationships between operating conditions and vibration parameters.

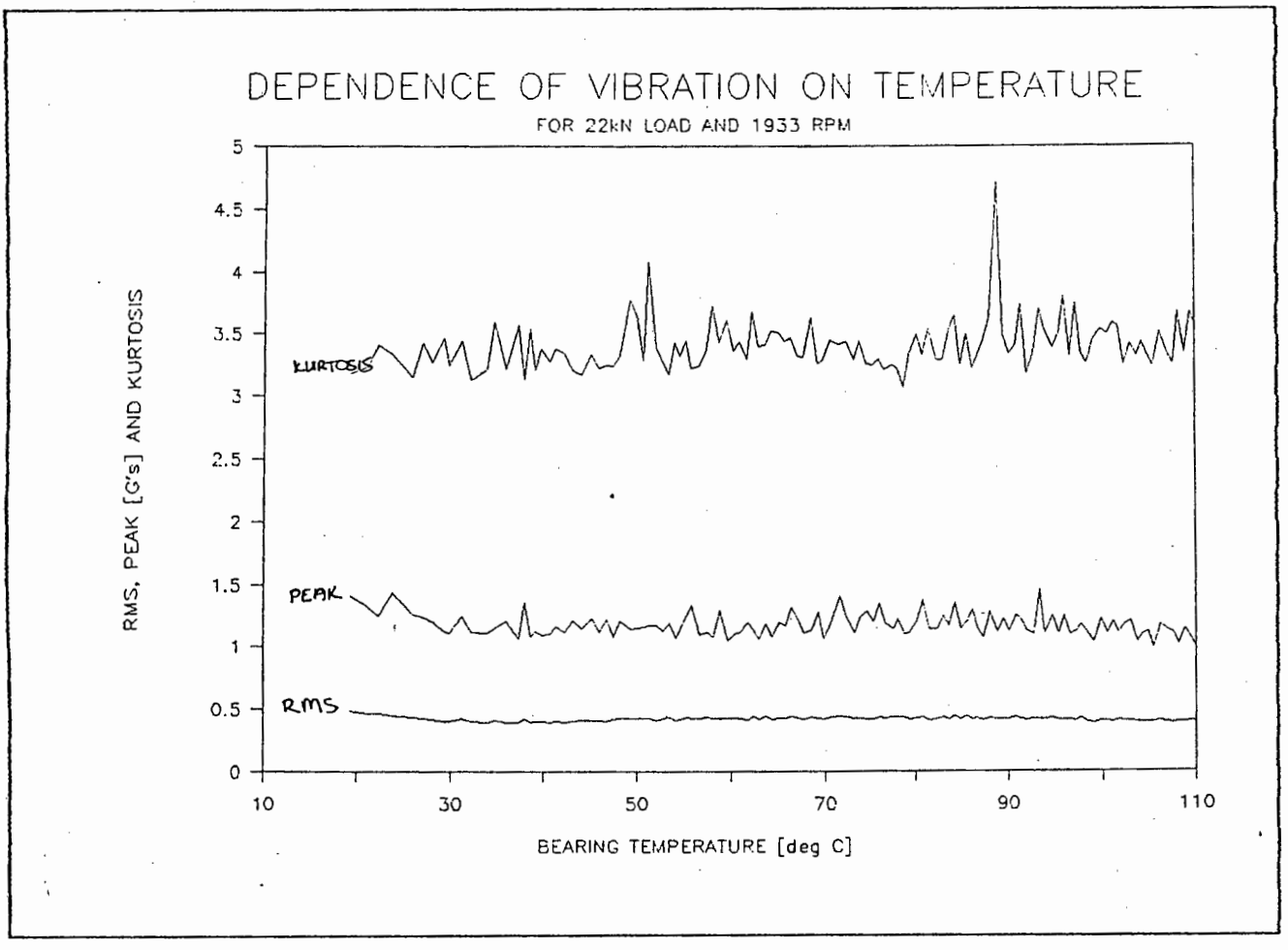

FIGURE 12.15 : THE DEPENDENCE OF VIBRATION PARAMETERS ON TEMPERATURE

12.4. THE DEPENDENCE OF BEARING CONDITION ON VIBRATIONS

The second group of experiments were used to define the inter-subset relationships between bearing condition and vibration parameters. This involved measuring the vibration patterns of a bearing experiencing fatigue failure.

The method was to systematically destruct test bearings under specific and constant operating conditions (the reference operating conditions). The various vibration parameters were collected for the duration of the useful life of the bearings. This data could then be used as a baseline for diagnosis of the condition of other bearings.

However it was difficult to measure the vibration parameters' dependence on bearing condition in the true sense of the word. For the bearing condition to be the independent variable it would have to be known accurately at all stages of the test. It would involve continuous stopping of the test to open the bearing up and give a subjective diagnosis of bearing condition by visual inspection. This would be overly time consuming and could affect the failure of the bearing.

Therefore it was necessary to infer the bearing condition from the measured vibration parameters. Seeing that they were designed for this specific purpose, the inference would be sufficiently reliable to enable future diagnosis. This was especially so under the idealized laboratory conditions.

It was particularly important that the bearing condition be measured almost continuously for the duration of the bearing test. This was to establish the earliest possible stage at which incipient failure could be detected. Continuous measurement would also provide better pattern recognition ability for future diagnosis. Therefore the measurement system was designed to provide updating of the vibration parameters approximately every two seconds.

The first experiment was used to check the operation of the data capture and processing system and to establish rough limits and ranges for the vibration parameters. The bearing was destroyed under a load of 22kN and shaft rotational speed of 1933 rpm.

The rated life of this bearing for 10% probability of failure was 3 hours [append. 42] and it lasted for about 4½ hours. Failure was by extensive fatigue spalling on the outer raceway. It was interesting to note that one of the balls showed signs of spalling as well. This was unusual [7] and appeared to have been caused by a large subsurface defect on the ball.

The second bearing was tested in detail to be used as the reference pattern for a bearing failing with fatigue spalling. This bearing was subject to the reference operating conditions of 22kN and 1800 rpm. The rated life was about 4 hrs. and 20 mins [append. 42]. However this bearing lasted for 10 hrs. and 50 mins.

The method for the experiment was to apply the reference operating conditions and then leave the bearing to run until failure. The operating conditions and vibration parameters were continuously measured and calculated. This provided updating of the values approximately every two seconds. These values were then stored on the hard disk every time there was a change in levels of various parameters.

The parameters used to initiate storage were the Temperature, RMS and Kurtosis. Storage was initiated if any of the parameters either increased or decreased by a certain amount from the values at the last storage. The change in values used to initiate storage are shown in table 12.3. Storage was also initiated if a certain amount of time had passed without storage.

TABLE 12.3 : VIBRATION PARAMETER CHANGES USED TO INITIATE STORAGE ON DISK

PARAMETER	VALUE
TEMPERATURE	3.0 °C
RMS ACCELERATION	0.2 G's
KURTOSIS	0.7
TIME	5.0 mins.

This method of storing the data was used so that the amount of data stored would not be excessive. It meant that if the parameters remained constant then only the first set of values for that constant period were stored. This constituted a large reduction of essentially redundant data and also served to emphasize the period where the bearing was showing signs of incipient failure.

However this meant that graphic display of the parameter values over time needed to be interpreted according to the method used to store the parameters. The graphs showed periods of flatness interspersed with periods of fluctuation during normal running (see for example figure 12.19).

This corresponded to periods when the parameters were varying by less and more than the change values respectively [table 12.3]. Without taking cognisance of this effect the graphs would be incorrectly interpreted as showing early incipient failure.

A graphic illustration of the storage frequency was developed to indicate how the parameters were stored to disk over time [figure 12.16]. For this illustration 'storage' was interpreted as 'sampling' because the stored parameters constituted samples of the continuously measured parameters. Thus sampling frequency in this context meant the frequency of storing samples of the parameters to disk.

The graph was produced by summing the inverse of the time interval between stored samples. Thus it illustrated how often and when storage was initiated during the life of the bearing. The most conspicuous feature of this graph was the steep rise during the final hour. This showed that the highest concentration of samples was taken from this period.

It was also interesting to note that there was a relatively quiet period during the second hour of testing. This may have corresponded to the bearing settling in after an initial running in period. Also apparent was a long quiet stretch between 9 and 10 hours, the stage just before incipient failure. This 'calm before the storm' would make incipient failure detectable only in the final stage.

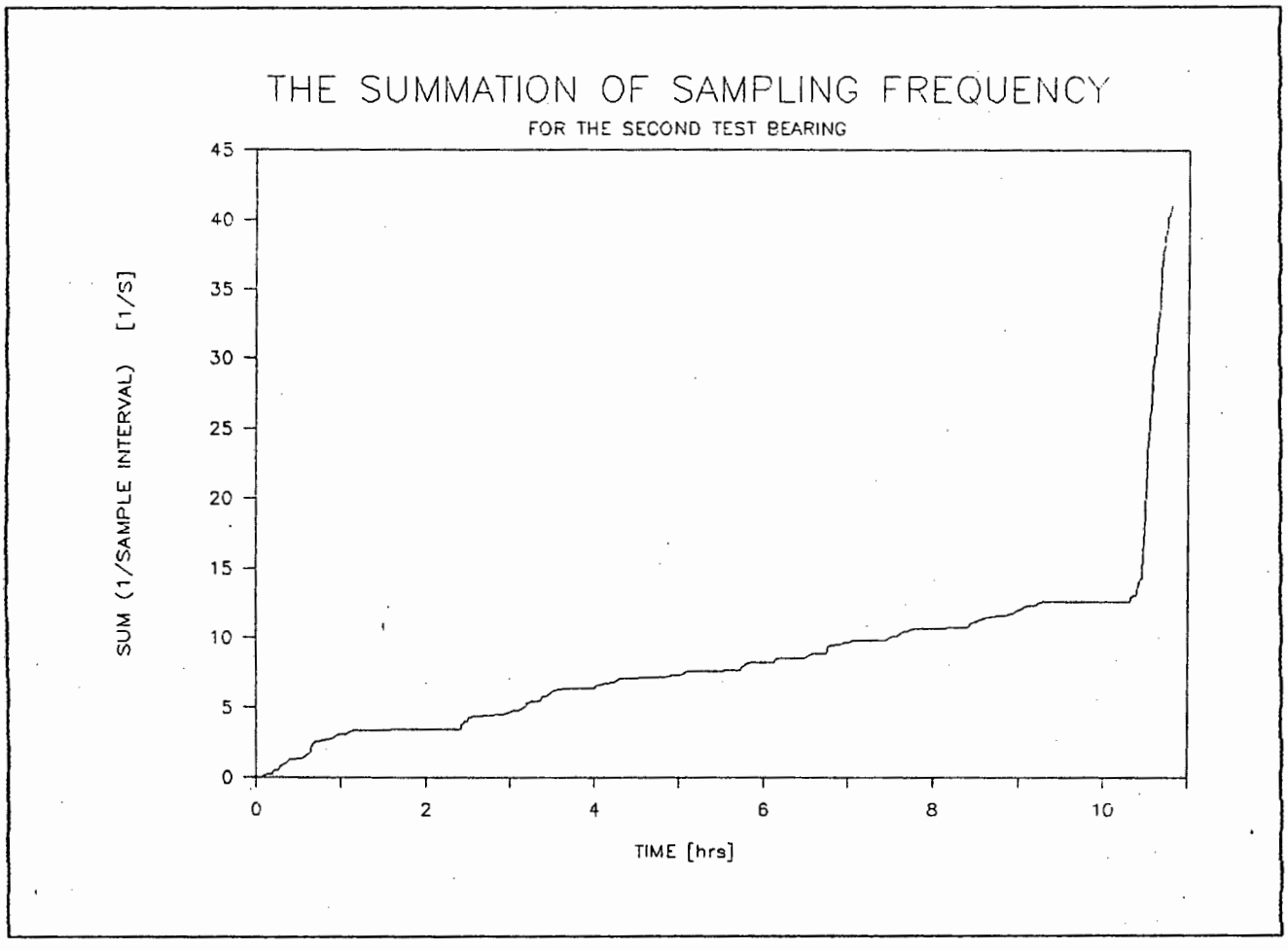

FIGURE 12.16 : THE SUMMATION OF SAMPLING FREQUENCY FOR THE 2nd TEST BEARING

The experimentation produced a number of graphs to illustrate the effect of failure on the various parameters. These were plots of the parameter magnitude over bearing running time. Graphs of the operating conditions Temperature an Load were plotted as well as each of the vibration parameters.

Firstly graphs of the operating conditions are discussed. Figure 12.17 shows the Temperature and Load on the upper and lower traces respectively. The speed was constant at 1800 and did not need to be displayed.

It can be seen that the load remained fairly constant at about 22 kN. However there was some minor adjusting between two and four hours. The test was stopped just before seven hours to check if any signs of fatigue were evident. To the naked eye the bearing surfaces were still in good condition at this stage.

The temperature trace showed a number of interesting effects. Firstly it was seen that any slight adjustment of the load had a noticeable effect on temperature. The temperature would rise or fall exponentially when the load was increased or decreased. It then stabilized out at the new operating temperature for that load.

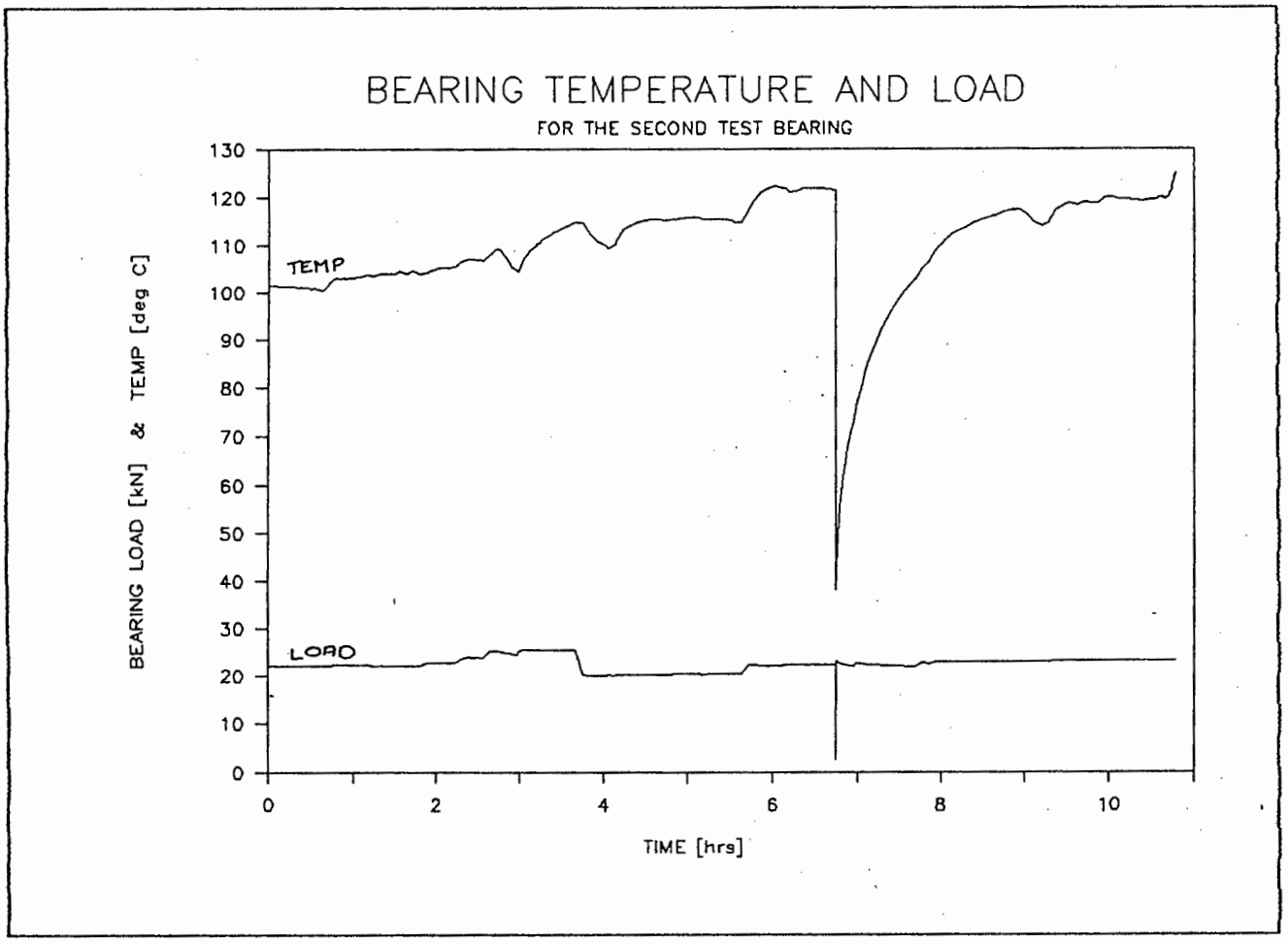

FIGURE 12.17 : OPERATING CONDITIONS FOR THE SECOND TEST BEARING

Secondly there was a slow upward trend of the temperature over the whole span of the test. The temperature started off at 102°C and rose to about 120°C. At the last stage, after about 10½ hours there was a sudden upswing to 125°C. This corresponded to the start of fatigue failure in the bearing.

The temperature for the last hour of the test is shown in figure 12.18. Superimposed on the temperature trace are 5 zones. These corresponded to various stages of degradation of the bearing and are explained later when discussing the vibration parameters.

Temperature was not a good parameter to use for direct diagnosis. This was because it was too dependent on the other operating conditions. Rather it was useful for estimating the lubricant properties. Thus no attempt was made to diagnose bearing condition directly from the temperature.

However it was interesting to note that in zone 4 corresponding to the start of spalling, the temperature showed signs of failure by increasing by about 1°C in the form of a hump. This was similar to some of the vibration parameters. The final stage of failure in zone 5 is also clearly visible.

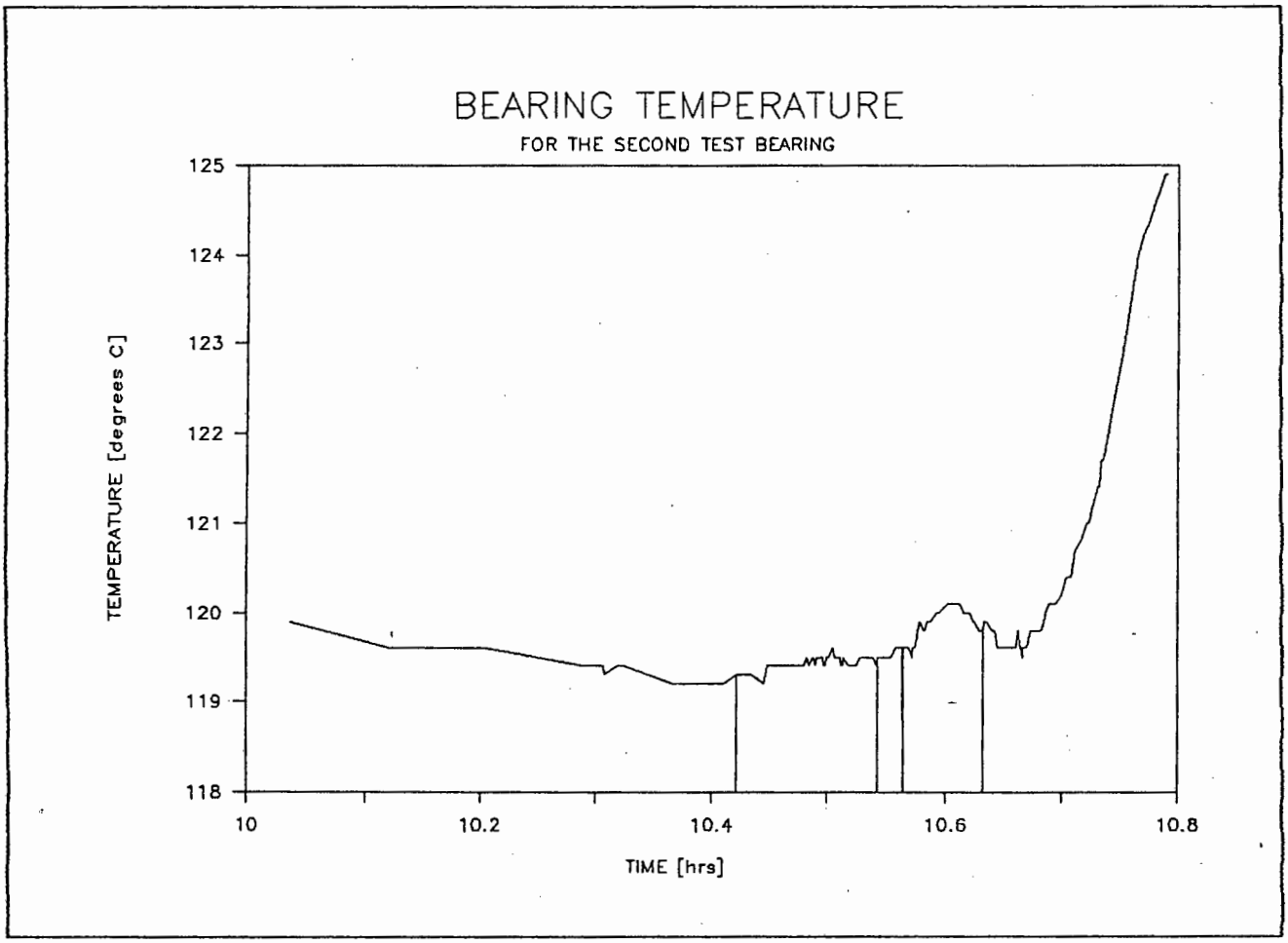

FIGURE 12.18 : TEMPERATURE FOR THE LAST HOUR OF THE SECOND BEARING

For the vibration parameters two graphs were produced in each case. Firstly one for the whole time period and secondly one for the last hour to show more detail of the final stages of failure.

The vibration RMS acceleration was the most important parameter to be measured. The graph of RMS for the whole time period is shown in figure 12.19. The RMS values were normalized by the the average value for normal running. This was done for two reasons.

Firstly it was desired to know the magnitude of the increase when failing as compared to the normal running level. And secondly the absolute values were lower than expected because of the attenuating effect of the filter.

The RMS values were seen to increase by a factor of ten over the normal running level. The failure occurred rapidly when compared with the ten hour life span of the bearing. The state of damage when the test was stopped was a severe pitting of one of the two tracks of the outer race. The bearing could possibly have run for longer, being supported by the less damaged track and its row of balls.

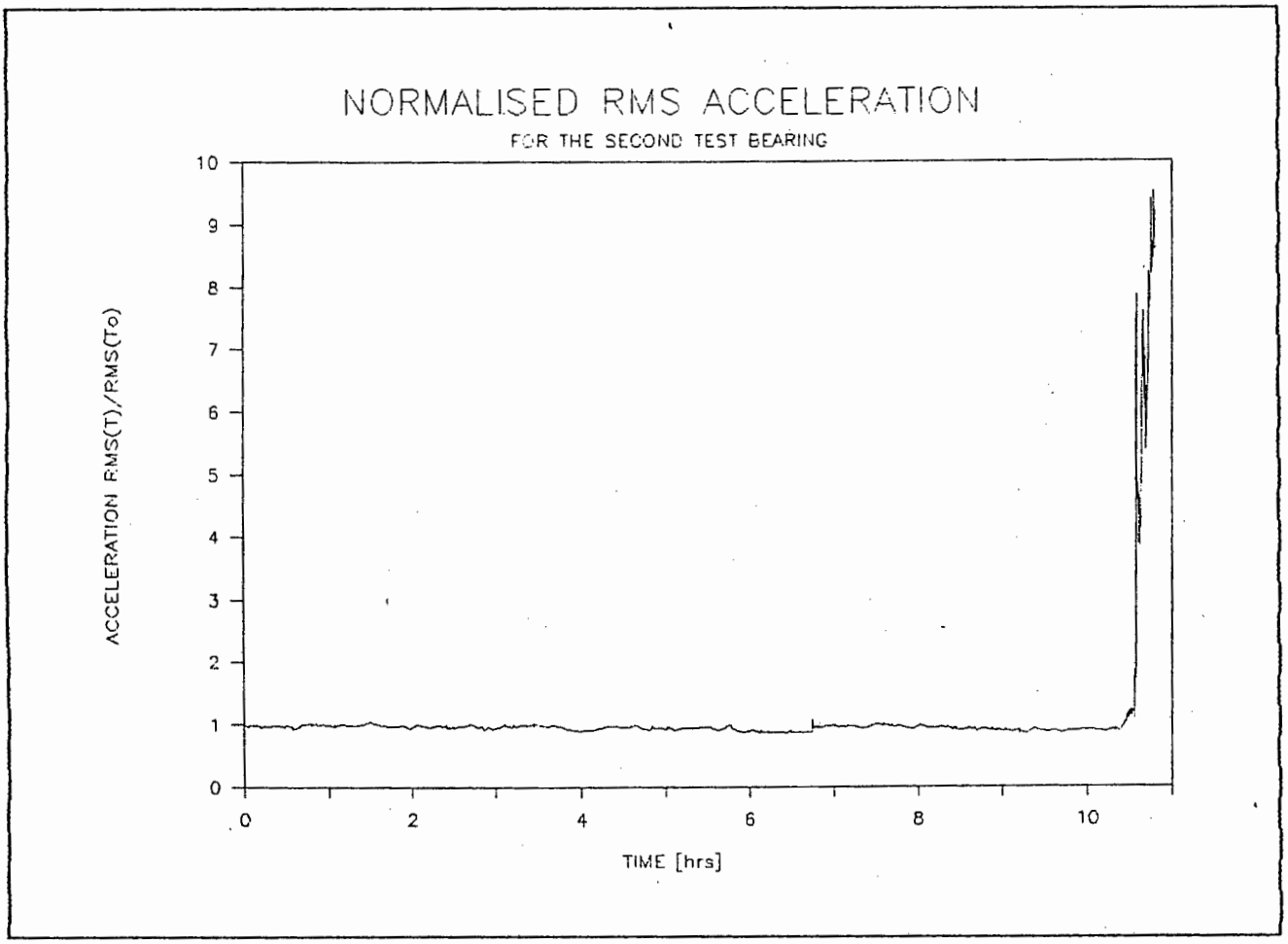

FIGURE 12.19 : NORMALIZED RMS ACCELERATION FOR THE SECOND TEST BEARING

However the test was stopped because it was decided that from an industrial point of view the useful life of the bearing had expired. The bearing may have seized or the outer race cracked very rapidly if an attempt was made to run the bearing further.

It was instructive to investigate the last hour of the test in more detail. This is shown in figure 12.20. It was found that there were five distinct groups of values of the RMS. These are shown as the five time zones in figure 12.20.

The zones were identified by close examination of the numerical values of RMS and Kurtosis as well as by knowing how fatigue of the bearing surfaces progressed. This was known by studying bearings in various stages of failure [7, 15]. It was found that failure progressed along the following path.

Firstly a smooth track or groove was worn into the bearing surface. This track usually showed discoloration due to the heating effect. During this time there would be subsurface cracks growing as each ball cycled over them. This would be particularly so in the loaded zone and also dependent on the positioning of subsurface defects [section 6].

At some stage flaking would start on the surface. This consisted of small flat flakes (approx. 10 to 50 micron diam.) as shown by the lubricant analysis [44]. These flakes would then be followed by the sudden occurrence of a single spalling pit on one of the ball tracks on the outer race. This pit would be worn smooth again by the continuous passage of the balls.

Growth of the defect would then be concentrated around this spalling pit. Usually this consisted of further pits growing directly after the single pit to form a continuous trough. This troughing effect was observed in bearings opened up at this stage of failure.

Growth of the damage then consisted of the trough widening and lengthening. The spalling flakes then contributed to similar pits and troughs growing on other parts of the race. This would be accelerated when compared with the growth of the first pit. Eventually the individual troughs would combine to form a large and continuous trough. This type of failure was used to identify the different zones on the graphs.

Zone 1 was for normal running of the bearing where no sign of incipient failure could be detected. This was recognized by the relatively small variations of the RMS with occasional small excursions returning rapidly to the normal running level. The average value of this zone was used to normalize figure 12.19.

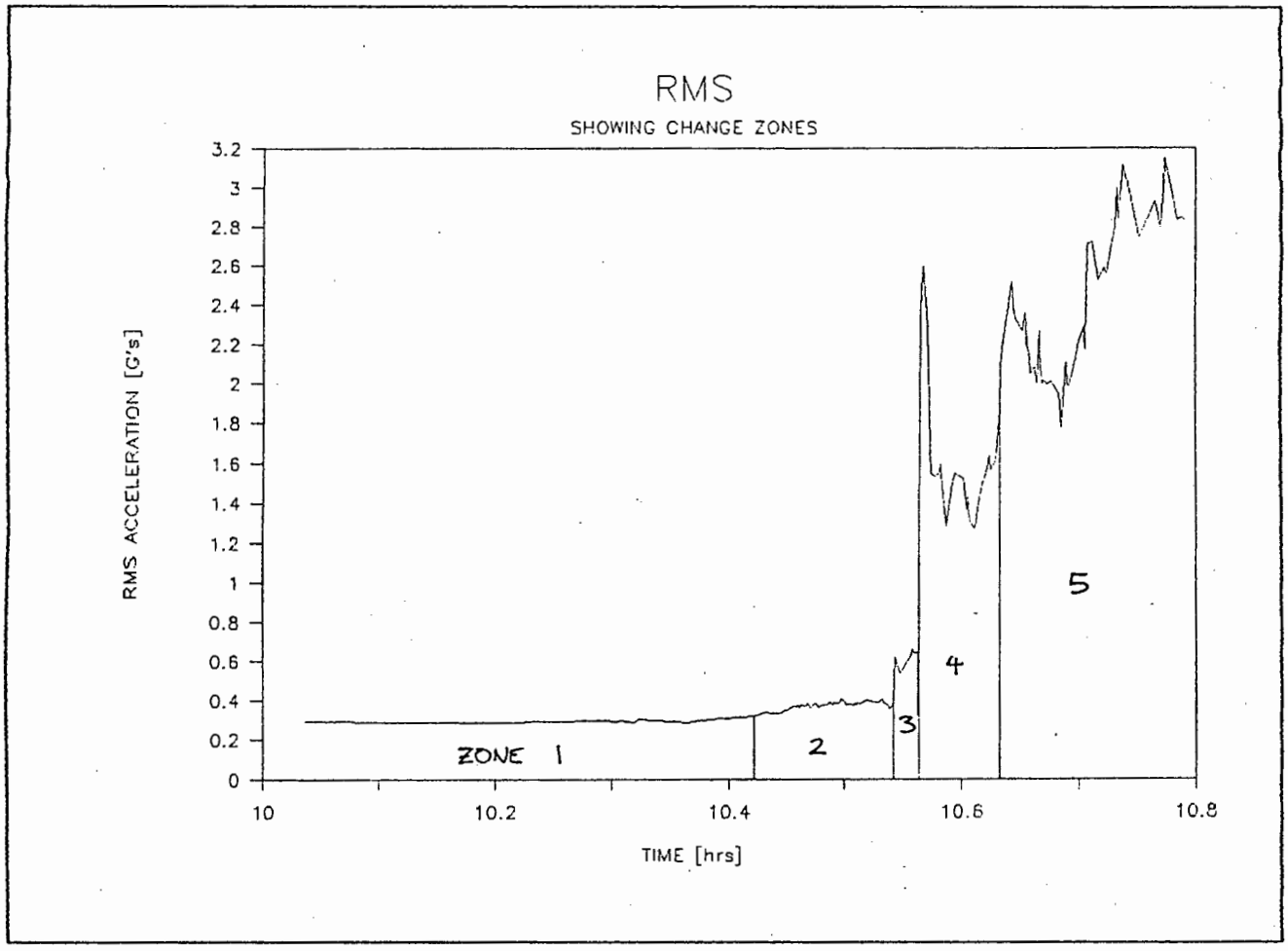

FIGURE 12.20 : RMS FOR THE LAST HOUR OF THE SECOND BEARING TEST

The first conceivable stage at which incipient failure could be detected was marked by the beginning of zone two. This was more evident in the Kurtosis [fig. 12.22]. As far as degradation of the bearing was concerned zone 2 was probably identified by scoring marks on the surface of the outer raceway and the start of small flakes breaking away from the surface as in the typical fatigue failure pattern.

Zone 2 was strongly defined by the Kurtosis showing consistently higher than normal values. However these values were not excessively high which indicated that the impacts were not as severe as they would be if spalling pits were present. This probably corresponded to the the bearing surface being roughened by initial flaking.

The start of zone 3 was clearly identified by a sudden increase in the RMS with the Kurtosis falling less suddenly. This indicated that the small flakes accelerated the production of further flaking, thereby increasing the overall vibration level but reducing the impulsiveness somewhat. Zone 3 was relatively short lived.

Zone 4 was introduced by a sharp and extreme increase in the RMS as well as the Kurtosis. This indicated strong impulsiveness and an increase in the overall vibration level. This indicated that the first large spalling pit had occurred.

Both the Kurtosis and the RMS were then seen to drop significantly after the initial peak. This corresponded to the flakes being ejected from the raceway and the spalling pit then being worn smoother. The RMS then remained fairly high for the rest of zone 4 while the Kurtosis dropped to about one or two points above normal running.

Zone 5 was indicated by a notable increase in both RMS and Kurtosis. However neither of them was as extreme as that for zone 4. Zone 5 was characterized by the sudden and continuous enlarging of the spalling pits to form the spalling trough discussed earlier.

The RMS remained continually high as well as showing a general upward trend as the trough lengthened and the vibrations became increasingly severe. The Kurtosis remained high for a relatively long period and then became variable as further spalling pits were formed and worn out or the flakes were expelled from the bearing. This growth pattern would continue until the outer race of the bearing cracked and the system failed dramatically.

The time taken for each zone along with the number of samples stored for that zone are shown in table 12.4. The time is given firstly as a percent of the time taken for the bearing to fail and secondly as a percent of the rated life of the bearing. Calculating the percent of the rated life for each zone was important for diagnostic purposes. This was because the rated life of the bearing was constant for any given operating conditions.

It can be seen that the failure zone (2,3,4 & 5) times were short when compared with the rated life of the bearing. This would mean that measurements would have to be taken often to ensure that incipient failure was detected.

TABLE 12.4. ENDURANCE TIME FOR BEARING FAILURE ZONES

ZONE	TIME				N° OF SAMPLES
	START	END	%FAIL	%RATED	
1	0.000	10.422	96.580	240.526	267
2	10.422	10.543	1.121	2.793	59
3	10.543	10.564	0.195	0.485	7
4	10.564	10.633	0.639	1.592	30
5	10.633	10.791	1.464	3.646	55

For example if a bearing had a rated life of one year then in theory zone 2 would begin 31 days before the end of the bearing life (if the bearing failed in the same way as bearing 2). Therefore measurements would have to be taken at least every month to have a chance of predicting complete failure before it occurred.

Table 12.5 shows the statistics of the RMS values for each time zone. This includes the standard deviation added to and subtracted from the average value and also the minimum and maximum values. These were used to calculate acceptable values for the RMS in each zone. All values have been normalized by the average RMS in zone 1.

TABLE 12.5 : NORMALIZED RMS STATISTICS FOR THE FIVE TIME ZONES

ZONE	AVERAGE	SDEV	AV-SDEV	AV+SDEV	MIN	MAX
1	1.000	0.042	0.958	1.042	0.887	1.130
2	1.210	0.054	1.156	1.265	1.069	1.306
3	1.970	0.125	1.845	2.095	1.706	2.116
4	5.364	1.270	4.094	6.634	3.838	8.313
5	7.711	1.198	6.513	8.910	5.685	10.080

The largest jump was clearly between zones 3 and 4. In fact for a simplified analysis the zones 2 and 3 could be grouped together as one zone and similarly for zones 4 and 5. However for the sake of this analysis and implementing expert system diagnosis, the zones were kept separate. It was also noted that there was some overlapping of the value ranges in zones 4 and 5.

The Kurtosis was considered along with the RMS when dividing the bearing running time into the various zones. The Kurtosis for the whole test is shown in figure 12.21. It can be seen that the Kurtosis clearly indicated the damage to the bearing once it started degrading, showing maximum values greater than 13.

The statistics for the five zones are shown in table 12.6. Note that an acceptable value for Kurtosis is generally regarded as being between 3 and 4.

TABLE 12.6 : KURTOSIS STATISTICS FOR THE FIVE TIME ZONES

ZONE	AVERAGE	SDEV	AV-SDEV	AV+SDEV	MIN	MAX
1	3.347	0.264	3.002	3.676	2.790	4.250
2	5.788	1.062	4.726	6.850	3.640	7.710
3	4.307	0.400	3.907	4.707	3.780	4.950
4	6.337	2.944	3.393	9.281	3.850	13.150
5	5.151	1.607	3.544	6.758	3.280	9.860

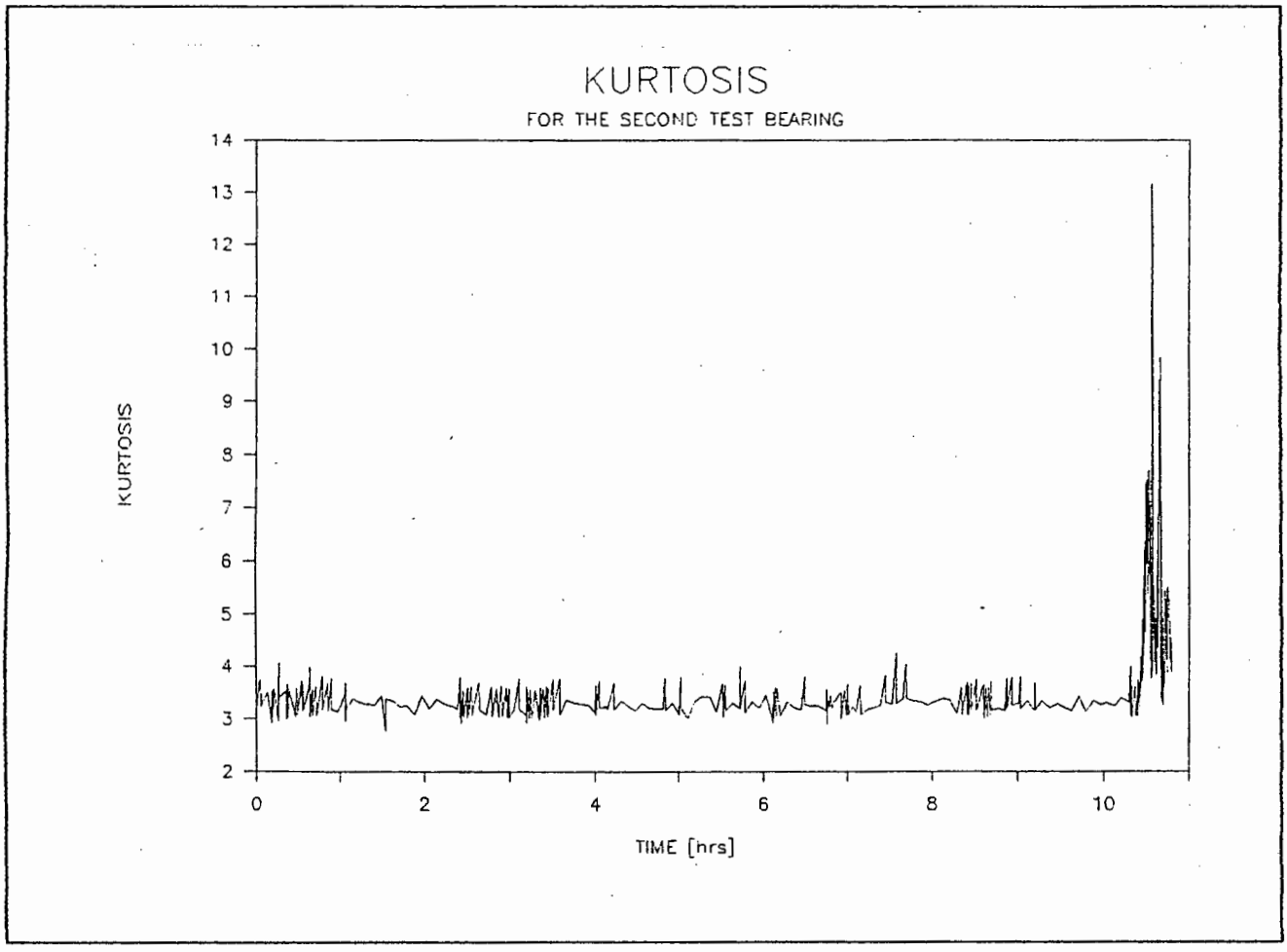

FIGURE 12.21 : KURTOSIS FOR THE SECOND TEST BEARING

Note that in zone 1 shown by the main part of figure 12.21, the fluctuating sections interspersed by the relatively flat sections were due to the storage algorithm and not sudden changes in the bearing condition (as discussed earlier).

The expanded graph of the last hour of the test is shown in figure 12.22. This was a most interesting graph with a number of features needing discussion. Firstly it was clear that the Kurtosis gave a clearer and more definite warning of incipient failure than RMS. This can be seen by the consistently high values in zone 2.

In fact the average value for zone 2 was 5·788 [table 12.6] which clearly indicated the incipient failure in this zone. This value represented a near doubling of the Kurtosis as compared to the small 1·2 times increase of the RMS in the same time zone [table 12.5]. It was this initial warning ability which made the Kurtosis particularly useful as a diagnostic parameter.

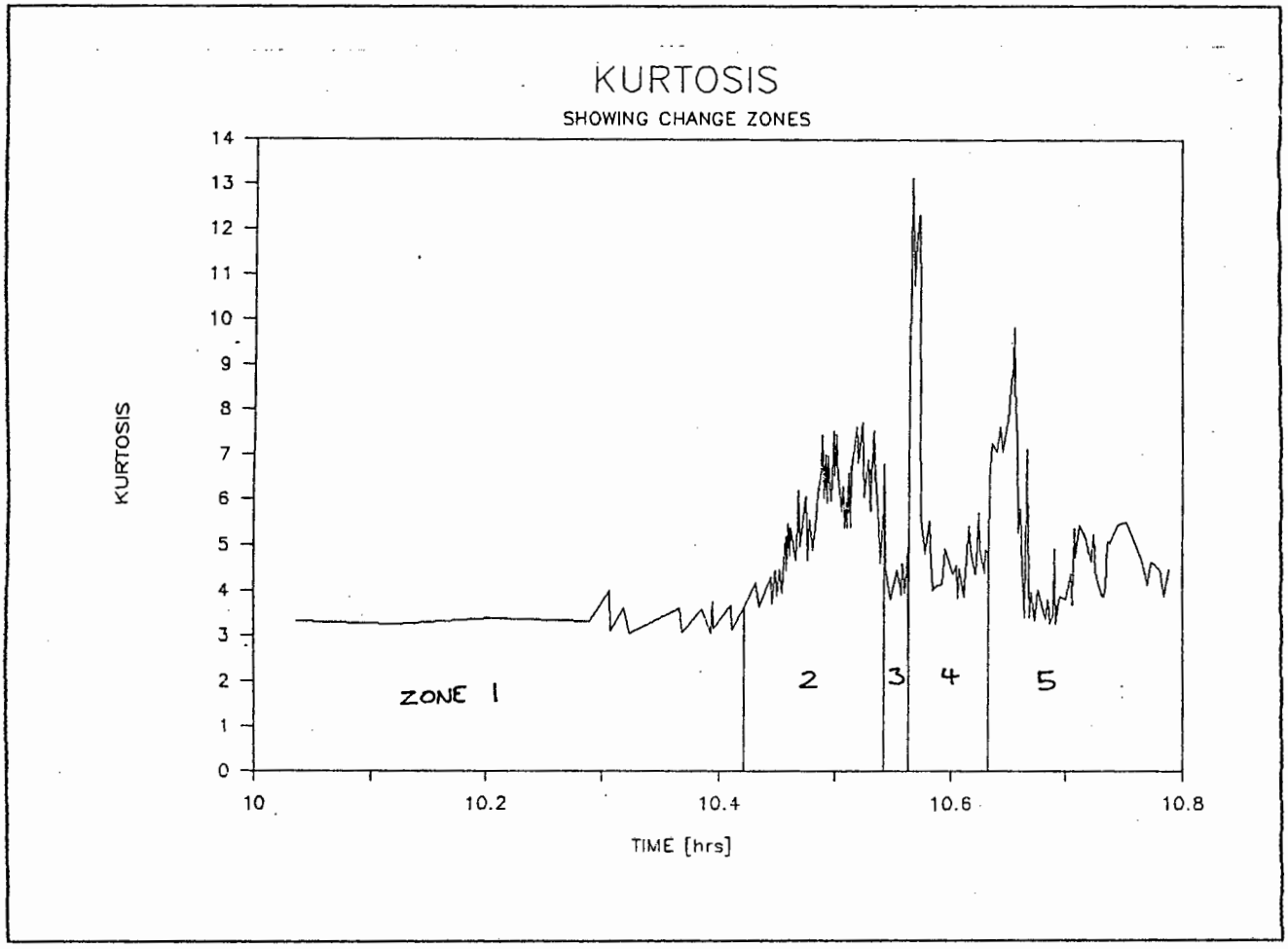

FIGURE 12.22 : KURTOSIS FOR THE LAST HOUR OF THE SECOND TEST

The second interesting and clearly distinguishable feature of Kurtosis as shown by figure 12.22 was the sudden increase at the beginning of each time zone. The peaks at the beginning of zones 4 and 5 clearly showed that the waveform was particularly impulsive at this stage. This led to the conclusion that these zones were introduced by the formation of new spalling pits on the raceway.

The rapid decrease of impulsiveness after this could be explained by two possible mechanisms. Firstly the large spalling flakes being ejected from the raceway would reduce the impulsiveness. And secondly the cycling of balls over the edge of the newly formed spalling pits would smooth them down rapidly due to the high stress concentrations at these points. The exact mechanism was not important and probably consisted of some combination of the two.

What was more important was the fact that in the zones of advanced damage the Kurtosis did not stay high for long and soon dropped to almost acceptable levels. This emphasized the fact that Kurtosis could not be used on its own for diagnosis but should nevertheless be used for early warning.

The next parameter to be considered was the Peak acceleration. The normalized values of Peak for the full test are shown in figure 12.23. Damage in the bearing was clearly indicated by a sharp rise in the peak values towards the end of the test. It was also noted that during normal running the Peak was more variable than the RMS.

The most interesting portion of the test was the last hour and this is expanded out in Figure 12.24. The peak showed a very similar pattern to RMS for this period, the main difference being the greater random variability of Peak compared to the RMS. Another clear difference was the less significant increase at the start of each zone, particularly for the fourth and fifth zones.

The normalized statistics for Peak in the five zones are shown in table 12.7. These statistics confirmed the similarity between the Peak and RMS acceleration. It was noted that in general the Peak values did not increase as dramatically as the RMS values and that the upward trend was more continuous than for RMS. This could be seen in the fourth and fifth zones in particular.

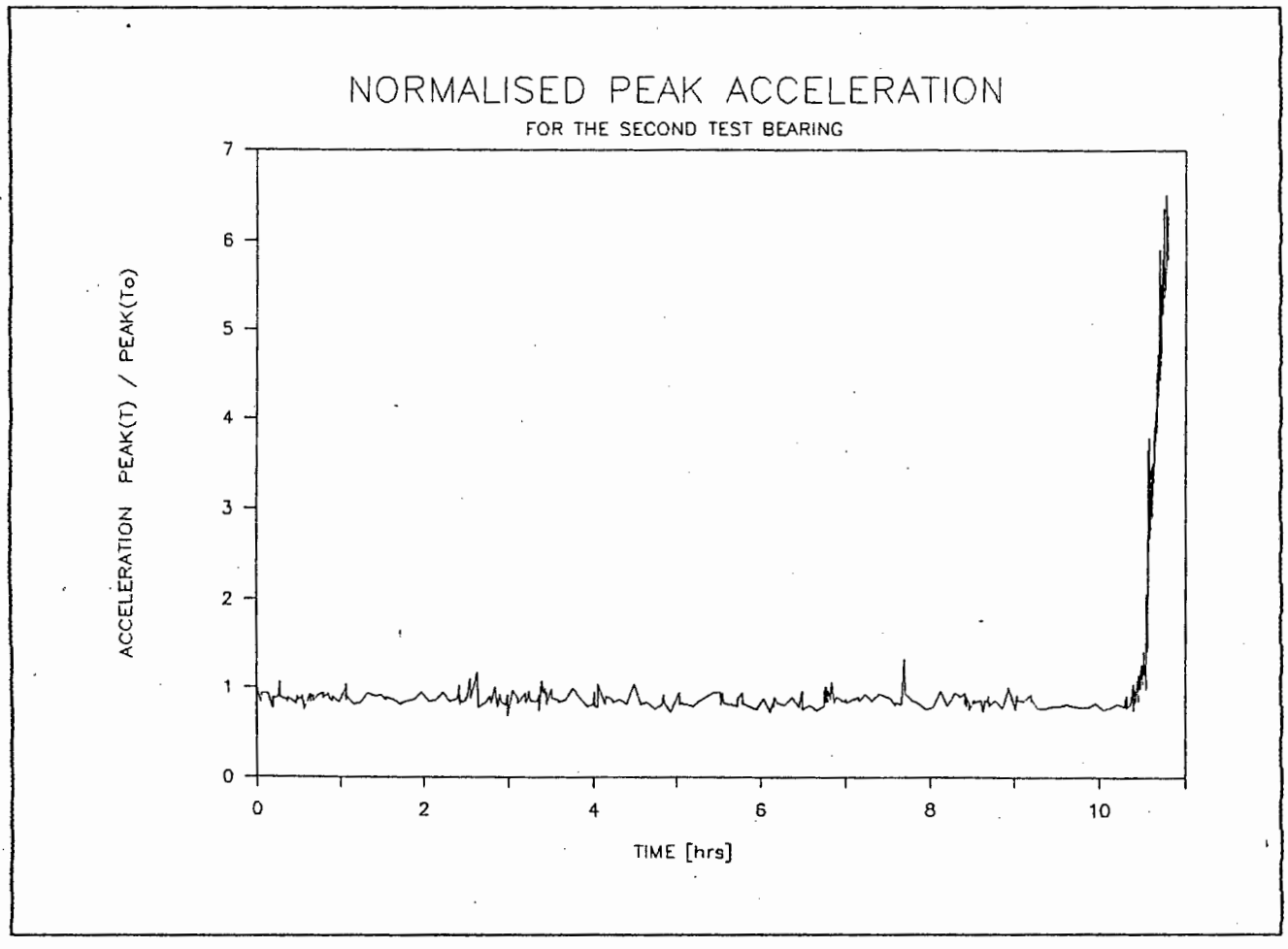

FIGURE 12.23 : PEAK ACCELERATION FOR THE SECOND TEST BEARING

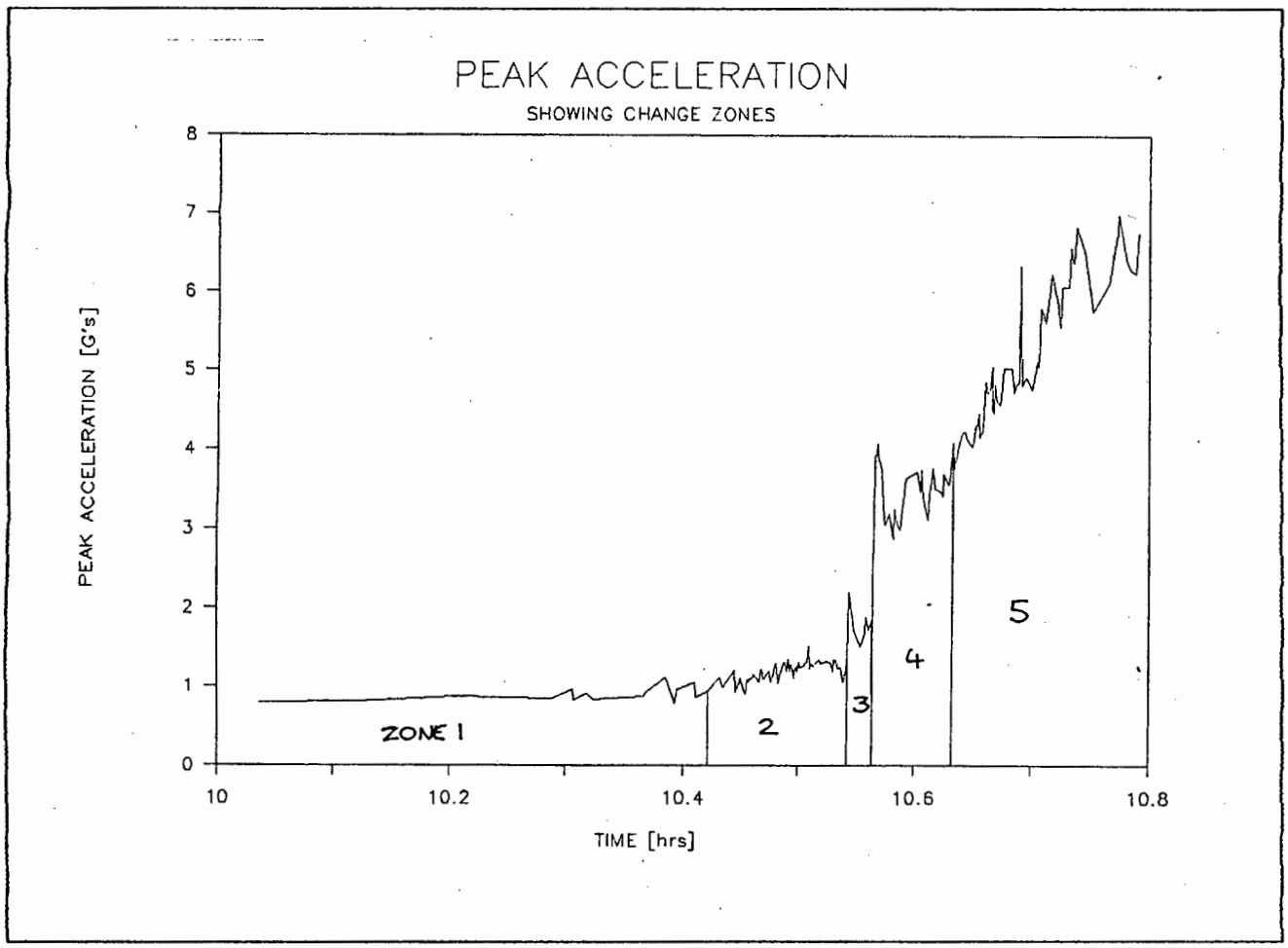

FIGURE 12.24 : PEAK ACCELERATION FOR THE LAST HOUR OF THE SECOND TEST

The similarities between RMS and Peak acceleration led to the question as to whether it was redundant to calculate the Peak for diagnosis. However it was decided to use the Peak value for two reasons. Firstly it was theoretically independent of RMS. Therefore it served to increase the confidence level of the diagnosis if it confirmed the RMS results. Secondly the Peak value was easily ascertained with a simple algorithm and would have little effect on processing time.

TABLE 12.7 : NORMALIZED STATISTICS OF PEAK FOR THE FIVE TIME ZONES

ZONE	AVERAGE	SDEV	AV-SDEV	AV+SDEV	MIN	MAX
1	1.000	0.090	0.910	1.090	0.801	1.516
2	1.292	0.128	1.164	1.420	0.977	1.628
3	1.927	0.214	1.713	2.142	1.634	2.368
4	3.677	0.513	3.163	4.190	1.969	4.396
5	5.664	0.983	4.681	6.648	4.029	7.527

The final parameter to be considered was the Crest Factor. The graph of this parameter for the full test period is shown in figure 12.25. The only indication of failure given by the Crest Factor was a slight drop of magnitude in the final hour (note the suppressed zero on the ordinate). This was further obscured by the large variation in the parameter. In fact the diagnostic potential of Crest Factor was severely limited by the large variation of the parameter over the whole range.

The drop in Crest Factor during the failure stage of the bearing life was predictable. This was because the RMS was seen to increase by more than the Peak in the failure zones 4 and 5 [figures 12.20 & 12.24]. This effect was ascribed to the peaks of the time waveform increasing in frequency of occurrence more than in magnitude, thereby increasing the RMS more than the Peak.

The last hour of the test is shown in more detail in figure 12.26.

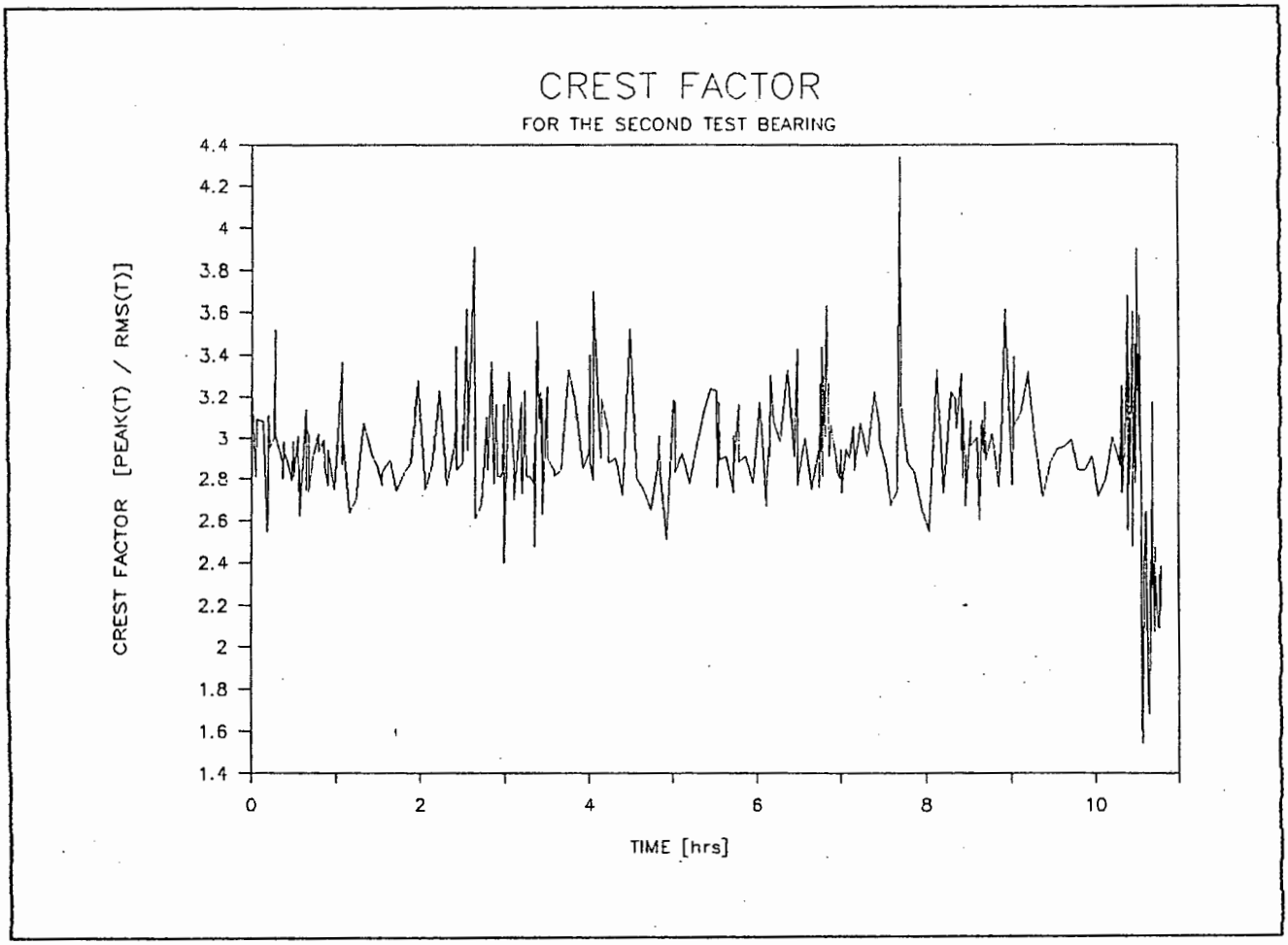

FIGURE 12.25 : CREST FACTOR FOR THE SECOND TEST BEARING

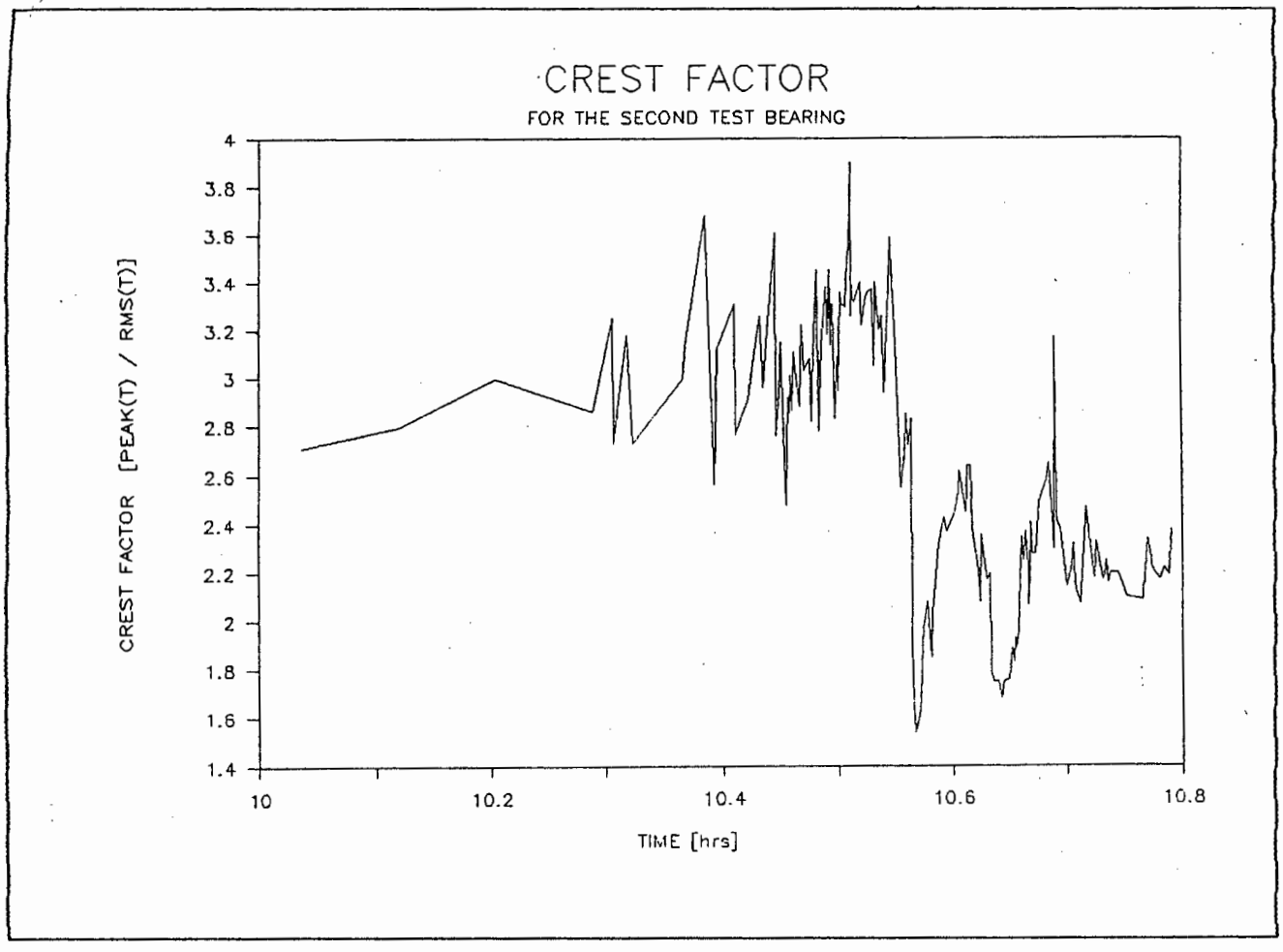

FIGURE 12.26 : CREST FACTOR FOR THE LAST HOUR OF THE SECOND TEST

In previous work the Crest Factor was seen to follow a similar trend to the Kurtosis [7]. However this work provided a much less detailed examination of the parameters than the test of bearing 2. The new test showed that the Crest Factor did not follow Kurtosis when examined in more detail and captured with more realistic time samples.

A slightly detectable similarity was seen in zone 2 (between about 10.4 and 10.6 hours) which was the incipient failure zone. The Crest Factor showed a hump similar to that of Kurtosis, but the hump was obscured by the large variation in the values of this zone. As for the failure zones 4 and 5 the two parameters almost showed opposite trends to each other, the one falling as the other rose.

These factors led to the discarding of Crest Factor as a useful diagnostic tool. For this reason it was recommended that Crest Factor not be used in industrial vibration monitoring programmes. In any case potential diagnostic information in the Crest Factor was contained in and more easily extracted from the RMS and Peak parameters.

Thus the experiments to define the relationships between the various subsets of the solution space were completed. These relationships were used in various ways to analyze the running condition of test bearings, given the measurements of the various parameters.

The use of these parameters and relationships for analysis of bearing condition is discussed in the next section.

SECTION 13

DEVELOPMENT OF THE EXPERT SYSTEM TO ANALYSE BEARING CONDITION

13. DEVELOPMENT OF THE EXPERT SYSTEM TO ANALYZE BEARING CONDITION

The expert system was developed according to the three levels of bearing condition analysis namely, detection, diagnosis and prognosis. A separate rulebase was developed for each level to allow for selection according to the users requirements. Where applicable the rulebases provided recommendations for more in-depth analysis and/or maintenance action.

This section discusses the preparation of the vibration parameters for analysis as well as the development of the three rulebases.

13.1. PREPARATION OF THE VIBRATION PARAMETERS

Some of the vibration parameters had to be prepared before they could be interpreted by the expert system. The preparation consisted of two separate stages.

Firstly the vibration parameters had to be adjusted for the measured operating conditions. This was done to make the vibrations independent of these operating conditions. Secondly the adjusted parameters were normalized to make the interpretation generic and adaptable to analyzing other bearings.

The stages of adjusting and normalizing are discussed in the following sections.

13.1.1. ADJUSTING THE VIBRATION PARAMETERS

The vibration parameters needing adjustment were the RMS and Peak acceleration. These had to be adjusted for the operating conditions speed and load. The adjustment was done by multiplying the relevant vibration parameter by an adjustment factor.

The adjustment factors were derived from the relationship between the parameter and the specific operating condition being accounted for. These relationships were derived in section 12.3 and are summarized in table 13.1.

TABLE 13.1 : SUMMARY OF RELATIONSHIPS BETWEEN OPERATING CONDITIONS AND VIBRATION PARAMETERS

PARAMETER	RMS	PEAK
SPEED	RMS = C·(Spd)^M C = 3.88 E-6 M = 1.594	Pk = C·(Spd)^M C = 1.647 E-5 M = 1.554
LOAD	RMS = M·(Load) + C M =-6.11 E-3 C = 0.567	Pk = M·(Load) + C M =-1.13 E-2 C = 1.528

The adjustment factors were defined as the ratio of the vibration parameter at the measured operating condition to the vibration parameter at the reference operating condition. Both were the values obtained from the equations in table 13.1, i.e. for a good bearing in each case. For example the adjustment factor of RMS for Speed is shown in equation 13.1.

$$RSF = RMS_{REF} / RMS_{MSR} \qquad \text{.....13.1}$$

Where :

RSF = RMS Speed Factor
RMS_{REF} = RMS at reference operating conditions
RMS_{MSR} = RMS at measured operating conditions

And both RMS values were calculated from the equation for RMS and speed in table 13.1.

Inserting the RMS-Speed relation [table 13.1] into equation 13.1 and taking the reference speed to be 1800 rpm, produced the following equation [append. 43] :

$$RSF = 1800^M / (SPEED)^M \qquad \text{.....13.2}$$

Where :

SPEED = Measured shaft speed in rpm.
M = 1.594 [table 13.1]

The value of C from table 13.1 was common to numerator and denominator.

The measured RMS value was then multiplied by this Speed Factor to enable it to be compared with acceptable RMS values at the reference operating conditions.

The adjustment factor for the Peak-Speed relation used the same equation with corresponding values for the exponent [table 13.1] as shown in the following equation [append. 43] :

$$PSF = 1800^{M} / (SPEED)^{M} \qquad \text{.....13.3}$$

Where :

PSF = Peak Speed Factor
SPEED = Measured shaft speed in rpm.
M = 1.554 [table 13.1]

The adjustment for Load was obtained by a similar process to that for Speed. In this case the equations relating the vibration parameters to load were used [table 13.1]. The adjustment factor for RMS and load is shown in the following equation, derived using 22 kN as the reference load [append. 43] :

$$RLF = (M * (22) + C) / (M * (Load) + C) \qquad \text{.....13.4}$$

Where :

RLF = RMS Load Factor
Load = Measured Load in kN.
M = -6.11 E-3 [table 13.1]
C = 0.567 [table 13.1]

The adjustment factor for Peak and Load used the same equation with the respective values for the exponent and constant. The equation was as follows [append. 43] :

$$PLF = (M \cdot (22) + C) / (M \cdot (Load) + C) \qquad \text{.....13.5}$$

Where :

PLF = Peak Load Factor
Load = Measured Load in kN.
M = -1.13 E-2 [table 13.1]
C = 1.528 [table 13.1]

The values of RLF and PLF were used to adjust the vibration parameters to compensate for the effects of varying load. This completed the adjustment of the vibration parameters for the operating conditions.

The second stage of preparation was to normalize the parameters. This is discussed in the following section.

13.1.2. NORMALIZING THE VIBRATION PARAMETERS

Once the vibration parameters had been adjusted for load and speed they were normalized. This was to make the parameters dimensionless for easier interpretation by the expert system and to simplify the task of modifying the expert system to diagnose other types of bearings.

The normalizing process was simply achieved by dividing the adjusted vibration parameter by its average running value at the reference conditions when the bearing was in good condition. For example the normalizing value for RMS was obtained from zone 1 of the second bearing test [Table 12.5].

Thus the overall RMS preparation equation consisting of adjusting and normalizing, was as follows [append. 44] :

$$NRMS = RMS \cdot RSF \cdot RLF \;/\; RMS_{Av} \qquad \ldots\ldots 13.6$$

Where :

NRMS = Normalized & adjusted RMS
RMS = RMS value to be tested
RSF = Adjustment factor for Speed [eqn. 13.2]
RLF = Adjustment factor for Load [eqn. 13.4]
RMS_{Av} = Average RMS for good bearing (zone 1) at reference operating conditions [table 12.5]

The overall Peak preparation equation was similar to equation 13.6 as follows.

$$NPeak = Peak \cdot PSF \cdot PLF \;/\; Peak_{Av} \qquad \ldots\ldots 13.7$$

Where :

NPeak = Normalized & adjusted Peak
Peak = Peak value to be tested
PSF = Adjustment factor for Speed [eqn. 13.3]
PLF = Adjustment factor for Load [eqn. 13.5]
$Peak_{Av}$ = Average Peak for good bearing (zone 1) at reference operating conditions [table 12.7]

The virtue of using NRMS and NPeak was that for any speed and load a value of one would represent a bearing in good condition. Thus the expert system could be developed to diagnose the overall and maximum levels of vibration without having to account for the load applied to the bearing or the shaft speed.

The load and speed would still have to be known to obtain NRMS and NPeak, but accounting for these outside of the expert system greatly simplified the task of it's development. In addition the expert system could use the load and speed as separate diagnosis parameters in their own right.

At this stage the parameters were ready to be used to develop the three expert system rulebases. These are discussed in the following sections.

13.2. THE DETECTION RULEBASE

Typically, the task of detection in an industrial situation was to provide simple analysis of a large amount of machinery. This usually consisted of a single RMS measurement and was therefore rapidly accomplished. As discussed in section 6 this type of analysis was widely used but sometimes lacking in reliability.

Thus the detection rulebase was designed to enable rapid processing combined with more reliability than a single RMS measurement. It also acted as a primary filter to enable selective use of the more time consuming diagnosis and prognosis.

In effect the detection rulebase only performed the relatively simple task of deciding whether a bearing fault was **suspected** or not. However this task was all important because, acting as the primary filter, it would often be the only form of analysis done on machinery until a fault was suspected.

Therefore detection would have to be as reliable as possible. It would err on the safe side so that if there was any suspicion of a fault it would be detected. Confirmation would be received by running the diagnosis rulebase which would perform a more in-depth analysis.

In contradiction to the requirements of reliability the fault detection system would have to avoid recommending diagnosis when possible. If the detection rulebase often recommended further testing (say more than 50% of cases) because of over-sensitivity of the system then the detection stage would be obsolete. It would then become more effective, but nevertheless more time consuming to skip the detection stage and proceed directly to diagnosis.

Therefore it was essential to select carefully the parameters and their values to be used as the failure detection criteria. In addition the detection levels would have to be adjusted until the the above constraints were fulfilled with the best compromise.

The problem of detection was represented as a simple decision tree. This is shown in figure 13.1

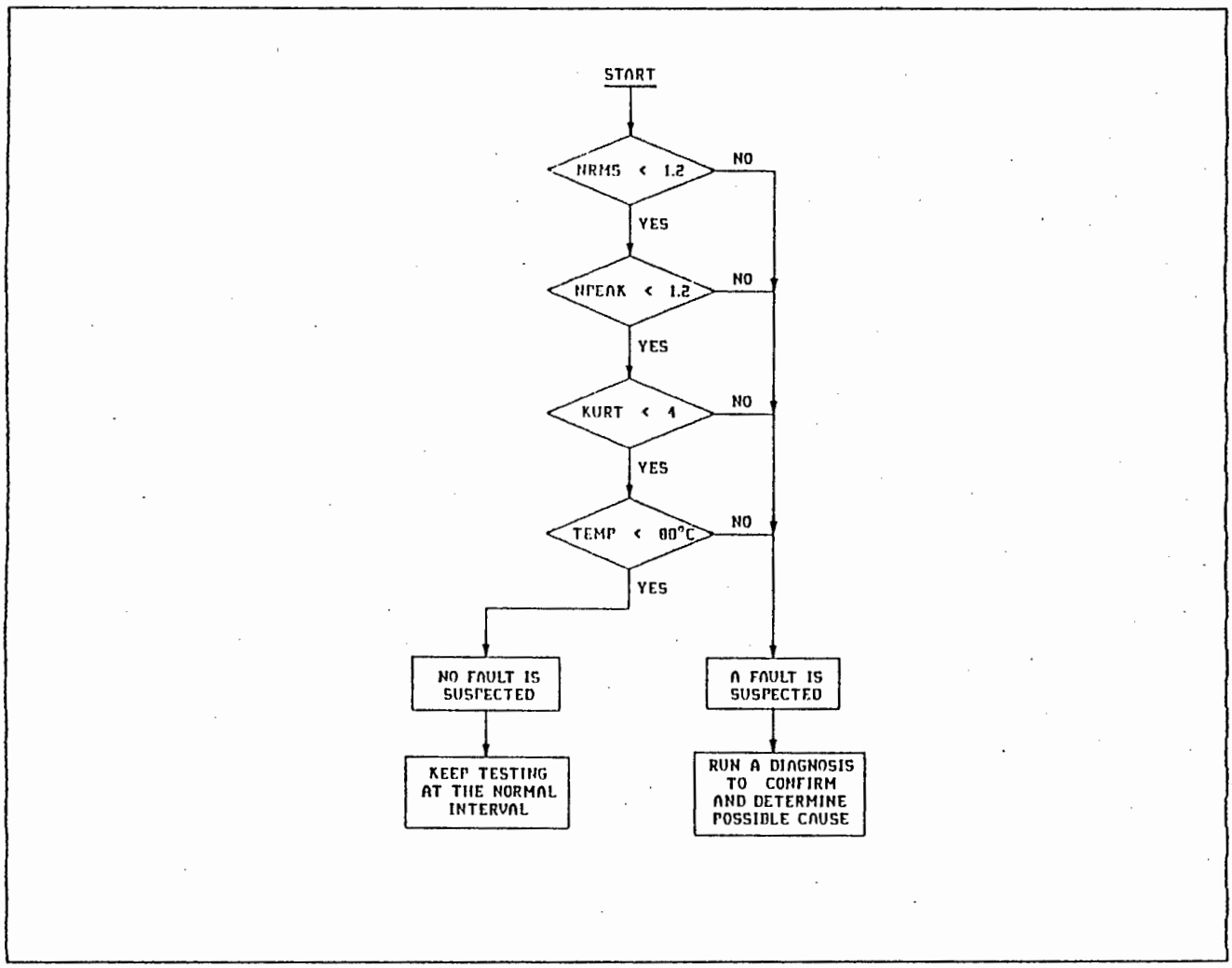

FIGURE 13.1 : DECISION TREE FOR THE DETECTION RULEBASE

This decision tree was used to derive the detection rulebase. It was condensed into a single rule as follows:

IF:

NRMS < 1.2
and KURT < 4
and NPEAK < 1.2
and TEMP < 80

THEN:

No fault is suspected in the bearing. Keep testing at the normal interval. - Probability = 1
and To determine a suitable testing interval run a prognosis.

ELSE:

A fault is suspected in the bearing. Run a diagnosis to confirm this and to determine the possible cause. Probability = 1.

This rule was sufficient for diagnosis and was designed to execute as rapidly as possible. The following aspects of the structure of the rule were important.

Firstly the 'IF' part consisted of four parameters providing the ability to detect failure. These were arranged in order of decreasing importance because the expert system inference engine tested each of the conditions in sequence.

This meant that the 'ELSE' part 'fired' immediately a parameter above its limit was detected, obviating the need to test the rest. Thus the 'THEN' part would only fire if all the parameters were below their critical limit.

The values of the critical limits given in the rule were selected as a first estimate using the values derived in zone 1 of [tables 12.5, 6 & 7]. In any application these would have to be refined by trial and error. This was clearly a potential case for applying some form of 'self learning' to the expert system [section 6].

The probability mode for the choices was selected as the 0/1 system [section 6]. This was to emphasize the 'GO/NO-GO' aspect of detection. Either a fault was suspected or it was not suspected. This would be confirmed by running the diagnosis rulebase if necessary.

Note that the final part of both choices were recommendations for further action. Thus the 'recommendation stage' was included in the detection rulebase.

The detection rule did not invoke the diagnosis rulebase directly. This was because it may have been desired to do this only after testing a large batch of machinery. However the automatic invoking of diagnosis could quite easily have been incorporated into the rule if this was desired.

The next rulebase to be developed was that for diagnosis. This is discussed in the following section.

13.3. THE DIAGNOSIS RULEBASE

The task of the diagnosis rulebase was to determine the running condition of the bearing and attempt to diagnose the cause of failure. In addition it was required to make recommendations on further action to be taken.

One of the main criteria for developing the diagnosis rulebase was that it should be as reliable as possible and provide consistent diagnosis. This could be done at the expense of processing time because the detection rulebase operated as a primary filter to reduce the amount of machinery to be diagnosed. Therefore the diagnosis rulebase consisted of many more rules, each analyzing individual parameters and combinations of parameters.

The diagnosis rulebase was divided into two distinct sections. The first analyzed the bearing condition using the vibrations, while the second determined the possible cause of predicted failure using the operating conditions and derived parameters.

Before the rules for diagnosing the vibration parameters could be developed, the vibration parameters had to be interpreted. This is discussed in the following section.

13.3.1. INTERPRETING THE VIBRATION PARAMETERS

The first purpose of diagnosis was to determine the running condition of the bearing being tested. This was achieved by interpreting the three vibration parameters NRMS, NPEAK and Kurtosis.

Interpretation involved relating the vibration parameter amplitudes to specific stages of bearing failure. It was for this reason that the five zones of bearing condition were established [section 12.4.]

However the problem still remained of defining what parameter amplitudes represented which zone of failure. What was needed were exact values for the lower and upper limits of the parameter amplitudes for each time zone.

In addition the limits for adjoining zones would have to be the same value so as to establish dividing values between the zones. This would enable the expert system to detect which time zone the bearing was operating in and thereby diagnose it's condition.

Ideally the limits should have been obtained by considering the probabilities of the amplitudes for each zone. The dividing values would then be where there was equal probability of the amplitude belonging to either the zone above or below it.

In other words the dividing values would be given by the intersection of the amplitude probability functions for each of the adjoining zones. This would also account for any possible parameter amplitude presented for diagnosis, because the zones would be adjoining each other.

However, due to some of the zones having insufficient number of samples to develop accurate probability functions an alternative method was used. This involved a statistical analysis of the amplitude values of each parameter in each zone. These statistics were then used to estimate the amplitude limits for each zone.

In particular there were a number of possible values which provided an estimate for the limits. Considering that the upper limit of one zone would be the lower limit of the next zone, these values could be provided by any one of the following statistics :

- The mean of the lower zone plus the standard deviation for that zone
- The mean of the upper zone minus the standard deviation for that zone
- The maximum value of the lower zone
- The minimum value of the upper zone

These statistics are given in normalized form in section 12.4 for RMS and Peak [tables 12.5 & 12.7]. Because the samples were taken at the reference operating conditions, these normalized statistics were representative of NRMS and NPeak.

In addition to these statistical parameters there was another potential method for establishing the limits. This was discovered when analyzing the results of two cases of vibration research work [7, 45]. It was found that the amplitude limits of RMS for categories of failure corresponded closely to a geometric progression.

Therefore it was decided to incorporate a geometric progression into the process of establishing the limits. The geometric progression factor was determined from [7 & 45] in [appendix 45.]

The results of the various analyses were then combined by taking the average of the five possible values contributing to each limit. This is shown in [table 13.2] for NRMS and [table 13.3] for NPeak. Note that there were six values corresponding to the five zones.

TABLE 13.2 : STATISTICAL SELECTION OF NRMS LIMITS

LIMIT	MEAN-SD	MEAN+SD	MIN	MAX	G.PROGR.	AVERAGE
1	0.958	-	0.887	-	0.599	0.815
2	1.156	1.042	1.069	1.130	1.099	1.099
3	1.845	1.265	1.706	1.306	2.017	1.628
4	4.094	2.095	3.838	2.116	3.699	3.168
5	6.513	6.634	5.685	8.313	6.786	6.786
6	-	8.910	-	10.080	10.480	10.480

Selection of values for the lower limit of the first zone and the upper limit of the final zone, i.e.. the first and sixth values was not as critical as the four values between the zones. This was because they defined the outer boundaries of the whole system. These could be as little as zero for the lower limit and as large as need be for the upper limit to cater for all possible vibration levels.

However these two outer limits were included in the analysis because they yielded possible diagnostic information on the functioning of the data capture system. For example if the vibration levels were very low or inexplicably high then it was possible that the data capture system was malfunctioning.

The average values in the final column were used as the dividing values for the five time zones. These values could easily be modified in the expert system rulebase if they were found to be producing incorrect diagnosis.

The same analysis was done to define the limits for the amplitudes of the five NPeak zones. The results of this analysis are shown in [table 13.3]. Again the average values in the final column were used as the dividing values for the Peak zones

TABLE 13.3 : STATISTICAL SELECTION OF NPEAK LIMITS

LIMIT	MEAN-SD	MEAN+SD	MIN	MAX	G.PROGR.	AVERAGE
1	0.910	-	0.801	-	0.771	0.827
2	1.164	1.090	0.977	1.516	1.187	1.187
3	1.713	1.420	1.634	1.628	1.826	1.644
4	3.163	2.142	1.969	2.368	2.810	2.490
5	4.681	4.190	4.029	4.396	4.324	4.324
6	-	6.648	-	7.527	6.653	6.9430

It was interesting to compare the average values with the geometric values to see how good an estimate the geometric progression provided. The comparison was done for both NRMS and NPeak as the difference between the average and geometric progression estimates [append. 46]. This was expressed as a percentage of the geometric progression as shown in [table 13.4.]

Note that for the 2nd and 5th limits the difference was zero. This was because the geometric progression was based on these two values.

TABLE 13.4 : PERCENTAGE DIFFERENCE BETWEEN THE AVERAGE AND GEOMETRIC PROGRESSION ESTIMATES

LIMIT VALUE	NRMS %DIFF	NPEAK %DIFF
1	35.967	7.307
2	0.000	0.000
3	-19.282	-9.956
4	-14.350	-11.374
5	0.000	0.000
6	-15.820	4.354

It can be seen that the geometric progression provided a better estimate for Peak than for RMS. The reason for this was apparent when considering figures 12.20 & 12.24. The Peak showed a continuing upward trend while the RMS showed more variation. In a case where no other information was available the geometric progression was recommended as a good starting point for establishing the limits for bearing analysis.

The Kurtosis was more difficult to use for determining the bearing running condition by zone. This was because Kurtosis did not show a continuously rising amplitude for worsening condition. In fact there was a lot of overlap in the values for the five zones [table 12.6]

However the real strength of Kurtosis was to determine the onset of failure by distinguishing zone 2 from zone 1. Therefore Kurtosis was incorporated into the rules of the expert system so as to achieve this. It was also used to confirm the condition zones by a less rigorous application of the amplitude statistics for each zone.

Having established the amplitude limits for the zones it was necessary to attach diagnostic labels to each. The labels were chosen according to a characteristic description of the bearing condition in each zone. These were obtained from the experiment described in section 12.4 and are shown in table 13.5.

TABLE 13.5 : DESCRIPTION OF BEARING CONDITION IN THE TIME ZONES

ZONE	CHARACTERISTIC DESCRIPTION OF BEARING CONDITION
1	Satisfactory and running normally.
2	Showing incipient fatigue failure in the form of surface scoring and light flaking.
3	Showing advanced surface flaking with spalling and pitting imminent.
4	Showing failure in the form of early pitting and spalling
5	Showing advanced fatigue failure with well defined pitting and spalling. Total failure is imminent.

The characteristic descriptions in table 13.5 were used as a 'Fuzzy Logic Lookup Table' to develop the rules relating to the vibration parameters. This is discussed in the following section.

13.3.2. THE VIBRATION PARAMETER RULES

The purpose of the vibration parameter rules was to determine which condition zone the bearing was operating in. This was done by assigning probabilities to the zones according to the levels of the three vibration parameters NRMS, NPeak and Kurtosis.

The probabilities for the zones determined by each parameter were then combined in such a way that the diagnosis was either confirmed or rejected. For example if the same diagnosis was given by all parameters then the probability of the diagnosis being correct was high and vice versa.

Therefore the first stage in developing the vibration parameter rules was to select a probability mode and the method of combining probabilities. The choices available were discussed in section 6.

For the probability mode the '0% - 100%' method was selected. This was so that the probabilities could be combined to form fractional probabilities instead of just integers.

For the combination of probabilities the independent mode was selected [sect. 6]. In other words the probabilities were combined according to the following formula.

$$P_c = 1-((1-P_1) * (1-P_2)) \qquad \text{.....13.8}$$

where:

P_c = Combined probability
P_1 = First probability
P_2 = Second probability

This meant that the individual probabilities could be used to confirm each other by increasing the overall probability of the particular zone diagnosed.

The next stage was to assign probabilities to the various zones according to the zone diagnosed. The same probabilities were used for the diagnosis provided by NRMS and NPeak while those for Kurtosis were developed differently.

It was obvious that the probability of the diagnosed zone being the actual operating zone should be highest. However it was apparent that the adjacent zones also had some probability of being correct. Therefore some probability distribution was needed to describe the situation.

It was decided to use the binomial distribution formulated from the binomial expansion [46]. The expansion involved seven zones, namely the 5 condition zones and the two zones below and above these.

The probabilities formulated in this way were found to correspond closely to subjective values assigned to the zones. Hence the binomial distribution was used as a first estimate. These values could be modified later if testing indicated the need.

The calculation of the probability distributions is developed in [appendix 47] while the results are given in table 13.6.

TABLE 13.6: DIAGNOSIS ZONE PROBABILITIES ACCORDING TO THE BINOMIAL DISTRIBUTION

DIAG-NOSED ZONE	ZONE PROBABILITIES (%)				
	1	2	3	4	5
1	80	40	11	2	0
2	53	66	44	16	3
3	19	47	63	47	19
4	3	16	44	66	53
5	0	2	11	40	80

There were a few notable aspects about these values. Firstly they were symmetrical about the central value of zone three with the larger probabilities being assigned to the outer zones. Also the probabilities did not add up to 100 for two reasons as follows.

Firstly they were multiplied by a common factor to increase their value to be aligned with subjective values. And secondly the values for the two external zones were not included. These modifications were permissible because the probabilities were essentially used to rank the zones in order of likelihood. In addition the form of the binomial distribution was unaffected by the modifications.

These probabilities were used to construct a set of five rules for each of the parameters NRMS and NPEAK. The rules used the zone limits and measured values of the parameters to attach probabilities to each of the zones. An example of one of these is illustrated in the following rule.

```
IF:
          NRMS >= RMS_LIMIT_3
   and    NRMS <  RMS_LIMIT_4

THEN:

          BEARING CONDITION: Zone 1   Probability = 19/100
          BEARING CONDITION: Zone 2   Probability = 47/100
          BEARING CONDITION: Zone 3   Probability = 63/100
          BEARING CONDITION: Zone 4   Probability = 47/100
          BEARING CONDITION: Zone 5   Probability = 19/100
```

In the 'IF' part of this rule the value NRMS would be the RMS value measured on the bearing to be tested, adjusted for speed and load, normalized and presented for diagnosis. The values RMS_LIMIT_3 and RMS_LIMIT_4 were the lower and upper limits respectively of zone 3. These values were as derived in table 13.2.

In the 'THEN' part of this rule the zones 1 to 5 related to the characteristic descriptions given in table 13.5. The connection to the descriptions was made on the final display of results by the expert system. The associated probabilities were taken from table 13.6 where the diagnosed zone was zone 3.

The only exception to this format was for the two zone 1 rules. These contained an additional clause in the IF part which incorporated Kurtosis. The clause was as follows.

```
IF:       KURTOSIS < 4.25
   and    NRMS     > RMS_LIMIT_1
            ......etc.
```

The value 4.25 was the maximum measured value of Kurtosis for zone 1 [table 12.6].

This Kurtosis clause made the performance of the rules for zone 1 severe in their diagnosis. This was because if Kurtosis was greater than 4.25, zone 1 would receive a very low probability. In other words emphasis was placed on Kurtosis to distinguish between zone 1 as the 'Running normally' zone and the other zones as the failure zones.

In addition another set of 5 rules was developed to confirm the distinction between all zones. These rules used a less rigorous method for Kurtosis than those for NRMS and NPeak in that the boundaries between zones were not distinctly defined.

The method was to use the statistics of Kurtosis to define two ranges for each zone. The first range was between the Maximum and Minimum Kurtosis values for each zone while the second range was bounded by the Mean ± Std. Deviation of Kurtosis values in each zone. This is shown in figure 13.2.

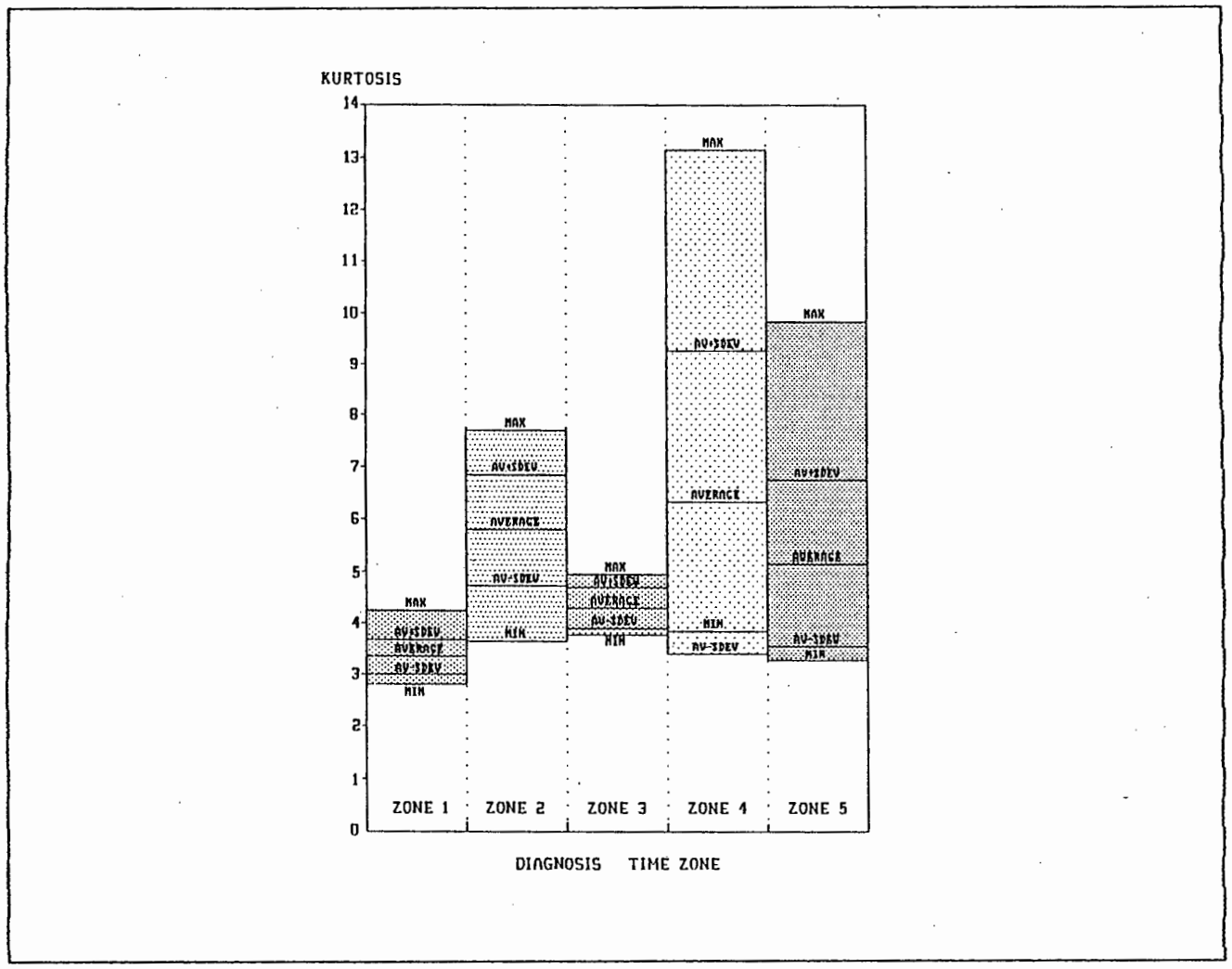

FIGURE 13.2 : THE KURTOSIS PROBABILITY RANGES FOR DIAGNOSIS

The values for the Average, Max, Min and Std. Deviation of Kurtosis as shown in figure 13.2 were derived in table 12.6.

If the measured Kurtosis fell within the Max-Min range then that zone received a probability of 40% and if it fell within the Mean±Sdev range then that zone received a probability of 60%.

This was equivalent to using a simplified probability distribution for each range. The skewness of the distribution would be taken into account if the Mean±Sdev range was not centralized on the Max-Min range. In this way the Kurtosis provided a general confirmation of the resulting zone diagnosed.

This concluded the development of the vibration rules for diagnosis. The next set of rules to be developed was the operating condition rules. This is discussed in the next section.

13.3.3. THE OPERATING CONDITION RULES

The purpose of using the operating conditions for diagnosis was to establish the possible cause of bearing failure. This could then be used to make maintenance recommendations.

Taking a simplistic view, the operating conditions providing potential diagnostic information were the Load, Speed and Temperature. However on closer examination it was a complex task to assess which combinations of these parameters would be the potential cause of the bearing failure.

Thus it was necessary to re-structure the task by grouping combinations of the parameters to assess their combined effect. These combination parameters were already inherent in the system, viz. the lubricant viscosity index 'K' and the the life rating in hours L_{10h}. Both were discussed in section 12.2.

The viscosity index K provided information about the combination of speed and temperature, while the life rating L_{10h} provided an overview of the combination of all three measurable operating conditions. Thus it was possible to divide the task of diagnosis into manageable portions.

Once it was known that some combination of operating conditions posed a particular threat to the bearing, it was possible to diagnose which condition in that group was most probably responsible. Thus each parameter could be considered for the diagnostic information it yielded.

This was then used to develop rules stating the probability of the bearing failure being attributed to that particular parameter. In addition if the bearing was not failing then warnings were given about extreme conditions leading to failure.

The first step was to categorize the values of each parameter into fuzzy logic lookup tables to enable the construction of rules. This is discussed as follows.

CATEGORIZING THE DIAGNOSTIC PARAMETERS

The range of values for each parameter were categorized into fuzzy logic lookup tables. This involved dividing the parameters into ranges and assigning a characteristic description to each range.

It was essential to economise on the number of categories used for each parameter. This was because the number of rules grew exponentially with the total number of categories for all parameters. For example having 5 parameters with 5 categories each would mean a possible $5^5 = 3125$ rules. However not all the combinations would necessarily make sense.

In contradiction to the above requirements, more categories for each parameter would mean a more realistic diagnosis. In other words it was necessary to step further than a simple 'High/Low' or 'Yes/No' type of category. Thus these two requirements had to be balanced when developing the parameter categories.

The first parameter to be considered was the bearing rated life L_{10h}. The categorization was done subjectively according to experience [7] and information gleaned from bearing literature [13, 15, 47]. The limits were first established using months and years to gain a feel for the time periods. These were then converted to rounded values of hours. The stages of this process are illustrated by table 13.7.

TABLE 13.7 : CATEGORY LIMITS FOR BEARING LIFE

L10h LIMIT	TIME PERIOD	APPROXIMATE HOURS
1	1 Month	1 000
2	6 Months	5 000
3	2 Years	20 000
4	10 Years	100 000

It was interesting to note that the limits corresponded closely to a geometric progression. These limits were then used to form the categories of the bearing rated life. The categories and their characteristic descriptions are shown in table 13.8.

TABLE 13.8 : A FUZZY LOGIC LOOKUP TABLE FOR THE CATEGORIES OF BEARING RATED LIFE

LIFE RANGE [Hours]	CHARACTERISTIC DESCRIPTION. The Rated Bearing Life is...
< 1 000	Totally inadequate
1 000 - 5 000	Very short
5 000 - 20 000	Satisfactory
20 000 - 100 000	Very Long
> 100 000	Too long (Over designed)

The next parameter to be considered was the lubricant viscosity index 'K'. The diagnostic purpose of this parameter was to determine the sufficiency of the lubricant viscosity.

This parameter effectively analyzed the combination of speed and temperature by comparing it with the required conditions for the particular lubricant being used. The temperature and speed were theoretically the operating conditions affecting the lubricant performance [15].

The limits, ranges and characteristic descriptions were developed in the discussion on the parameter in section 12.2. These are summarized in table 13.9.

TABLE 13.9 : THE LUBRICANT VISCOSITY INDEX CATEGORIES

K RANGE	CHARACTERISTIC DESCRIPTION. The Lubricant Viscosity is ...
< 0.4	Totally inadequate
0.4 - 1.0	Barely sufficient
1.0 - 4.0	Satisfactory
> 4	Very good

The next parameter to be considered was the bearing load. This parameter was the most influential on bearing life as discussed in section 12.2. Therefore the division into categories was closely related to the bearing rated life. Again the categories were subjectively chosen with guidelines being given by [15].

The first load to be considered was the dynamic load rating of the bearing. This was defined to be that load under which the bearing had a 90% chance of running for 1 million cycles. However for speed ranges of 1000 to 2000 rpm, 1 million cycles represented only between 8 and 16 hours of continuous running time. Clearly this was inadequate for any bearing design.

Therefore the load limits had to be selected on a more realistic basis. A first estimate was obtained by selecting 'normal' running conditions and then selecting the loads according to the rated life in table 13.7. By applying a bearing speed of 1500 rpm and temperature of 60°C to equation 12.6, the load limits shown in table 13.10 were obtained [append. 48]. It was noted that the values of load were well below the rated loading of the bearings being considered (15.9 kN).

TABLE 13.10 : THE BEARING LOAD LIMITS

RATED LIFE [Hours]	LOAD [kN]
1 000	4.82
5 000	2.82
20 000	1.78
100 000	1.04

Because the load was an important parameter it was decided to use the four limits to form 5 categories describing the severity of the load as follows.

TABLE 13.11 : THE BEARING LOAD CATEGORIES

LOAD [Kn]	CHARACTERISTIC DESCRIPTION The Bearing Load is ....
0 - 1.04	Very low
1.04 - 1.78	Fairly low
1.78 - 2.82	Intermediate
2.82 - 4.82	Fairly high
> 4.82	Very High

The next parameter to be characterized was the bearing temperature. It was decided that for simplicity this parameter only needed three categories. These were set from experience as follows.

TABLE 13.12 : THE BEARING TEMPERATURE CATEGORIES

TEMPERATURE [°C]	CHARACTERISTIC DESCRIPTION The Bearing Temperature is...
< 80	Satisfactory
80 - 100	Very Hot
> 100	Overheating

Finally the running speed was the least important of the operating conditions. Therefore it was only given 3 categories as follows.

TABLE 13.13 : THE BEARING SPEED CATEGORIES

SPEED [RPM]	CHARACTERISTIC DESCRIPTION The Bearing Speed is...
< 200	Very Slow
200 - 800	Fairly Slow
> 800	Intermediate

The last category was described as being intermediate because much higher speeds than the test rig was capable of, would be classified as fast.

The categorizations as developed in tables 13.7 to 13.12 were used to develop the diagnosis rules for the operating conditions as discussed in the following section.

BUILDING THE OPERATING CONDITION DIAGNOSIS RULES

The rules for diagnosing the bearing operating conditions were developed by considering the Life Rating (L_{10h}) and the Lubricant Viscosity index K. A set of rules was developed for each, and included the probability that some specific operating condition was at fault.

Firstly considering the Life Rating, the diagnostic flow chart shown partially in figure 13.3 was developed. The number of rules for this set would be given by the combination of options for each level as follows.

$$\begin{aligned} \#\ \text{Rules} &= (O_{FAIL}) \cdot (O_{LIFE}) \cdot (O_{LOAD}) \cdot (O_{VISC}) \\ &= 2 \times 5 \times 5 \times 4 \\ &= 200 \end{aligned}$$

Where : O_{FAIL} = Number of options for bearing failing
O_{LIFE} = Number of options for Life Rating
O_{LOAD} = Number of options for Load
O_{VISC} = Number of options for Lubricant Viscosity

Only the more interesting and obvious combinations were used to develop a set of trial rules.

The number of rules would have been increased if all the categories of bearing running condition were included. However it was decided that it was sufficient to categorise the bearing as failing or not failing. Thus if the probabilities of any of the running zones 2 to 5 was higher than zone 1 [table 13.5] then the bearing was failing.

The next stage was to establish the cause of failure if the bearing was failing. This was done by testing the calculated life rating for the measured operating conditions. The test was essentially used to establish if the general combination of applied conditions was unsatisfactory and if so, how severe.

If the life rating was 'totally inadequate' or 'very short' then one of the operating conditions was clearly at fault. This would be due to some combination of load and K (i.e. a combination of speed and temperature).

In particular figure 13.3 illustrated the following rule

```
IF      : The bearing is failing
   AND  : The rated life is very short
   AND  : The load is fairly low
   AND  : The lubricant viscosity is barely sufficient

THEN    : The cause of failure is incorrect Lubrication
                                    Probability = 60%
   OR   : Overloading               Probability = 30%
   OR   : Some other cause          Probability = 10%
   RECOMMEND: Use a lubricant with higher viscosity
```

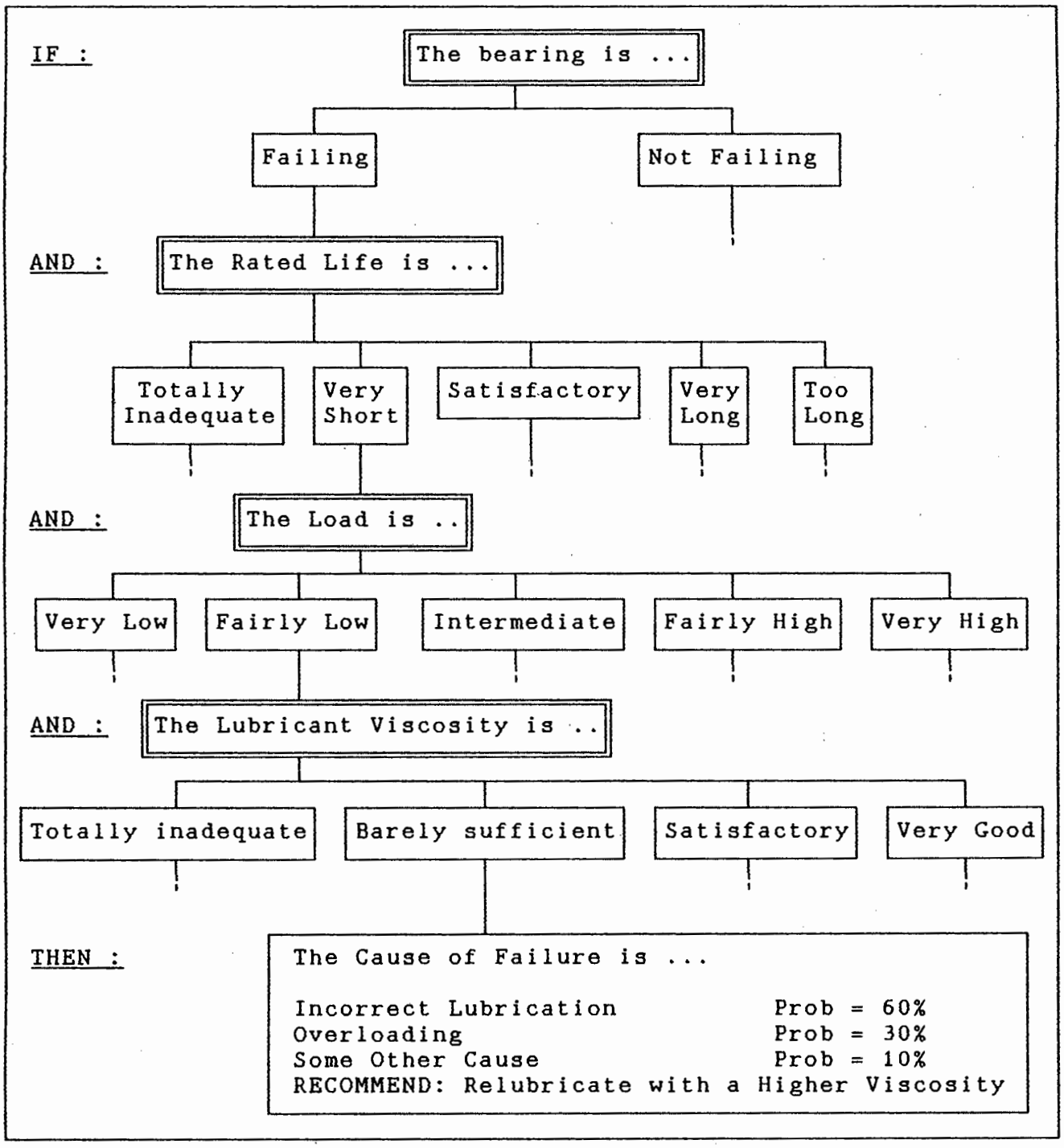

FIGURE 13.3 : DIAGNOSTIC FLOW CHART FOR THE LIFE RATING RULES

On the other hand if the bearing was not failing then a similar rule was used to warn that the bearing could fail prematurely due to incorrect lubrication, etc.

Also if the bearing was failing and the rated life was very long, etc. then the failure was more likely to be some other cause such as ingression of dirt due to seal failure. If the life rating was too long then a recommendation to use a cheaper bearing (lower load rating) was included.

To prevent the rules from becoming too clumsy the conditions associated with the lubricant viscosity were diagnosed separately as shown in the diagnostic flow chart of figure 13.4. The only categories of the viscosity index to be included were 'totally inadequate' and 'barely sufficient' because the others did not need diagnosing.

The diagnostic flow chart illustrated the following rule.

IF : The lubricant viscosity is totally inadequate
AND : The temperature is satisfactory
AND : The running speed is very slow

THEN : The lubricant viscosity is totally inadequate because the wrong lubricant is being used for such a slow running speed Probability = 60%
RECOMMEND: Replace lubricant with one of a higher viscosity

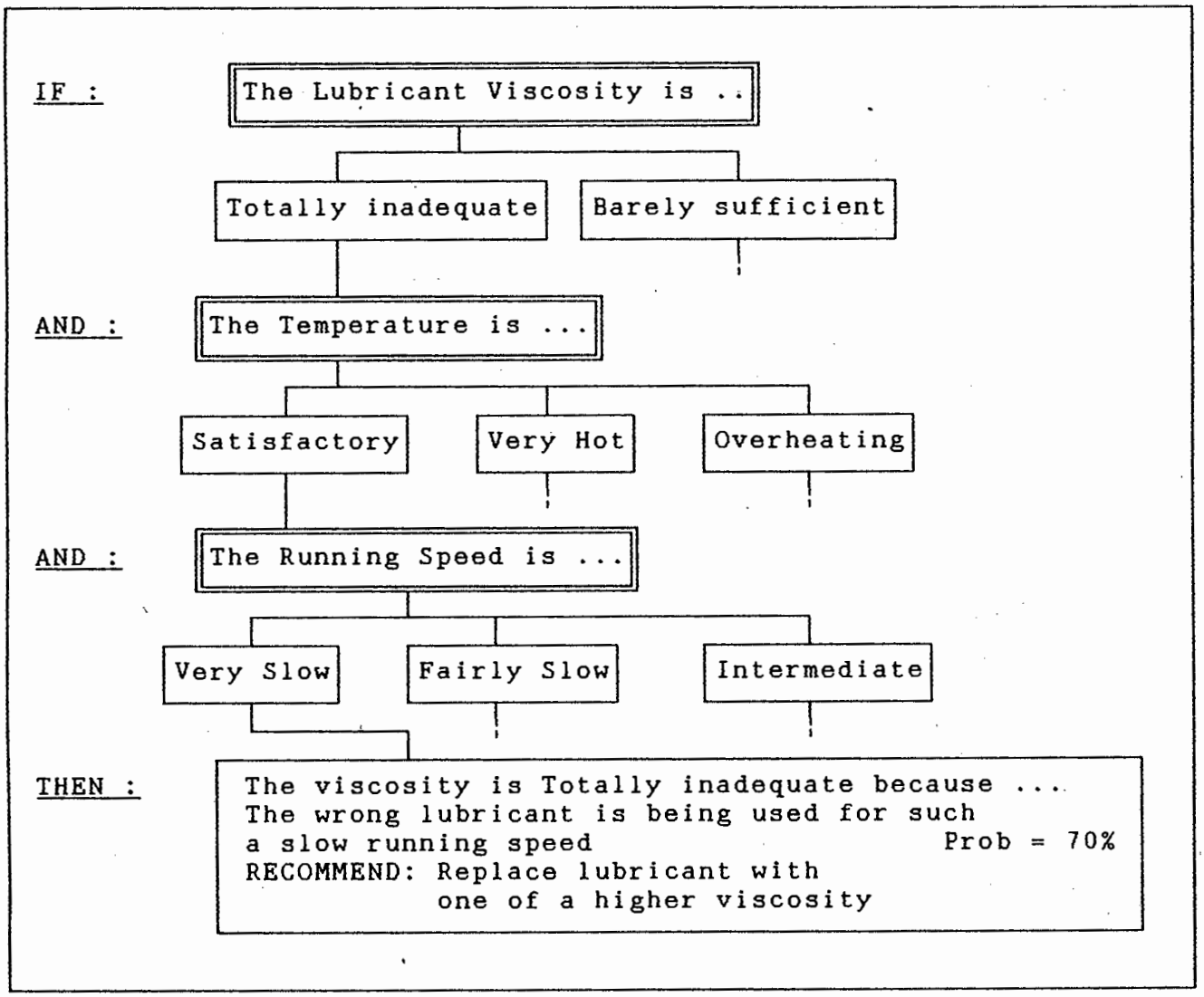

FIGURE 13.4 : DIAGNOSTIC FLOW CHART FOR THE LUBRICANT VISCOSITY

Other similar rules were developed for the lubricant viscosity such as :

IF : The lubricant viscosity is barely sufficient
AND : The temperature is overheating
AND : The running speed is intermediate

THEN : The lubricant viscosity is insufficient due to overheating Probability = 60%
RECOMMEND: Circulate the lubricant through a cooling system or if one exists then increase the cooling effect.

The rules were tested for functioning with various trial values. The results were predictable and tended to be a simplification of the real task of diagnosis.

The final rulebase to be developed was that for prognosis. This is discussed in the following section.

13.4. THE PROGNOSIS RULEBASE

The task of the prognosis rulebase was to provide a simple estimate of the remaining life of the bearing. This was done by estimating the time to failure given the rating life under the current operating conditions, and the diagnosis of the running condition of the bearing. In addition it provided recommendations on the testing interval.

The prognosis was based on the time duration of the 5 running zones as given in table 12.4. It was recognized that these values would not necessarily be statistically representative of bearing failures in general. Rather they served to illustrate the concept of developing a prognosis. Thus it was assumed that these values were statistically representative.

Once the operating time zone was known from the vibration diagnosis [section 12.4], the remaining life was estimated as follows. It would not be known how far the current operating zone had progressed. However the bearing should last at least as long as the time taken for the following zones. Thus the bearing life could be prognosed with a certain associated probability.

Thus if the bearing was operating in zone N, then the remaining life could be calculated by :

$$L_R = \sum_{i=N}^{5} (L_i) \times L_{10h} \qquad \text{.....eqn. 13.1}$$

Where : L_R = Remaining life of the bearing
L_i = Percentage of rated life taken by zone 'i' [table 12.4]
L_{10h} = Bearing rated life [hrs] from equation 12.6.
N = Zone number diagnosed as being operated in.

For example if the bearing was diagnosed as operating in zone 2 (i.e. 'showing incipient fatigue in the form of surface scoring and light flaking') and the rated life was calculated as being 365 days, then the remaining life would be given by :

$$L_R = (0.5\% + 1.6\% + 3.6\%) \times 365 \text{ days}$$

$$= 21 \text{ days}$$

In addition a suitable testing interval could be deduced from the same source. For example if the bearing was running in zone 1 and the life rating was given as 365 days, then a suitable testing interval would be given by :

$$L_R = (2.8\% + 0.5\% + 1.6\% + 3.6\%) \times 365 \text{ days}$$

$$= 31 \text{ days}$$

In other words the bearing would have to be tested at least once a month to ensure a high probability of diagnosing failure before it was catastrophic.

Thus the rules were structured as follows.

```
IF:     BEARING CONDITION: Zone 2

THEN:   [LR] = [L10H] * 0.057 / 24
        [RETEST] = [LR] * 0.8
        The bearing will last for '[LR]' days
                                                Prob = 70%
        RECOMMEND:  Retest the bearing within the next
                    '[RETEST]' days.
```

Note that in this rule the retest interval was calculated as 80% of the remaining life.

Similar rules were developed for the other four testing zones.

Thus separate rulebases were developed for each of the stages of bearing condition analysis, viz. detection, diagnosis and prognosis.

SECTION 14

CONCLUSIONS REACHED

14. CONCLUSIONS REACHED

Through testing and development, the following conclusions were reached about the use of vibrations and operating conditions to analyze bearing running condition within the framework of an expert system.

- The analysis of bearing running condition could be divided into the three stages of detection, diagnosis and prognosis.
- Detection of bearing faults was more successfully accomplished by using a combination of condition monitoring parameters.
- The progression of bearing failure could be divided into five zones representing further stages of damage.
- The combination of the parameters RMS, Peak and Kurtosis performed well together to distinguish these five zones of damage.
- The operating conditions load, speed and temperature provided simple but effective diagnosis of the cause of bearing failure once failure had been diagnosed.
- The severity of combinations of operating conditions was more easily analyzed by grouping them into two parameters, viz. the 'rated bearing life' and the 'lubricant viscosity index'.
- The development of the expert system to analyze bearing condition was best accomplished by separate rulebases for the three tasks of detection, diagnosis and prognosis.
- The solution space of bearing analysis could be fully covered by dividing each parameter in the space into categories from which fuzzy logic lookup tables and rules could be constructed.
- Recommendations for further action (maintenance etc.) could be included at each stage of the expert system.
- The expert system required many rules before it could begin to perform analysis in any way close to that of a human expert.

The next section discusses recommendations made for further testing.

SECTION 15

RECOMMENDATIONS FOR THE PROJECT AND FURTHER TESTING

15. RECOMMENDATIONS FOR THE PROJECT AND FURTHER TESTING

Recommendations were made for the overall proposed condition monitoring project and for further testing on bearings. These were as follows.

THE OVERALL PROJECT

It was recommended that a project be established to develop an portable expert system based package to analyze overall condition of typical industrial machinery. The package could be developed by researching the monitoring of individual components which make up the machinery. These were listed in section 8.3.

Typically the research would be at the undergraduate or masters level and the overall system could possibly be developed as a Ph.D.. In addition the project could include a combined effort by the departments of Computer Science, Mechanical Engineering and Electrical Engineering.

As the specific components researched for this thesis were bearings, specific recommendations were made for these as follows.

FURTHER TESTING OF BEARINGS

Recommendation for the further testing of bearings included the following.

The destruction of more bearings to increase the statistical reliability of the methods developed in this thesis.

The testing and fine-tuning of the expert system rulebases developed in this thesis.

The testing of bearings under other simulated industrial failures such as ingression of lubricant contaminants.

The incorporation of automated lubricant debris analysis into the expert system.

The testing of other types of rolling element bearings.

The testing and incorporation into the expert system of other techniques such as spectrum and cepstrum analysis, enveloping, ultrasonics and acoustic emission.

SECTION 16

CONCLUSION

16. CONCLUSION

In conclusion it was felt that a successful bearing test and diagnosis system was developed. From a mechanical view the test rig worked well in the loading and driving of the test bearings. In addition it could easily be adapted for the testing of other bearings and even other components such as gears.

The data capture system attached to the rig also performed its various tasks accurately speedily and consistently. One of the more significant contributions of the system was that it measured the vibration parameters every few seconds for the life of the bearing. This gave a new depth of insight into the performance of these parameters.

One of the most interesting aspects of the project was the combination of different types of measurable information to form a more complete analysis of bearing condition. In particular the operating conditions provided a significant contribution to the tasks of diagnosis and prognosis.

The work also revealed that it was a complex task to achieve meaningful and real expert-like analysis with an expert system. The expert provided no stunning or unexpected conclusions and the performance was predictable. However, to a person less skilled in diagnostics the system could provide new insight and relieve an otherwise difficult task. To the more experienced user it could relieve the routine and free up time for the more interesting diagnostic problems.

The future of this type of system looks bright with a large amount of development being done internationally. It is hoped that work in this country will keep abreast of this trend and that the recommended research and development will come to fruition.

SECTION 17

REFERENCES

17. LIST OF REFERENCES

1. "Condition Monitoring of Machinery and Plant". Papers presented at a seminar organized by the Fluid Machinery Committee of the Power Industries Division of the Institution of Mechanical Engineers, London (June 1985).

2. Noyes R.B. and Jongens A.W.D. Machinery Vibration Analysis Course presented to Koeberg Nuclear Power Station (1987).

3. Leggat S.B. "The Development of a Computerized Vibration Diagnostics Package Using Expert Systems to Determine Machinery Running Condition and Provide Maintenance Recommendations". University of Cape Town, 1988.

4. Noyes R.B. "Application For Equipment Grants - 1990 : Multi Channel Spectrum Analyzer", Internal Memorandum, Department of Mechanical Engineering, University of Cape Town, August 1989.

5. Leggat S.B. A presentation given to SA Nylon Spinners Management on the development of vibration monitoring at the University of Cape Town, SANS, March 1988.

6. Leggat S.B. "Proposed Vibration Project", Koeberg Performance Monitoring, December 1988.

7. Leggat S.B. "The Condition Monitoring of Rotating Machinery using Vibration Analysis", Undergraduate Thesis, UCT Mechanical Engineering, 1987.

8. Oertel C. "The Use of Vibration Measurements to Determine Bearing Condition", Undergraduate Thesis, UCT Mechanical Engineering, 1988.

9. Newland D.E. "An Introduction to Random Vibrations and Spectral Analysis". Second Edition, Longman 1984.

10. Braun S. (Ed). "Mechanical Signature Analysis - Theory and Applications". Academic Press, 1986.

11. SKF "Bearing Installation and Maintenance Guide". Application Note 140-710, April 1986.

12. ISO 281/I-1977. Recommendations for the Determination of Fatigue Life in Rolling Element Bearings.

13. SKF "General Catalogue - 3200/I E", SKF 1981.

14. Shigley J.E. "Mechanical Engineering Design - First Metric Edition". McGraw-Hill 1986.

15. Eschman P, Hasbargen L, Weigand. "Ball and Roller Bearings; Theory, Design and Applications". 2nd Edition, J Wiley & Sons.

16. SKF "Bearing Maintenance and Replacement Guide". Catalogue 3600E, SKF 1986.

17. Dyer D, Stewart R.M. "Detection of Rolling Element Bearing Damage by Statistical Vibration Analysis". Journal of Mech. Design, April 1978, Vol 100.

18. Bannister R.H. "A Review of Rolling Element Bearing Monitoring Techniques". Paper presented at a seminar organized by the Institute of Mechanical Engineers, London, 1985.

19. Stewart Hughes Ltd. "Bearing Fault Diagnosis Technology Brochure - SHL 135". August 1984.

20. Levine R.I, Drang D.E, Edelson B. "A Comprehensive Guide to AI and Expert Systems". McGraw-Hill 1986.

21. Exsys, Inc. Exsys Expert System Shell Users Guide - Version 3.2.5 Exsys, Inc. 1985.

22. Stewart R.M. "The Way Ahead for Machinery Health Monitoring as a Subset of Plant Control - Part Two". Noise and Vibration Control Worldwide, March 1985.

23. Leanord-Barton D, Sviokla J. "Putting Expert Systems to work". Harvard Business Review, March 1988, Vol 66, No 2.

24. Exsys, Inc. Exsys Expert System Shell and Software.

25. Leggat S.B. "An investigation Into Using Condition Monitoring on The SANS Polymer Plant to Reduce Maintenance Costs and Lost Production". UCT, 1988.

26. See ref 15.

27. Bohler Steel (Pty) Ltd, Cape. Advice on and supply of case hardened M120 steel shaft. 1988.

28. Data Translations, Inc. "Single Board Analog and Digital I/O systems for the IBM personal Computer, DT 2801 Series User Manual". 1983.

29. Vos J. Notes From a Postgraduate Software Development Course, Department of Civil Engineering, University of Cape Town, 1988.

30. "Computer Science Study Guide 1 for COS112-V, Programming : Practical". Department of Computer Science and Information Systems, University of South Africa, 1987.

31. Study Curriculum, Department of Computer Science, University of Cape Town, 1989.

32. See ref 9.

33. Holman J.P. "Experimental Methods For Engineers". Fourth Edition, McGraw-Hill, 1984.

34. Gryzagorides J. Notes from a Final Year Course in Experimental Methods. UCT, 1987.

35. Analog Devices, Inc. "The AD594 Monolithic Thermocouple Amplifier with Cold Junction Compensation". Product description and Specifications.

36. See appendix 50.

37. Rockland "FFT Spectrum Analyzers" 1988 Catalog. Rockland Scientific Corporation.

38. "Rion VA-10 Vibration Analyzer". Specifications and Product description. Andersen and Hurley Instruments. CC., Johannesburg, 1988.

39. "CSI Wavepack". Specifications and Product Description. Andersen and Hurley Instruments. CC., Johannesburg, 1988.

40. Horowitz P, Hill W. "The Art of Electronics". Cambridge University Press, 1982.

41. Course Notes for a Final Year Course In Electronic Circuit Design, Department of Electrical Engineering, University of Cape Town, 1988.

42. See ref 16.

43. Penter A.J. "Vibration Monitoring of low-speed machinery as found in the mining industry". Paper presented at a seminar organized by the Institute of Mechanical Engineers, London, 1985.

44. Grantham R. "Lubricant Wear Debris analysis in Rolling Element Bearings". Undergraduate thesis, UCT Department of Mechanical Engineering, 1989.

45. See ref 18.

46. Humphrey D, Topping J. "A Shorter Intermediate Mechanics". Longman, 1971.

47. SKF "General Catalogue - 4000 E", SKF 1989.

48. "Dynamic Signal Analyzer Applications - Effective Machinery Maintenance Using Vibration Analysis". Hewlett Packard Application Note 243-1, 1983.

49. "SKF Bearing Detector : TMED 1 Handbook". SKF Maintenance Products.

50. Mathew J, Alfredson R.J. "The Condition Monitoring of Rolling Element Bearings Using Vibration Analysis". Journal of Vibration, Acoustics, Stress and Reliability in Design, Vol 106, July 1984.

51. Broch J. T. "Mechanical Vibration and Shock Measurements". Bruel & Kjaer, 1984.

52. "Generic Control Blackboard - A Batch Queueing System for the PC" Users Manual, Meta (Inference) Services, Inc. 1987.

53. "The Generic Data Blackboard - Release 2.5". Users Manual, Meta (Inference) Services, Inc. 1987.

54. van Niekerk F, Sunder R. "COMOS - An Online System for Problem-Orientated Vibration Monitoring", Geseblschaft f.r Reaktorsicherheit (GRS) mbH, D-8046 Garching, Federal Republic of Germany.

55. A demonstration of the Toshiba T5200 Laptop Computer, Joffe and Associates, Cape Town, 1988.

56. A demonstration of the CSI "Wavepak" given at the Mount Nelson Hotel, Cape Town, by Vibracon (Vibration Consultants) in association with Andersen and Hurley Instruments.

57. David K. "The Use of Vibration Measurements to Determine Cavitation in Pumps". Undergraduate Thesis, UCT Mech. Eng. 1988.

58. Wighton D. "The Use of Vibration Measurements to Determine Gear Condition". Undergraduate Thesis, UCT Mech. Eng. 1989.

APPENDICES

APPENDIX 1 : AN EXAMPLE OF BEARING STRESS CYCLING LEADING TO HIGH CYCLE FATIGUE

In typical operation a bearing raceway can experience rapidly accumulating stress cycling with resulting fatigue damage (if high stresses are present). For example the bearings used for testing had a Ball Pass Frequency on the Outer race of 6.285 times for each revolution of the shaft [append. 22].

In other words the stress on the outer race would be cycled 6.285 times for every rotation of the shaft. Therefore if the shaft was rotating at 1800 rpm, the stress cycles would occur at the following hourly rate :

$$\text{Stress cycle rate} = 6.285 * 1800 \text{ rpm} * 60 \text{ mins}$$

$$= 678\ 780 \text{ cycles/hour}$$

Or alternately :

$$\text{Hours for } 10^6 \text{ Cycles} = 1 / 0.678780$$

$$\text{approx.} = 1\tfrac{1}{2} \text{ hours}/10^6 \text{ cycles}$$

APPENDIX 2. : A SYSTEMATIC METHOD FOR MONITORING MACHINERY CONDITION

While working on the SANS project [5] a systematic method was developed for monitoring machinery condition using vibration analysis. The outline of this system is presented in simplified form as follows :

- Obtain drawings and equipment specifications, such as number of teeth on gears, where possible.
- Calculate expected frequencies such as rotating speed and harmonics, gear mesh, vane pass, bearing element passage, etc.
- Study bearing positioning to find shortest transmission path of vibration to machine outer surface. Ensure minimum mechanical interfaces in this path.
- Determine maximum load directions of components such as bearings and gears.
- Choose Accelerometer mounting positions according to the above criteria.
- Use a spectrum analyser to investigate frequency content of signal. Use exponential averaging or peak hold where vibrations are transient.
- Select frequency ranges which cover expected frequencies and add any other observed frequencies of interest. Usually 200 Hz, 1kHz & 10kHz ranges will cover rotating freq, gears and vanes and bearings respectively.
- Select windowing and averaging techniques with overlap processing according to needs.
- Collect overall vibration levels and spectra as 'baselines' to be used for future comparisons.
- Monitor periodically ensuring the following :
    - Repeatability of measurements.
    - Correct accelerometer mounting.
    - Short enough interval between monitoring to ensure incipient failure detection.

Note that while each item is presented briefly here it contains extensive detail in implementation.

APPENDIX 3. : SPECIFICATIONS OF THE BEARINGS AND RELATED EQUIPMENT

The following table presents the SKF codes and load ratings of the bearings and related equipment [13].

ITEM DESCRIPTION	CODE	LOAD RATING
Test bearings :		
Self aligning ball bearings	1207K	15,9 kN (Dynamic) 6,7 kN (Static)
	1207EK	19,0 kN (Dynamic) 6,0 kN (Static)
Support bearings :		
Spherical roller bearings	2207CCK	63,3 kN (Dynamic) 40,5 kN (Static)
Tapered adaptor sleeves		
- Test bearings	H207F	
- Support bearings	H307F	
Plummer blocks	SNH507	25 kN (Max load) 75 kN (Yield)
"V-ring" seals	TSNA 507A	

APPENDIX 4. : DESIGN CALCULATIONS FOR THE BEARING SHAFT

The bearing shaft was recognised as a critical component from previous research [7]. A shaft was recommended for further testing done on the same test rig. It was required to have a UTS (ultimate tensile strength) of greater than 1900 MPa using a safety factor of approximately 6 to account for high cycle fatigue. This was for a maximum applicable load of approximately 17 kN.

An M120 steel shaft was recommended by BHOLER steel [27]. This was to be case hardened to a depth of 0.4 mm. The UTS of the shaft was given as being in excess of 2000 Mpa at the case hardened surface. This was the strongest steel which could be practically obtained.

However with the new test rig used for research, the maximum applicable load was 25 kN. Thus a new required UTS had to be calculated. In addition the dimensions of the bearing spacings were reduced with the re-design of the test rig in an attempt to decrease the maximum applied moment. The new dimensions are shown in figure A2.1. while the maximum applied stress is calculated as follows.

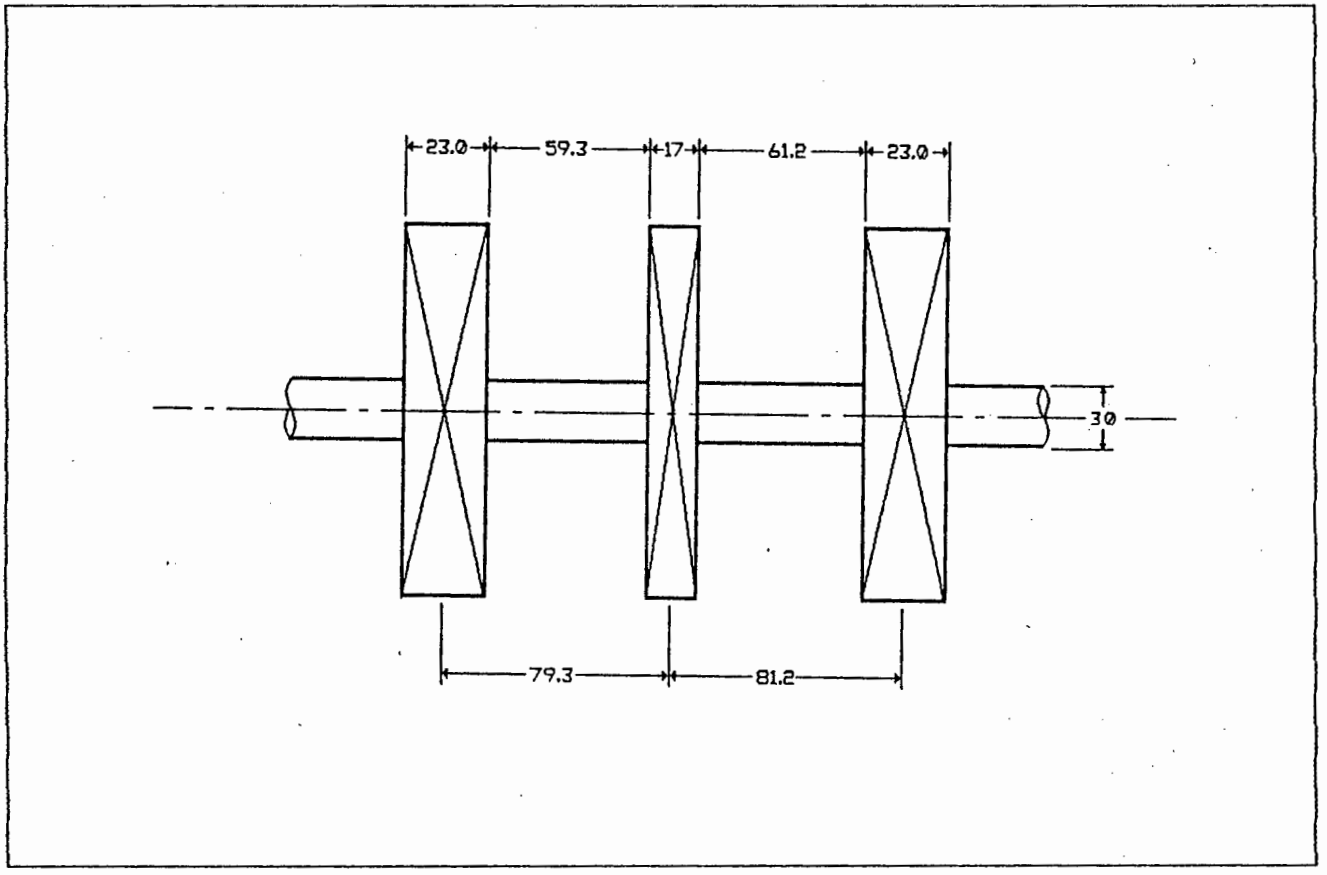

FIGURE A3.1 : DIMENSIONS OF THE SHAFT AND BEARINGS FOR CALCULATING THE MAXIMUM SHAFT STRESS

CALCULATION OF THE MAXIMUM STRESS APPLIED TO THE BEARING SUPPORT SHAFT :

Maximum applied moment :

$$M = 12.6 \text{ kN} * 0.0793 \text{ m}$$
$$= 1.0 \text{ kNm}$$

For shaft diameter d = 0.03 m the second moment of area I will be:

$$I = \frac{\pi * d^4}{64}$$

$$I = \frac{\pi * (0.03)^4}{64}$$

$$= 3.976 \times 10^{-8} \text{ m}^4$$

The maximum stress will ocurr at the surface of the shaft, giving:

$$\sigma_{max} = \frac{M\ y}{I}$$

$$\sigma_{max} = \frac{1 \text{ kNm} * 0.015 \text{ m}}{3.976 * 10^{-8} \text{ m}^4}$$

$$\sigma_{max} = 377.256 \text{ MPa}$$

CALCULATION OF THE ENDURANCE LIMIT OF THE BEARING SUPPORT SHAFT USING HIGH CYCLE FATIGUE ANALYSIS

This analysis is derived from the method used by J E Shigley [14]. All page numbers refer to this reference.

$$S_e = K_a * K_b * K_c * K_d * K_e * K_f * S'_e$$

Where :

S_e	=	endurance limit of shaft
S'_e	=	endurance limit of rotating beam specimen
K_a	=	surface factor
K_b	=	size factor
K_c	=	reliability factor
K_d	=	temperature factor
K_e	=	stress concentration factor
K_f	=	miscellaneous effects factor

For this application :

S'_e	=	0.4 * UTS (conservative est.)	[Fig. 7-4 p.232]
K_a	=	0.7 for machined shaft	[Fig. 7-8 p.244]
K_b	=	$1.189 * d^{-0.097}$ for 8mm < d $\leq$ 250 mm	
	=	$1.189 * 30^{-0.097}$	
	=	0.855	[eqn. 7-16 p.246]
K_c	=	0.753 for 0.999 reliability	[Tbl. 7-16 p.246]
K_d	=	1 for T < 350°C	[eqn. 7-26 p.253]
K_e	=	1 for no stress concentration	
K_f	=	0.9 for fretting corrosion	[p.258]

Combining these factors gives :

$$K = K_a \ldots . K_f = 0.7 * 0.855 * 0.753 * 0.9 = 0.41$$

Now starting with $S_e = \sigma_{max} = 378$ MPa gives :

$$S'_e = 378/0.41 = 922 \text{ Mpa}$$

and $$S_u = 922/0.4 = 2305 \text{ Mpa}$$

Although this required S_u was higher than the quoted 2000 MPa, the shaft was not run at a loading of 25kN for more than about 10^4 cycles. Thus it lasted the duration of all bearing testing without failure.

APPENDIX 5. : SPECIFICATIONS OF THE ANALOG TO DIGITAL CONVERSION CARD

The following specifications apply to the DT2801-A analog to digital conversion card (typical at 25°C and rated voltage unless otherwise specified):

ANALOG INPUTS:

Number of Analog Inputs	16 Serial 8 Direct
Input Ranges	Jumper selectable, unipolar or bipolar;
Unipolar Input Ranges	0 to +10V, 0 to +5v, 0 to +2.5V, 0 to +1.25V full scale, depending on programmable gain setting
Bipolar Input Ranges	±10V, ±5V, ±2.5V, ±1.25V full scale depending on programmable gain setting
Output data codes	Straight binary (unipolar) Offset binary (bipolar)
Software programmable Gain Range	1, 2, 4, 8
Input impedance	Off Channel : 100 MΩ, 10pF On Cannel : 100 MΩ, 100pF
Bias Current	±20nA
Common Mode Input Voltage, Maximum	±11V
Common Mode Rejection Ratio (CMRR), Gain = 1	80dB at 60kHz, 1kΩ unbalanced
Maximum Input Voltage Without Damage	±30V Power on ±20V Power off
Amplifier input noise	10^{-6}V rms ±1 LSB (bipolar, Gain = 8) Note: Where gain > 1, input noise is multiplied by gain.
Channel-to-channel Input Voltage Error	$\pm 10^{-6}$V

ACCURACY :

Resolution	12 Bits
Differential Nonlinearity	less than ± ½ LSB
Inherent Quantizing Error	less than ± ½ LSB
System Accuracy	to within ±0.05% FSR
Channel Crosstalk	-80dB at 1kHz
Sample & Hold Droop Rate	0.1 mV/mS
Gain Error	Adjustable to 0
Zero Error	Adjustable to 0

DYNAMIC PERFORMANCE :

Channel Acquisition Time to within ½ LSB	$1.5 * 10^{-5}$ Seconds
A/D Conversion Time	$1 * 10^{-5}$ Seconds
A/D Throughput to System Memory	27 500 samples per second (using direct memory access)
Sample & Hold Aperture Uncertainty	$1 * 10^{-8}$ Seconds
Sample & Hold Aperture Delay	$5 * 10^{-8}$ Seconds
Sample & Hold Feedthrough Attenuation	80dB at 1kHz

THERMAL CHARACTERISTICS :

A/D Zero Drift	$\pm 2 * 10^{-5}$ V/°C (unipolar) ±20 ppm of FSR/°C (bipolar)
Amplifier Zero Drift	$\pm 2.5 * 10^{-5}$ V/°C $\pm((3 * 10^{-6}$ V/°C) * Gain)
Gain Drift	±35 ppm of FSR/°C
Differential Linearity Drift	±3 ppm of FSR/°C
Monotonicity	Monotonic, 0 to +50°C

APPENDIX 6. : CALCULATION OF THE DYNAMIC RANGE AND RESOLUTION OF THE ANALOG TO DIGITAL CARD

The dynamic range and resolution of the A/D card are related as follows (see specs. in append. 5) :

12 bit A/D => 2^{12} possible integers representing discrete digital voltage amplitudes

= 4096 integers

= 20 * Log (4096) dB

= 72.25 dB Dynamic Range

The voltage resolution 'VR' will depend on the FSR (full scale voltage range) for the various gain settings as follows :

Gain 1 : VR = FSR / 4096

VR = 10V / 4096

= 2.44 mV

Gain 2 : VR = FSR / 4096

= 5V / 4096

= 1.22 mV

Gain 4 : VR = FSR / 4096

= 2.5V / 4096

= 0.61 mV

Gain 8 : VR = FSR / 4096

= 1.25V / 4096

= 0.305 mV

APPENDIX 7. : DERIVATION OF THE EQUATION TO CONVERT THE ANALOG TO DIGITAL CARD BINARY VALUES TO VOLTS

The A/D Card was Jumper Configured for use in the Bipolar Mode to accept the ± Voltage range produced by the Accelerometer. Hence all voltages were represented by an Offset Integer on a linear scale as follows :

For A/D Integer = 0
Output Voltage = - Full Scale Volts

For A/D Integer = 4095
Output Voltage = + Full Scale Volts

For A/D Integer = 4096 / 2
= 2048
Output Voltage = 0 Volts

Therefore the Output Voltage may be calculated from the A/D Integer using the FSR (Full Scale Range) and the Software Configurable Gain as follows :

$$\text{Output Voltage} = \frac{\text{FSR} * (\text{A/D Integer} - 2048)}{4096}$$

The output voltage may then be converted to the measured parameter if the conversion factor from volts is known.

APPENDIX 8. : CALCULATION OF THE ANALOG TO DIGITAL CARD SAMPLING FREQUENCY

The clock speed of the card was given as 0.8 MHz [28], which related to a `Clock Base' of $1.25*10^{-6}$ seconds (1/clock speed). The sampling frequency of the Card was then defined by some integer multiple of the clock base as follows :

Sampling Frequency = 1/(Integer * Clock Base)

The lowest useable integer multiple was limited by the combined Channel Acquisition Time and A/D Conversion Time [append. 5] as follows :

Channel Acq. Time + A/D Conv. Time

$= (10 + 15) * 10^{-6}$ Secs

$= 25 * 10^{-6}$ Secs

= 20 * Clock Base

In addition it was needed to account for DMA (Direct Memory Access) Time and a safety margin to ensure data integrity. Thus a minimum integer multiple of 29 was recommended by the manufacturers [28]. This corresponded to a maximum sampling frequency as follows :

Max Sampling Freq = 1/(29 * Clock Base)

$= 1/(29 * 1.25 * 10^{-6})$

= 27 586 Hz [cf. append. 5]

APPENDIX 9. : SELECTION OF THE NUMBER OF SAMPLES AND SAMPLING FREQUENCY FOR DIGITALLY MEASURED PARAMETERS

It was necessary to select the number of samples and the sampling frequency for each parameter that was measured digitally. These parameters were the operating conditions Load, Speed and Temperature and the Vibration parameters RMS and Peak Acceleration, Kurtosis and Crest Factor. The selection is discussed as follows :

THE OPERATING CONDITIONS

Firstly the two parameters Load and Temperature were represented by DC values which were regarded as constant for each sampling. Theoretically, one sample should have been sufficient to characterise the parameter at that instant.

However the signals were contaminated with some amount of noise from various sources [section 10]. This was so even after using capacitive smoothing and other noise reduction techniques.

Therefore it was necessary to use statistical averaging to obtain repeatable measurements. The amount of samples necessary to obtain repeatability was selected by experimentation. It was found that 100 samples were sufficient in each case.

In addition it was found that the maximum sampling frequency (27.5 kHz) could be used for measuring Temperature, while a sampling frequency of 20 kHz provided more reliable results for Load. This was because of periodic noise on the Load measuring system.

Secondly, the resolution with which shaft speed could be calculated was proportional to the amount of samples captured and the sampling frequency. It was found that 6000 samples taken at the maximum sampling frequency (27.5 kHz) provided the required resolution [append. 23].

THE VIBRATION PARAMETERS

The number of samples taken for the calculation of vibration parameters was determined by two aspects. These were the sampling frequency and the length of time required to capture enough data to characterise the bearing condition accurately.

Firstly the sampling frequency was determined by two factors. These were the response of the accelerometer and the frequency range of typical damaged bearing noise. From previous experience [7] it was known that bearings produced a characteristic high frequency noise occuring in 'broadband humps'. These started at about 5 kHz and continued up to the maximum response range of the data capture system.

The main limitation to the response of the system was therefore the range of the accelerometer. This was known to be in the region of 8 kHz [6]. Therefore to prevent aliasing effects the sampling frequency would have to be at least 2 x 8 kHz to fulfil the Nyquist sampling criterion [append. 49].

In addition, to take account of the deviation from ideality of the antialiasing filter, the sampling frequency would usually be set to 2.56 * 10 kHz [append. 49], giving a sampling frequency of 25.6 kHz.

The second approach to selecting the sample frequency was to consider the number of samples captured in one revolution of the shaft at full speed. This would determine whether an impact such as a ball passing over a small defect on the bearing surface could be characterised accurately enough. The higher the sampling frequency the better the characterisation of the vibration by the sampled time waveform.

For example if the highest sampling speed of 27.5 kHz was used on a bearing running at 1800 rpm, then the following would apply to the sampled waveform :

```
          Shaft speed =  1800 rpm
                      =    30 rps
                      =>  1/30 Secs/rev

    For Sampling Speed =  27.586 kHz
                      => 27 586 / 30  samples/rev
                      =    919.5  samples/rev
                      =>  2.55 samples per degree of rotation
```

In addition the higher the sampling frequency the less the Biasing error [section 6]. Therefore it was decided to use the highest sampling frequency possible (27.586 kHz)

As for the number of samples collected it was desired to capture as long a sample as possible. It was known from previous work [7] that too few samples caused unreliable and erratic statistical parameters. This was due to a decrease in the possibility of capturing single events in the bearing operation. Such events would be for example a ball defect occasionally impacting on one of the bearing races.

Therefore it was necessary to select the maximum number of samples possible. The only real limitation to the number of samples was the time required to process them. It was found by trial and error that approximately 6000 points provided repeatable statistical parameters while allowing for a total processing time of about 2 seconds.

The 6000 points sampled at 27.586 kHz represented 6.525 revolutions of the shaft at maximum speed. This was sufficient to capture most of the transient events as discussed above.

The following table provides a summary of the number of samples and related sampling frequencies :

TABLE A9.1 : SUMMARY OF SAMPLES AND SAMPLING FREQUENCIES USED TO COLLECT TEST DATA

PARAMETER SAMPLED	SAMPLING FREQUENCY	SAMPLES PER RECORD
LOAD	20.000 kHz	100
SPEED	27.586 kHz	6000
TEMPERATURE	27.586 kHz	100
VIBRATIONS	27.586 kHz	6000

APPENDIX 10 : THE COOLEY-TUKEY RADIX 2 FAST FOURIER TRANSFORM ALGORITHM

The Fast Fourier Transform was used to transform the vibration signals from the Time domain to the Frequency domain. It can be shown that the Fourier Transform of a finite time sequence x_R of length N is another finite sequence X_K in the frequency domain, where the two are related by the following equation [9].

$$X_K = \frac{1}{N} \sum_{R=0}^{N-1} x_R e^{-i(2\pi KR/N)} \qquad \text{.....eqn. A10.1}$$

In Particular the Cooley-Tukey Radix 2 implemenation of the transform was used. This algorithm operates by sucessively partitioning the original sequence into two half sequences of odd and even terms. The partitioning continues until the sequence consists of N partitions of one term each. The transform of each term will then be itself, whereupon the results are combined to obtain the whole transform.

The algorithm starts with a sequence of length N, where N must be a power of 2. This is so that the original sequence can be partitioned until N partitions are obtained. Figure A10.1 shows a flow diagram of the algorithm.

The process of partitioning results in the reordering of the initial array into a 'bit-reversed' array. The position of a term in the reordered sequence can be determined by reversing the order of the bits in a binary representation of the original term position. This reordering is achieved by a bit reversed counter algorithm applied to the sequence before the transform [append. 11].

There are various levels of increasing sophistication of the FFT algorithm as implemented in real situations. These become too complex for discussion in this appendix.

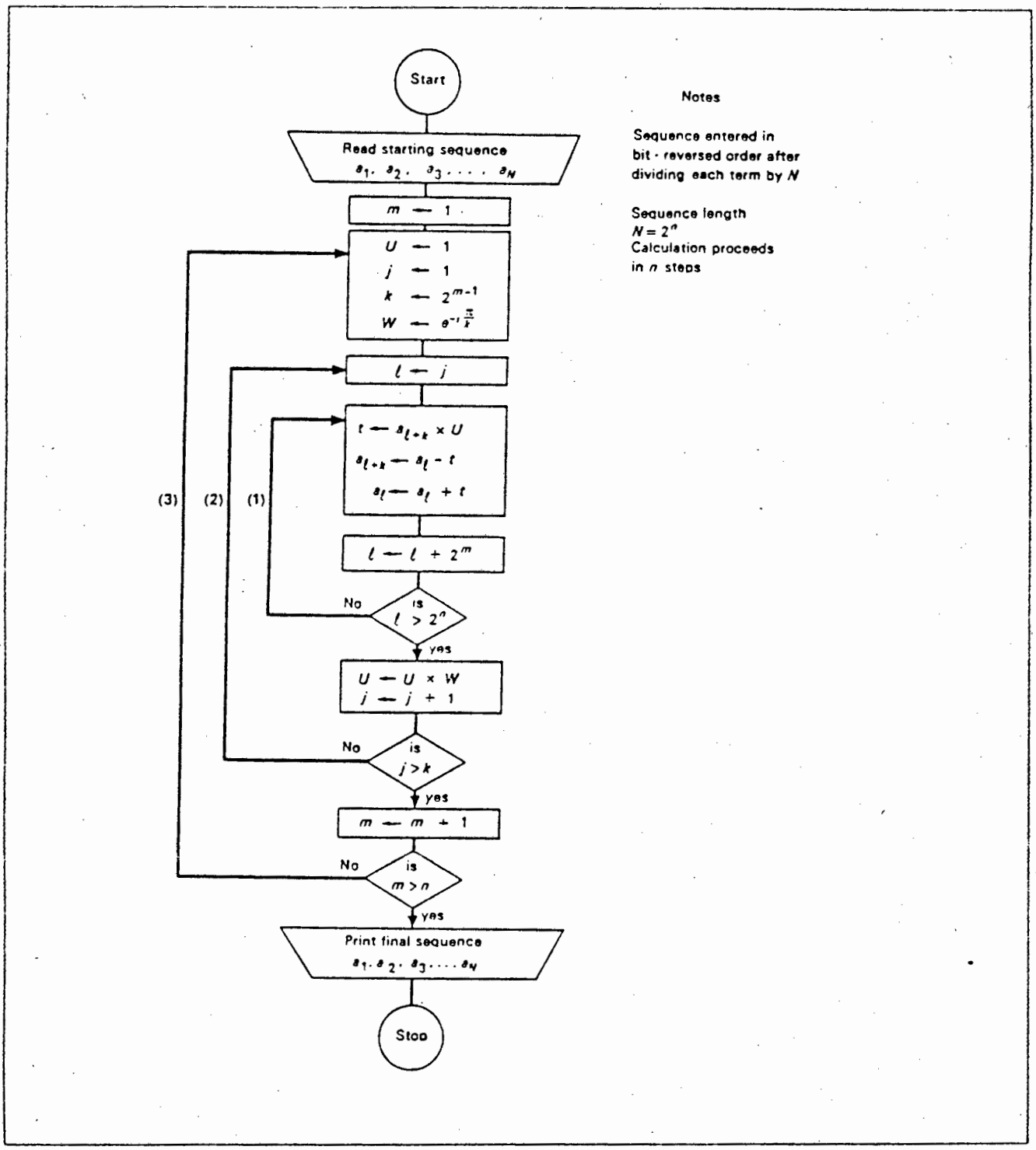

FIGURE A10.1: FLOW CHART OF THE COOLEY-TUKEY RADIX 2 FAST FOURIER TRANSFORM ALGORITHM

APPENDIX 11. : THE BIT REVERSED COUNTER ALGORITHM

The use of a bit-reversed array is required as the input for the FFT algorithm in appendix 10. The algorithm used to achieve this is illustrated in figure A11.1 by a programming flow chart.

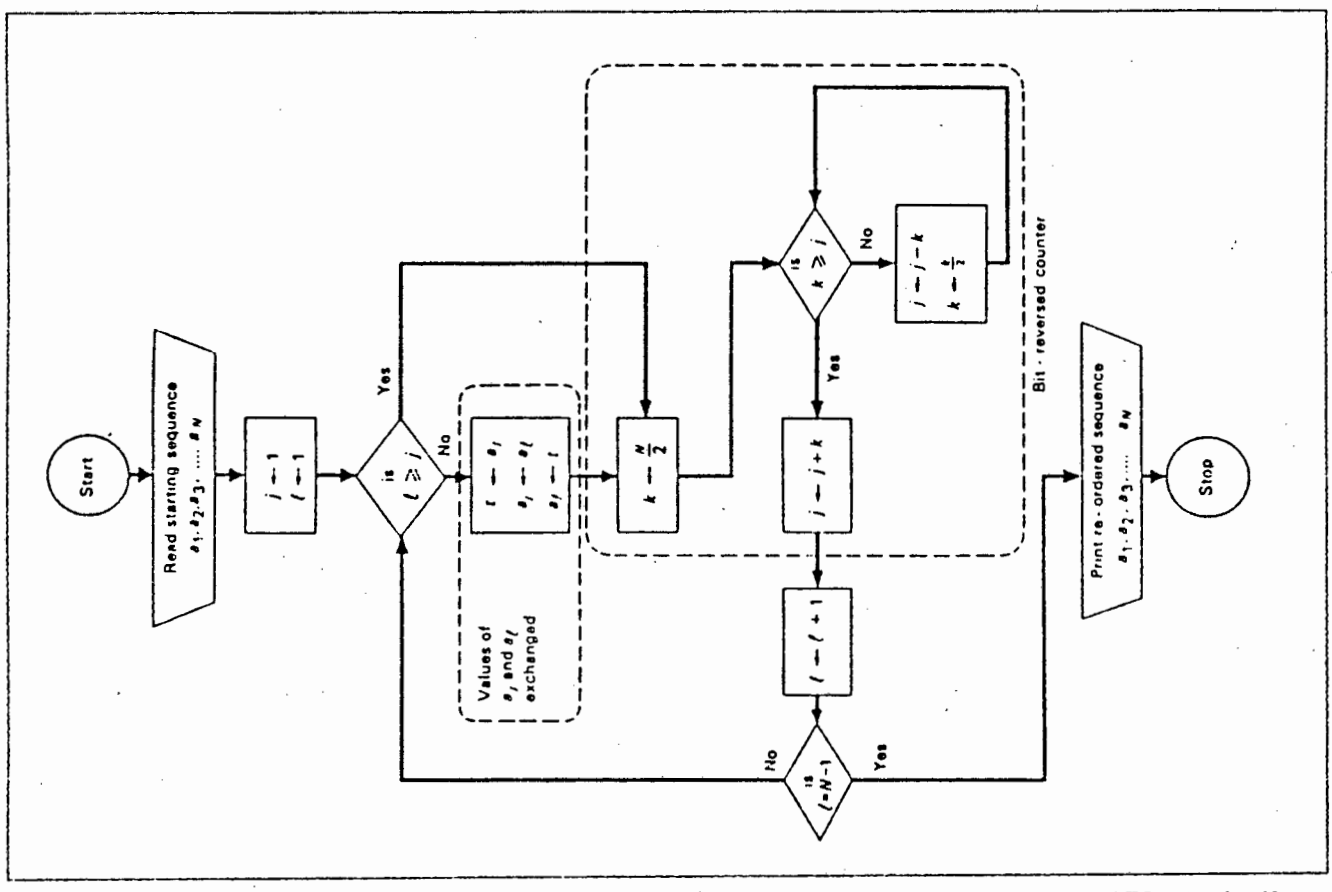

FIGURE A11.1: PROGRAMMING FLOW CHART OF THE BIT REVERSED ARRAY ALGORITHM.

APPENDIX 12. : SELF DESIGNED TIME DATA WINDOW

The use of time domain windowing on a time sequence before it Fourier transformed is common practice [9,10]. One of the reasons for performing windowing is the phenomenon of 'leakage' in the Fourier spectrum. This is observed as a leakage of amplitude to adjacent terms in the frequency sequence [append 10] when the time sequence is not periodic within the sample.

Thus the purpose of the time domain window is to force the time sequence to zero at both ends, whereupon it effectively becomes periodic within the time sample. The peaks of the spectrum are then seen to become more narrow, and the measurement of frequency supposedly more accurate. The better frequency accuracy is traded off with a reduction in amplitude accuracy.

A self designed window was experimented with to observe these effects. The window made use of $Sin^2\Phi$ and $Cos^2\Phi$ functions with Φ varying between 0 and $\pi/2$. This provided the smooth transitions at the beginning and end of the time sequence, while the top was flat [figure A12.1].

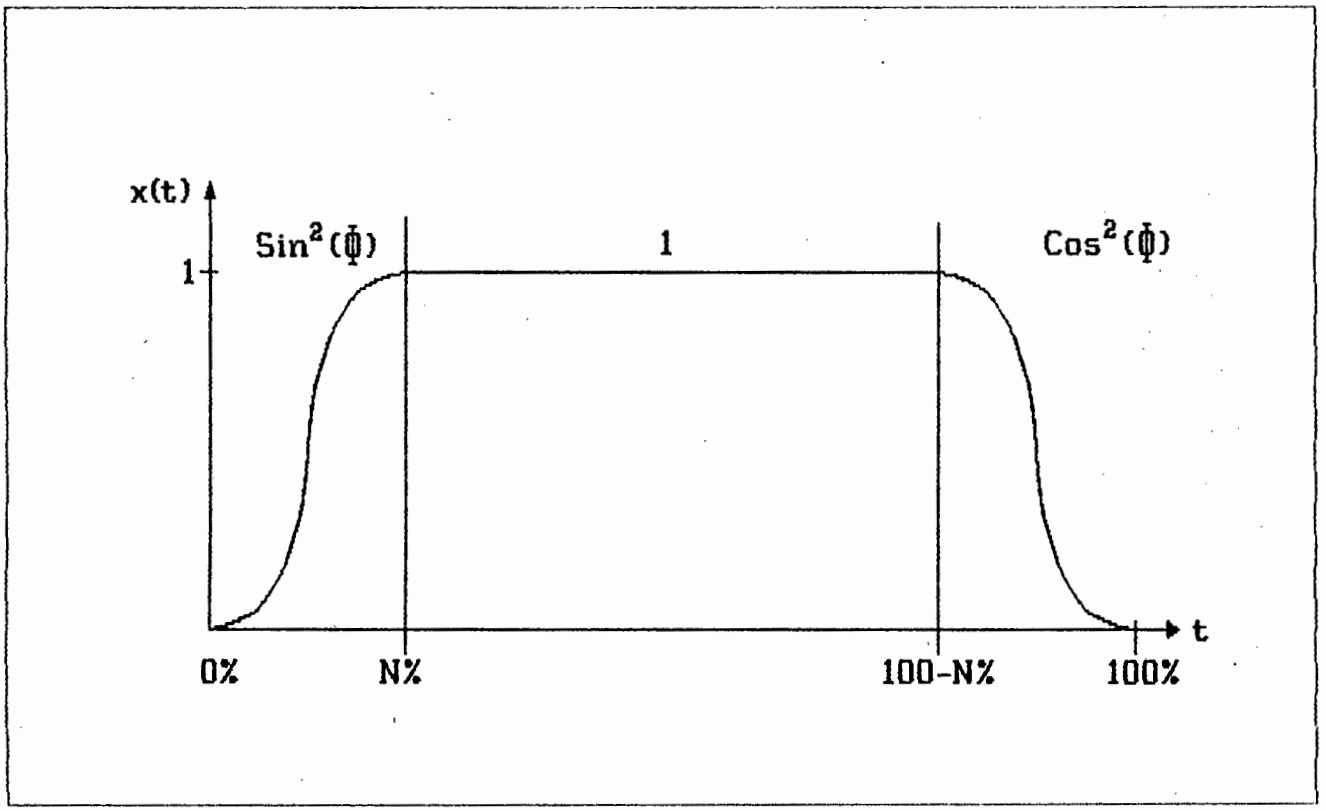

FIGURE A12.1: SELF DESIGNED TIME DOMAIN WINDOW FOR FFT ANALYSIS

A feature of the windowing algorithm was that the width of the flat top could be varied as a percentage of the length of the sequence. For example the $Sin^2\Phi$ and $Cos^2\Phi$ portions could be set to take up 10% each and the top the remaining 80% or any other combination totalling 100%.

The effect of varying the width of the top of the window was as follows. As the top was made wider, the spectral peaks widened and the amplitude was reduced. Narrowing the width of the top caused the peaks to narrow and the amplitude to increase. When the top was increased to take up more than about 95% of the sequence length, the leakage was clearly evident.

The windowing algorithm was computationally intesive and contributed to making the total process of frequency analysis on th IBM XT computer very slow.

APPENDIX 13. : STRAIN GAUGE MOUNTED SPECIFICATIONS

The following specifications applied to the strain gauges used for the load measuring system :

Adoptable Thermal Expansion	ppm/°C
Gauge Length	2 mm
Gauge Factor Change with Temp.	0.015%/°C
Lot Number	Y1781-520
Gauge Resistance	120Ω ± 0.4Ω
Gauge Factor	2.08 ± 1%
Thermal Output	± 1.8 uε/°C
Temperature Compensation For Steel	

APPENDIX 14. : RESULTS OF THE PRESSURE GAUGE CALIBRATION TEST

GAUGE PRESSURE [kPa]	LOAD [kN]					
	USING PUMP		USING INSTROM		STATISTICAL	
	UP	DOWN	UP	DOWN	AVERAGE	REGRESS
0	0.00	0.00	0.00	0.00	0.00	0.46
500	1.21	1.81	1.74	1.27	1.51	1.56
1000	2.22	3.08	2.94	2.23	2.62	2.65
1500	3.30	4.22	4.13	3.31	3.74	3.75
2000	4.30	5.29	5.34	4.39	4.83	4.85
2500	5.39	6.58	6.45	5.53	5.99	5.95
3000	6.55	7.66	7.65	6.64	7.12	7.04
3500	7.55	8.55	8.79	7.94	8.21	8.14
4000	8.67	9.50	9.81	9.12	9.27	9.24
4500	9.74	10.41	11.08	10.16	10.35	10.34
5000	10.72	11.23	12.05	11.50	11.38	11.43
5500	11.79	-	13.14	-	12.47	12.53

APPENDIX 15. : RESULTS OF STRAIN GAUGE BRIDGE CALIBRATION TEST

The following table shows the test results for calibration of the load cell. The units are Load in kN and strain in microstrain.

GAUGE STRAIN	INSTROM LOAD 1U	INSTROM LOAD 2U	INSTROM LOAD 3U	INSTROM LOAD 1D	INSTROM LOAD 2D	INSTROM LOAD 3D	AVERAGE ALL LD	REGRESS LINE
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.50
50	1.11	1.21	0.93	1.02	0.89	0.82	1.00	0.79
100	2.18	2.31	2.03	1.91	1.78	1.73	1.99	2.08
150	3.25	3.41	3.20	2.97	2.89	2.85	3.10	3.37
200	4.35	4.54	4.35	4.14	4.02	3.96	4.23	4.67
250	5.64	5.91	5.72	5.57	5.43	5.42	5.61	5.96
300	7.08	7.29	7.10	7.04	6.91	6.86	7.05	7.25
350	8.69	8.76	8.53	8.46	8.35	8.28	8.51	8.54
400	10.32	10.17	11.08	10.01	9.95	9.83	10.23	9.84
450	11.61	11.52	11.30	11.48	11.32	11.30	11.42	11.13

APPENDIX 16. : RESULTS OF STRAIN GAUGE AMPLIFIER CALIBRATION TEST

The strain gauge amplifier was calibrated by manually adjusting the output of the strain gauge bridge to the values shown in column 1 of table A16.1. The equivalent a to d value was then read from the computer and is shown in column 2. This was then converted into an equivalent piston load using the results of the calibration in appendix 15 (see appendix 17). The piston load was then doubled to give the load on the bearing.

TABLE A16.1 : STRAIN GAUGE AMPLIFIER CALIBRATION RESULTS

Bridge [Micro Strain]	Digital A to D Value	Piston Load [kN]	Bearing Load [kN]	Regression Analysis [kN]
0	1486	-0.503	-1.005	-0.96841
50	1737	0.790	1.580	1.57141
100	1992	2.082	4.165	4.15172
150	2247	3.375	6.749	6.73203
200	2503	4.667	9.334	9.32245
250	2759	5.960	11.919	11.91287
300	3016	7.252	14.504	14.51342
350	3270	8.544	17.089	17.08360
400	3526	9.837	19.674	19.67403
450	3783	11.129	22.259	22.27457

TABLE A16.2 : REGRESSION ANALYSIS FOR TABLE A16.1

Regression Output:	
Std Err of Y Est	0.0175
R Squared	1.0000
No. of Observations	10
Constant	-16.0050
X Coefficient	0.010119
Std Err of Coefft:	0.000008

APPENDIX 17. : CALIBRATION OF THE OVERALL LOAD MEASURING SYSTEM

The calibration of the overall load measuring system was achieved as follows. Firstly the the strain gauge bridge was calibrated using the Instrom tester [append. 15]. The results related the value of microstrain on the bridge display to the applied load on the piston.

The following relationship was derived from a regression analysis of the results of the calibration:

$$P = MS + C \qquad \text{.....eqn. A17.1}$$

where:

P = Piston Load [kN]
S = Gauge Strain [microStrain]
M = 0.025849 [x coefft. from append. 15]
C = -0.50254 [const from append. 15]

Secondly the strain gauge bridge value was varied while reading the A to D value on the computer. This amounted to calibrating the strain gauge amplifier [append. 16]. Eqn. A17.1 was then used to convert the strain gauge bridge value to piston load. The piston load was then multiplied by 2 to obtain the bearing load. Thus the bearing load could be related directly to the A to D value on the computer.

The following relationship could therefore be derived from a regression analysis of the results of these two parameters.

$$B = MA + C \qquad \text{.....eqn. A17.2}$$

where:

B = Bearing Load [kN]
A = A to D value (dimensionless integer)
M = 0.010119 [x coefft from append. 16]
C = -16.0050 [const from append. 16]

Equation 10.2 is equivalent to eqn. A17.2

APPENDIX 18. : RESOLUTION OF THE LOAD MEASURING SYSTEM

The resolution of the load measuring system was calculated as follows.

From appendix 16 :

A to D value of 1486 => -0.96841 kN
And A to D value of 3783 => 22.27457 kN

The A to D range will be (3783 - 1486) = 2297

i.e. the dynamic range is 20 Log (2297) = 67.22 dB

The equivalent load range is (22.2746 + 0.9684) = 23.243 kN

Therefore the resolution will be :

23243 N / 2297 = 10.12 N per A to D integer.

APPENDIX 19. : SIGNAL TO NOISE RATIO OF THE LOAD MEASURING SYSTEM

The signal to noise ratio of the load measuring system was calculated before and after the implementation of noise reduction techniques. The method was to measure the background noise (Volts) from the strain gauge bridge without any load applied. Then full load was applied to the bridge and the output voltage measured again. The results were as follows.

Background Noise : 27.5 mV
Full Load : 240.0 mV

SN ratio = 20 Log (240/27.5)
= 18.81 dB

After capacitive smoothing and averaging techniques :

Background Noise : 8 mV
Full Load : 240 mV

SN ratio = 20 Log (240/8)
= 29.54 dB

APPENDIX 20. : THERMOCOUPLE VOLTAGE-TEMPERATURE CONVERSION TABLE

CHROMEL vs. ALUMEL THERMOCOUPLE

Degrees Centigrade — Reference Junction 0° C.

°C	0	1	2	3	4	5	6	7	8	9
	Millivolts									
0	0.00	0.04	0.08	0.12	0.16	0.20	0.24	0.28	0.32	0.36
10	0.40	0.44	0.48	0.52	0.56	0.60	0.64	0.68	0.72	0.76
20	0.80	0.84	0.88	0.92	0.96	1.00	1.04	1.08	1.12	1.16
30	1.20	1.24	1.28	1.32	1.36	1.40	1.44	1.49	1.53	1.57
40	1.61	1.65	1.69	1.73	1.77	1.81	1.85	1.90	1.94	1.98
50	2.02	2.06	2.10	2.14	2.18	2.23	2.27	2.31	2.35	2.39
60	2.43	2.47	2.51	2.56	2.60	2.64	2.68	2.72	2.76	2.80
70	2.85	2.89	2.93	2.97	3.01	3.05	3.10	3.14	3.18	3.22
80	3.26	3.30	3.35	3.39	3.43	3.47	3.51	3.56	3.60	3.64
90	3.68	3.72	3.76	3.81	3.85	3.89	3.93	3.97	4.01	4.06
100	4.10	4.14	4.18	4.22	4.26	4.31	4.35	4.39	4.43	4.47
110	4.51	4.55	4.60	4.64	4.68	4.72	4.76	4.80	4.84	4.88
120	4.92	4.96	5.01	5.05	5.09	5.13	5.17	5.21	5.25	5.29
130	5.33	5.37	5.41	5.45	5.49	5.53	5.57	5.61	5.65	5.69
140	5.73	5.77	5.81	5.85	5.89	5.93	5.97	6.01	6.05	6.09
150	6.13	6.17	6.21	6.25	6.29	6.33	6.37	6.41	6.45	6.49
160	6.53	6.57	6.61	6.65	6.69	6.73	6.77	6.81	6.85	6.89
170	6.93	6.97	7.01	7.05	7.09	7.13	7.17	7.21	7.25	7.29
180	7.33	7.37	7.41	7.45	7.49	7.53	7.57	7.61	7.65	7.69
190	7.73	7.77	7.81	7.85	7.89	7.93	7.97	8.01	8.05	8.09
200	8.13	8.17	8.21	8.25	8.29	8.33	8.37	8.41	8.46	8.50
210	8.54	8.58	8.62	8.66	8.70	8.74	8.78	8.82	8.86	8.90
220	8.94	8.99	9.02	9.06	9.10	9.14	9.18	9.22	9.26	9.30
230	9.34	9.38	9.42	9.46	9.50	9.54	9.59	9.63	9.67	9.71
240	9.75	9.79	9.83	9.87	9.91	9.95	9.99	10.03	10.07	10.11
250	10.16	10.20	10.24	10.28	10.32	10.36	10.40	10.44	10.48	10.52
260	10.57	10.61	10.65	10.69	10.73	10.77	10.81	10.85	10.89	10.93
270	10.98	11.02	11.06	11.10	11.14	11.18	11.22	11.26	11.30	11.34
280	11.39	11.43	11.47	11.51	11.55	11.59	11.63	11.67	11.72	11.76
290	11.80	11.84	11.88	11.92	11.96	12.01	12.05	12.09	12.13	12.17

APPENDIX 21. : CALIBRATION OF THE TEMPERATURE MEASURING SYSTEM

A linear regression analysis was performed on the data from the thermocouple calibration table in appendix 20. This was done for the temperature range 0°C to 200°C. Figure A21.1 shows the results of this analysis graphically. Also shown is the % of full scale error for each value of the table.

Because the values of the table were rounded it can be seen that there was a stepwise error compared to the regression line. This error was less than about 0.5% in all cases.

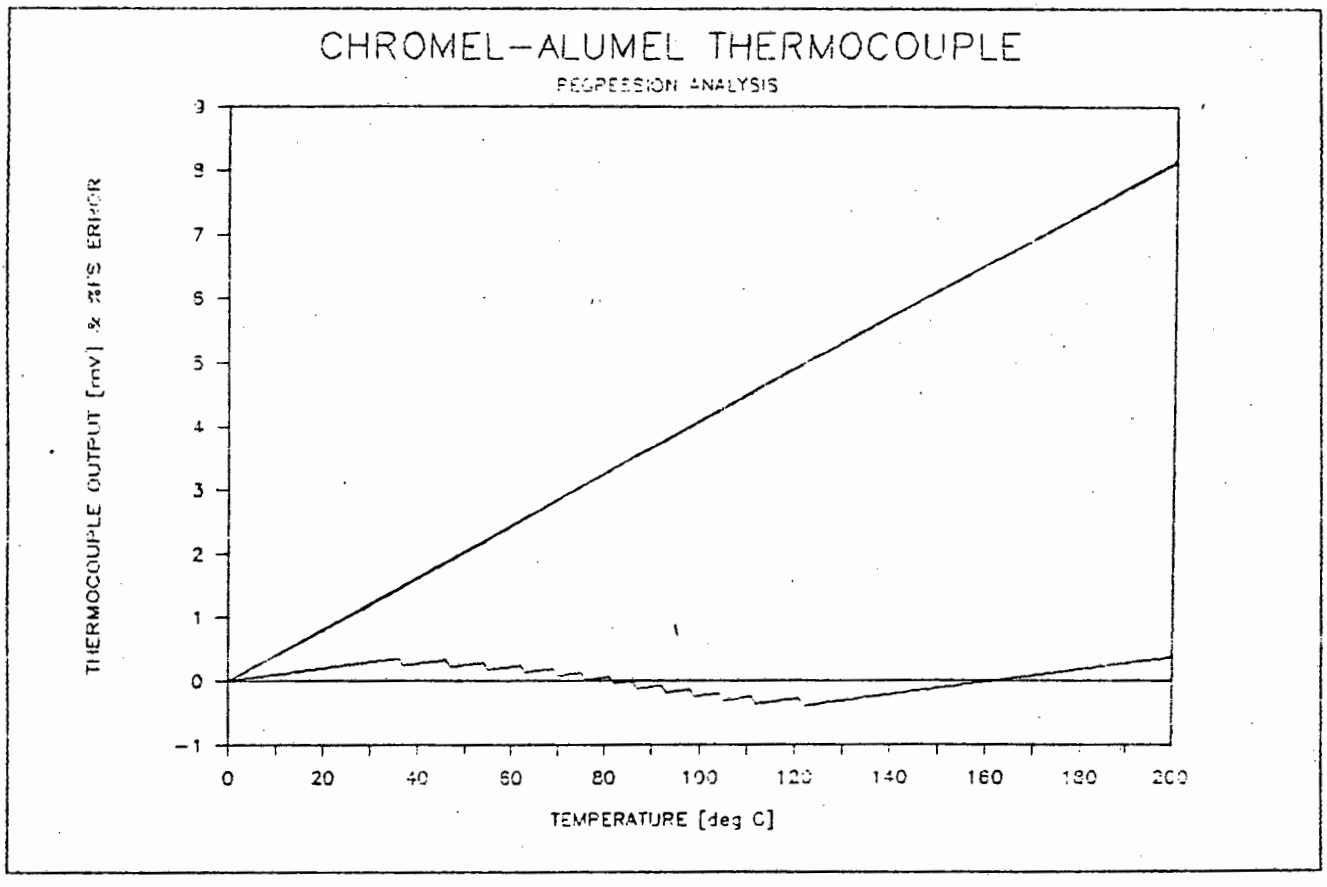

FIGURE A21.1: THERMOCOUPLE CALIBRATION REGRESSION LINE

The integrated circuit thermocouple amplifier was designed to have an output of 10mV/°C [35]. However it was decided to calibrate the amplifier using ice water and boiling water to establish the output for 0°C and 100°C (sea level) respectively.

It was found that the output differed from the expected values as follows:

TABLE A21.1 : THERMOCOUPLE AMPLIFIER CALIBRATION TEST

EXPECTED VALUES	0.0 mV 0.0 °C	1000.0 mV 100.0 °C
MEASURED VALUES	25.6 mV 2.56 °C	1038.8 mV 103.88 °C
ERROR	2.56 °C	3.88 °C

Thus it was found that the system was reading about 3°C too high. This was compensated for by reducing the final value of displayed temperature by 3°C.

The resolution of the temperature was calculated as follows:

For temperature range of 0°C to 100°C

A to D values range from 2048 to 2876
=> 2876 - 2048 values
= 828 values

i.e. 100°C => 828 values

Therefore 1 value = 100/828
= 0.12°C

APPENDIX 22. : CALCULATION OF BEARING ELEMENT PASSAGE FREQUENCIES

The following equations were used to calculate the expected bearing element passage frequencies :

Using :

R = Shaft Running Frequency
BPFO = Ball Pass Frequency on the outer race
BPFI = Ball Pass Frequency on the inner race
FTF = Fundamental Train Frequency (1 cycle of cage)
BSF = Ball Spin Frequency

n = N° of balls
= 15 balls * 2 races
= 30 balls

Bd = Ball diameter
= 8.735 mm

Pd = Pitch diameter
= 53.26 mm

B = Bd/Pd
= 8.735/53.26
= 0.164

Φ = Contact angle [48]
= 9°

C = Cos Φ
= Cos 9°
= 0.988

In the following equations [48] :

$$BPFO = n/2 * (1 - B*C) * R$$
$$BPFI = n/2 * (1 + B*C) * R$$
$$FTF = 1/2 * (1 - B*C) * R$$
$$BSF = 1/(2*B) * (1 - B^2*C^2) * R$$

Gives :

BPFO = 6.285 * R
BPFI = 8.715 * R
FTF = 0.419 * R
BSF = 2.969 * R

APPENDIX 23. : DERIVATION OF THE EQUATION TO CALCULATE SHAFT SPEED AND RESOLUTION

The speed of the shaft was measured using a slotted disk and shaft encoder as described in section 10.7. The equation to calculate speed from this system was derived by measuring the frequency of the triangular wave it provided. This was done as follows.

The triangular wave was sampled with a frequency F_s high enough to ensure that on average 4.5 samples were collected per cycle of the wave at the fastest speed of the shaft. An algorithm was then developed to count the number of times (C) that the wave passed over the mean value of the wave in a negative or positive direction.

Then by knowing the number of wave cycles in one revolution of the shaft (C_R) it was possible to derive the number of revolutions.

i.e. Revolutions = Total Counts / Counts per Rev.

$$= C/C_R. \qquad \text{.....eqn. A23.1}$$

Similarly the time to collect the wave would be given by the sampling frequency F_s (Hz) and the total number of samples collected S_T.

i.e. Time [secs] = Total Samples / Samples per sec.

$$= S_T/F_s \qquad \text{.....eqn. A23.2}$$

Combining these two equations would give the shaft speed.

i.e. RPS [revs/sec] = Revolutions / Time

$$= (C/C_R) \;/\; (S_T/F_s) \qquad \text{.....eqn. A23.3}$$

To obtain speed in RPM, eqn. A23.3 was multiplied by 60.

$$RPM = \frac{C * F_s * 60}{S_T * C_R} \qquad \text{.....10.3}$$

The resolution of the speed measuring system was derived from equation A23.3 as follows.

A better resolution would be obtained by using more slots in the encoder disk and the highest sampling speed. However the maximum number of slots (i.e. C_R) was limited by practical design constraints and the response of the shaft encoder circuit to 180 slots [append. 25].

Further the number of counts per sampled wave (C) would depend on the speed of the shaft.

Thus equation 10.3 could be given by :

$$RPM = C * K / S_T \qquad \text{.....eqn. A23.4}$$

where K was a constant representing $F_s * 60 / C_R$.

Therefore the speed resolution given by delta RPM would be :

$$\text{delta RPM} = K / S_T \qquad \text{.....eqn. 10.4}$$

i.e. the more samples collected, the smaller would be delta RPM and therefore the better the resolution of speed measurement.

APPENDIX 24. : USE OF THE SPEED EQUATION TO OBTAIN A RESOLUTION OF 1 RPM

Equation 10.4 was use to calculate the number of samples required to obtain a speed resolution of 1 RPM or better. This equation gave the speed resolution as follows:

$$\text{delta RPM} = K / S_T \qquad \text{.....eqn. 10.4}$$

where K was a constant representing $F_s * 60 / C_R$ [append. 23].

Thus the more samples collected, the smaller would be delta RPM and therefore the better the resolution of speed measurement.

If the speed resolution was required to be 1 RPM then the required number of samples could be calculated from eqn. 10.3 as follows.

$$\text{i.e. } S_T = K = F_s * 60 / C_R \qquad \text{.....eqn. A24.1}$$

The sampling frequency was set at the maximum of 27586 Hz [append. 8] and the counts per revolution as 2 * 180 = 360. [append. 25]

$$\text{Therefore } S_T = 27586 * 60 / 360$$

$$= 4598 \text{ samples}$$

This was increased to 6000 samples to ensure a resolution of better than 1 RPM.

APPENDIX 25. : CALCULATION OF THE MAXIMUM NUMBER OF SLOTS USABLE ON THE SHAFT ENCODER DISK

The maximum number of slots on the encoder disk was limited by the maximum sampling frequency of the A to D card, the number of samples per slot and the maximum speed of the shaft.

Firstly the output voltage of the shaft encoder was a triangular wave formed by the encoder disk slots interrupting the encoder beam. This wave had a DC offset value which was equivalent to the mean of the wave.

The number of samples per slot was set so that at least two samples would fall on either side of the mean of the wave per cycle. In other words at least 4 samples were required per cycle of the wave. This was set to 4.5 to allow a margin of safety.

Then the maximum speed of the shaft was set at approximately 33 Hz (1980 RPM) and the maximum sampling speed was 27586 Hz.

Therefore the Maximum number of samples per revolution at full speed would be:

$$\text{Max Samples per rev.} = 27586 \text{ Hz} / 33 \text{ Hz}$$
$$= 835.94 \text{ samples/rev}$$

And for 4.5 samples per slot, the maximum number of slots allowed on the encoder disk would be :

$$\text{Max Slots per rev.} = 835.94 / 4.5$$
$$= 185.8 \text{ slots per rev.}$$

The closest multiple of 60 to this value was 180 slots per rev. Therefore this was chosen for the final design of the encoder disk.

APPENDIX 26. : DESIGN OF THE SHAFT ENCODER DISK

The maximum allowed number of slots on the encoder disk was calculated as 180 [append 25]. It was decided that for the slots to have sufficient strength and to enable more accurate machining the slots should be at least 2mm wide. Thus the divider between the slots would also have to be 2mm wide to produce an even waveform at the encoder output. In other words 4 mm were required for each slot.

Thus the diameter of the disk could be calculated as follows.

disk circumference = 4mm * 180 slots

= 720 mm

Thus disk diameter = 720 / π

= 229.18 mm (nominal)

And for a slot depth of 3 mm the outer disk diameter would be approximately = 230.5 mm.

These values were used to make the encoder disk.

APPENDIX 27. : COMPARISON OF SPEED MEASUREMENTS INDICATED BY VARIOUS INSTRUMENTS

The shaft encoder system was compared with three other speed measuring systems. These were a frequency counter and an oscilloscope connected to the output of the encoder and a 'Jaquets' mechanical speed indicator applied to the end of the rotating shaft.

The oscilloscope was used to calculate the speed by measuring the wavelength and inverting. The value on the frequency counter was divided by 3 to get RPM. The results were as follows.

METHOD	MAX RPM	MEDIUM RPM	MIN RPM
JAQUETS	1939	983	114
OSCOPE	1933.45	982.46	115.26
COUNTER	1933.31	980.67	114.68
ADBOARD	1934	982	114

Thus the system correlated well with other methods and was accurate enough for the requirements of diagnosis.

APPENDIX 28. : CALIBRATION OF THE KROHN-HITE FILTER

The Krohn-hite filter was used as the bandpass filter for vibration measurements. It was set for a high pass frequency of 18Hz and a low pass frequency of 8kHz. Calibration was performed by applying a sine wave of varying frequency to the input. Then the input and output volts were measured using RMS voltmeters.

Thus an amplitude response was obtained for various frequencies. In particular the response for a frequency range of 6Hz to 20kHz was measured. More detailed results were collected around the low and high cutoff frequencies.

The amplitude response (A) was converted into decibels by the following equation.

$$A = 20 * \text{Log} (V_{OUT}/V_{IN})$$

The results are presented in table A28.1. and displayed in figure 10.12.

TABLE A28.1 : CALIBRATION RESULTS OF THE KROHN-HITE FILTER

Freq[Hz]	V in	V out	Atten[dB]
6	1.770	0.072	-27.813
8	1.770	0.132	-22.548
10	1.785	0.192	-19.367
12	1.785	0.231	-17.761
14	1.790	0.281	-16.083
16	1.792	0.349	-14.210
18	1.793	0.423	-12.545
20	1.794	0.470	-11.634
22	1.795	0.505	-11.015
24	1.795	0.525	-10.678
26	1.795	0.532	-10.563
28	1.795	0.536	-10.498
30	1.795	0.537	-10.482
35	1.796	0.537	-10.487
40	1.796	0.538	-10.470
45	1.797	0.538	-10.475
50	1.797	0.538	-10.475
75	1.797	0.543	-10.395
100	1.797	0.554	-10.221
200	1.797	0.632	-9.077
400	1.797	0.808	-6.943
600	1.797	0.927	-5.749
800	1.797	0.999	-5.100
1000	1.797	1.045	-4.709
1500	1.797	1.103	-4.239
2000	1.797	1.128	-4.045
2500	1.798	1.140	-3.958
3000	1.798	1.146	-3.912
3500	1.798	1.148	-3.897
4000	1.798	1.147	-3.905
4500	1.798	1.141	-3.950
5000	1.798	1.129	-4.042
5500	1.799	1.105	-4.233
6000	1.799	1.065	-4.554
6500	1.799	1.001	-5.092
7000	1.799	0.919	-5.834
7500	1.799	0.819	-6.835
8000	1.799	0.708	-8.100
8500	1.799	0.603	-9.494
9000	1.799	0.505	-11.035
9500	1.799	0.424	-12.553
10000	1.799	0.353	-14.145
11000	1.799	0.248	-17.212
12000	1.799	0.177	-20.141
13000	1.799	0.129	-22.889
14000	1.799	0.095	-25.546
15000	1.799	0.073	-27.834
16000	1.799	0.056	-30.137
17000	1.799	0.043	-32.431
18000	1.798	0.034	-34.466
19000	1.798	0.027	-36.469
20000	1.798	0.021	-38.651

APPENDIX 29. : DESIGN OF THE BUTTERWORTH 5-POLE LOW PASS FILTER

The design of the Butterworth filter was obtained from ref 41. Only passive components were used. The circuit diagram for a 5 Pole low pass filter was as shown in figure A29.1

According to ref 41 the values of the components for a 10 kHz cutoff were as follows :

$R_1 = R_2 = 1$ kΩ

$L_1 = L_2 = 25.75$ mH

$C_1 = C_3 = 9.83$ nF

$C_2 = 0.0318$ microFarad

However these values were ideal and had to be approximated by the use of available components [append. 33].

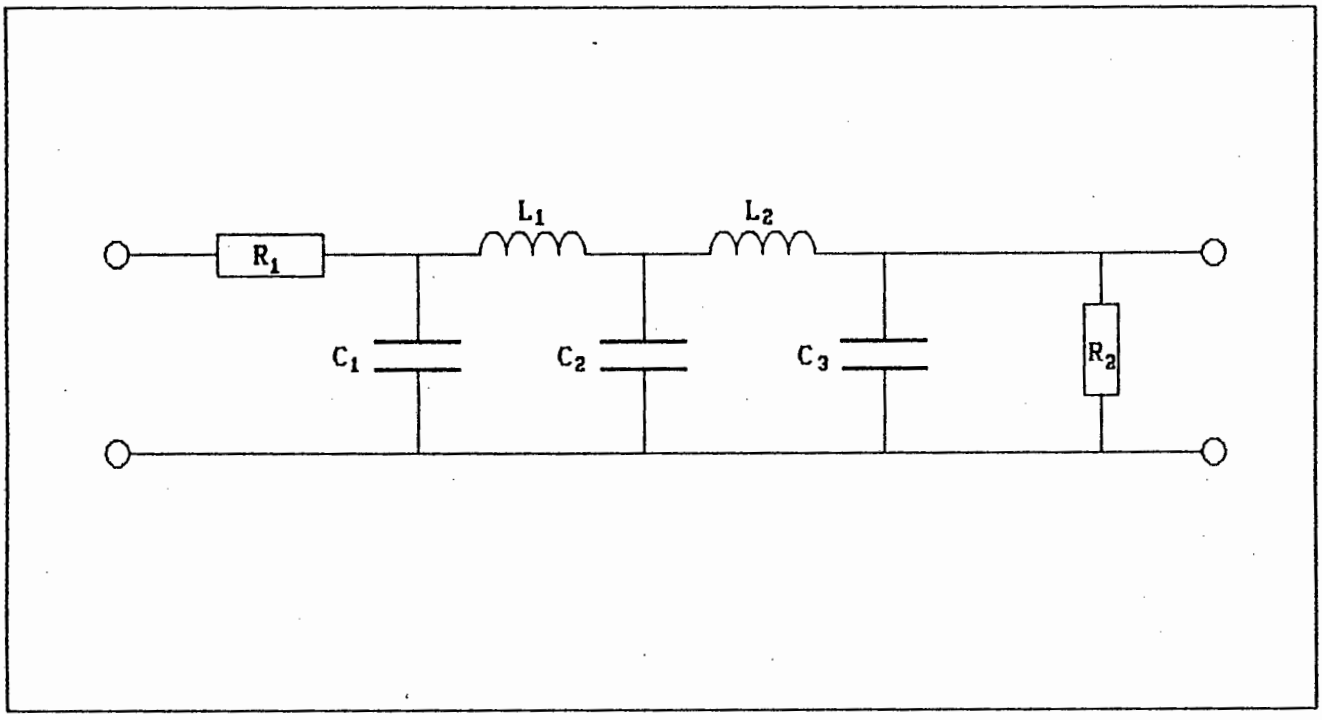

FIGURE A29.1: CIRCUIT DIAGRAM FOR THE BUTTERWORTH 5 POLE LOW PASS FILTER.

APPENDIX 30. : RESULTS OF THE BUTTERWORTH FILTER CALIBRATION TEST

The Butterworth filter was calibrated in the same way as the Krohn-Hite filter [append. 28]. The results are shown in table A30.1 and given graphically in figure 10.13. Note that the input voltage was kept constant at 1.476 volts.

FREQ [Hz]	VOLTS OUT	ATTEN [dB]
1000	1.472	-0.024
2000	1.452	-0.142
3000	1.412	-0.385
4000	1.361	-0.705
5000	1.289	-1.177
6000	1.202	-1.784
7000	1.102	-2.538
8000	0.997	-3.408
9000	0.892	-4.374
10000	0.788	-5.451
11000	0.691	-6.592
12000	0.603	-7.775
13000	0.526	-8.962
14000	0.458	-10.164
15000	0.398	-11.384
20000	0.198	-17.448
30000	0.038	-31.786

APPENDIX 31. : THEORETICAL RESPONSE CURVE OF A BUTTERWORTH 5-POLE LOW PASS FILTER

The theoretical response of various Butterworth filters are shown in figure A31.1. Comparing the response for the 5 pole filter with figure 10.3 shows that the built filter had a poor response in comparison with the ideal.

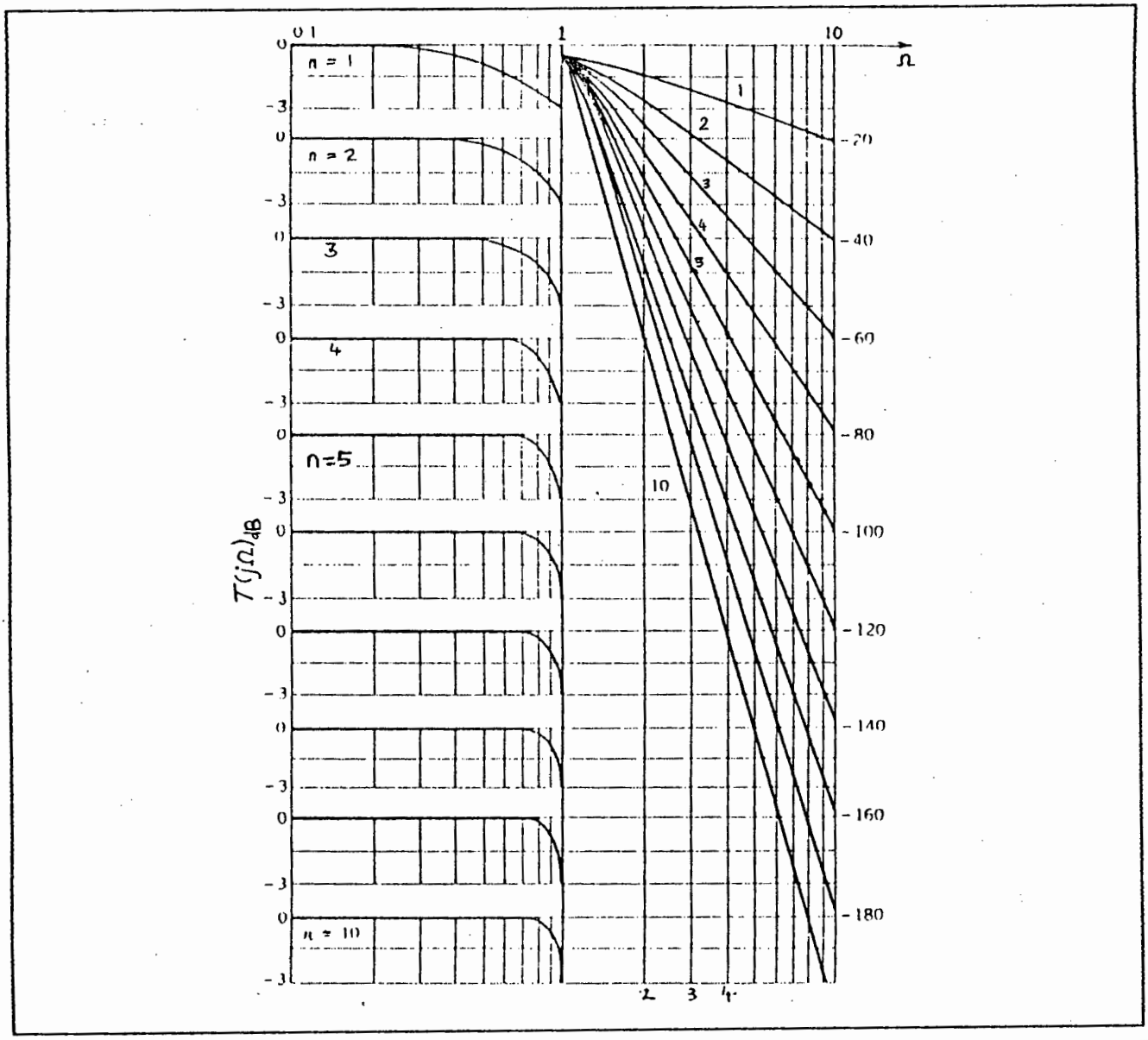

FIGURE A31.1: THEORETICAL RESPONSE CURVES FOR BUTTERWORTH FILTERS

APPENDIX 32. : CALCULATION OF THE PASSBAND FLATNESS AND ROLL-OFF FOR THE BUTTERWORTH FILTER USED

The initial rolloff of the Butterworth filter was calculated from the attenuation at 10kHz (cutoff freq.) and 20kHz to get the rolloff per octave. These values were taken from table A30.1 [append. 30].

Initial Rolloff :

10 kHz	:	-5.451 dB
20 kHz	:	-17.448 dB
		12.00 dB / Octave

Passband flatness :

Assuming that the attenuation tended to zero at DC, which was clearly indicated by figure 10.13, then the passband flatness would be given by the value at the cutoff frequency.

i.e. Flatness = 5.45 dB [append. 30]

APPENDIX 33. : COMPARISON OF BUTTERWORTH FILTER COMPONENTS WITH IDEAL VALUES

The design of the Butterworth filter design provided ideal values for the components. However these had to be rounded to available components. This had a negative effect on the performance of the filter [append. 32].

The comparison of values was as follows [see fig A29.1]. Note that the two inductors used had approximate values as shown.

TABLE A33.1 : COMPONENT VALUES FOR BUTTERWORTH FILTER

COMPONENT	IDEAL VALUE	ACTUAL VALUE
R_1 & R_2	1 kΩ	1 kΩ
L_1 & L_2	25.75 mH	27 & 28 mH
C_1 & C_3	9.83 nF	10 nF
C_2	31.8 pF	33 pF

APPENDIX 34. : DESIGN OF THE CHEBYCHEV 8-POLE LOW PASS FILTER

A Chebychev 8 pole low pass filter was designed with the intention of using it for low frequency vibration analysis. The method used was after Horowitz and Hill [40] for a cutoff of 2 kHz. However the filter was not built due to its complexity and number of components. This appendix briefly describes the design.

The filter consisted of a cascade of four 2-pole voltage controlled voltage source (VCVS) filter sections. One section is shown in figure A34.1. Similar to the Butterworth filter, each section had a prescribed gain K, resistors $R_1=R_2=R$ and capacitors $C_1=C_2=C$. However unlike the Butterworth, the RC products for each section were different and had to be scaled by a factor f_N given in the relevant tables.

The method was to select resistors in the typical range of 10 kΩ to 100 kΩ and avoiding small values. The values of the capacitors were then calculated by the following equation :

$$RC = 1/(2\pi f_N f_c) \qquad \text{.....eqn. A34.1}$$

where f_c = cutoff frequency (2 kHz)
f_N = normalizing factor (tabulated)

The capacitance value was then rounded off and a new resistance value calculated. This was done because it would be easier to select components. The process was repeated for each section.

The required gain of each section (K) was also tabulated with the normalizing factor for each section. For the 8 pole filter the values for the four 2-pole sections were as follows (see fig. A34.1):

TABLE A34.1 : SELECTION OF CHEBYCHEV FILTER COMPONENTS

STAGE	R [Ω]	K	f_N	C [pF]	NEW C	NEW R	(K-1)R
1	100000	1.522	0.297	535.88	540	496181	259006
2	100000	2.379	0.599	265.70	270	492039	678522
3	100000	2.711	0.861	184.85	180	513469	878546
4	100000	2.913	1.006	158.21	160	494393	945774

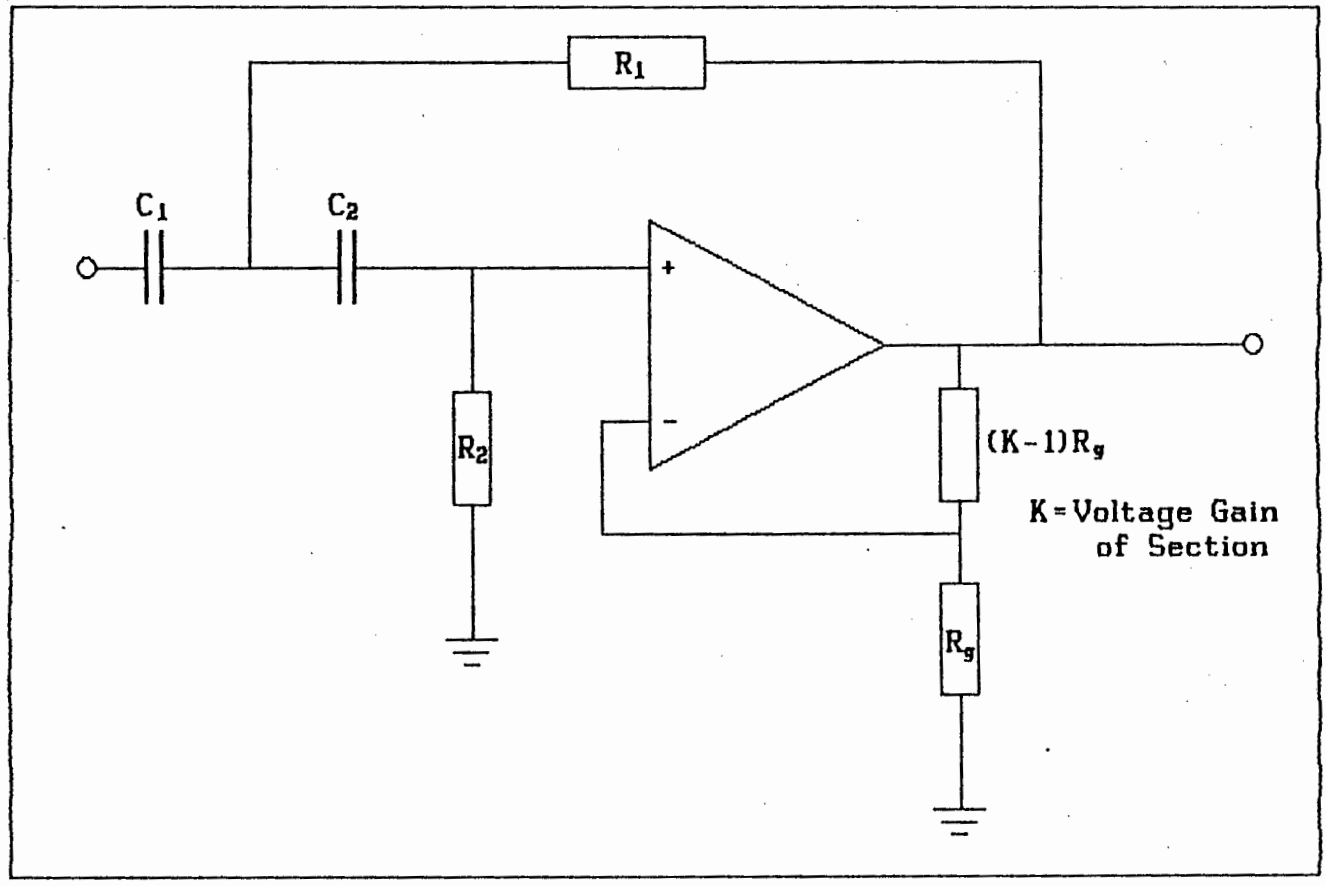

FIGURE A34.1: TYPICAL VCVS SECTION OF THE CHEBYCHEV FILTER

APPENDIX 35. : THE LUBRICANT FILM THICKNESS EQUATION

It was initially attempted to calculate the lubricant film thickness using an empirical equation [15]. It was desired to use this equation to determine if the combination of operating conditions namely load, speed, temperature and lubricant viscosity were satisfactory. In addition the lubricant film thickness would give a direct indication of the protection of the bearing surfaces by the lubricant.

The equation was as follows :

$$h_o = \frac{0.145\ a^{0.6}\ (nv)^{0.7}}{(2(1/r+1/R))^{0.43}\ (Q/l_{EFF})^{0.13}}$$

Where :

h_o = Lubricant film thickness [microns]
a = pressure viscosity coefft. [mm²/N] (usually 0.01 - 0.02)
n = operating viscosity [MPa S]
v = Velocity of ball
r = Radius of ball
R = Internal radius of race.
Q = Load on individual ball [N]
l_{EFF} = Effective 'roller' length [mm]

One of the hindrances to using the equation was that the effective roller length was not defined for a ball.

It was interesting to note the similarities between this equation and the life rating equation [append. 40]

APPENDIX 36. : LOG REGRESSION ANALYSIS FOR THE LUBRICANT VISCOSITY EQUATION

The operating viscosity of the lubricant was required to be known for all temperatures which would be encountered on the test bearing. In other words the operating viscosity had to be calculated from the temperature measured on the bearing. Thus it was required to formulate a simple relationship between the two parameters.

This formulation was achieved by performing a regression analysis on data taken from a lubricant viscosity chart for the multigrade oil used. The values of viscosity at various temperatures are shown in table A36.1.

TABLE A36.1 : LUBRICANT VISCOSITY FOR VARIOUS TEMPERATURES

TEMP [°C]	VISCOSITY [mm²/s]
10	900
20	450
30	250
40	150
50	95
60	65
70	45
80	33
90	25
100	19
110	15
120	12
130	9.5
140	8.25

Because these values were related logarithmically, the linear regression analysis was performed on the log of the values instead of the values themselves. i.e. a normal linear regression was used on the log values to obtain the x coefficient (M) and the constant (C) such that :

$$(\text{Log } Y) = M\,(\text{Log } X) + C \qquad \text{.....eqn. A36.1}$$

Where Y is the dependent variable and X the independent variable.

Thus $Y = 10^{M \log X + C}$ .....eqn. A36.2

$= C'X^M$ .....eqn. A36.3

Where $C' = 10^C$

The regression analysis as discussed in this section was used to derive the relationships between all parameters related logarithmically.

In particular the results of the regression analysis on the data in table A36.1 provided the following values :

$C' = 920668$

$M = -2.345$

Thus $V = C'T^M$ .....eqn. A36.4

For example if Temp = 100°C

Then Operating Viscosity = $920668 * 100^{-2.345}$

$= 18.8 \text{ mm}^2/\text{s}$

APPENDIX 37. : GENERALIZED CHART FOR RELATING SHAFT SPEED TO REQUIRED LUBRICANT VISCOSITY IN ROLLING ELEMENT BEARINGS

The chart in figure A37.1 [13] was used to find the required lubricant viscosity for various bearing speeds. To use the the chart the mean diameter of the bearing had to be known. This was calculated for the test bearings as follows.

$$d_m = (d + D) / 2 \qquad \text{.....eqn. A37.1}$$

Where d and D are the inner and outer diameters of the bearing respectively. The values for the test bearings were as follows.

d = 35 mm
D = 72 mm

$d_m = (35 + 72) / 2$
$= 53.5$ mm

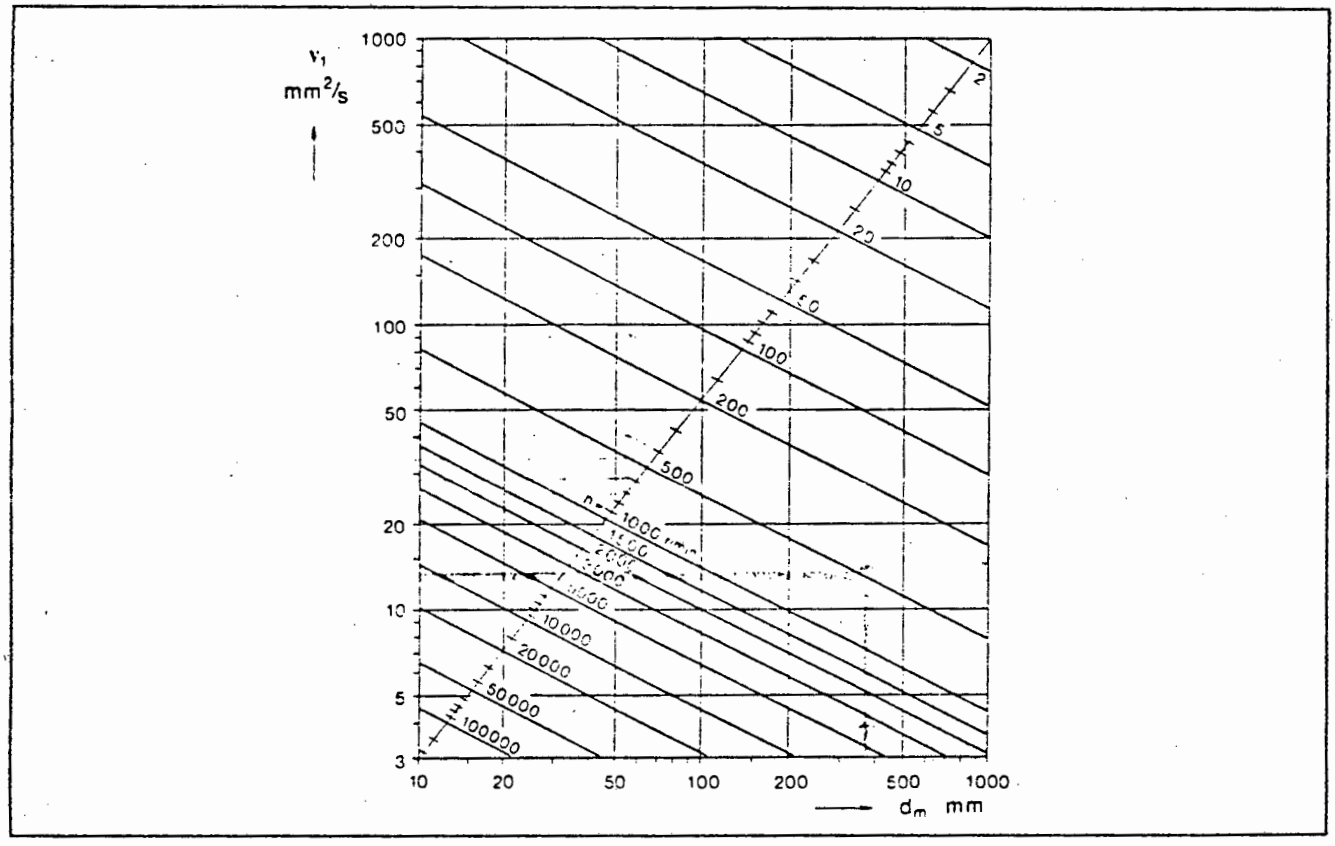

FIGURE A37.1: GENERALIZED SPEED VISCOSITY CHART FOR BEARINGS

APPENDIX 38. : DERIVATION OF THE RELATIONSHIP BETWEEN BEARING SPEED AND REQUIRED LUBRICANT VISCOSITY.

The relationship between the required lubricant viscosity and bearing speed was formulated by performing a log regression analysis [method in append.36] on data taken from the chart in appendix 38. This data is shown in table A38.1

TABLE A38.1 : REQUIRED LUBRICANT VISCOSITY FOR VARIOUS SPEEDS

SPEED [RPM]	VISCOSITY [mm^2/s]
100	130
150	90
200	73
300	51
400	40
500	33
600	28
700	25
800	23
900	20.5
1000	18.0
1500	15.5
2000	13.5

The results of the log regression analysis were as follows [append. 36] :

in equation A36.3 : $Y = C'X^M$

$C' = 5164.168$

$M = -0.80706$

i.e. $V_i = C'S^M$ .....eqn. A38.1

For example if speed = 1000 RPM

Then required viscosity = $5164.168 * 1000^{-0.8706}$

= 19.58 mm^2/s

APPENDIX 39. : DERIVATION OF THE BEARING LIFE ADJUSTMENT FACTOR 'a_{23}' FROM THE LUBRICANT VISCOSITY INDEX 'k'

The lubricant viscosity index 'k' was required to obtain the life rating factor 'a_{23}. The two parameters were related by the chart shown in figure A39.1.

Firstly 'k' was found from the values of operating viscosity (v) and required viscosity (v_1) derived in appendices 36 and 38. This was done as follows.

$$k = v/v_1 \qquad \text{.....eqn. A39.1}$$

This value of 'k' was then used to find a_{23} from the chart below.

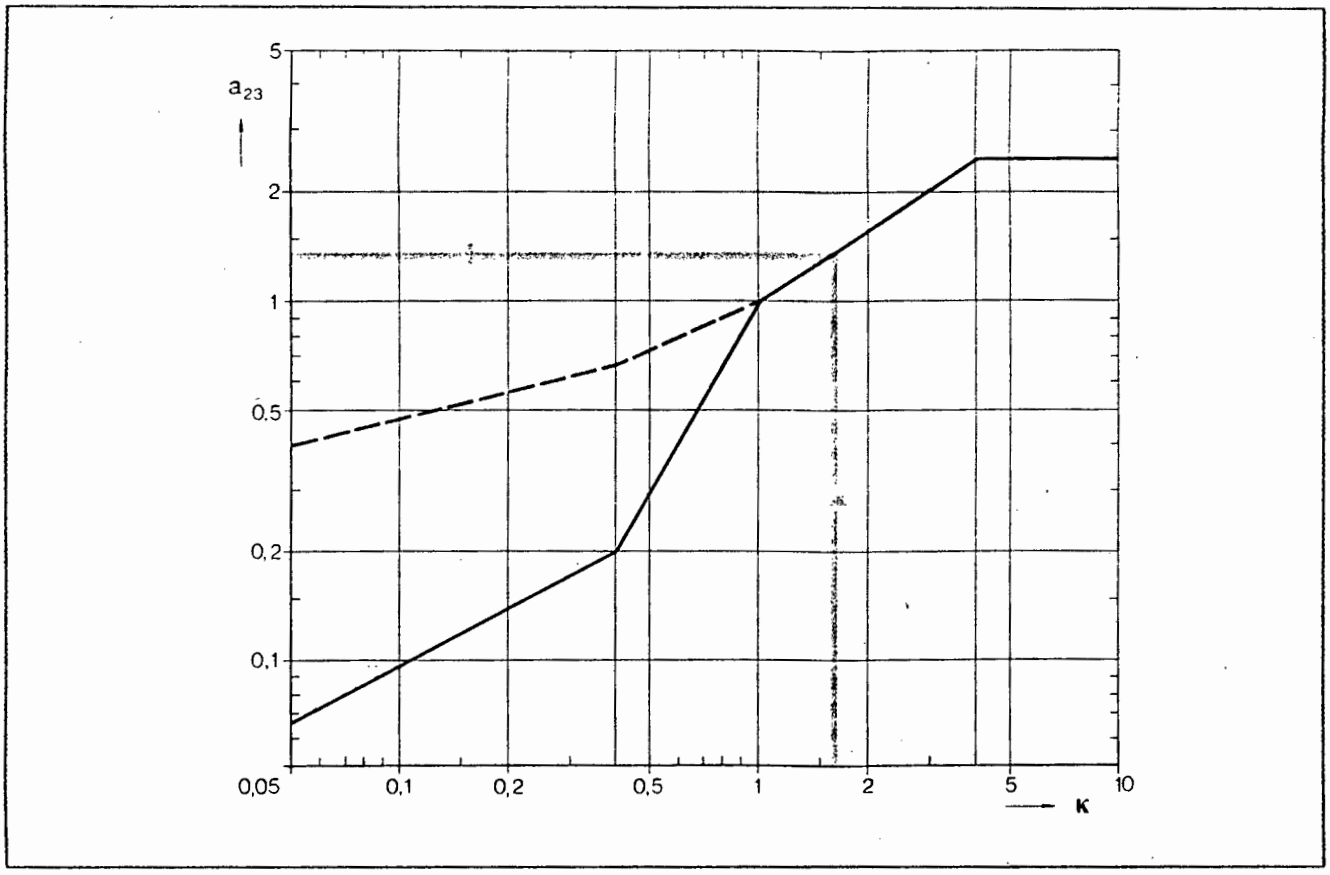

FIGURE A39.1: CHART RELATING LUBRICANT VISCOSITY INDEX 'k' TO THE LIFE ADJUSTMENT FACTOR 'a_{23}'.

As can be seen from the chart there were four distinct zones of 'k'. Thus four different equations had to be derived from the chart. These were obtained from a log regression analysis of data derived from the chart according to the method of appendix 36. The results for the four zones were as follows :

In equation A36.3 : $Y = C'X^M$

i.e. $a_{23} = C' * k^M$ .....eqn. A39.2

Where C' and M were selected according to the K Zone being operated in. The values obtained from the regression analysis are shown in table A39.1

TABLE A39.1 CONVERSION FACTORS FOR 'k' TO 'a_{23}

K ZONE	EXPONENT M	CONSTANT C'
1	0.534045	0.32933883
2	1.756313	1.00008692
3	0.662682	1.00000697
4	0.000000	2.50600000

APPENDIX 40. : DERIVATION OF THE OVERALL BEARING LIFE EQUATION

The equations derived from appendices 36 to 39 were combined to form one overall life equation as follows.

Firstly the general life rating equation for a reliability of 90% was given in [13] as :

$$L_{10h} = \frac{1\ 000\ 000}{60\ N}\ a_{23}\ (P_R/P)^3 \qquad \text{.....eqn. A40.1}$$

Where : N = Shaft speed [rpm]
a_{23} = Combined adjustment factor for operating conditions
P_R = Basic dynamic load rating [N]
P = Equivalent dynamic bearing load [N]

this equation was simplified to :

$$L_{10h} = C\ \frac{a_{23}}{N\ P^3} \qquad \text{.....eqn. A40.2}$$

Where C was a combined constant of all the constant values in equation A40.1

The value of a_{23} was then derived according to equation A39.2 and the k zone being operated in. Thus replacing a_{23} with equation A39.2 gave :

$$L_{10h} = C\ \frac{k^{Mka}}{N\ P^3} \qquad \text{.....eqn. A40.3}$$

Where : k = Lubricant viscosity coefficient
Mka = Exponent for conversion from k to a_{23} [table A39.1]

and the combined constant C was expanded to include the constant from equation A39.2

Then k was derived from the required viscosity and operating viscosity as in equation A39.2. Thus substituting eqn. A39.2 into eqn. A40.3 gave :

$$L_{ioh} = C \frac{(V/V_i)^{Mka}}{N\ P^3} \qquad \text{.....eqn. A40.4}$$

This was rearranged to :

$$L_{ioh} = C \frac{V^{Mka}}{N\ P^3\ V_i^{Mka}} \qquad \text{.....eqn. A40.5}$$

The two viscosity parameters were then replaced by their respective equations derived in appendices 36 and 38. Thus substituting equations A36.4 and A38.1 into A40.5 gave :

$$L_{ioh} = C \frac{T^{Mtv\cdot Mka}}{P^3\ N^{Msv\cdot Mka+1}} \qquad \text{.....eqn. A40.6}$$

Where : Mtv = conversion exponent for temp. to operating visc.
Msv = conversion exponent for speed to reqd. visc.
(values in appendices 36 & 38)

and the constant C was expanded to include the constants from equations A36.4 and A38.1

This equation was then simplified by combining the exponents to form a new set as follows :

$$L_{ioh} = C \cdot T^{Mt} \cdot N^{Mn} \cdot P^{Mp} \qquad \text{.....eqn. A40.6}$$

Where Mt = Mtv·Mka
Mn = -(Msv·Mka + 1)
Mp = -3

and the combined constant C was given by :

$$C = Cka \cdot Ctv^{Mka} \cdot Csv^{-Mka} \cdot P_R^3 \cdot 10^6/60 \qquad \text{.....eqn. A40.7}$$

Where Cka = Conversion const. for k to a_{23}
Ctv = Conversion const. for Temp. to Operating viscosity
Csv = Conversion const. for Speed to Reqd. Viscosity

Equation A40.6 was the same as equation 12.6. An example of the use of equation A40.6,7 is shown in appendix 42 and the values of the exponents and constants are given in table 12.2.

APPENDIX 41. : THE EFFECT OF CONVERTING RATED LIFE FROM CYCLES TO HOURS ON THE SPEED EXPONENT

An apparent paradox was observed on the change of rated bearing life with varying operating conditions. This was the effect of increasing bearing life by increasing the running speed under particular operating conditions. The effect was observed when considering the rated life in terms of hours instead of running cycles.

In particular the effect was observed when considering the exponent M_w of running speed for K-Zone 2 in equation 12.6. Intuitively an increase in running speed should have had a decreasing effect on the rated bearing life. However the conversion from cycles to hours was achieved by multiplying by the ratio of $10^6/N$, where N was running speed [eqn. A40.6]. This amounted to increasing the exponent for speed by one.

Thus in the case of K-Zone 2 the exponent became positive [table 12.2] and therefore by increasing speed, the bearing rated life was increased.

APPENDIX 42. : RATED LIFE FOR THE TEST BEARINGS USED

This appendix illustrates the calculation of the rated life of the test bearings destroyed on the test rig. It also illustrates how equation 12.6 was used to calculate the bearing rated life directly from the measured operating conditions of Temperature, Load and Speed. This life rating was then used for diagnosis and prognosis in the expert system.

The first test bearing was run under the following operating conditions:

Load = 22 kN
Speed = 1933 rpm
Temp = 125 °C (approx)

Firstly it was necessary to calculate the K zone being operated in. This was done using equation A39.1 as follows :

$$K = v/v_t \qquad \text{.....eqn. A39.1}$$

$$= Ctv \cdot T^{Mtv} / Csv \cdot N^{Msv}$$

$$= \frac{920667.93 \cdot (125)^{-2.345}}{5164.168 \cdot (1933)^{-0.80707}}$$

$$= 0.97$$

Therefore the K zone was 2 and the exponents for equation 12.6 could therefore be selected from table 12.2. These were as follows :

$C = 602170.249 * 10^6$
$M_t = -4.11856$
$M_n = 0.41746$
$M_p = -3$

Thus the life rating could be calculated for the bearing using equation 12.6 as follows :

$$L_{10h} = C \cdot T^{Mt} \cdot N^{Mn} \cdot P^{Mp} \qquad \text{.....eqn. A40.6}$$

$$= 602170.249 \cdot (125)^{-4.11856} \cdot (1933)^{0.41746} \cdot (22)^{-3}$$

$$= 3.1 \text{ Hours}$$

APPENDIX 44. : DERIVATION AND APPLICATION OF EQUATION 13.6

To enable vibration levels to be diagnosed at all operating conditions a system was devised to adjust and normalize them. This enabled comparison with acceptable levels at the reference operating conditions. This was achieved as follows.

Firstly the adjustment factor for speed was applied to the measured vibration parameter. For illustration, the RMS level could be adjusted for speed as follows.

$$RMS_{ADJ} = RSF \cdot RMS \qquad \text{.....eqn. A44.1}$$

Where :

RMS_{ADJ} = Adjusted value of RMS
$RSF = RMS_{REF} / RMS_{MSR}$ [eqn. 13.1]

The RMS level could also be adjusted for load as follows :

$$RMS_{ADJ} = RLF \cdot RMS \qquad \text{.....eqn. A44.2}$$

Where :

$$RLF = (M * (22) + C) / (M * (Load) + C) \text{ [eqn. 13.4]}$$

Finally the RMS value could be normalized by the average value at normal operating conditions (RMS_{AV}), so that a dimensionless value (NRMS) of 1 would be acceptable.

i.e. $$NRMS = RMS / RMS_{AV} \qquad \text{.....eqn. A44.3}$$

Equations A44.1, 2 and 3 were combined into one equation as follows :

$$NRMS = RMS \cdot RSF \cdot RLF / RMS_{AV} \qquad \text{.....eqn. 13.6}$$

The following example serves to illustrate the use of equation 13.6.

Assume current operating conditions of Load = 10 kN, Speed = 400 rpm and measured RMS value = 0.045 G's. The reference conditions are known to be Load = 22 kN, Speed = 1800 rpm and RMS_{av} = 0.312 G's. Then the adjustment factors will be :

$$RSF = 1800^{n} / (Speed)^{n} \qquad \text{.....eqn. 13.2}$$
$$= 1800^{1.594} / 400^{1.594}$$
$$= 10.996$$

$$RLF = (M \cdot (22) + C) / (M \cdot (Load) + C) \qquad \text{.....eqn. 13.4}$$
$$= (-6.11 \cdot 10^{-3} \cdot 22 + 0.567) / (-6.11 \cdot 10^{-3} \cdot 10 + 0.567)$$
$$= 0.4326 / 0.5059$$
$$= 0.8551$$

And the normalized and adjusted RMS value will be :

$$NRMS = RSF \cdot RLF \cdot RMS / RMS_{av} \qquad \text{.....eqn. 13.6}$$

$$= 10.996 \cdot 0.8551 \cdot 0.045 / 0.312$$

$$= 1.356$$

In other words an RMS value of only 0.045 G's at 400 rpm and 10 kN load indicated a slightly more damaged bearing than a value of 0.312 G's at 1800 rpm and 22 kN load.

Of course at this low speed the diagnosis would become less reliable because of the greater statistical variance of vibration measurements compared with higher speeds.

APPENDIX 45. : GEOMETRIC PROGRESSIONS IN THE CATEGORIZATION OF VIBRATION LEVELS

Data taken from various literature sources [e.g.. 7, 45] indicated that categorization of vibration levels corresponded reasonably well with a geometric progression. In other words successive boundary values of the categories increased by some constant factor.

For example ref 45 categorized vibration severity levels as follows.

TABLE A45.1 : CATEGORIZATION OF VIBRATION SEVERITY

ACCELERATION [G's]	BEARING CONDITION
2	Normal
2 - 4.5	Light Surface Marking
4.5 - 9	Heavy Marking to Visible Damage
9 - 15	Shutdown

A geometric progression between the first (V_{MIN}) and final (V_{MAX}) values and with 4 values (N) in total was obtained by finding the Geometric Progression Factor (GPF) as follows.

$$GPF = \sqrt[N-1]{(V_{MAX}/V_{MIN})}$$

$$= \sqrt[3]{(15/2)}$$

$$= 1.957$$

i.e. Value 2 = 1.957 * Value 1 etc.

The comparison of the two sets of boundary values are shown in table A45.2.

TABLE A45.2 : COMPARISON OF VIBRATION CATEGORIZATION METHODS

VALUES FROM TABLE A45.1	GEOMETRIC PROGRESSION
2	2
4.5	3.9
9	7.7
15	15

APPENDIX 46. : SAMPLE CALCULATION FOR TABLE 13.4

Table 13.4 illustrated the percentage difference between a geometric progression estimate and an estimate based on the average of various techniques.

For example the percentage difference between the first values of NRMS was calculated as follows.

Average Estimate = 0.815
Geometric Estimate = 0.599

Difference = 0.216

Percentage difference = 0.216/599

= 35.9 %

APPENDIX 47. : USE OF THE BINOMIAL THEOREM TO ESTIMATE PROBABILITIES FOR THE EXPERT SYSTEM DIAGNOSIS ZONES

The binomial distribution was used to estimate the probabilities for the diagnosis zones of bearing condition. The general form of the binomial expansion is [46] :

$$(p + q)^n = p^n + np^{n-1}q + n(n-1)/2!\ p^{n-2}q^2 + \ldots.$$

$$\ldots + [n(n-1)\ldots(n-m+1)]/m!\ p^{n-m}q^m + \ldots. + q^n$$

For example if there are 5 zones of bearing condition then n = 5. And for the mean to fall on say the 2nd value then the mean = 2. In addition p + q = 1 and mean = nq = 2.

Thus $q = 2/n = 2/5$

and $p = 1 - q = 1 - 2/5 = 3/5$

i.e. $(3/5 + 2/5)^n = 1/5^5\ [3^5 + 5\cdot3^4\cdot2 + 10\cdot3^3\cdot2^2 +$

$$10\cdot3^2\cdot2^3 + 5\cdot3\cdot2^4 + 2^5]$$

$$= 0.08 + 0.26 + 0.36 + 0.23 + 0.08 + 0.01$$

These values were doubled before being used as the probability percentages for zone 2.

APPENDIX 48. : CALCULATION OF THE LOAD LIMITS FOR TABLE 13.10

The bearing load limits of table 13.10 were established using the rated life hours of table 13.7 in equation 12.6. This equation was as follows.

$$L_{10h} = C \cdot T^{m_t} \cdot N^{m_n} \cdot P^{m_p} \qquad \text{.....eqn. 12.6}$$

This was rearranged for P as :

$$P = \sqrt[3]{(C \cdot T^{m_t} \cdot N^{m_n} / L_{10h})} \qquad \text{.....eqn. A48.1}$$

Then using a speed of 1500 rpm and a temperature of 60°C, the values of life rating hours in table 13.7 were inserted into equation A48.1. For example using a life rating of 5000 hours, the equivalent load was given by :

$$P = \sqrt[3]{(C \cdot (60)^{m_t} \cdot (1500)^{m_n} / 5000)}$$

$$= \sqrt[3]{(167.89E6 \cdot 60^{\circ} \cdot 1500^{-1} / 5000)}$$

$$= 2.82 \text{ kN}$$

The other values were calculated using a similar process.

APPENDIX 49. : AVOIDANCE OF THE ALIASING EFFECT IN DIGITIZED WAVEFORMS USING ANTI ALIASING FILTERS AND THE NYQUIST SAMPLING CRITERION

The aliasing effect can occur when an analog waveform is digitized by sampling at some constant time interval (T_s). In particular it occurs when the sampling frequency ($F_s=1/T_s$) is not high enough to reproduce the high frequencies of the waveform. The result is that components appear in the frequency spectrum which were non existent in the original waveform.

The Nyquist sampling theorem states that to avoid the aliasing effect, the sampling frequency must be at least 2 times higher than the highest frequency component of the waveform. This can be conceived intuitively by considering a sine wave of a ceratin frequency (F_w).

If this wave is sampled exactly twice per cycle (i.e. twice the wave frequency) then the wave could be reconstructed using the samples, as long as they occurred at the maximum amplitude of the wave. If the wave was sampled slightly faster than twice per cycle then one of the samples would eventually occur at the maximum frequency and the wave could be fully reconstructed.

However if the wave was sampled at slightly less than twice the wave frequency then consecutive samples would occur at slightly increasing or decreasing amplitudes. Reconstruction of the wave from the samples would therefore indicate a superimposed sine wave of lower frequency clearly not in the original wave.

Aliasing is avoided by filtering out components in the analog waveform which are greater than ½ F_s. However real filters do not have infinitely steep skirts and therefore the cutoff frequency (F_c) can not be placed at ½ F_s. In fact it is usually specified that the filter should have an attenuation of say 60 dB at F_c.

This means that the spectrum amplitudes between F_c and ½ F_s will be affected by the filters attenuation. In other words the spectrum will only be linear up to F_c. Thus the sampling frequency is usually set higher than twice the the required frequency to be displayed.

For example, if a waveform is sampled with 1024 (2^{10}) samples then the Fourier transform will consist of 512 unique values. Some of these values will be affected by the rolloff of the filter skirt and therefore typically the first 400 values are displayed.

i.e. F_s = 1024/400 = 2.56 x Displayed frequency.

APPENDIX 50. : SPECIFICATIONS OF THE 'PCB' ACCELEROMETER USED TO MEASURE THE BEARING VIBRATIONS

Model Number (see optional models below specifications)		308B04
Sensitivity Peak[3]	mV/g	100 ±2%
Range (for ±5V out)	±gpk	50
Range (for ±10V out)[11]	±gpk	100
Resolution	gpk	.001
Resonant Freq (mtd)	>kHz	25
Freq Range ±5%[7]	Hz	2 to 5000
Freq Range ±10%[7]	Hz	1.3 to 7000
Overload Recovery	μs	10
Discharge Time Constant[2] @ 70°F	>s	0.5
Amplitude Linearity (zero based best straight line)	%	1
Output Impedance★	<ohm	100
Transverse Sens (max)	%	5
Strain Sens	g/μin/in	01
Temperature Range[4]	°f	-100 to +250
Temp Coefficient	%/°F	03
Vibration (max)	±gpk	500
Shock (max)	gpk	5000
Structure (iso-compression)		inverted
Size (hex × height)	in	.75 × 2.0
Weight	gm	87
Connector		5/16-32 Con-hex
Case Material	SS	SS
Sealing		hermetic
Ground Isolation[6]		no