

University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND THE ENVIRONMENT

NATIONAL CENTRE FOR ADVANCED TRIBOLOGY SOUTHAMPTON (nCATS)

**ROLLING CONTACT FATIGUE FAILURES IN SILICON NITRIDE AND THEIR
DETECTION**

by

Robert J. F. Hanzal

Thesis for the degree of Doctor of Philosophy

September 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND THE ENVIRONMENT

Thesis for the degree of Doctor of Philosophy

ROLLING CONTACT FATIGUE FAILURES IN SILICON NITRIDE AND THEIR DETECTION

Robert J. F. Hanzal

The project investigates the feasibility of using sensor-based detection and processing systems to provide a reliable means of monitoring rolling contact fatigue (RCF) wear failures of silicon nitride in hybrid bearings. To fulfil this investigation, a decision was made early in the project to perform a series of hybrid rolling wear tests using a twin disc machine modified for use on hybrid bearing elements.

The initial part of the thesis reviews the current understanding of the general wear mechanisms and RCF with a specific focus to determine the appropriate methods for their detection in hybrid bearings. The study focusses on vibration, electrostatic and acoustic emission (AE) techniques and reviews their associated sensing technologies currently deployed with a view of adapting them for use in hybrids. To provide a basis for the adaptation, an understanding of the current sensor data enhancement and feature extraction methods is presented based on a literature review.

The second part describes the test equipment, its modifications and instrumentation required to capture and process the vibration, electrostatic and AE signals generated in hybrid elements. These were identified in an initial feasibility test performed on a standard twin disc machine. After a detailed description of the resulting equipment, the thesis describes the calibration tests aimed to provide base data for the development of the signal processing methods.

The development of the signal processing techniques is described in detail for each of the sensor types. Time synchronous averaging (TSA) technique is used to identify the location of the signal sources along the surfaces of the specimens and the signals are enhanced by additional filtering techniques.

The next part of the thesis describes the main hybrid rolling wear tests; it details the selection of the run parameters and the samples seeded with surface cracks to cover a variety of situations, the method of execution of each test run, and the techniques to analyse the results.

The research establishes that two RCF fault types are produced in the silicon nitride rolling element reflecting essentially different mechanisms in their distinct and separate development; i) cracks, progressing into depth and denoted in this study as C-/Ring crack Complex (CRC) and ii) Flaking, progressing primarily on the surface by spalls. Additionally and not reported in the literature, an advanced stage of the CRC fault type composed of multiple and extensive c-cracks is interpreted as the result of induced sliding in these runs. In general, having reached an advanced stage, both CRC and Flaking faults produce significant wear in the steel counterface through abrasion, plastic deformation or 3-body abrasion in at least three possible ways, all of which are described in details.

The vibration and electrostatic sensor-produced signals are dominated by eccentricity components, presenting a noisy background from which the fault-related components are difficult to discern. The filtering methods combined with TSA are able to reduce this background to a level from which the fault profiles can be identified at a particular stage of their development. Flaking presents more prominent vibration profiles than CRC. In electrostatic signals, most faults are detectable earlier than in vibration with CRC being more prominent. In addition, the electrostatic detection methods are able to sense direct contacts between the two specimens or silicon nitride debris attached to the steel counterface.

The AE profiles, despite the major work involved in the development of the signal enhancement methods and obtaining promising results from the preparatory tests, were poor on features, both pre- and post-filtering. The study reasons that the major cause of this is the complex profile patterns produced by the test machine, presenting complex transmission paths for the fault-generated signals. Nevertheless, crack propagation events from indentation tests showed that the AE fault signatures in silicon nitride, produces high amplitude signal profiles that could be extracted using the signal enhancement techniques developed in this work.

The study concludes that both vibration and electrostatic monitoring methods are able to detect and track RCF development in hybrid bearings. As each of these methods is better suited to the detection of a particular fault type, a combination of the two methods would provide a basis for a robust detection system.

Contents

ABSTRACT.....	i
Contents	iii
List of Tables	xii
List of Figures.....	ix
DECLARATION OF AUTHORSHIP	xv
Dedication	xvi
Acknowledgements	xvi
Definitions and Abbreviations	xvii
Chapter 1. Introduction and Overview.....	1
1.1 Background	1
1.2 Project Aim	2
1.3 Objectives in Detail.....	3
1.3.1 Wear Generation Aspects	3
1.3.2 Signal Processing Aspects	3
1.3.3 Objectives	4
1.4 Project Approach	5
1.5 Structure of the Document	6
Chapter 2. Review of Background Tribology.....	7
2.1 Rolling Element Bearing Tribology.....	7
2.1.1 Introduction.....	7
2.1.2 Hertzian Contacts.....	7
2.1.3 Dynamic Contacts.....	10
2.1.4 Friction.....	10
2.1.5 Lubrication.....	11
2.1.6 Wear.....	12
2.2 Hybrid Bearings	15
2.2.1 Introduction.....	15
2.2.2 Silicon Nitride.....	15
2.2.3 Wear Characteristics	22
2.3 Summary	31
Chapter 3. Review of Signal Processing.....	33
3.1 Overview of Signal Processing	33
3.2 RE Bearing Sensing Strategies	35
3.2.1 Introduction: Sensing Techniques Used in Bearing Monitoring	35
3.2.2 Vibration	35
3.2.3 Electrostatic Charge	39
3.2.4 Acoustic Emission	49
3.3 Data Enhancement and Feature Extraction	57
3.3.1 Time - Frequency Analysis.....	57
3.3.2 Time Synchronous Analysis	70
3.4 Summary	73

Chapter 4. Hardware Preparation.....	77
4.1 Indentation Machine	77
4.1.1 The Standard Machine.....	77
4.1.2 Configuration and Instrumentation.....	77
4.1.3 Data Acquisition System	78
4.2 Twin Disc Machine	79
4.2.1 The Standard Machine.....	79
4.2.2 Modification Details.....	81
4.2.3 The Modified Machine.....	85
4.3 Physical Inspection Tools.....	90
4.3.1 Talysurf Profilometer	90
4.3.2 Alicona Infinite Focus	90
4.1.1 Digital Anglemeter.....	90
4.3.3 Scanning Election Microscope.....	90
4.3.4 Computer Tomography Analysis	90
Chapter 5. Signal Processing Development.....	91
5.1 Overview of Process	91
5.2 Signal Generation Tests.....	92
5.2.1 Vibration Sensing.....	92
5.2.2 WSS Sensing.....	94
5.2.3 AE Sensing.....	98
5.2.4 Healthy Run.....	109
5.3 Signal Analysis Development	115
5.3.1 Vibration.....	115
5.3.2 Electrostatic Charge	117
5.3.3 AE.....	118
5.4 Summary.....	123
Chapter 6. Production Runs Planning and Execution	125
6.1 Overview of Process.....	125
6.2 Runs Planning.....	125
6.3 Execution.....	127
6.3.1 Preparation of Specimens.....	127
6.3.2 Pre-Run Inspection	127
6.3.3 Setting Up the Twin Disc Machine	127
6.3.4 Monitoring of the Run and the Data Collected	129
6.3.5 Post-Run Inspection	130
6.4 Overview of the Analysis Process	130
Chapter 7. Visual Inspection Analysis.....	133
7.1 Introduction and Overview	133
7.2 Post-Run Fault Snapshots.....	133
7.3 Fault and Indentation Correlations	136
7.4 Basic Model of Fault Propagation.....	140
7.5 CRC Propagation Dynamics.....	144
7.5.1 EFD→ CRC(a) → CRC(b)	144
7.5.2 CRC(b) → CRC(c).....	144
7.5.3 CRC(c) → CRC(d).....	145
7.5.4 The Contribution of the Steel Counterface.....	147
7.6 Flaking Propagation Dynamics	149
7.6.1 EFD→ F(a) → F(b).....	149
7.6.2 F(b) → F(c)	150
7.6.3 F(c) → F(d)	151
7.6.4 3-body Abrasion.....	153
7.6.5 The Contribution of the Steel Counterface.....	155
7.7 Summary.....	159

Chapter 8. Sensor Data Analysis	163
8.1 Introduction and Overview	163
8.1.1 Analysis Tools	163
8.1.2 Analysis Approach.....	164
8.2 Runs at a Glance: Signal-Time Profiles	166
8.3 Vibration Sensing.....	169
8.3.1 Horizontal Bands	170
8.3.2 Raw Burst Profiles - Contact Eccentricity Bands	175
8.3.3 Frequency Analysis.....	179
8.3.4 Vibration Results Summary	182
8.4 Electrostatic Sensing.....	184
8.4.1 Eccentricity Bands	185
8.4.2 Raw Burst Profiles- Contact Eccentricity Bands	188
8.4.3 Vertical Strips	190
8.4.4 Frequency Analysis.....	194
8.4.5 Summary	198
8.5 Fault Histories Deconstruction.....	199
8.6 Acoustic Emission Sensing.....	201
8.7 Summary	204
Chapter 9. Conclusions and Recommendations	207
9.1 Project Overview	207
9.2 Conclusions.....	209
9.3 Recommendations.....	212
Appendices.....	215
Appendix A: Hertzian (Elliptical) Contact Equations.....	215
Appendix B: Feasibility Test.....	217
Appendix C: Test Cell Modification Details.....	225
Appendix D: WSS Charge Modelling	228
Appendix E: Indentation Tests	231
Appendix F: AE Filter Development.....	236
References.....	241

List of Tables

Table 2.1: Hertzian contact equations for spherical and elliptical contacts [3, 4].	8
Table 2.2: The properties of the three materials used in the study by Lee, Wuttiphan et al [33]	17
Table 2.3: Material properties of three bearing grade silicon nitride materials typically used by the industry (data provided by the manufacturers)	18
Table 3.1: Work functions of a range of metals [131].	41
Table 4.1: Material properties of the specimens (information provided by the manufacturers i.e. Toshiba for silicon nitride and SKF for steel).	87
Table 4.2: Charge amplifier settings for the vibration sensor and WSSs.	88
Table 5.1: Loads applied between the specimens for vibration impulse response measurements	92
Table 5.2: Indentation loads for Indentation Tests	99
Table 5.3: Healthy Run settings.	110
Table 5.4: Prominent fault and noise peaks and the filtering efficiency	122
Table 5.5: Summary of the advanced signal processing techniques	124
Table 6.1: Production Run parameter configurations.	126
Table 6.2: Summary of the test sequence programme implemented by Compendx for Production Runs	128
Table 6.3: SRR values and their corresponding applied specimen rotating speeds.	128
Table 6.4: Calculated values for various contact parameters corresponding to the run conditions. Unless otherwise stated, calculations were performed using the equations in Appendix A.	129
Table 6.5: Actual run durations.	129
Table 7.1: Post-run fault snapshots (taken using the AIF). Fault 1 images taken at 20x magnification, remaining faults taken at various magnifications and fault dimensions are instead indicated. The rolling and traction directions are indicated for each run in column 2. The centres of the seeded indents are marked with a red dot. Coloured boxes are referred to in the sections following.	134
Table 7.2: Intermediate fault snapshots (taken using the AIF) at run interruption for Run 2-1. The rolling and traction directions are indicated in column 2.	135
Table 7.3: Specimen faults and secondary wear mechanisms identified from post-run fault snapshots and the intermediate snapshots from Run 2-1. The numbers indicate fault numbers.	136
Table 7.4: Pre-run fault snapshots (taken using the AIF under polarised light filter). Fault 1 images taken at 20x magnification, remaining faults taken at 10 x magnifications. The rolling and traction directions are indicated for each run in column 2. The centres of the seeded indents are marked with a red dot	137
Table 7.5: Pre-cracks and corresponding fault categories (F= Flaking, CRC, npf =no prominent features) and wear categories (3B= 3body abrasion and A= adhesion).	139
Table 7.6: Identification of pre-cracks and faults with the model stages.	143
Table 7.7: EDX measurements of the corresponding locations from Fault 3 in Run 2-4 shown in Figure 7.9.	148
Table 8.1: Runs selected for detailed analyses.	165
Table 8.2: Signal-time profile descriptions	166
Table 8.3: Signal-time profiles for the runs selected for detailed analyses.	167
Table 8.4: Run periods of the three selected runs (EoR- end of run)	168
Table 8.5: Eccentricity band locations as identified in the raw vibration angular-time profiles	170
Table 8.6: Fault snapshots and prominent vibration events ('FS' = first sight, subsequent event peaks labelled by letters).	172

Table 8.7: Snapshots of the steel wear tracks.	175
Table 8.8: Comparison between resonant frequency bands (*indicate the resonant frequencies found in the Healthy Run).....	182
Table 8.9: Eccentricity band locations as identified in the raw WSS1 angular-time profiles.....	185
Table 8.10: Summary of WSS1 fault-relevant features (event labels are referred to in section 8.5).	187
Table 8.11: Comparisons of the first sight points between WSS1 and vibration.	187
Table 8.12: Deconstruction of the fault developments in Run 1-2	199
Table 8.13: Deconstruction of the fault developments in Run 2-1	200
Table 8.14: Deconstruction of the fault developments in Run 2-2	200
Table F.1: Wavelet bases and their corresponding MATLAB index range.	236

List of Figures

Figure 2.1: Cross sections of a RE bearings a) front on view and b) Side on views of two typical bearing designs.....	7
Figure 2.2: The profile of principal stresses response for a circular elastic contact a) σ_1 principal tensile stress and b) principal shear stress τ_{12} . Both have been normalised using p_0 :the mean contact pressure [5].....	8
Figure 2.3: a) Subsurface profile of a cone crack b) concentric ring cracks in silicon nitride [11].....	10
Figure 2.4: The Stribeck curve.....	12
Figure 2.5: SEM image of a silicon nitride material consisting of both Alpha and Beta grains [27].....	17
Figure 2.6: Snapshot of a Star crack under fluorescent light [43].....	19
Figure 2.7: An illustration of the typical structure of sectioned Vickers indentations produced in brittle materials with a) Palmqvist radial cracks and b) median radial cracks.	20
Figure 2.8: SEM snapshot of a C-crack found on silicon nitride [53], b) illustration of a subsurface profile of a c-crack [9]	21
Figure 2.9: Development of grain pull-outs a) fracture of the interphase due to cyclic fatigue, b) removal of grains as a result of traction.	22
Figure 2.10: Transgranular surface film with microcracks developed after 100hrs of testing [61].....	23
Figure 2.11: An illustration of various material damages in silicon nitride in a sliding contact. Random distributed Si ₃ N ₄ grains near the surface are either fractured; pulled out as a whole/partial grain or plucked out as a whole grain [63].	23
Figure 2.12: SEM of Silicon nitride wear surface a) pits on the worn surface b) debris on the surface and c) close up of debris (a cluster of finer/sub-micron particles) [63]	24
Figure 2.13: An illustration of the stages of spallation described by Kida [46]	27
Figure 2.14: Illustrations of the testing configuration performed in [52] showing: (a) Machine and loading configuration; (b) geometry of a ring crack on the contact track; (c) 12 tested crack locations. The three failures are highlighted in red with their respective millions of cycles to failure.....	28
Figure 2.15: Illustration of the two crack orientations in the investigation by Wang and Hadfield [81]. a) crack orientation acts to trap lubricant in the crack, whilst b) acts to squeeze lubricant from the crack.	29
Figure 2.16: The subsurface branching cracks after 3.2×10^7 cycles in separate tests [46].....	29
Figure 2.17: An illustration of the pre-crack with a shallow incline opening under mode 1 stress b) a representation of the cantilever effect	30
Figure 2.18: Post-test SEM analysis of the failed surfaces a) radial crack propagation and delamination; b) lateral crack delamination; c) surface bulging Hadfield, Stolarski et al [55].	30
Figure 3.1: Signal processing stages	33
Figure 3.2: Illustration of an electrical double layer [130].....	40
Figure 3.3: Charge measurements from a steel rotating disc with an aluminium insert [132].	41
Figure 3.4: Charge intensities during dry pin on disc scratching tests. a) Negatively charged particles and corresponding friction coefficient on various solids [136]. b) Negatively charged particles (upper) and positively charged particles with CoF during scratching Si ₃ N ₄ [137].	42
Figure 3.5: The behaviour of (a) a conducting plate in the vicinity of (b) charged particle (c) and with a connection to the earth.	44
Figure 3.6: Illustration of two types of electrostatic sensors available for wear monitoring. a) WSS. b) OLSs [14].....	45
Figure 3.7: The layout of the FZG gear testing rig with the corresponding ES sensor locations [146]	47
Figure 3.8: The locations of the WSSs in respect to the bearings. Harvey et al [147]	48

Figure 3.9: Sensor responses from a bearing rig test leading to spalling failure, WSS3 (upper), Vibration (middle) and Thermocouple (lower), during the final 13hrs of a bearing test [147].....	48
Figure 3.10: An illustration of a simplified signal arising from an AE event with the associated hit-based features.	51
Figure 3.11: AE Hit count rate at the incipient of damage located at segment 19 where transverse cracks formed.[164].....	52
Figure 3.12: Illustration of the ‘zoning’ technique proposed by Faisal et al [155].....	54
Figure 3.13: Analysis of a AE signal from a defective bearing [168]- the raw signal (upper), with added Gaussian noise (middle) and EI applied (the red circles correspond to peaks of the fault components).	55
Figure 3.14: Frequency-time profile of AE signals recorded from a pin on disc machine produced using Welch PSD method [160].....	58
Figure 3.15: Examples of various wavelet functions a) Morlet (with two different β values), b) Gaussian (order 2 - <i>Mexican Hat</i>), c) Laplace and d) One sided Morlet wavelet.	61
Figure 3.16: Frequency-time profile (scalogram) comparisons of a signal decomposed using CWT using a wavelet producing a) minimal entropy b) higher entropy values. Figures reproduced using the data provided in [201].	63
Figure 3.17: Illustration of WPA decomposition tree to level 2.	65
Figure 3.18: Illustration of the iterative sifting process proposed by Huang [222].	67
Figure 3.19: Frequency-time profiles of the Hilbert-Huang technique applied to vibration data from a RE bearing a) healthy, b) with an outer race defect [226].	69
Figure 3.20: Illustration of the TSA technique.	70
Figure 3.21: Angle-time profiles, for the two test types a) gradual increasing load, b) debris seeding.	71
Figure 3.22: A life cycle model of a RE in a hybrid bearing indicating dominant physical and their signal-generation features.....	74
Figure 4.1: A photograph illustrating the experimental hardware for Vickers indentation tests with a steel ball sample.	78
Figure 4.2: A photograph of the Twin Disc machine with the original unmodified test cell.....	80
Figure 4.3: An illustration of the side on view of the original test cell and the specimen locations.	80
Figure 4.4: Schematic of the instrumented test cell in the original machine configuration.	81
Figure 4.5: A photograph of the test cell of the modified machine	85
Figure 4.6: Schematic of the test cell of the modified machine. The rotational angle between the contact and WSS1 for the upper specimen is illustrated and indicated in red, and similarly in orange for WSS2.	85
Figure 4.7: Sample pair geometry for modified machine showing a) steel, b) silicon nitride	86
Figure 4.8: An illustration of the two WSS sensor head geometries	87
Figure 5.1: Organisation of the main tasks in the signal analysis development	91
Figure 5.2: Recorded vibration responses of the four impact tests impulse measurements at various loads (indicated in the figure labels)	93
Figure 5.3: PSD Estimates of the vibration impulse responses at various loads (highlighted in the legend entries (in kN). The main resonant vibration modes are labelled and their respective frequencies are given.	94
Figure 5.4: Illustrations of the Impulse Disc specimen a) photograph b) schematic showing the weld geometry.....	95
Figure 5.5: Normalised TSA responses of the calibration disc for both sensor sizes (indicated in the legend), a) in the angular domain b) a close-up of the weld location and c) in the frequency domain (PSD).	97

Figure 5.6: An illustration of Indentation Start, First Contact and AE0 (introduced in the following subsection). The start button represents the starts of the indentation tip, the stopwatch and AE recording	100
Figure 5.7: An indent with radial cracks and its measurements and calculations.	101
Figure 5.8: a) A SEM image of indent 20-A, b) average diagonal surface lengths ($2a$) with load for steel indentations.	101
Figure 5.9: a) An AIF image of indent 20-A, b) Average indent geometries with load for silicon nitride indentations.	102
Figure 5.10: Measured first contact times and first AE responses (AE0) for both steel and silicon nitride indentations	103
Figure 5.11: The smoothed AE profiles for steel indentation tests, at a) 15, b) 20 and c) 30 kgF loads. The PSDs of the tests at 30 kgF loads are illustrated in d).	104
Figure 5.12: Graphs of mean duration and mean amplitude responses for the steel Indentation Tests	105
Figure 5.13: a) to c) The truncated and smoothed AE waveforms for all silicon nitride indentation tests for each of the three test loads from a) to c) 15, 20, 30 kgF. The subsequent zones have been indicated for each test and their respective boundaries highlighted d) PSD estimates of the three zones for 15-A.	106
Figure 5.14: Graphs of a) mean duration and b) mean amplitude responses of each of the three zones of the AE waveforms in silicon nitride indentation tests.	107
Figure 5.15: Talysurf wear track profile of the steel specimen from the Healthy Run.	111
Figure 5.16: Vibration PSD profile for the Healthy Run	112
Figure 5.17: WSS RMS profiles for the Healthy Run	112
Figure 5.18: WSS signal profiles distributed across the rotation of the specimens a) silicon nitride and b) steel	113
Figure 5.19: WSS signal profiles distributed across the rotation of the specimens a) silicon nitride, b) steel, cross-timed (lower and upper tachometer signals respectively)	113
Figure 5.20: WSS1 frequency profile of the silicon nitride specimen.....	114
Figure 5.21: Frequency profiles: a) time profile, b) the frequency spectra at the beginning (5 hrs) and the end (70 hrs) of the run.	115
Figure 5.22: The raw (TSAa) and the decomposition mode profiles IMF1 & IMF2 showing a) angular profiles and b) frequency profiles (PSD)	117
Figure 5.23: WSS1 profiles taken at various stages of the data processing development.	118
Figure 5.24: Time (a) and frequency (b) patterns of Indentation 30-A (Zone 2) and Healthy Run (end of test).....	119
Figure 5.25: a) combined signal profiles: pre- (top) and post- (bottom) WP filter, b) Indentation frequency profiles: pre- and post- WP filter.....	121
Figure 5.26: EI of AE waveforms with the fault-related peaks indicated in red	122
Figure 6.1: Sketch of the contact shape on a specimen, illustrating also the motion of the contact and specimen rotating direction. Leading and trailing directions refer to the rolling contact	129
Figure 6.2: Flow chart illustrating the stages for the analysis of results	131
Figure 7.1: Example pre-crack snapshots, illustrating the symmetry (or asymmetry) characteristics a) symmetrical, b) leading and c) trailing edge skewness. The red dot indicates the centre of the indentation (i.e. the tip of the indenter). Snapshots taken from Runs: 1-1 (Fault 2), 1-2 (Fault 3), 2-3 (Fault 3) respectively.	138
Figure 7.2: left snapshot: pre-crack Fault 3 snapshot before Run 2-2 illustrating the lateral crack skewness was relatively small and subsequently led to a CRC failure. The right snapshot shows a more typical skewed crack for comparison (Fault 3 of Run 1-2).	139
Figure 7.3: Stages of fault development.....	141

Figure 7.4: a) indentation process made at an oblique angle b) AIF snapshot from Run 1-1 (Fault 1) showing a prominent skewness in the pre-crack.....	143
Figure 7.5: AIF Snapshots of Fault 2 from Run 2-1: a) intermediate b) end-run fault surface profile measurements c) corresponding depth profile measurements.....	145
Figure 7.6: a) Post-run snapshot of Fault 3 from Run 2-2 b) corresponding surface profile measurements.	145
Figure 7.7: Rendered CT images of Fault 3 Run 2-2 combined with the surface AIF image in the surface view (the subsurface cracks shown in variants of yellow colouring with darker colours indicating shallower crack regions, surface cracks are shown in black).....	146
Figure 7.8: Analyses of the steel wear track from Run 2-2 a) AIF image b) profile measurement.	147
Figure 7.9: Post-run SEM analyses of Fault 3 from Run 2-4 indicating (with yellow markers) the locations where steel has adhered (confirmed by EDX analysis). Image locations are indicated in the relevant image of Table 7.1 with the respective box colours. a) Cracks and material loss at the trailing edge of the fault), b) Cracks located at the leading side of the fault.....	148
Figure 7.10: a) AIF image of the steel wear track from Run 2-4 b) corresponding profile measurement	149
Figure 7.11: AIF snapshots of Fault 2 (<i>M</i> pre-crack geometry) from Run 2-2 a) pre-run (polarised light), b) post-run c) post-run profile measurement.....	150
Figure 7.12: Post-run AIF snapshot of Fault 2 (<i>M</i> pre-crack geometry) from Run 2-5.....	150
Figure 7.13: AIF snapshots of Fault 3 from Run 2-1 a) intermediate stage, b) post-run c) post-run colour map showing the surface topography obtained by the 3D analysis tool of the AIF.....	152
Figure 7.14: AIF snapshots of Fault 3 from Run 1-2 a) pre-run indentation, b) post-run c) colour map showing the surface topography. The markers are positioned for reference of the corresponding steel wear track shown in Figure 7.17 and discussed in Subsection 7.6.5.	152
Figure 7.15: a) AIF image of Fault 3 and surrounding areas from Run 2-4 b) AIF image of the steel wear track and c) its profile measurement.	153
Figure 7.16: SEM images of the silicon nitride sample detailing the areas within the boxes shown in Figure 7.15 a) portion of the upper band b) lower band c) trailing edge of the fault area.	154
Figure 7.17: a) AIF post-run snapshot of the steel sample from Run 1-2 b) profile measurement. Corresponding markings of the ceramic fault are shown in Figure 7.14.....	155
Figure 7.18: Cross-sectional view illustrating a) the imprinting and gouging of the steel by central and peripheral zones of a fault b) the smooth surface emphasises the central area (the specimen rotate in direction perpendicular to the page).	157
Figure 7.19: Run 1-1 steel wear rack analyses, SEM images a) 35x and b) 250x at the corresponding box colour in (a); c) a Talysurf profile measurement across the wear track.....	157
Figure 7.20: SEM image of the upper part of Fault 3 from Run 1-1 (refer to Table 7.1 for its position), and in the table, the corresponding EDX measurements at the locations indicated.....	157
Figure 7.21: SEM images of the two spalls found in the steel of Run 1-1 a) <i>Large Spall</i> , b) <i>Small Spall</i>	158
Figure 7.22: Topography images from Run 1-1 of a) the Large Spall, b) Small Spall and c) Fault 3 of the ceramic specimen.....	158
Figure 8.1: Vibration angular-time profiles for TSaA-raw (left column), TSaA-EMD (right column)....	169
Figure 8.2: Vibration angular-time profiles for final 20 hrs Run 2-2 for TSaA-raw (left column), TSaA-EMD (right column).	170
Figure 8.3: Vibration angular-time profiles with respect to the steel specimen for TSaA-raw (left column), TSaA-EMD (right column).....	174
Figure 8.4: Raw burst vibration data for Run 1-2 with respect to silicon nitride (left column) and steel (right column).	176

Figure 8.5: Raw burst vibration data for Run 2-1 with respect to silicon nitride (left column) and steel (right column).	177
Figure 8.6: Raw burst vibration data for Run 2-2 with respect to silicon nitride (left column) and steel (right column).	178
Figure 8.7: Vibration frequency-time profiles for raw-TSaA (left column) and EMD-TSaA (right column).	181
Figure 8.8: WSS1 angular-time profiles for TSaA-raw (left column) and TSaA-filtered (right column).	184
Figure 8.9: WSS1 angular-time profiles for the last 20 hrs of Run 2-2 for TSaA-raw (left column) and TSaA-filtered (right column).	185
Figure 8.10: WSS1 angular-time profiles for raw burst data timed against the silicon nitride (left column) and steel (right column) tachometers.	189
Figure 8.11: WSS2 angular-time profiles for raw burst data of Run 2-2 timed against the steel (left column) and silicon nitride (right column) tachometers.	189
Figure 8.12: WSSs angular-time profiles for TSaA-raw timed against the steel tachometer showing WSS2 (left column) and WSS1 (right column).	191
Figure 8.13: WSS1 burst profile at 17.7 hrs of Run 1-2 timed against a) silicon nitride, b) steel.	192
Figure 8.14: A photograph showing Fault 2 post-Run 1-2 before cleaning.	192
Figure 8.15: a) WSS2 angular-time profile for Run 2-2 from 75 hrs to end of run b) Steel specimen rotating frequency.	194
Figure 8.16: WSS1 frequency-time profiles pre- (left column) and post- filtered (right column).	195
Figure 8.17: WSS1 frequency-time profiles for the last 20 hrs of Run 2-2 pre- (left column) and post-filtered (right column).	196
Figure 8.18: Illustration of the charge smearing effect on the silicon nitride via the steel specimen a) primary transfer and b) consecutive secondary transfer.	197
Figure 8.19: AE signal-time (left column) and frequency-time (right column) profiles.	202
Figure B.1: Schematic of the instrumented test cell in the original disc-on-disc configuration.	217
Figure B.2: Samples for the feasibility test a) steel, b) silicon nitride.	219
Figure B.3: SEM images of the steel disc wear track at magnifications (a) x75, (b) x750, (c) x1,500. (b): two parallel interacting cracks forming a pit. (c) shows a 1,500 a typical pit.	221
Figure B.4: SEM images of the SN disc surface at magnifications (a) x75, (b) x1,000 (c) x1,500. (a) the damage of the wear track and polishing damage, (b) untested surface- damage from polishing. (c) typical damage at the wear track	221
Figure B.5: RMS graphs of sensor responses split in to a): initial phase of test (0-9hrs) and b): final phase of test (42.5-51.5hrs). Both plots show the graphs of (from top to bottom) AE, WSS1, WSS2, Vibration.	222
Figure B.6: TSA angular-time profile of WSS2 over the entire test. Blank vertical bands are due to raw data loss by the processing system.	223
Figure C.1: Assembly of the modified upper shaft and its modified upper bearing housings attached to the Load Arm. a) front-on-view sectioned vertically at the shaft axis shown in (b). b) & c) the side-on-views.	225
Figure C.2: Drawings of the upper specimen shaft assembly in its exploded and assembled forms.	226
Figure C.3: Drawings of the lower specimen assembly in its exploded and assembled forms.	227
Figure D.1: Illustration of a charge source passing the WSS sensing face	228
Figure D.2: Transfer profiles for WSS with various sensing face lengths, in the spatial domain (a) and in the normalised frequency domain (b).	229

Figure D.3: Modelled defects of various sizes in the a) spatial and b) frequency domains. The signal profiles of the modelled 0.3 mm length WSS in the c) spatial and d) frequency domains	230
Figure E.1: Alicona optical images (20x magnification) of indentations made at 15 kgF for steel tests.	231
Figure E.2: Alicona optical images (10x magnification) of indentations made at 20 kgF for steel tests.	231
Figure E.3: Alicona optical images (10x magnification) of indentations made at 30 kgF for steel tests.	231
Figure E.4: SEM of indentations <i>15-A</i> and <i>20-A</i>	231
Figure E.5: Alicona optical images (20x magnification) of indentations made at 15 kgF for steel tests.	232
Figure E.6: Alicona optical images (20x magnification) of indentations made at 20 kgF for steel tests.	232
Figure E.7: Alicona optical images (20x magnification) of indentations made at 30 kgF for steel tests.	232
Figure E.8: a) Recorded AE waveforms for Background tests 1 (upper) and 2 (lower). b) PSD estimated of the two waveforms (indicated in the legend).	232
Figure E.9: a) Recorded AE waveforms for Steel indentation tests 15-A (upper) and 15-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).	233
Figure E.10: a) Recorded AE waveforms for Steel indentation tests 20-A (upper) and 20-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).	233
Figure E.11: a) Recorded AE waveforms for Steel indentation tests 20-A (upper) and 20-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).	233
Figure E.12: a) Recorded AE waveforms for Silicon nitride indentation tests 15-A (upper) and 15-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).	234
Figure E.13: a) Recorded AE waveforms for Silicon nitride indentation tests 20-A (upper) and 20-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).	234
Figure E.14: a) Recorded AE waveforms for Silicon nitride indentation tests 30-A (upper) and 30-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).	234
Figure E.15: The smoothed AE waveforms incorporating the total durations of Zones 1 -3 for all silicon nitride indentation tests at each of the three test loads (labelled). Up arrows indicate the full retraction of the indenter tip from the sample	235
Figure F.1: KER results for various applied mother wavelets, with the wavelets corresponding to the maximum and minimum KER values indicated.	237
Figure F.2: The smoothed response of <i>30-A</i> (upper) and the CWT time-frequency representation using the db43 mother wavelet.	238
Figure F.3: Illustration of the decision tree applied to wavelet decomposition for recorded AE signals.	239
Figure F.4: Illustrations of the WP selection analyses of the raw AE waveform (zone 2 and partial zone 3) a), the waveform with added noise, b) filtered with wavelet db43 at packet (4,6), filtered with wavelet db1 at packet (6,25). Kurtosis (K) values for each of the waveforms are indicated in the bottom right hand corner of their respective graph	240
Figure F.5: The PSD characteristic of the optimal WP filter demonstrated on the original signal waveform (see Figure F.4a)	240

DECLARATION OF AUTHORSHIP

I, ROBERT HANZAL

declare that the thesis entitled

ROLLING CONTACT FATIGUE FAILURES IN SILICON NITRIDE AND THEIR
DETECTION

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

Health Monitoring of Hybrid Bearings. Proceedings of UK-China Symposium 2009

Condition Monitoring of Hybrid Rolling Contacts. . Proceedings of STLE/ASME 2010

Condition Monitoring of Hybrid Rolling Contacts. Proceedings of CM2011 / MFPT2011

Charge and Vibration Monitoring of Silicon Nitride RCF Failures in Hybrid Rolling-Sliding Contacts. Proceedings of BINDT 2013

Signed:

Date:.....

Dedication

I dedicate this work to my father who has been a considerable mentor and role model to me.

Acknowledgements

I would like to express my gratitude to all those who gave help for my research work from the University of Southampton. A huge thanks to Dr Ling Wang for supervising and supporting me throughout the project, I realise this has been a particularly difficult and frustrating role one that has required tremendous patience. Many thanks to Professor Robert Wood for his assistance in supervision; Dr Terry Harvey for continued support and advice in various matters. Mr Steve Pilcher for the advice on technical matters including the Twin Disc machine; Dr Mark Mavrogordato for CT scanning and advice in these matters, Dr Jack Lambert for advice in CT image enhancement. Dr Andy Franz Robinson for advice in many practical aspects of the project. Mr Collin Waylen and Mr David Beckett for time spent assisting in the cutting of samples. Mr Peter Sellen and Mr Chris Williams for their technical advice and support.

EPSRC and nCATS for funding this project. I would like to thank SKF for supplying the silicon nitride samples.

I have received some tremendous encouragement and support from those who have gone beyond the typical roles of duty; who have allowed me to burden them with my moaning and groaning and self-centred conversations/monologues over a prolonged period. These people include:

Dr Nina Mauland Mansoor, in particular for her emotional support and continued patience with me,

Dr Andy Franz Robinson – ‘Dr Bobinson’ (and of course ‘the Bobinsons’), in particular his (their) continued assistance and mothering,

Mr Alan Berzolla, in particular his unhesitating support and advice in various matters,

Dr Camilla Tooley, Mr Phill Husband, Mr Josh Hartland, Dr Jack Lambert (and Miss Kim Jarvis – so sorry things didn’t work out);

my family, who on top of this have been ever patient with me and provided me with unwavering and unconditional support. I am incredibly lucky to have them and include:

My grandparents: Pauline and James Gilmartin.

My grandma and Aunt: Jana Hanzalova and Marta Knoblockova.

My sister, brother-in-law (and niece): Bex and Jim Rowlands, Eva;

and of course those who have supported me most throughout, my parents: Anne and Jan Hanzal.

Definitions and Abbreviations

AE:	acoustic emission
AIF:	Alicona Infinite Focus (-machine/microscope)
CoF:	coefficient of friction
CRC:	c-/ring crack complex
CWT:	continuous wavelet transform
DWT:	discrete wavelet transform
EDX:	electron dispersive x-ray
EFD:	early fault development
EI:	energy index
EMD:	empirical mode decomposition
HHT:	Hilbert-Huang decomposition
IMF:	intrinsic mode functions
npf:	non-prominent fault (i.e. indeterminate)
OLS:	oil line sensor
RE:	rolling element
RMS:	root mean square
SEM:	scanning electron microscope
SRR:	slide-roll ratio
TSA:	time synchronous analysis
TSaA:	time synchronous average analysis
WP:	wavelet packet
WPA:	wavelet packet analysis
WSS:	wear site sensor
WVD:	Wigner-Ville decomposition

Chapter 1. Introduction and Overview

1.1 Background

Almost all modern mechanical engineering applications involve some rotating functions enabled by the use of bearings. Rolling element bearings are extensively used due to their low friction, high load capacity and longevity. However, because of the critical role they play, their breakdown can lead to failures of the components in which they are embedded, potentially leading to catastrophic consequences for the entire machine and beyond. The detection and avoidance of these failures are therefore of paramount importance.

The essential principle behind such bearings is that a rolling contact allows movement of one component with respect to another with lowest achievable friction and high stability. This is achieved through rolling elements which act as intermediaries between the two components. Depending on the application, these can be subject to high stresses exerted on them not only by the surrounding components but also arising internally from their motion itself. These stresses persist over the entire operation of the machine, subjecting each part of the bearing. The persistent, intensive and repetitive nature of these forces leads, over time, to material fatigue. Materials are being developed that can better resist these forces; however, new applications inevitably bring in conditions in which these forces are also increased.

One such material is silicon nitride, possessing significantly higher strength, greater stiffness, lower density, higher temperature tolerances than steel and can also be manufactured into components with smoother surfaces and to higher tolerances. Because of these properties, silicon nitride is being used more extensively in modern bearings.

These bearings utilise silicon nitride as a material for the rolling elements, but keeping the remaining parts in steel. This is because the use of silicon nitride there achieves the greatest performance gain whilst steel offers a near-perfect counterface to it due to its relative softness, and the combination of the two materials achieving significantly reduced friction. Other reasons for hybrid bearings are that silicon nitride rolling elements are considerably easier to manufacture than would be for other bearing parts (i.e. raceways) and that the thermal expansion characteristics of steel are in line with the surrounding components of the machine.

Such measures achieve both improved performance and increased lifetimes, but do not avoid the possibility of breakdown. Replacing broken down components is an inefficient and expensive way of maintaining machinery and, equally so, replacing parts that have a significant lifetime remaining. Thus modern approaches to maintenance are based on monitoring the on-going capabilities of critical components and scheduling their replacement at appropriate times before their degradation or even failure.

Because of their critical role in the functioning of the machines, a major focus for monitoring is the bearings and, more specifically, their rolling elements. A major cause for the reduction in their performance or breakdown is material fatigue.

Current failure prevention and detection systems typically rely on debris analysis utilising the magnetic properties of steel and are therefore restricted to use in magnetic materials; also, they become sensitive only at advanced stages of the failure development. In order to extend failure prevention and detection to hybrid bearing systems, novel sensing techniques must be developed. These would require sensors adapted to monitor fatigue progression in silicon nitride as its wear characteristics are different to those in steel. Associated with this is the need for new signal data processing because of the different fault signatures.

Although hybrid bearings are already produced and deployed industrially, the understanding of their RCF failures is not yet sufficiently developed; in particular, the interactions between the silicon nitride and the steel counterfaces. Recent research on hybrids focuses on physical examinations and development of models to capture the details of the fault development. The principal method of their verification is based on visual inspections taken at various points in the fault progression requiring interrupts to the test runs and extraction of the test specimens for the inspection.

A major objective of the research is therefore to improve the understanding of the signals generated in the course of a fault development and thus provide a means for their detection and continuous monitoring in a non-destructive manner. Of particular interest to this project were vibration, electrostatic and acoustic emission (AE) media, none of which have been previously applied to the monitoring of developing RCF faults in hybrids.

1.2 Project Aim

The project was set up by nCATS in order to determine whether sensor-based monitoring can provide a reliable means of detecting rolling contact fatigue (RCF) failures in silicon nitride rolling elements of hybrid bearings. Such a method, if developed into a practical application, would potentially provide huge savings to industrial machine maintenance, prevent catastrophic

failures, reduce machine downtime and potentially lead to improvements in the quality of manufactured bearings.

This would require, first of all, a detailed understanding of the various fault generating mechanisms in hybrid bearings - in particular, their signal producing characteristics that could be detected by sensors and recognised in subsequent processing.

Therefore a significant part of this project was dedicated to the selection and development of appropriate sensing techniques and for finding suitable methods to analyse and interpret the sensor data acquired.

1.3 Objectives in Detail

1.3.1 Wear Generation Aspects

Rolling contact fatigue is a type of wear generated in a material by the application of repeated and variable stress produced by the rolling motion by a stress-applying counter-surface. Such a motion can also produce other types of wear for example abrasion, however RCF poses a particular problem for bearings placing a serious limitation to their lifetimes.

The exact wear types produced depends on many different contributing factors, for example friction, and also depends crucially on the properties of the component material. A precise knowledge of these conditions would enable identification of those that are relevant to hybrid bearings.

The wear generating mechanisms produce phenomena that can be detectable over distance; a detailed knowledge of these mechanisms will be required to, for example, distinguish them from each other. Some of these phenomena are already well established - for example, vibration is widely used as a means in condition monitoring of machinery. Of particular interest to nCATS were also the acoustic and electrostatic charge generating phenomena.

The emulation of RCF generation under laboratory conditions is performed on machines involving rotating surfaces. In order to accelerate run times, defects are pre-generated using machines employing non-rolling processes (for example indentation). A detailed knowledge of these defect generating mechanisms will be required in order to determine how they fit into RCF progression.

1.3.2 Signal Processing Aspects

Signal processing in general is concerned with identifying the patterns in a stream of data that could be linked with a particular source or process of interest – a developing RCF fault in this project.

In practice it is divided into several functional stages each reflecting a different underlying type of process. Following its generation, such as in a fault, the signal is transmitted as a disturbance through a medium and detected in a sensor which converts it in to an electrical form. This is subsequently converted to digital format as it allows subsequent processing using digital techniques. The major issues encountered in this are that i) the generated signal in its pure form is unknown, ii) the medium transmits a vast range of disturbances originating from multiple sources (including background ‘noise’) and therefore producing complex patterns in the detecting sensors and iii) consequently vast amounts of data to be analysed.

To address these issues, various techniques have been developed that can be divided into those that focus on signal selection and enhancement (choosing some features or making them more prominent), and those that focus on actual signature extraction. In order to determine the most suitable techniques consideration must be given primarily to the character of the medium and the source.

In particular, for a successful signature extraction, as much information must be used as is possible on the characteristics of the source of interest. For example, a particular source could be distinguished by its time or frequency characteristics. These then in turn determine the detailed experimental conditions, set up to isolate the particular features as much as possible, or make them more pronounced.

1.3.3 Objectives

With the background presented in the subsections above, more detailed project objectives can now be stated. These are:

- To become familiar with the current understanding of RCF development and its distinguishing characteristics in hybrid bearings;
- in particular, the characteristics with which developing faults in silicon nitride disturb the surrounding media i.e. vibration, electrostatic and acoustic emission.
- To review available sensor devices for their capabilities and assess their potentials for the use on silicon nitride. If necessary (as expected), perform suitable sensor adaptations.
- To develop a prototype set up for the generation and sensing of RCF faults in hybrid bearings based on existing laboratory equipment, suitably adapted and enhanced.
- To select and further develop appropriate signal data processing methods for the analysis of faults generated in hybrid bearings.
- To develop concepts for the interpretation of the sensor data collected and perform overall assessment of the viability of the detection methods employed.

1.4 Project Approach

The approach was based on setting up a series of experiments in the nCATS laboratories for the emulation of RCF failures in hybrid bearings, which would collect vibration, electrostatic and acoustic signals for detailed analysis.

Before this could be done, the issues highlighted above had to be addressed, by an in-depth review on the research done to date on the wear mechanisms, the latest sensing technologies and sensor data analyses methods - with particular focus on hybrid contacts.

To perform the RCF fault emulation experiments, an in-house twin disc machine was chosen. This is a relatively simple machine, consisting of one contact region between two specimens (referenced henceforth as *contact*), whilst also replicating some of the complexity of an industrial machine with various noise generating components in close proximity of one another.

The machine available in the nCATS laboratories had to be first fitted with the vibration, electrostatic and acoustic sensors and corresponding electronic processing equipment. Following this a feasibility test was performed on this equipment using a hybrid contact. This identified the need for multiple enhancements required for the machine and equipment to better handle the hybrid conditions.

To accelerate the test runs, specimens were initially seeded with defects. For their controlled generation, an in-house Vickers machine was used following its review in the literature.

The modified hardware, including the twin disc and indenter machines, were then used in a series of tests to provide data that could be used for the development of the signal processing methods. These included tests to simulate faults and provide corresponding signals separately for each of the media of interest (i.e. vibration, electrostatic and acoustic). In addition, a test was performed to provide signal corresponding to normal running conditions; this was named '*Healthy Run*' and was aimed to provide background data against which the fault signatures were to be contrasted by the signal processing methods. The resulting data were used in the development of the processing of the signal processing systems for each sensor type, based on several data processing and filtering techniques and producing advanced processing techniques.

With the hardware and software systems ready, and some practical experience with their use as a result of the preparatory testing, it was possible to plan in detail and subsequently execute the main experiments. In order to distinguish from the proceeding preparatory tests, these were named as '*Production Runs*'. The data collected were then processed using the developed techniques and then analysed in several graphical data presentations set up specifically for this

project to reveal the information of interest. The results of the analyses were followed by an evaluation of the project as a whole and presenting recommendations for future work.

1.5 Structure of the Document

This thesis is structured in nine chapters, detailed as follows:

Chapters 2 & 3 concern the reviews of the literature on the two main topics, respectively the relevant aspects of hybrid contacts (general wear mechanisms and specific rolling contact fatigue) and the sensing techniques (the technologies and the processing/analyses methods).

Chapter 4 presents the details of the hardware used in this study. This concerns the instrumentation of the indentation and Twin Disc machines with sensors and the electronic data processing systems, and subsequent modifications that were performed to them following an initial feasibility test. It also describes the physical inspection tools that were used for direct inspections of the specimens.

Chapter 5 presents the details of the signal processing development, based on the initial scoping tests and subsequent Healthy Run.

Chapter 6 presents the planning of the Production Runs, including parameter settings that were used to produce the RCF data for subsequent analysis.

Chapter 7 presents the results from a detailed analysis from test samples using physical inspection tools.

Chapter 8 presents a detailed analysis of the sensor signal data produced during the Production Runs.

Finally, **Chapter 9** provides the conclusions of the project and presents recommendations for further work.

Less critical information relevant to this study has been provided in the **Appendices** – for example, details of the initial equipment test, details of calculations and computations, etc.

Chapter 2. Review of Background Tribology

This chapter provides an in-depth review of the background tribology and hybrid bearing systems. Each of these topics is presented in a separate section.

2.1 Rolling Element Bearing Tribology

2.1.1 Introduction

A rolling element (RE) bearing assembly consists of four elements: outer race, inner race, rolling elements and a cage. Of the many types available (for example cylindrical RE bearings), ball bearings are a common choice as they are able to provide sufficient load bearing capacity to a wide range of applications and are relatively inexpensive. Figure 2.1 illustrates the components of two ball bearing designs: a deep groove bearing (supporting dominant radial loads) and an angular contact bearing (supporting both radial and axially applied loads).

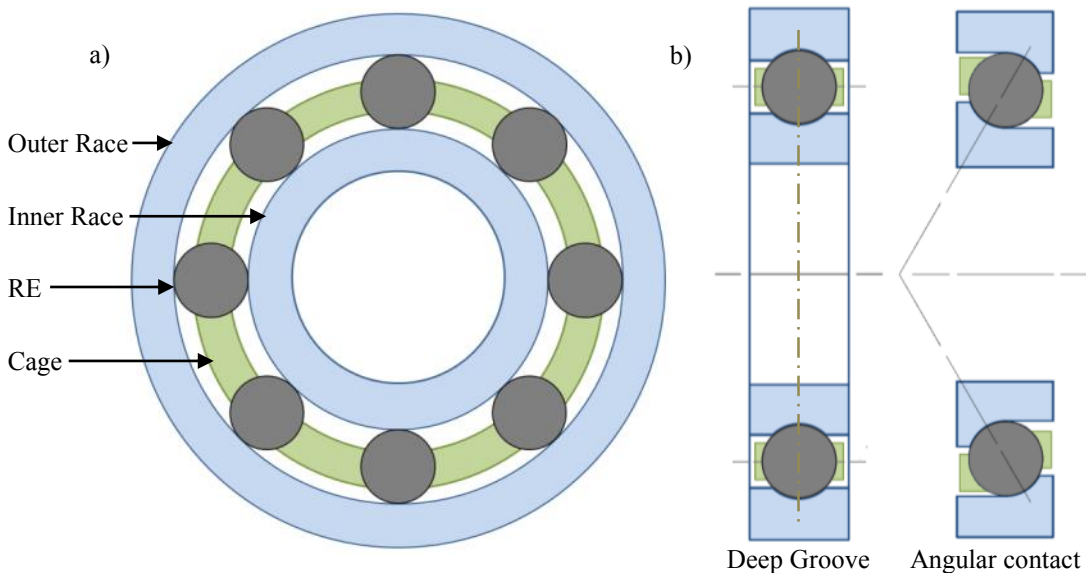


Figure 2.1: Cross sections of a RE bearings a) front on view and b) Side on views of two common bearing designs

Although relatively simple in their componentry, they represent a variety of complex underlying physics characterised by cyclic stresses, friction, surface deformation and wear. This section is split into two subsections; the first part introduces some of the fundamentals of Hertzian contact stress analysis; the second introduces the fundamentals of dynamic contacts including friction, lubrication and wear.

2.1.2 Hertzian Contacts

When two curved surfaces are brought into contact they form a concentrated area of contact. The application of load causes an elastic deformation over a finite area and a stress distribution. In 1881, Hertz described the dimensions and pressures of contacting bodies with a force applied

between them. His analysis made some critical assumptions, notably: that the contact area is small compared to the curved surface, the bodies are perfectly smooth and frictionless and each undergoes ideal elastic deformation. In bearings whose components are generally smooth and which operate under loads within the elastic confines of the component materials, the contact stress fields follow closely the Hertzian conditions [2].

The Hertzian equations describing the contact dimensions and pressures for both spherical and elliptical contact shapes are summarised in Table 2.1 below; their meanings are explained in Appendix A.

Parameter	Spherical Contact	Elliptical Contact
Contact radius a Equivalent contact radius c	$a = \left(\frac{3PR_c}{4E^*} \right)^{\frac{1}{3}}$	$c = \sqrt{ab} = \left(\frac{3PR_c}{4E^*} \right)^{\frac{1}{3}} F$
Maximum contact pressure p_{max}	$p_{max} = \frac{2P}{2\pi a^2}$	$p_{max} = \frac{3P}{2\pi c^2}$
Mean contact pressure p_0	$p_0 = \frac{2}{3} p_{max}$	

Table 2.1: Hertzian contact equations for spherical and elliptical contacts [3, 4].

The contact and areas and pressures are used to calculate the resulting stresses in the material. The stress vector field are based on suitable coordinate systems (typically cylindrical) from which other stresses can be derived, in particular: σ_1 (tensile), σ_2 (hoop), σ_3 (normal). Their advantages are that $\sigma_1 > \sigma_2 > \sigma_3$ is true in most practical situations [5]. Of particular relevance to bearing RCF failures are the principal tensile σ_1 and the shear stress $\tau_{12} = 0.5(\sigma_1 - \sigma_3)$. These stress fields are illustrated in Figure 2.2 below.

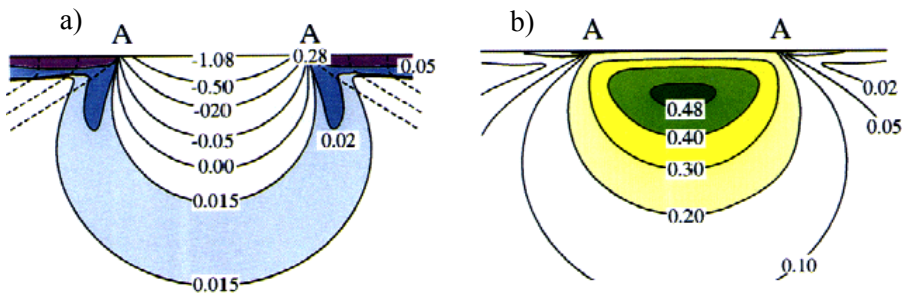


Figure 2.2: The profile of principal stresses response for a circular elastic contact a) σ_1 principal tensile stress and b) principal shear stress τ_{12} . Both have been normalised using p_0 :the mean contact pressure [5].

The tensile stresses concentrate at a shallow surface region located at the periphery of the contact (indicated A in Figure 2.2a). The maximal value σ_m can be calculated using the following equation:

$$\sigma_m = 0.5 (1 - 2\nu)p_0, \quad 2.1$$

where ν the Poisson's ratio of the material.

In contrast to the above, the maximum shear stress occurs at a region below the contact centre at a distance of approximately $0.4a$ from the surface with a magnitude of $0.48p_0$.

At a critical load, the stresses induced by the contacting bodies will result in irreversible material changes. Two dominant general phenomena are described: plastic deformation and fracture.

2.1.2.1 Plastic Deformation

For ductile materials, such as those typically used in bearing steels, plastic deformation will initiate beneath the contact at the location of maximum shear stress¹. At this onset, the surrounding region remains elastic and subsequently the changes in the contact shape are therefore small. If the applied load is increased further, the plastic zone grows until eventually it extends to the free surface, reaching a state of full plasticity.

For a material with a yield stress Y positioned in a state of uniaxial load (i.e. $\sigma_1 = Y$ and $\sigma_2 = \sigma_3 = 0$), the critical shear stress τ_m for the onset of plastic deformation has been described by Tresca (see e.g. [6]) to occur at half the difference between the principal stresses. This equates to:

$$\tau_m = \frac{1}{2}Y \quad 2.2$$

The maximum contact pressure at this onset is therefore calculated as follows [3]:

$$p_{max} = 1.6Y \quad 2.3$$

The yield stress can be estimated from the material hardness determined from Vickers hardness tests using the relationship $\text{Hardness} \cong 2.7Y$ [3], and subsequently yield will occur at a mean contact pressure described in the following equation [3]:

$$p_{0yield} \cong 0.4 * \text{Hardness (GPa)} \quad 2.4$$

Bearing steels typically exhibit hardness values around 800 Hv (approximately 7.85 GPa²), corresponding to a p_{0yield} value of 3.1 GPa.

2.1.2.2 Fracture

For a brittle material, such as silicon nitride, fracture occurs at the maximum tensile regions at the periphery of the contact (see Figure 2.2a). Ichikawa et al [8] found experimentally that the

¹ The minimum stress required to reach the condition of plastic deformation is denoted as the *yield* stress.

² To convert Vickers hardness (Hv) to GPa multiply by 0.009807 [7].

(maximum) contact pressure to induce cracking for circular contacts was approximately 16 GPa for a typical bearing-grade silicon nitride. Using equation 2.1 the corresponding tensile stress for fracture is calculated as follows [9]:

$$\sigma_{mT} = \frac{(1 - 2\nu)p_{max}}{3} = \frac{(1 - 2(0.28)) * 16}{3} \approx 2.3 \text{ GPa}$$

For circular contact geometry, a surface ring crack forms following the contours of the circumference of the contact resulting in a '*ring crack*'. With increasing load, the crack propagates into the material flaring outward into a truncated cone configuration (following a tensile *mode I* stress field), with a profile similar to that illustrated in Figure 2.3a. This is known as a '*cone crack*'.

If the load continues to increase the circumference of the contact engulfs the initial ring crack forming secondary ring cracks with larger radii [10] as shown in Figure 2.3b.

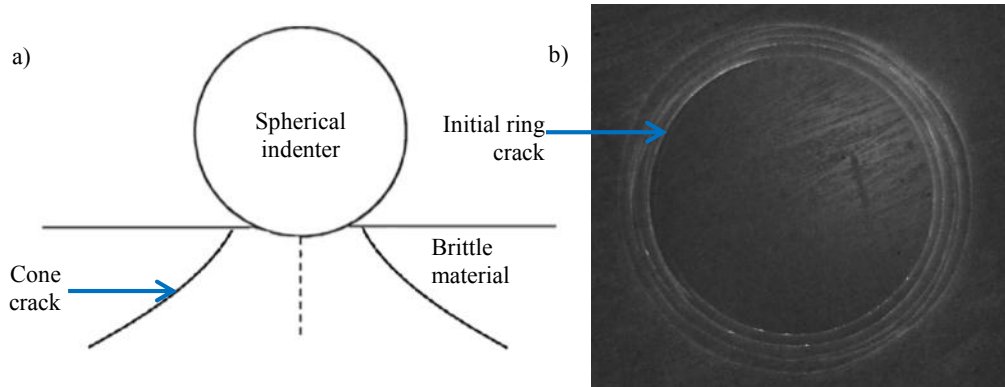


Figure 2.3: a) Subsurface profile of a cone crack b) concentric ring cracks in silicon nitride [11].

2.1.3 Dynamic Contacts

Contacting bodies with a relative motion between them give rise to the phenomena of friction and wear and the use of a lubricant to reduce them becomes important. These are briefly described in turn in the following subsections.

2.1.4 Friction

Friction is the resistance force encountered when surfaces in contact move relative to one another. For rolling contacts, friction can be dominated by processes including micro-slip, elastic hysteresis, plastic deformation and adhesion within the contact [12].

For RE bearings under normal operation friction is dominated by micro-slip [13]. This is produced as raceways elastically distort to accommodate the REs, resulting in a differential movement between the surfaces of the REs and raceways - known as *Reynolds slip*. For

spherical and elliptical contacts (i.e. ball bearings), slip is further induced due to the differences in the distance travelled between various points of a RE [14].

Significant slip can be generated during transients such as start-ups and variable loads. For angular contact RE bearings, the slip between the raceways and REs increases during high speed operations where centrifugal and gyroscopic effects on REs become more prominent.

Slip can have a significant detrimental effect on the component interfaces, by increasing wear rates and promoting surface-initiated failures. The contribution of slip in a contact for bodies with surface speeds U_1 and U_2 can be calculated by the slide-roll ratio (SRR) as follows:

$$SRR = \frac{(U_1 - U_2)}{(U_1 + U_2)} \quad 2.5$$

The magnitude of the SRR produced at RE-raceway interfaces is difficult to measure in practice. For steel angular contact ball bearings, an estimate of 2% was provided in an experimental set up [14] using an applied maximum contact pressure of 2.5 GPa and a rotating inner race speed of 1,500 RPM.

2.1.5 Lubrication

To minimise friction, bearing surfaces are separated by imposing an intervening layer of lubricant that also forms a protective film. Additionally, a lubricant can assist in thermal regulation of the contact and removal of wear debris.

A measure of the degree of separation of the two interfaces by the lubricant is provided by the Lambda (λ) ratio. Using Equation 2.6 below, this parameter compares the surface roughness (R_{a1} , R_{a2}) of the two opposing interfaces and the minimum film thickness of the lubricant between them h_{min} :

$$\lambda = \frac{h_{min}}{\sqrt{(R_{a1}^2 + R_{a2}^2)}} \quad 2.6$$

The minimum thickness h_{min} can be calculated using Equation 2.7 below, derived famously by Dowson and Higginson [15]:

$$\frac{h_{min}}{R} = 1.79(2\alpha E^*)^{0.49} \left(\frac{U\eta}{2E^*R} \right)^{0.68} \left(\frac{W}{2E^*R^2} \right)^{-0.067} \quad 2.7$$

(for the meaning of the parameters refer to [15])

The influence of the λ ratio on the coefficient of friction (CoF) is shown in Figure 2.4 below.

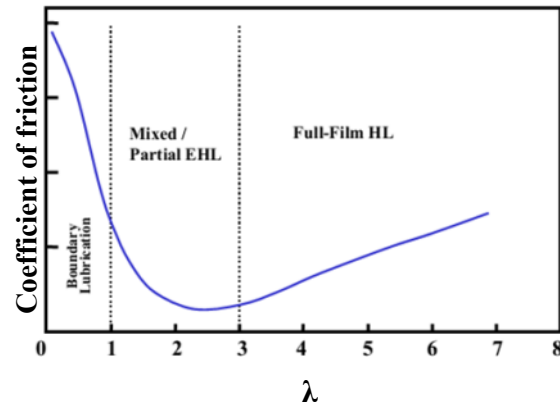


Figure 2.4: The Stribeck curve.

Three elastohydrodynamic lubrication regimes are identified based on the λ value and are now described:

Full-film ($\lambda > 3$): surfaces are fully separated by a hydrodynamic lubricating film as the pressures in the film carry the load; correspondingly, wear is minimal.

Mixed (EHL) ($1 < \lambda < 3$): some interaction between the interfaces occurs as asperities protrude beyond the minimum film thickness, resulting in fracturing and plastic deformation of asperity tips.

Boundary ($\lambda < 1$): severe contact between the interfaces occurs producing high CoF, forming localised stress concentrations at the surface which can lead to the onset of severe wear regimes. RE bearings can experience boundary lubrication during start-ups and shut downs, and at the onset of failure.

Typically bearings operate at λ values between 2.5 and 4 [14] where CoF exhibits a minimal value.

2.1.6 Wear

Wear is the process of material loss from two surfaces that have relative motion between them. There are generally four basic mechanisms of wear namely: abrasion, adhesion, tribochemical and RCF. Bearing systems can exhibit more than one mechanism at the same time, particularly at the onset of failure. The following sections discuss some of the characteristics of these mechanisms in turn.

2.1.6.1 Abrasion

Abrasive wear involves the removal or displacement of material from a surface by hard particles or protuberances present on the counterface. The method of abrasion can generally be separated into: two-body, and three-body abrasion. For two-body, the resultant abrasive wear is caused by hard surface abrading a softer counterface. For three-body, entrainment of hard particles

between the contacting faces can abrade both faces. Impingement of hard protuberances or particles into counter face(s) can result in various processes including: ploughing, wedge formation and cutting [16] and indentation.

2.1.6.2 Adhesion

Adhesive wear is the localised welding of asperities or protuberances from each of the contacting elements. The asperities of two mating surfaces in relative motion contact with each other and plastically deform. With high localised contact pressures and flash temperatures, asperities can weld together. Relative motion between welded asperities causes plastic shearing of the junctions, resulting in the tips of the softer asperities to be plucked off. A broken asperity tip can fully adhere to the harder counter-face or be ejected as wear debris.

For bearings operating in the EHL regime this wear mechanism is generally avoidable. However, as the lubricant film breaks down and asperities interact, contact flash temperatures can increase resulting in a further reduction in lubricant film thickness. This forms the basis of a positive feedback mechanism and *thermal runaway*. Initially, the combination of high temperature asperity interaction and lubricant quenching can lead to martensitic transformations at the surface forming ‘*white layers*’. Here, localised micro-hardened areas can temporarily reduce the wear rate, but the presence of micro-cracks eventually results in delamination and severe loss of material.

2.1.6.3 Tribochemical

Tribochemical wear involves a coupling between mechanical and chemical processes that occur at the interface and with the surrounding medium. Chemical reactions between the environment and a surface produce a surface film, producing a change in mechanical properties from the bulk material. If the film is more readily worn away than the bulk material, the layer can increase the wear of the surfaces through debris entrainment. The formation of the film can be accelerated by high contact temperatures and pressures.

In contrast to the above, a tribofilm can form which produces lower wear rates than that of the bulk.

2.1.6.4 Rolling Contact Fatigue

Rolling contact fatigue (RCF) can be defined as the mechanism of crack propagation caused by the near surface alternating stress field within the rolling contact bodies, which eventually results in material removal [16]. Various failures can be induced from RCF; however the mechanism generally involves the following characteristics:

- Stress concentrations either at the surface or subsurface

- Crack initiation either at the surface or subsurface due to stress concentrations
- Crack propagation due to cyclic loading
- Spalling or pitting - leading to the production of debris
- Gross material loss and failure of the contact interfaces

The exact mode of failure can differ substantially from failure to failure, as a number of wear mechanisms compete and are affected by complex local tribology.

Generally, the RCF failure modes are separated into subsurface or surface initiated failures.

i. Subsurface Initiated Failures

For bearing surfaces that are smooth, separated sufficiently by a lubricant film and loaded below the elastic limit, the stress field follows closely to that of Hertzian conditions [2]. Under these conditions, principle shear stress is maximal some depth below the surface (as discussed in Subsection 2.1.1). Stress raisers in the material will interact with applied stresses during over-rolling processes such that cracks initiate and propagate under mode 2 (shear) stresses.

Cyclic stresses from repeated over-rolling propagate cracks towards the contact surface, eventually leading to portions of the undermined material to be liberated from the surface and the formation of a spall.

ii. Surface Initiated Failures

Surface originated failures are initiated at regions of localised surface stress, which can include asperities, dents, pre-crack or pores. Cracks will initiate and propagate under mode 1 (tensile) stresses into the material. As mode 1 tensile stresses decrease with depth, subsurface shear stresses begin to dominate, turning the crack parallel to the surface. Eventually, a liberation of the undermined material will form a spall.

For surface originated fatigue, failure can be accelerated by penetration of the lubricant in the crack and a reduction of the friction between crack faces and development of hydrostatic pressures at the crack front through fluid entrapment.

For both initiated failure modes, if the liberated material is of geometry comparable to the film thickness, over rolling of the resultant debris can induce further damage downstream of the spall.

When a spall is produced, the changes in topography at the location lead to increases in contact stress to form a positive feedback mechanism and progressive wear. This can lead to rapid deterioration of the contact resulting in increased heat and vibratory loads on the bearing.

2.2 Hybrid Bearings

2.2.1 Introduction

Hybrid bearings are increasingly used in a variety of high-end applications providing high performance and lifetimes to extreme operating conditions. One of their major attractive properties are the high rotational operating speeds they can endure making them particularly attractive to the gas turbine [17-19], spacecraft [20], racing and high speed tooling [21] industries. Additionally, their low friction characteristics reduce energy losses [22] and enable the use of solid lubricants for high temperature and vacuum conditions [23] [24], where heat convection is restricted. This section discusses the properties of silicon nitride and reviews its wear characteristics in a variety of wear contacts.

2.2.2 Silicon Nitride

Silicon nitride provides a range of attractive properties for their replacement of steel REs, some notable examples include:

- low density (40% less than steel);
- low friction coefficients (0.15 silicon nitride-steel dry sliding [25] compared to 0.54 for steel-steel),
- high hardness (up to two times that of common bearing steels);
- mechanical strength at high temperatures (up to 1000°C [22]).

Due to a combination of material properties, silicon nitride has been chosen above that of other potential technical ceramics, such as zirconia, in particular: high fracture toughness, economic viability in component fabrication and the ability to avoid catastrophic modes of failure.

2.2.2.1 Fabrication

Two general methods are available to manufacture silicon nitride material namely: reaction bonding and sintering methods.

Production by reaction bonding (RB) involves the pressing of silicon powder into the desired shape then heating to around 1350 to 1450 °C in a nitrogen rich atmosphere. At these temperatures reaction bonding occurs as the compacted silicon powder is nitrided by the gaseous nitrogen. The method forms a sample which requires little or no further machining. However, silicon nitridation is a highly exothermic reaction which can cause localised melting of the silicon, reducing local surface area exposed to nitrogen producing a material with relatively high porosity rated at around 25-35% [3].

Sintering methods generally involve the binding of raw silicon nitride powder under high temperatures, through the addition of a few weight percent (typically up to 5%) of sintering

additives, commonly MgO, Y₂O and Al₂O₃. The sintering additives form an intergranular glassy phase between the grain structure, assisting in the production of a material with low void content and high density.

Pressure applied to the material during binding processes can also aid to achieve higher densities, whilst producing fewer sintering flaws. Hot isostatic pressing (HIPing) applies temperatures and pressures of around ~1700°C and 300 MPa respectively during binding - achieving a near fully dense material with minimal further machining required [26]. HIPing processes can produce components of high and consistent quality, achieving high strength above that of other sintering methods [27]. Subsequently, the method is used in preference above that of other techniques to produce bearing components.

The finishing of silicon nitride elements is an important aspect which can greatly affect the performance of the subsequent component. Due to the high hardness and relatively low fracture toughness of the material, the finishing method requires careful attention and alteration from the methods used in finishing equivalent steel components. Subsequently, material is removed at a very slow rate to avoid inducing damage [28] and imparting residual stresses which can significantly reduce the lifetime of bearing components [29]. Currently, diamond abrasives in a water or kerosene based solutions [30] are used to finish elements and the adoption of magnetic float polishing techniques have been suggested to further reduce inducing damages [31]. It is estimated therefore, that up to 50% of the cost of ceramic bearings are attributed to the finishing of the ceramic elements alone [32].

2.2.2.2 *Microstructure*

Silicon nitride is sintered from a powdered form consisting of silicon nitride predominately in the alpha crystallographic form. Subsequently, powder grains are typically spherical with sizes of approximately 0.5 microns in diameter. During the binding processes, the acircular alpha phases transform to the beta crystallographic phase producing elongated grains between 1 and 9 µm in length.

This transformation is dependent on both the sintering temperature and time, and subsequently the microstructure of the material can be tailored to accommodate both phases; producing a material with an advanced microstructure. A silicon nitride consisting of a both material phases is shown in Figure 2.5 below.

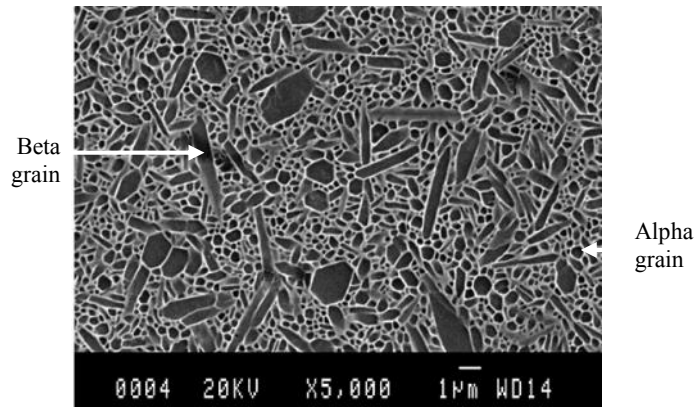


Figure 2.5: SEM image of a silicon nitride material consisting of both Alpha and Beta grains [27]

The grain size characteristics of the material can significantly influence its properties. This is demonstrated by the comparison of three silicon nitride materials with different grain size characteristics in Table 2.2 below.

Material label (reflecting its grain size)	Average grain size (μm)	Sintering temp ($^{\circ}\text{C}$)	Beta/ Alpha ratio	Fracture Toughness ($\text{MPa}\cdot\text{m}^{1/2}$)	Tensile Strength (MPa)	Hardness (GPa)
Fine	β : 5, 0.4 α : 0.4	1570	0.2	3.9 (± 0.3)	885 (± 85)	21.0 (± 0.7)
Medium	β : 4, 0.5 α : 0.5	1650	3.5	5.3 (± 0.4)	1084 (± 62)	17.9 (± 0.4)
Coarse	β : 9, 1.5	1800	100	4.5 (± 7)	792 (± 32)	15.8 \pm (1.1)

Table 2.2: The properties of the three materials used in the study by Lee, Wuttiaphan et al [33]

The investigation, performed by Lee, Wuttiaphan et al [33], demonstrates various pay-offs between the microstructure; for example, larger grain lengths produces a material with a high fracture toughness³, but a corresponding loss in hardness results. A microstructure that provides optimal properties for bearing performance is therefore an important consideration.

The influence of grain structure on wear characteristics has been investigated in a number of studies including for example [27, 35-37]. Generally, a finer grain structure comprising of grain sizes of approximately 2 μm in length has been found to demonstrate higher fatigue lives in rolling wear [35] Whilst the presence of the beta phases and an interlocking grain structure is thought to reduce grain pull-outs and impede crack development [37] [27].

The sintering additives used in fabrication also affect the material properties. To achieve a highly dense material, sintering additives will invariably be used forming a glassy interphase between the grains. This can play an important role on the material properties including wear performance [38, 39]. Chen, Cuneo et al [40] reported a visco-elastic response of the intergranular phase in contact fatigue tests on silicon nitride balls. During loading- unloading

³ The material demonstrated significant R-curve toughening behaviour producing a material forming micro-cracks under shear [34].

cycles, the grain boundary interphase was found to exhibit both an immediate (elastic) and delayed (viscous) response, producing a tensile stress field at the grain-interphase boundaries.

Despite this, the interphase is necessary to producing a material with high performance wear characteristics. Braza et al [41] demonstrated this by comparing the performance of two silicon nitride materials in lubricated rolling contact tests, one material was produced through reaction bonding, the other through sintering. The reaction bonded material, with limited interphase, produced significantly higher wear rates than the latter as a result of intergranular fracture and grain pull-outs.

The fabrication route for silicon nitride production is also an important aspect that can determine material wear properties, Zhou and Wu [42] demonstrated the significant reduction of fatigue life in rolling wear for a sintered silicon nitride in comparison to a HIPed material. This was owed heavily to the reduction of pores and sintering defects. The content of sintering elements has also been observed to affect the wear properties of the resultant material [38, 39].

Currently, three standard silicon nitride materials have been developed by bearing manufacturers for RE components, these are: *NBD200* (produced by Norton), *SN101C* (produced by Saint-Gobain) and *TSN03-NH* (produced by Toshiba). Their characteristics are summarised in Table 2.3 below.

Property	NBD200	SN101C	TSN-03NH
Density (g.cm^{-3})	3.16	3.21	3.22
Elastic Modulus (MPa)	322	307	309
Poisson's Ratio	0.27	0.26	0.28
Vickers Hardness (Hv)	1550	1600	1550
Tensile Strength (MPa)	1000	1000	900
Fracture Toughness k_{1c} ($\text{MPa.m}^{1/2}$)	5.5	6.5	6.7
Sintering Aids	MgO	Y_2O_3 , Al_2O_3	Y_2O_3 , Al_2O_3

Table 2.3: Material properties of three bearing grade silicon nitride materials typically used by the industry (data provided by the manufacturers)

Correspondingly, the RCF lives of silicon nitride REs typically surpass that of their steel counterparts. However, owing to the brittleness of the material, silicon nitride bearing elements are susceptible to fracturing on their surfaces. Hadfield [43] was one of the first to report the existence of surface cracks (*'pre-cracks'*) on silicon nitride bearing components - which are produced even before their insertion in the bearing assembly itself. These have been found to significantly reduce bearing lifetimes. Their characteristics are now described in the following section.

2.2.2.3 Material Defects- Surface Pre-Cracks

Surface pre-cracks are generated during the manufacturing and handling stages of ball production, for example during lapping phases where collisions between REs or other surfaces occur. They are difficult to detect during large scale production requiring specialist techniques not currently set up for mass production lines. Two methods currently being adapted for this role are based on ultrasonic methods [44] and optical coherence tomography [45].

Two main types of surface pre-cracks are reported, both of which can substantially reduce bearing lifetimes, namely:

1. ‘*Star cracks*’
2. ‘*C-cracks*’

The characteristics of each are discussed individually below.

i. Star Cracks

Star cracks are characterised by the presence of straight surface cracks emanating from a central region. A snapshot of star crack found on a silicon nitride RE is shown in Figure 2.6a below.



Figure 2.6: Snapshot of a Star crack under fluorescent light [43].

Because of their similarity to the crack structure produced by Vickers indentations in brittle ceramic materials, Vickers indentations have been used to replicate star cracks for RCF investigations (for example [46]). Therefore, the cracks produced from Vickers indentations are described below.

Figure 2.7 below shows the typical crack structures produced from Vickers indentations made in a brittle material. They consist of surface radial cracks emanating at the four corners of the indent and a lateral crack that resides below the surface.

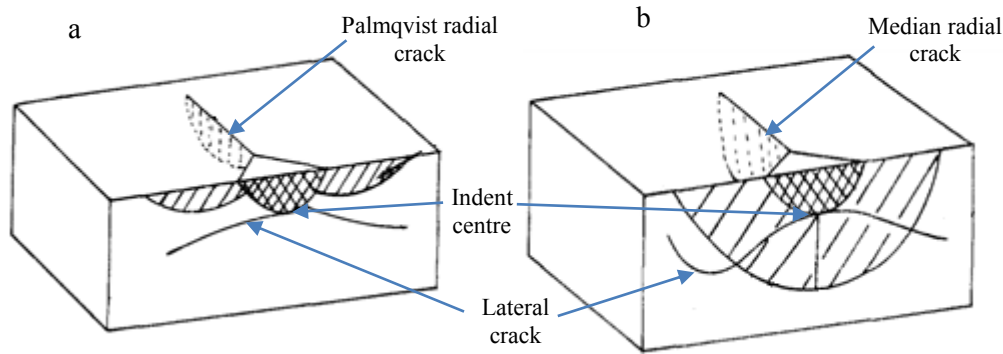


Figure 2.7: An illustration of the typical structure of sectioned Vickers indentations produced in brittle materials with a) Palmqvist radial cracks and b) median radial cracks.

The radial cracks are produced during indentation where at a critical load, a transition from elastic deformation to that of fracture and the formation of ‘*Palmqvist*’ radial cracks (see Figure 2.7a). This occurs simultaneously across each of the major diagonals of the indenter - corresponding to regions where tensile stresses are highest.

Progressive loading causes these cracks to extend both radially outward and into the material, eventually meeting its diametric counterpart producing a ‘*median crack*’ as shown in Figure 2.7b. Niihara et al [47] developed a technique to distinguish between Palmqvist and median cracks by a comparison of the average lengths of the radial cracks (l_a), and diagonals of the indent ($2a$). Palmqvist cracks were found to exhibit measurements described by the following equation:

$$l_a \leq 1.25a \quad 2.8$$

The lateral crack is produced during the unloading of the indenter, where a system of cracks form close to the tip of the indenter and propagate under the action of a residual stress field [48]. The lateral crack can extend toward surface of the material as illustrated in Figure 2.7b. For harder materials and higher indentation loads, the extent of the lateral cracking is larger and correspondingly its tendency to meet with the surface of the material increases [48]. This can lead to a flaking off of the undermined material.

For RCF failure investigations, Hadfield et al [49] emulated star cracks using an indentation load of 5 kg, whilst Kida et al [46, 50, 51] applied a force of 20 kgF, before removing the lateral crack through polishing. Subsequently, the radial lengths of the cracks produced ranged from 140 to 250 μm .

ii. C-Cracks

C-cracks are characterised by a semi-circular crack typically extending 1/4 to 1/3 of the circumference of a circle [52]. A snapshot of c-crack found on a silicon nitride RE is shown in Figure 2.8a below.

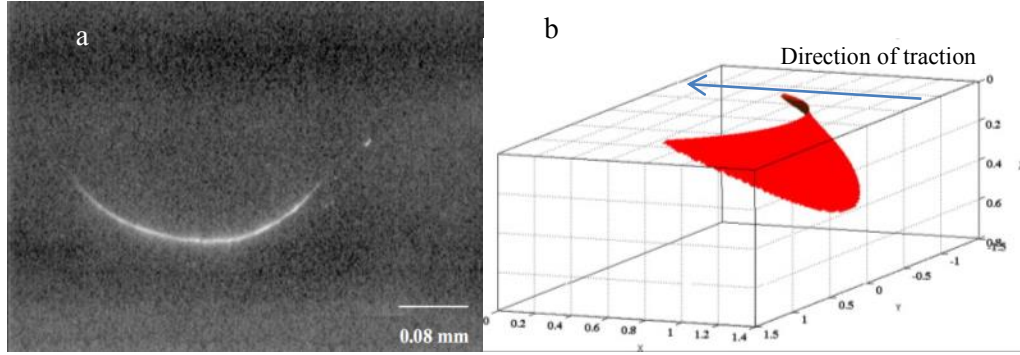


Figure 2.8: SEM snapshot of a C-crack found on silicon nitride [53]. b) illustration of a subsurface profile of a c-crack [9]

C-cracks are considered to be the most common pre-crack type found in silicon nitride [54] and therefore potentially most limiting to hybrid bearing lifetimes.

They have been emulated for RCF investigations (for example [55]) using a spherical indenter usually of the same material or harder impacted against the surface of the specimen at an oblique angle [56], [44]. The mechanics of generation is similar to that of the ring cracks described in section 2.1.3 with a crack forming at the periphery of the contact under high tensile stress. However, their semi-circular characteristic is owed to surface traction during impact/indentation which is produced from loadings made at an oblique angle [9]. The traction increases the tensile stress at one side of the contact, opposite to the direction of traction (as illustrated in Figure 2.8b), subsequently promoting the production of a crack at this region.

The sub-surface propagate characteristics are similar to cone cracks following a mode I (tensile) propagation path in the manner shown in Figure 2.8b. The maximum tensile stress of a contact with traction can be calculated using equation 2.9 below [9]:

$$\sigma_{t \max} = p_{\max} \left(\frac{\pi(4 + \nu)f}{8} + \frac{(1 - 2\nu)}{3} \right) \quad 2.9$$

Where f is the traction coefficient.

Typically, c-cracks are found to exhibit radii of approximately 190 to 270 μm or 1/40 of the diameter of the hosting ball and extending approximately 50 μm in to the surface [52].

2.2.3 Wear Characteristics

Chen et al [40] demonstrated that under cyclic loading silicon nitride undergoes three stages of wear and characterises the wear behaviour of silicon nitride in various systems. The stages were noted as follows:

1. Surface roughening, a process driven through combination of surface traction and tensile stresses at the edges of Hertzian contact, leading to intergranular fracture at the glassy interphase (illustrated in Figure 2.9).
2. Crack initiation and propagation at the periphery of the contact.
3. Flaking off of material from the surface as a result of cracks produced in (2) through the combined action of traction.

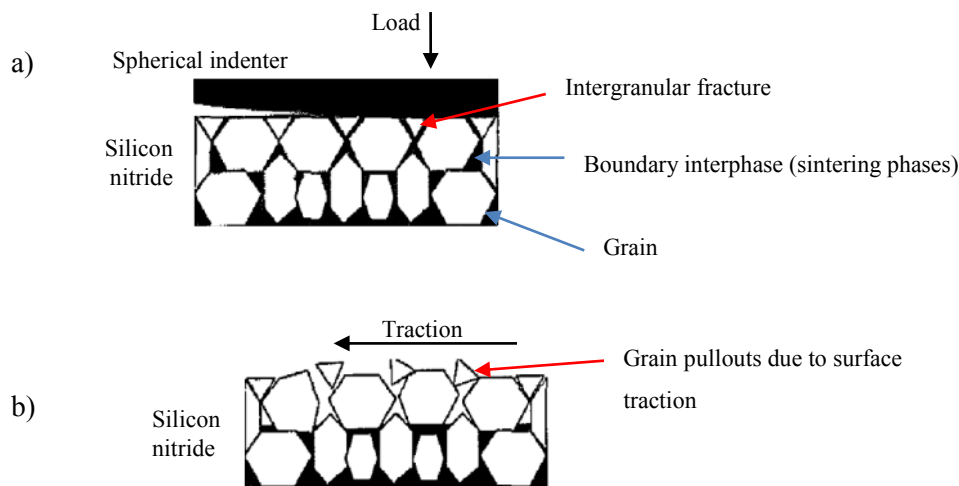


Figure 2.9: Development of grain pull-outs a) fracture of the interphase due to cyclic fatigue, b) removal of grains as a result of traction.

The characteristics of hybrid tribological systems are now reviewed for both sliding and rolling wear regimes.

2.2.3.1 Hybrid Sliding Wear

Two wear regimes are generally recognised in ceramics under sliding regimes: mild wear - producing wear rates below $10^{-6} \text{mm}^3/\text{Nm}$ [57], and severe wear - producing higher wear rates. During mild wear, tribo-chemical reactions take place to form a protective and smooth tribofilm consisting of oxides of silicon, an example is shown in Figure 2.10. The production and removal of this tribofilm is dependent on factors including humidity [58], temperature [39] sliding speed and contact pressure [59]. The tribofilm wears through micro-fracture [60] and abrasion producing fine wear debris smaller than the grain size by one-tenth or less [57].

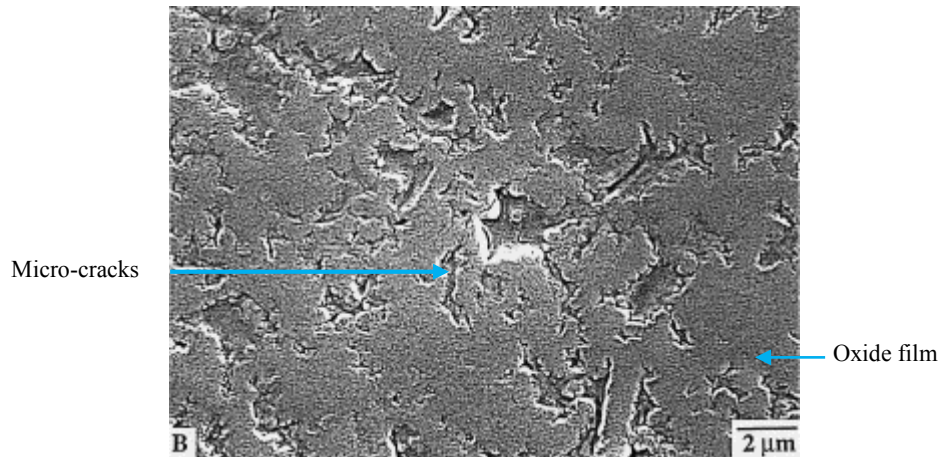


Figure 2.10: Transgranular surface film with microcracks developed after 100hrs of testing [61].

Bashansky [45] observed a transitional regime in hybrid dry contacts in which micro-abraded steel formed a protective layer of oxidised steel on the silicon nitride surface. Similarly this was produced under lubricated hybrid sliding by Akdogan and Stolarski [62], where a metallic tribofilm formed on the ceramic ball surface at low speeds (0.5 m.s^{-1}) and contact pressures ($<2 \text{ GPa}$).

Wang and Wood et al [63], investigated the performance of various steels against silicon nitride using a Pin on Disc machine at a high sliding velocity (7 ms^{-1}) and under a contact pressure of 2.5 GPa . Running in was characterised by polishing wear in which exposed silicon nitride grains were either fractured, pulled out as a whole or partial grain, these processes are illustrated in Figure 2.11. Transgranular cracking was also observed in the silicon nitride specimen, where clusters of grains were plucked out leaving pits in the surface of the silicon nitride pin as shown in Figure 2.12. These clusters were found to produce three-body abrasion, creating local plastic deformed regions in the steel specimen and cracks in the silicon nitride specimen.

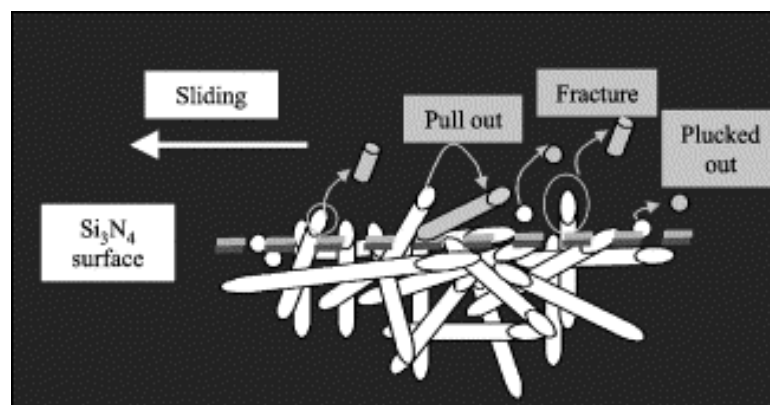


Figure 2.11: An illustration of various material damages in silicon nitride in a sliding contact. Random distributed Si_3N_4 grains near the surface are either fractured; pulled out as a whole/partial grain or plucked out as a whole grain [63].

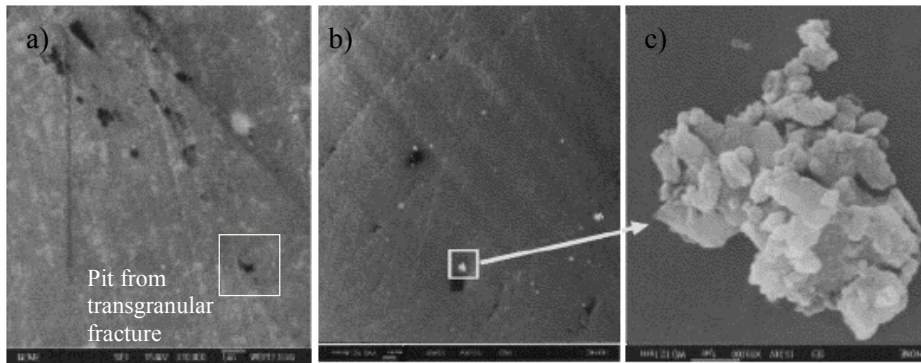


Figure 2.12: SEM of Silicon nitride wear surface a) pits on the worn surface b) debris on the surface and c) close up of debris (a cluster of finer/sub-micron particles) [63]

Denape et al [64] investigated the friction and wear for hybrid lubricated sliding tests performed on a Pin on Disc machine. Using a roughened silicon nitride pin ($>0.3 \mu\text{m Ra}$) contacting a common bearing steel disc (AISI 52100), the lubrication regime was found to be initially in the boundary regime. Subsequently, prominent asperities of the ceramic specimen abraded the steel forming grooves in the steel track. Fracturing of ceramic asperities produced silicon nitride debris which subsequently acted as a polishing paste - smoothing both interfaces and reducing the wear rates substantially to a mild and steady regime.

2.2.3.2 Rolling wear

Rolling contact wear of silicon nitride was studied on a twin disc machine under dry contact by Akazawa [65] and Asada et al [66]. Mild and severe wear modes were observed with wear rates of the silicon nitride specimen in the order of 10^{-8} and $10^{-6} \text{ mm}^3.\text{Nm}^{-1}$ respectively. Using a similar test machine Braza et al [41], investigated hybrid lubricated contact with three different silicon nitride materials contacting a cast iron roller. Wear in the silicon nitride material produced by reaction bonding techniques was characterised by grain pull outs and intergranular fracture.

Chao and Lakshminarayanan et al [67] using a Ball-on-Rod machine performed lubricated (using Exxon Turbo 2380) hybrid rolling wear tests at high contact pressures between 3 and 6.5 GPa. Wear rates were found to be approximately linear with rolling distance for both steel (balls) and silicon nitride (rod) samples. For tests involving rough silicon nitride samples ($0.18 \mu\text{m Ra}$), plastic deformation was observed at the surfaces of the (AISI 52100) steel balls where metal was squeezed outward forming folded metal layers at the edges of the contact. Comparatively, steel samples contacting polished ($0.09 \mu\text{m Ra}$) silicon nitride exhibited smooth wear tracks, reducing the wear rates in the steel substantially (2.8×10^{-5} compared to $6.6 \times 10^{-7} \text{ mm}^3.\text{Nm}^{-1}$).

In a similar set-up Chao and Lakshminarayanan et al [61], investigated silicon nitride balls (R_a $0.005 \mu\text{m}$) rolling on an as-ground (R_a $0.18 \mu\text{m}$) silicon nitride rod. A high initial wear rate in

both elements was found which decreased exponentially with the number of loading cycles. For tests involving rods with polished surfaces (to R_a 0.05 μm), a delayed wear response in both elements was observed. Here, the wear rate was initially zero during an incubation period before increasing. The smooth oxide layer present at the beginning of the test acted as a solid-lubricant. The incubation period was attributed to the formation of micro-fractures before its disintegration producing wear.

Mittchel and Mercholsky et al [68] investigated the influence of seeded debris on silicon nitride-M50 steel rolling contact using a Ball-on-Rod machine under high contact pressures 5.14 – 5.91 GPa. Alumina (size 1-3 μm) and Arizona test dust (up to 100 μm), consisting mainly of SiO_2 , were seeded into the lubricant (Exxon Turbo Oil 2380) in separate tests. Alumina particles were found to have indented the silicon nitride surface resulting in roughened surfaces and the formation of cracks. Cutting and ploughing abrasive wear was observed in the M50 steel from alumina particles as they indented in to the silicon nitride surface. Though larger, the Arizona test dust was not found to produce significant wear on the ceramic specimen due to its relative low hardness.

2.2.3.3 Rolling Contact Fatigue

RCF failures in a hybrid rolling contact, is typically determined by the steel counterfaces (as found in for example [67-72]). As such, hybrid bearings life times have been estimated based on the assumption of failure occurring in the steel element(s) [73].

For RCF failures in the steel counterface, the failure occurs in much the same way in equivalent steel-steel rolling systems as described in Section 2.1. In three related studies by Rosado, Forster et al. [70] Arakere, Branch et al. [71] Forster, Rosado et al. [74] comparisons were made between hybrid and steel bearing spalls. Using a bearing test machine, hybrid angular contact bearings, incorporating M50 steel races, were compared to an equivalent all steel (AISI 52100) bearing in tests. Tests were oil lubricated and run under both high and reduced loads (3.1 GPa and ~2.4 GPa respectively). RCF failure was observed fastest in the steel bearing; whilst for hybrid bearings, all failures occurred in the steel races.

Spall propagation characteristics were further investigated under reduced load and the number of cycles to propagate a spall at its first detection to complete failure were analysed (spalling was detected through steel debris monitoring). Through the action of over-rolling, propagation of the initial spall was found to spread in a direction transverse to rolling for both bearing configurations- where shear stress was calculated to be maximum using finite element analysis (FEA) [71]. As the spall propagated to the width of the contact ellipse, progression occurred rapidly at its trailing edge as REs descended into the spall before impacting at the trailing edge.

The low density of the silicon nitride elements was found to impart less energy than their steel counterparts during these impacts, resulting in reduced propagation rates. Microstructural analysis [74] of the steel samples also demonstrated the importance of the steel counterface. Using etching techniques, the AISI 52100 steel races showed the presence of high residual subsurface stresses and phase changes in comparison to M50 steel races. Furthermore, Mosleh et al [75] confirmed the reduced fatigue lives of AISI 52100 steel races in comparison to M50 steel in hybrid bearing configurations at high contact pressures (>7 GPa).

Zaretsky and Parker [76] were some of the first to observe RCF failures in silicon nitride, reporting the similarities to RCF spalling in equivalent steel contacts. Failures from inclusions [42] and voids [77] have been observed leading to both surface and subsurface related spalling. However, high quality dense material achieved through HIPing techniques in the past three decades have led to the production of consistently reliable material [78] [25]. For recent investigations since, RCF failure from pre-cracks have been found to be the dominant cause for failure and subsequently the characteristic of these failures are discussed below.

The typical stages leading to failure from pre-cracks in rolling lubricated conditions have been described in-part by Kida et al [51] to occur in three general stages:

- i. The pre-crack propagates into the material through mode 1 dominated stresses;
- ii. Under the influence of mixed mode stresses (predominantly mode 2), the crack begins to turn or in some cases branches out in both directions to propagate parallel to the surface.
- iii. The extent of the subsurface crack leads to a relaxation of the contact around the defect, and increased tensile stresses at the trailing edge to form secondary c-cracks. These eventually propagate to meet the subsurface crack, leading to a flaking off of the undermined material.

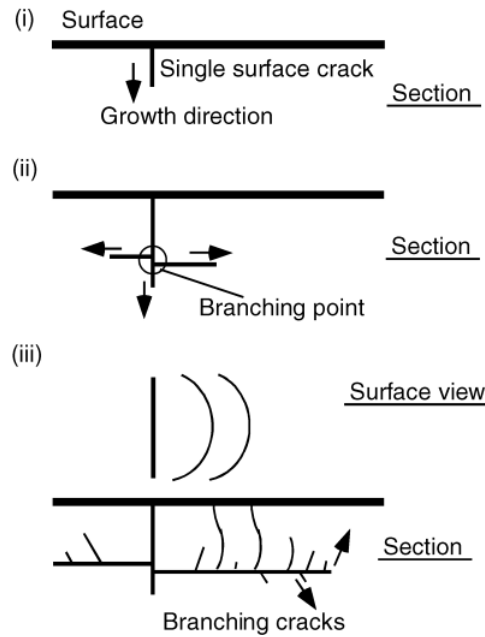


Figure 2.13: An illustration of the stages of spallation described by Kida [46]

The formation of c-cracks is an integral part of the spalling failure in silicon nitride; however the propagation characteristics of the spall are greatly affected by various parameters, some of which are discussed below and the research performed by various investigators to date.

Wang and Hadfield [52] first investigated the influence of the c-crack orientation in the wear path on the failures of the silicon nitride both experimentally and analytically [79], using finite element analysis (FEA) techniques. Tests were performed on a Four-Ball test machine with contact pressures up to 5.58 GPa and run to a maximum 100×10^6 cycles before suspending each test. Twelve crack positions illustrated in Figure 2.14 were tested under lubricated conditions (using Exxon Turbo 2380). Only three failures were observed and their respective number of cycles in millions is indicated in Figure 2.14c. FEA analysis confirmed the sensitivity to crack orientation through calculation of the stress intensity factors (SIFs) at the respective crack front. Notably, the alignment number 9, whose centre corresponded to the wear track centre (see Figure 2.14c), was found analytically to produce the highest SIFs, and correspondingly produced the fastest failure.

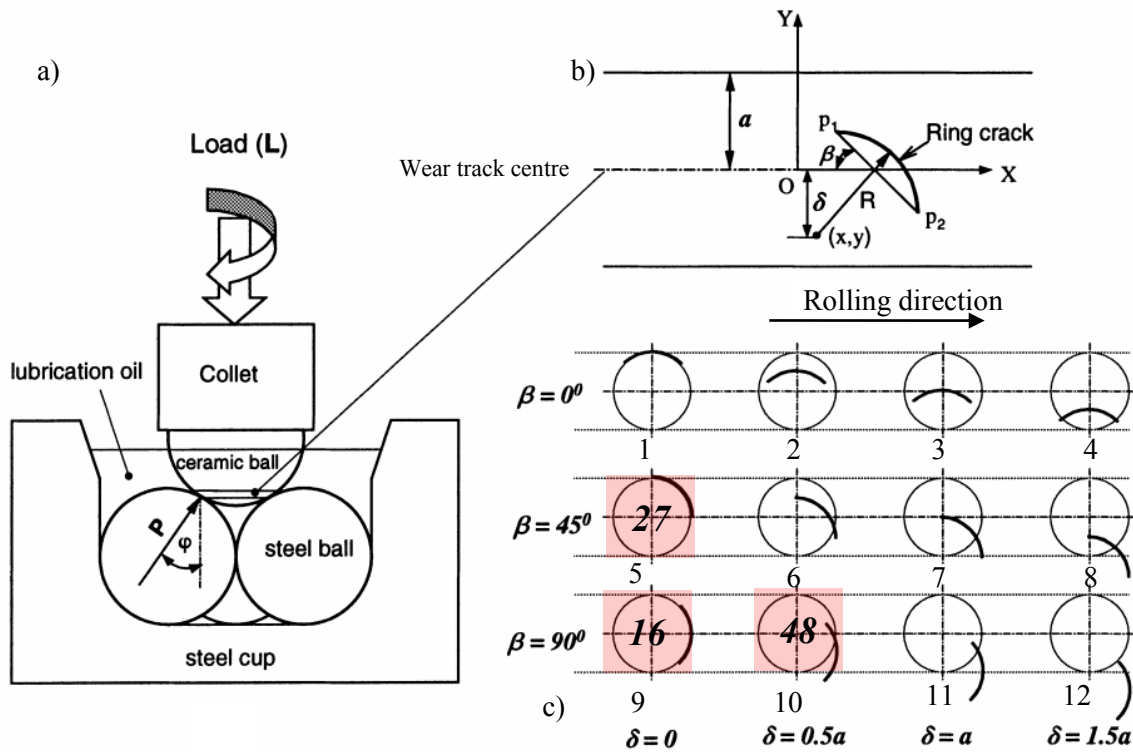


Figure 2.14: Illustrations of the testing configuration performed in [52] showing: (a) Machine and loading configuration; (b) geometry of a ring crack on the contact track; (c) 12 tested crack locations. The three failures are highlighted in red with their respective millions of cycles to failure.

Zhao, Hadfield et al.[80] investigated the influence of the lubricant in the propagation of c-crack defects in silicon nitride using a Four-Ball test machine. Using three lubricants with different viscosities, a substantial variation in the cycles to failure was observed between lubricants. Spalling was observed fastest with the traction oil for each of the loads used (ranging from 4 to 6.6 GPa) compared to the other lubricants (grease and gearbox oil). This was due to its relatively low viscosity and crack penetrating ability.

The effect of lubricant penetration was investigated by FEA by modelling the reduction in friction between crack faces as it is entrapped between the crack faces during over rolling. A substantial increase in maximum tensile stress at the surface was found with reduced crack face friction coefficients; whilst also an increased SIF under shear was noted at the subsurface crack front.

The influence of the hydrostatic pressure produced as lubricant becomes trapped in the crack was investigated using FEA techniques by Zhao, Hadfield et al. [81]. Two orientations of the crack were investigated (illustrated in Figure 2.15). Here, configuration *a*) showed larger hydrostatic pressures at the respective crack front, as the direction of roll acted to force lubricant to the crack tip.

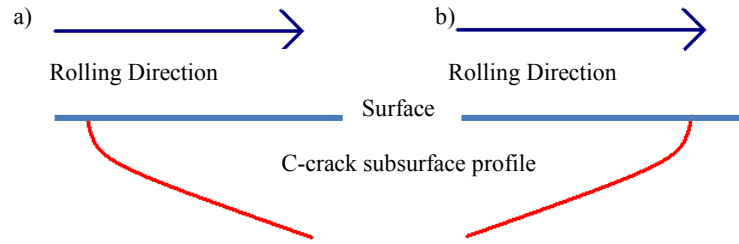


Figure 2.15: Illustration of the two crack orientations in the investigation by Wang and Hadfield [81]. a) crack orientation acts to trap lubricant in the crack, whilst b) acts to squeeze lubricant from the crack.

The influence of surface traction on the growth of c-cracks was investigated using FEA by Levesque and Arakere [54]. For modelled traction coefficients from 0 to 0.15, the mode I (tensile) SIF at the crack front substantially increased (1.1 compared to $4 \text{ MPa}\cdot\text{m}^{1/2}$).

Wang and Hadfield [82] investigated the RCF characteristics of a line defect - with near vertical inclination into the surface of the specimen and length of approximately $500 \text{ }\mu\text{m}$, orientated perpendicular to the contact path in a Four-Ball test machine. Adopting high contact pressures exceeding 5.5 GPa , failures were observed in the order of $>10 \times 10^6$ rolling cycles. Notably, the initial crack was observed to propagate parallel to the surface in both directions of the contact path and consequently forming secondary c-cracks at the surface either side of the initial crack. In a more recent study, Kida, Urakami et al. [51] [46] investigated the stages of flaking failures from similar cracks generated from Vickers indentations, exhibiting the near-vertical inclined cracks. The indents were made using an indentation load of 20 KgF before removing the lateral crack using grinding and polishing to a depth of $> 70 \text{ }\mu\text{m}$, leaving median cracks with depths and surface lengths and of approximately 75 and $290 \text{ }\mu\text{m}$ respectively. Tests were lubricated with a turbine oil (ISO VG 46) and performed at 3.98 GPa contact pressure on a Ball-on-Plate rolling tribometer. By sectioning the samples after 32×10^6 rolling cycles, Kida was able to determine the steps leading to failure. Here, subsurface cracks were observed to branch towards the surface simultaneously as secondary c-cracks formed at the surface, each propagate towards one another as shown in Figure 2.16 below.

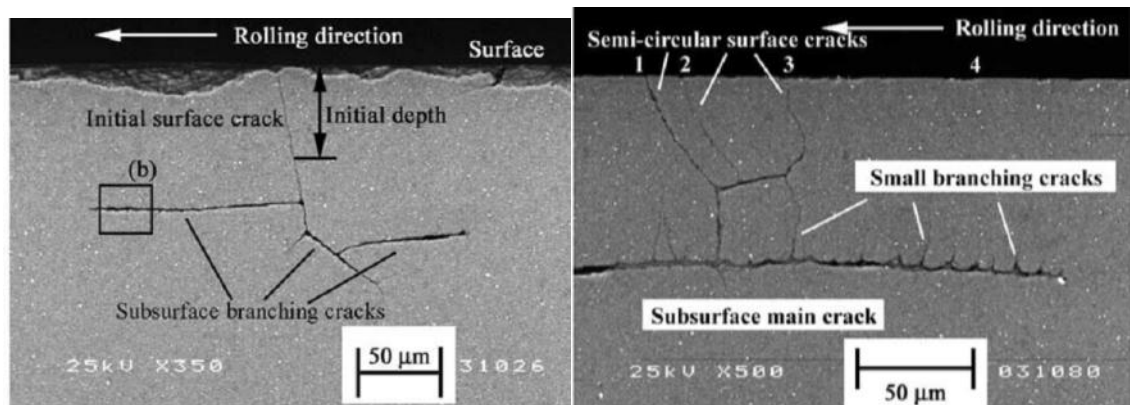


Figure 2.16: The subsurface branching cracks after 3.2×10^7 cycles in separate tests [46]

Kadin et al [1] investigated the influence of angle of incline of a pre-crack as it propagates into the material from the surface. Notably, for cracks with a shallow incline, crack propagation was promoted by increased mode 1 (tensile) SIFs as the contact approaches the leading edge of the crack. This mechanism is described as a ‘cantilever’ effect and is shown schematically in the figure below.

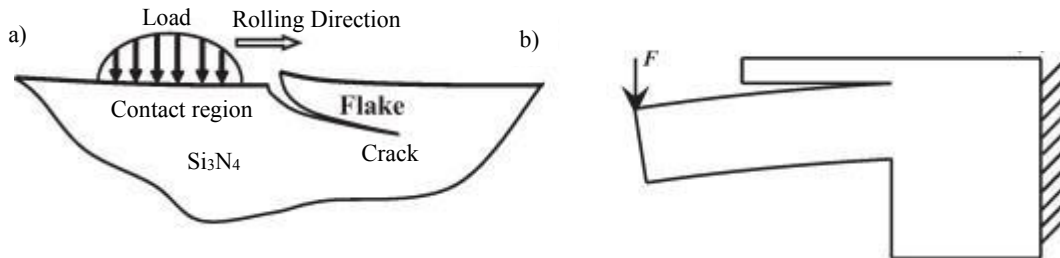


Figure 2.17: An illustration of the pre-crack with a shallow incline opening under mode 1 stress b) a representation of the cantilever effect

In a series of investigations, Hadfield and Stolarski et al [83] [84] [49] [55] used Vickers indentations to simulate star cracks in silicon nitride rolling tests in a Four-Ball test machine. Here indents were generated with 5 KgF load and various lubricants were used to lubricate subsequent tests at high contact pressures (6.4 GPa). Various types of failure were observed, notably delamination failures (see Figure 2.18); where shallow lateral subsurface cracks propagated to form shallow spalls of $\sim 5 \mu\text{m}$ in depth. In a test using kerosene as lubricant, surface bulging was observed at the indent location (see Figure 2.18c) caused by high hydrostatic pressures in the lateral crack from lubricant penetration. Natural ring cracks of depth $\leq 5 \mu\text{m}$ were also observed to delaminate [85] forming shallow spalls.

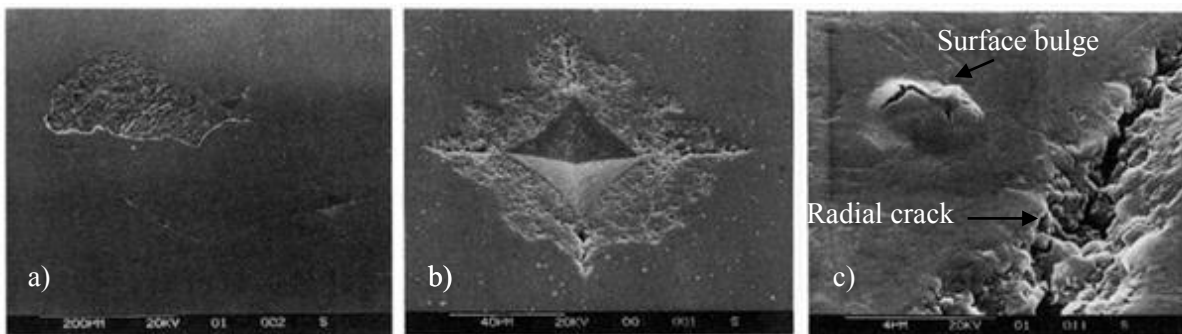


Figure 2.18: Post-test SEM analysis of the failed surfaces a) radial crack propagation and delamination; b) lateral crack delamination; c) surface bulging Hadfield, Stolarski et al [55].

The sensitivity of pre-crack geometry to RCF failures has been investigated using FEA techniques. Levesque and Arakere [9] found that increasing the pre-crack radius of a modelled c-crack from 0.3 to 0.4 mm, the respective cycles to failure decreased significantly from 0.3 to >10 million cycles under same test conditions. Additionally, the high sensitivity of cracks to parameters including lubricant type, crack orientation, crack geometry, traction and load conditions as discussed above, lifetimes of respective elements are difficult to predict [86].

2.3 Summary

The objective of this chapter was to learn about the wear processes in hybrid tribo-system with a particular focus on developing an experimental arrangement for simulating wear production in hybrid bearings.

The main findings can be categorised into i) material processes within steel and silicon nitride in the production of various faults; ii) external contributory factors to the production of faults, and are as follows:

Three stages of RCF development in silicon nitride have been identified:

1. Crack propagation to a depth
2. Crack branching below surface
3. Secondary cracks on surface assisted by (2).

In general, material loss is only produced following the last stage; therefore much of the fault development occurs without producing debris. Stage 1 can be initiated in static processes such as indentations or by the presence of imperfections e.g. induced during manufacture. Two crack types have been described: star cracks and c-cracks differing in the shape of the crack patterns (radial or semi-circular respectively) produced in Stage 1. Also, the exact angle of the crack advancement in stages 1 and 2 has been found to be critical in the fault propagation: shallow inclined tend to advance through stages more rapidly than near-vertical and lead to an early production of a flake.

The general RCF characteristics in silicon nitride have therefore been found to be broadly similar to steel i.e. localised material loss aided by crack propagation.

In general, silicon nitride exhibits a smooth tribofilm consisting of silicon oxides leading to low wear rates and providing low friction and highly wear resilient components, superior to steel. The exact way and rate of the fault development through stages above depend crucially on its microstructure. Therefore, to emulate a realistic practical situation, an experimental set up should utilise materials of bearing standard.

External factors that affect the origination and progression of faults include the indenter or counterface shape: a sharp indenter produces star cracks, a blunt counterface tends to produce c-cracks. In practice, a Vickers indentation machine is used to produce star cracks as a seed to accelerate subsequent RCF development. To induce c-cracks requires the application of traction in the indentation process.

RE ball bearings, deployed in high speed operations are subject to slip with estimated SRR to be approximately 2%. In order to provide a realistic hybrid testing configuration, test runs should incorporate sliding in conjunction with rolling.

Three lubrication regimes are described based on their Lambda value, assessing the degree of separation of the two contact interfaces. Typically RE bearings operate with a Lambda value between 2 and 4 which should therefore be reproduced in any realistic test configuration.

The hardness of silicon nitride produces wear in the softer steel counterface of hybrid bearings; therefore a realistic experimental situation will include devices monitoring both components.

Chapter 3. Review of Signal Processing

3.1 Overview of Signal Processing

Signal processing is based on the principle of detection of a particular recognisable pattern transmitted as disturbances through a physical medium from a source of interest. The major issue encountered in this process is that the medium transmits a vast range of disturbances from multiple sources and correspondingly the amount of detected data to be processed is vast.

Signal processing techniques are based on the recognition of several distinct functional elements outlined in the following figure and described briefly below.

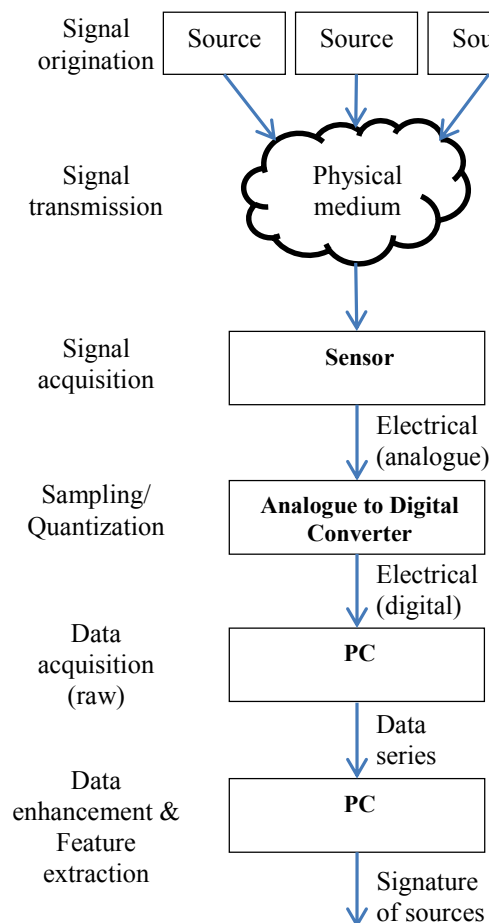


Figure 3.1: Signal processing stages

Source: indicates each contributory source that generates the disturbances. Specifically, three types of sources are of interest in this project: RCF faults in the silicon nitride, wear in steel and background/ noise (representing all other sources).

Physical medium: the environment through which the disturbances propagate from the source. Specifically, three types of media are of interest: vibration (mechanical waves), charge (electrostatic flux) and acoustics (stress waves).

Sensor: this is a physical device that transduces the disturbances in the medium into electrical signals. Sensors must be appropriate to both the media and the source of interest.

Analogue to digital converter: device that converts the electrical signals from the sensor to digital signals through sampling. In particular, a device is required that can provide appropriate sampling rates for the source of interest.

Data acquisition: This represents the storage of the digital signals for their subsequent analysis. A standard personal computer with appropriate computational capacity to perform the respective analysis is commonly used.

Data enhancement and feature extraction: This represents the processing that needs to be done on the data captured in order to discern patterns characteristic to each source and represents one of the key parts of the project. The choice of any particular methods depends on both the character of the source and the medium.

The following sections provide a comprehensive survey of the state-of-the-art signal processing strategies relevant to the research project, and are organised as follows:

Section 3.2: reviews how each of the chosen media are disturbed by various physical processes associated with wear - in particular, developing RCF faults. Additionally, the section reviews the signal processing strategies that are employed in order to identify them.

Section 3.3: reviews methods utilising the frequency-time or spatial-time distinctions to enhance a signal in order to extract the physical processes associated with wear.

Section 3.4: provides a summary of the main findings in the review

3.2 RE Bearing Sensing Strategies

3.2.1 Introduction: Sensing Techniques Used in Bearing Monitoring

There are a wide range of sensing strategies currently available for steel RE bearing wear monitoring, including oil debris analysis, temperature and thermal imaging, and vibration monitoring.

Degradation of oils can be monitored using various methods and described recently by Hamilton [87], while wear monitoring of bearings can be achieved through the analysis of wear debris entrained in the oil line. Depending on their working principles, wear debris analysis is classified into magnetic [88, 89], conductive [90], acoustic and ultrasonic [91, 92], and image based [93, 94] methods.

As highlighted in the previous section, RCF of silicon nitride is characterised by the development of surface and subsurface cracks before the liberation of material forming debris. Because of this, debris monitoring is unlikely to be an effective strategy in detecting RCF during their early stages of development. Temperature sensing methods, responding to increased friction as a result of increased wear rates and friction, are also unlikely to detect RCF during its infant stages. Therefore, these techniques were not selected for the investigation in this study.

Media that have been identified to be disturbed by the physical processes associated with the development of RCF failures in silicon nitride are as follows: vibration, electrostatics and acoustic emission (AE), and are reviewed in turn in the subsections following.

3.2.2 Vibration

Vibration monitoring is one of the most widely used online analysis methods for rotating machinery and one of the most established to monitoring bearing health due to its high sensitivity in detecting a wide range of machinery faults. It has been particularly useful in conventional RE bearing monitoring in identifying RCF failures and is thus chosen for this study.

3.2.2.1 Principles of Vibration Monitoring

Under EHL and hydrodynamic conditions, bearing surfaces are kept apart by a lubricating film and the vibration levels from the bearings are therefore minimal and consistent. Defected surfaces prohibit functionality of the EHL film and can lead to impacts between the rolling elements and races, inducing more vibration to the surrounding structure. When placed accordingly, a vibration sensor can provide information to the type of fault in the machine,

including faults such as bearing misalignment, imbalance, defect location and potentially the severity of the defect(s).

The technique has been extensively employed to determine spalls or pits in bearings produced from RCF (for example [95-115]). The location of a defect (whether it be on the races, RE or cage), can be determined by the periodicity of the impulsive vibration produced from their repetitive over-rolling. The corresponding periodicities for each defect location can be calculated using the bearing geometry and rotating speed using the following equations:

$$\text{Outer Race} = z_1 \omega_c = \frac{z_1 \omega_s}{2} \left(1 - \frac{d_r}{D_b} \cos \alpha_1 \right) \quad 3.1$$

$$\text{Inner Race} = z_1 (\omega_s - \omega_c) = \frac{z_1 \omega_s}{2} \left(1 + \frac{d_r}{D_b} \cos \alpha_1 \right) \quad 3.2$$

$$\text{Rolling Element} = 2 \omega_b = \omega_s \left(\frac{d_r}{D_b} \right) \left(1 - \left(\frac{d_r}{D_b} \cos \alpha_1 \right)^2 \right) \quad 3.3$$

$$\text{Cage} = \frac{\omega_s}{2} \left(1 - \frac{d_r}{D_b} \cos \alpha_1 \right) \quad 3.4$$

Where ω_s is the shaft speed in Hz, d_r the diameter of the rolling element in meters, D_b the pitch diameter in metres, z_1 the number of rolling elements and α_1 , the contact angle in radians. In practice, the characteristic frequencies of bearing defects can differ slightly from the theoretical due to the effects of slip [109, 116].

3.2.2.2 Vibration Sensors

Three types of transducers are available for sensing vibrations, measuring: displacement, velocity and acceleration. An important consideration in the choice of sensor type is the frequency measurement range, as various machine faults can produce vibrations with a characteristic frequency range. For example, shaft misalignment can produce vibrations in the range of up to 3.5 kHz [117] and in particular, bearing faults produced from RCF can resonate high structural frequencies of the surrounding structure, therefore necessitating a sensor with a large frequency sensitivity range. Displacement and velocity type sensors are typically limited to 1.5 and 3.5 kHz respectively [118] whilst accelerometers can provide measurements to vibrations exceeding 20 kHz and are therefore considered appropriate for this research.

Accelerometers monitor the rate of velocity change of a mass using a piezoelectric transducer which produces an electrical output proportional to the mechanical force applied.

3.2.2.3 Vibration Analysis Methods

A range of conventional techniques has been established to analyse vibration signals from rotating machines in order to diagnose an impending failure. Two notable time domain methods are described below, namely statistical methods and enveloping techniques⁴.

i. Statistical Methods

Most notable of the statistical analysis techniques applied to vibration monitoring are the calculations of the root mean square (RMS), and kurtosis - the fourth central moment statistic about the mean. They are used to quantify the average amplitude of a signal and ‘peakedness’ respectively.

The RMS and kurtosis can be calculated for a signal x (in digital format) using the following equations:

$$RMS = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N}} \quad 3.5$$

$$kurtosis = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{(N-1)s^4} \quad 3.6$$

Where N is the number of signal data points, s the standard deviation and \bar{x} the mean value of the signal analysed.

These techniques are particularly useful for providing a trend or ‘*history data*’ of the signal in the course of the machine’s operation. Prominent increases in these values generally indicate signs of distress in the machinery and an impending failure.

Williams Ribadeneira et al [112] were able to highlight the generation of spalls on RE bearings with seeded defects using RMS trends of the vibration response. The RMS amplitude increased as surface defects became evident. After the initial increase in RMS at the onset of failure, the RMS values were found to decrease before rising sharply again. The drop in RMS was attributed to the smoothing of sharp edges of the initial spalled area by the action of the continued over-rolling. The RMS signal increased again as the spall enlarged, encompassing a greater area.

A signal demonstrating a Gaussian probability density function has a kurtosis value of 3. Dyer and Stewart [119], first reported the Gaussian like nature of a vibration signal from a healthy bearing. Thereafter, kurtosis values of vibration signals are often quoted with a value of 3

⁴ Analysis methods in the frequency-time domain are discussed in Section 3.3.

subtracted as an indication of bearing health. When a spall or surface defect is formed on a bearing surface, impulses generated through over-rolling will increase the peakedness of the signal, and correspondingly the kurtosis value increases.

However, these methods can suffer from poor sensitivity to fault features with low signal power which are masked by background machine noise. Analysis is therefore combined with pre-processing to enhance fault features and reduce machine noise. A commonly applied technique is ‘*envelope analysis*’ and discussed in the following subsection.

ii. Envelope Analysis

Characteristic defect frequencies are often masked by noise with overlapping frequencies produced by rotating harmonics, and nearby machinery and components. To enhance fault signatures, enveloping is often used before defect frequency analysis is performed.

The envelope method, is a common technique applied in the monitoring of RE bearings, and often accompanies the *shock-pulse method* where impulses or shocks generated as a result of component deterioration. These impulses often exhibit themselves at characteristic frequencies governed by the resonant oscillatory modes of the machine. The excitation range of the machine can be determined by monitoring the vibration signal from a series of impact tests at the vicinity of the RE bearing monitored. Subsequent vibration signal analyses are then tuned to the resonant oscillatory modes of the machinery using band-pass filters where the signal-to-noise ratio can be significantly higher.

Amplitude demodulation is then performed to envelope the response, where an envelope is placed around the pulse like response from the over-rolling of a defect. A common demodulation technique is performed by the applying the Hilbert transform to filtered signal x , generating a 90° phase shifted representation of the signal y . The envelope a , can be calculated by the equation:

$$a = \sqrt{(x^2 + y^2)} \quad 3.7$$

Frequency analysis can be often performed to determine the periodicities of the shocks and compare them to the defect frequencies described in equations 3.1 to 3.4.

The band-pass filtering method applied is an important consideration to extracting faults. Methods reliant on the application of analogue filters, or their digital implementation results in payoffs between filtering characteristics [101], for example the application of the commonly applied Butterworth filter can result in the addition of contaminating ringing components. Similarly, the choice of bandwidth will influence fault extraction too - too wide and noise

components can contaminate the envelope, too narrow can reduce the time sensitivity of the envelope. $\pm 5\%$ of centre-band has been adopted for a bandwidth by Broderick et al [120] and subsequently by Darlow et al [121] for the monitoring of bearings. However, McFadden and Smith [101] recommend a bandwidth at least four times the highest defect frequency component, to accommodate modulation sidebands.

Further signal methods are adopted to avoid the payoffs between filter parameters, to extract fault features more concisely; these techniques based on time-frequency analysis are discussed in Section 3.3.

3.2.3 Electrostatic Charge

Electrostatic monitoring was first developed to monitor debris ingested by jet engines and gas turbines [122, 123], as ingested debris were found to exhibit a measurable charge. Furthermore, the method was applied to identify debris present in the exhaust path, produced by the deterioration of engine components [124]. A charge was found to be induced during the deterioration and subsequently the method has since been applied to monitor wear, in order to provide monitoring capability to a range of tribological applications.

In a series of experimental investigations, the method has been found to be able to detect wear failures over conventional monitoring techniques including vibration [125-128] and temperature [126-129].

3.2.3.1 Principles of Charge Generation

The charge mechanisms generated during a tribological contact are complex and are affected by various competing mechanisms, for examples the chemical and physical properties of the lubricant, the materials of the contact pair and wear regimes.

The charge mechanisms thought to be dominant in tribological contacts are: *tribocharging*, *contact potential difference*, *triboelectrification*, *triboemission* and *debris generation*. The following subsections discuss each in turn.

i. Tribocharging

Tribocharging describes the mechanism of charging of a low-conductivity fluid by its relative motion over a surface.

An electrical double layer forms naturally at a solid-liquid interface, where a charged surface attracts opposite charges and repels like charges. Currently, two theories exist for the charge current generated in a flowing medium (and have been investigated by [130]), '*Double Layer Stripping theory*' and '*Differential Rate theory*'.

The first theory is based on the relative motion at the fluid-solid interface being sufficient to strip a portion of the charge double layer and entrain it into the main body of the lubricant; this effect is illustrated in Figure 3.2 below. The second theory is that charge is created by a differential rate of either diffusion of charge species to and from the interface or a differential rate of adsorption/desorption of these charged species.

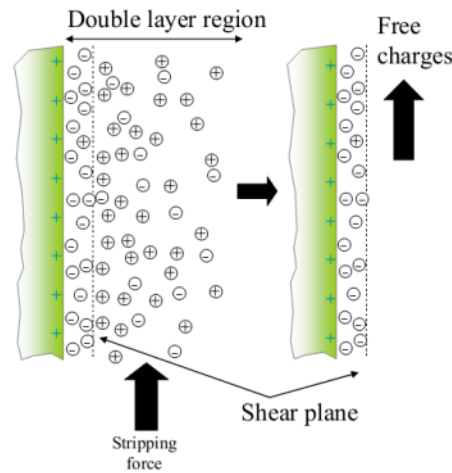


Figure 3.2: Illustration of an electrical double layer [130].

The charging properties of a system is dependent on various factors including the properties of the lubricant, temperature, shear rate and surface area of the interface have been identified.

Harvey et al [130] detected increased charge at higher temperatures as a result of the decreasing viscosity of the lubricant. This effect was valid only to a threshold temperature above which the charge was found to decrease. Below the threshold, increasing temperatures reduced the oil viscosity, promoting ionic mobility, resulting in an increase in charge magnitudes. At higher temperatures however, the greater ionic mobility increased the rate of mixing, promoting the recombination of charges.

Increased surface roughness was also found to promote tribocharging, with an increased surface area for solid-fluid interface facilitating the double layer region [130]. Also, an increased roughness promotes greater micro-turbulence which facilitates greater charge removal from the double layer.

ii. Contact Potential Difference

Work function is defined as the minimum work required, in a vacuum, to extract an electron from the Fermi level of a conducting phase from its surface. Different metals have different Fermi levels and hence different work functions. Examples of work function values are summarised in Table 3.1 below.

Metal	ϕ_M (eV) (Photoelectric)
Zinc	3.63
Silver	4.26
Aluminium	4.28
Tin	4.28
Chromium	4.44
Tungsten	4.55
Iron	4.40
Copper	4.65
Gold	5.10
Nickel	5.15

Table 3.1: Work functions of a range of metals [131]

When two metals, are brought into contact their Fermi energies equalize. This equalisation is achieved by electron transfer and results in the generation of an electrical charge on the respective surfaces.

In order to investigate the sensitivity of an electrostatic sensor and provide a basis for sensor calibration, Morris et al. [132] inserted copper, aluminium, En31 and carbon steel inserts in a bearing steel disc. The disc was subsequently rotated using a pin on disc machine with the sensor applied directly above its surface at the radial location of the insert. The experiments produced peaks corresponding to the rotational position of the insert. The polarity of the peak was dependant on the relative work function of the insert material with respect to the steel disc; correspondingly, inserts with work functions values greater than the steel disc produced negative peaks (see Table 3.1) and vice versa. The aluminium subsequently produced a positive peak and is illustrated in the figure below.

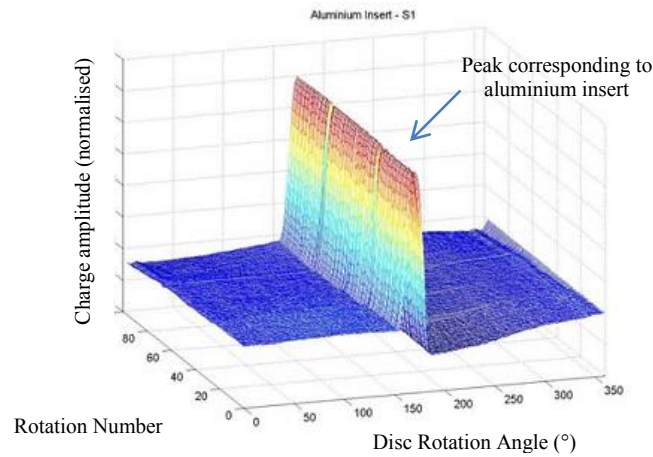


Figure 3.3: Charge measurements from a steel rotating disc with an aluminium insert [132].

The formation of fresh surfaces as oxide layers are stripped off during wear have been reported to produce a charge as a result of the difference in the work function from that of the bulk [128].

iii. Triboelectrification

Triboelectrification occurs when two solid material interfaces are contacted together with relative motion between them. Electrons are transferred from one interface to the other.

Which material becomes negative and which becomes positive depend on the materials involved and the charge polarity is indicated by their relative positions in the triboelectric series. This ranks materials according to their tendency to gain or lose electrons (a variety of materials are ranked in [133]) and subsequently, silicon nitride charges positive when rubbed against steel. Also, an insulator-conductive pair are found to generate charge magnitudes above that of a conductive-conductive pair as the insulator is able to maintain a high charge density [125].

iv. Triboemission

Triboemissions are the emission of photons and particles released during various wear processes e.g. deformation, abrasion and fatigue cracking (*fractoemissions*).

Nakayama et al [134-137], investigated the phenomena using a diamond stylus contacting various disc materials on a pin on disc machine. Emissions of burst charge and photons increased in intensity with higher loads and speeds. The charge intensity of negatively charged particles emitted was also found to increase in the order of conductor < semiconductor < insulator as illustrated in Figure 3.4a below.

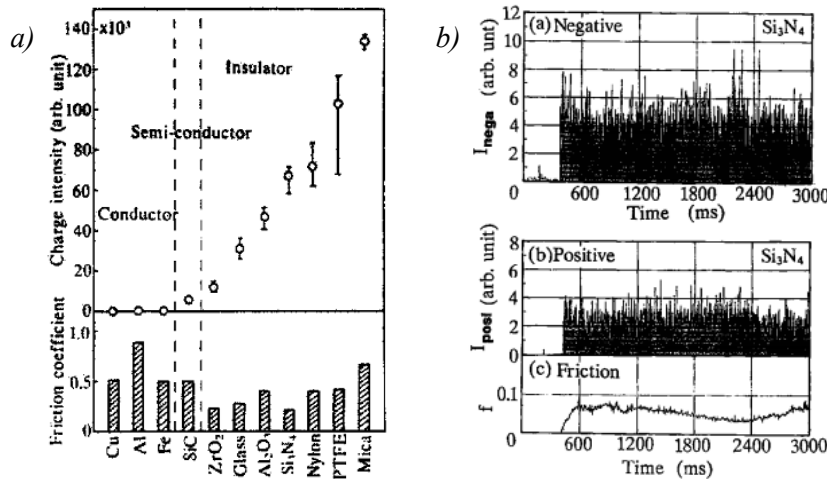


Figure 3.4: Charge intensities during dry pin on disc scratching tests. a) Negatively charged particles and corresponding friction coefficient on various solids [136]. b) Negatively charged particles (upper) and positively charged particles with CoF during scratching Si₃N₄ [137].

The intensity of positive charged particles emitted from an insulator was found to vary little compared to the intensity of negatively charged particles, as shown in Figure 3.4b above. In contrast for conductors and semiconductors, the intensity of positively charged particle

emissions was a lot smaller than the intensity for negatively charged particles. This was attributed to the greater supply of bulk electrons from more electrically conductive solids.

Nakayama also found that the humidity, atmosphere type and surrounding gas pressure affected the intensity of triboemission. In hydrocarbon atmospheres the intensity of emissions decreased as the hydrocarbon chain length was increased. In a branched hydrocarbon atmosphere, the intensity is slightly less than that of a straight chain of the corresponding number of carbon atoms. For example, the intensity of triboemissions in iso-butane was measured to be slightly less than that in n-butane.

Enomoto et al [138] also demonstrated the high charging ability of various ceramics (including Si_3N_4), through measuring exo-electrons generated through fracturing of ceramic in air. The resulting fracture surfaces of the specimens were found also to be electrified, reaching potentials of the order of a few hundred volts, ionising the surrounding gas molecules.

v. Debris Generation

Debris generation involves the breakage of numerous bonds producing triboemissions, but can also render the debris and parent surfaces charged. This phenomenon can be exploited to monitor the debris entrained in the oil-line downstream of the contact zone, assuming the entrainment fluid is insulating and the debris has not discharged by a contact made with an earthed surface [132]. The debris itself can become tribocharged as it is entrained in the lubricant. Furthermore, if the debris contacts a surface it can form a contact potential difference [132].

3.2.3.2 *Sensors*

If a conducting plate is placed close to a charged source as illustrated in Figure 3.5, the free electrons within the plate will be drawn towards or repelled from the point charge depending on the polarity of its charge. If the plate is earthed, electrons can flow to or from the earth with minimal resistance. By measuring this electron flow, it is possible to monitor the magnitude and polarity of the charge.

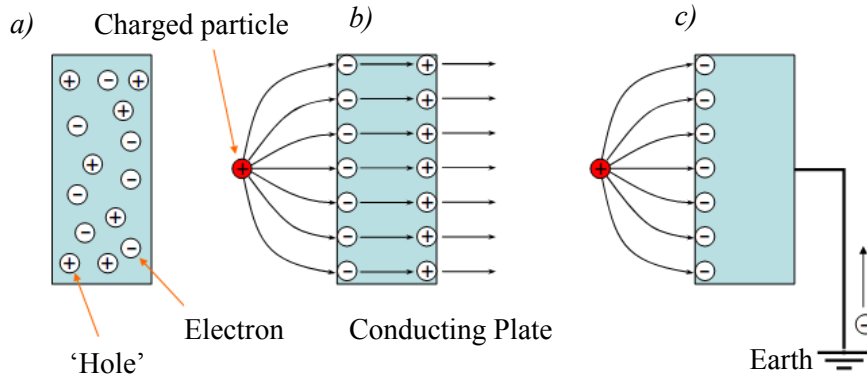


Figure 3.5: The behaviour of (a) a conducting plate in the vicinity of (b) charged particle (c) and with a connection to the earth.

The ability to create a potential difference between the source charge Q and the sensor surface Q_A is determined by the capacitance of the system approximated by equation 3.8.

$$Q_A \sim \frac{Q \cdot A}{r} \quad 3.8$$

Where A is the area of the sensing face and r the distance between Q_A and Q .

Two types of sensors are currently available to detect charge produced as a result of wear, namely:

- Wear site sensor *WSS*- uses a flat sensing face positioned directly above or near wearing interfaces.
- Oil line sensor *OLS* – uses a ring type sensing face where the lubricating medium passes through and is placed downstream of the wearing faces.

An example of each type is shown in Figure 3.6 below.

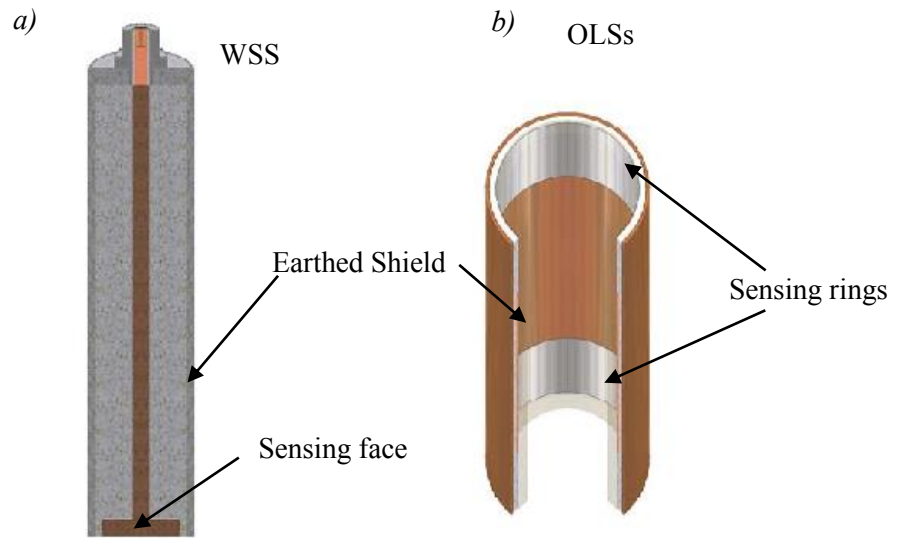


Figure 3.6: Illustration of two types of electrostatic sensors available for wear monitoring. a) WSS. b) OLSs [14].

The effect of sensor geometry for both ring [130] [132] [139] and button type [140] sensors have been investigated. Harvey et al [130] discussed the effects of the earthing shield surrounding the sensor face which competes with the sensing face for terminating electrostatic flux lines. Subsequently, the earthing shield can reduce the peak magnitude of the detected charge, reducing the sensor's efficiency. However, Morris [132] used the surrounding earth shield to reduce the field of view of the OLS sensing rings to improve the spatial resolution for separating highly localised charge sources.

Hammer and Green [139] and more recently Zhe et al [140] investigated the properties of sensor geometry to various charge sources using a model based on linear time invariant (LTI) system theory.

3.2.3.3 *Electrostatic Monitoring Investigations*

i. Sliding Wear Investigations

Investigations in electrostatic monitoring have been focused predominantly on sliding wear regimes conducted using a Pin on Disc machine and monitored using a WSS placed directly above the wear track of the disc [63, 126-128, 130, 141-143]. Some of the notable findings of these investigations are now discussed.

For steel-steel lubricated tests, Tasbaz et al [128] was able to identify scuffing before significant increases in friction were detected. A prominent increase in the charge magnitude (measured

using RMS) was found corresponding to the formation of white layers as a result of phase changes in the steel disc

Harvey et al [126] investigated the charge response from debris by monitoring charge debris centrifuged radially outward from the wear track. This was performed using a second sensor position above the disc at radial distance greater than the wear track position. Charge detected by the primary WSS increased with both positive and negative features at a threshold load where severe wear resulted in the generation of significant wear debris. As debris was centrifuged away from the wear track, the secondary sensor detected intense dynamic positive charge features as debris itself became tribocharged. Also, a bipolar, high magnitude signal response was detected during the run in stages attributed to high tribocharging phenomena. After the running in period, the magnitude dropped significantly at the onset of steady state wear.

Morris et al [129] also reported a high magnitude charge response during the run-in and a decrease at the onset of steady state for a dry sliding steel-steel contact. This was attributed to the smoothing of asperities by fracture and plastic deformation [144]. Steady state wear was characterised by transient charge response as a result of the generation of nascent surfaces from asperity removal - producing a contact potential difference between the bulk surfaces. The return of charge magnitudes at positions of nascent surfaces was due to oxidation returning the contact potential difference to bulk surface positions, this phenomena was referred to as '*charge recovery*'.

Wang et al [142] [145], used the investigated the charge produced during lubricated hybrid contacts, using a silicon nitride ball loaded against various steel discs. Initially, high RMS charge magnitudes were measured than that of similar steel-steel wear tests due to the non-conductive properties of silicon nitride. The reduction of the RMS magnitudes was found to be related to the plateauing of the wear rate during steady state. The plateauing was attributed to anti-wear additives activated when the contact reached a threshold temperature.

Other sliding wear monitoring investigations using charge sensors include the use of the following test rigs:

- Phoenix Tribology TE77 reciprocating tribometer by Tasbaz et al [128];
- TU3 engine cam follower test cell by Booth et al [125];
- FZG gear testing machine by Powrie et al [141, 146].

During oil lubricated tests, Tasbaz et al measured high charge signal amplitudes at the formation of white layers prior to friction increases. Booth et al, found positive charges were associated

with full film lubrication whilst film breakdown and adhesive wear resulted in a negative charge signal.

The FZG wear tests performed by Powrie et al were monitored using three WSS and two OLS sensor rings in a set up illustrated in Figure 3.7

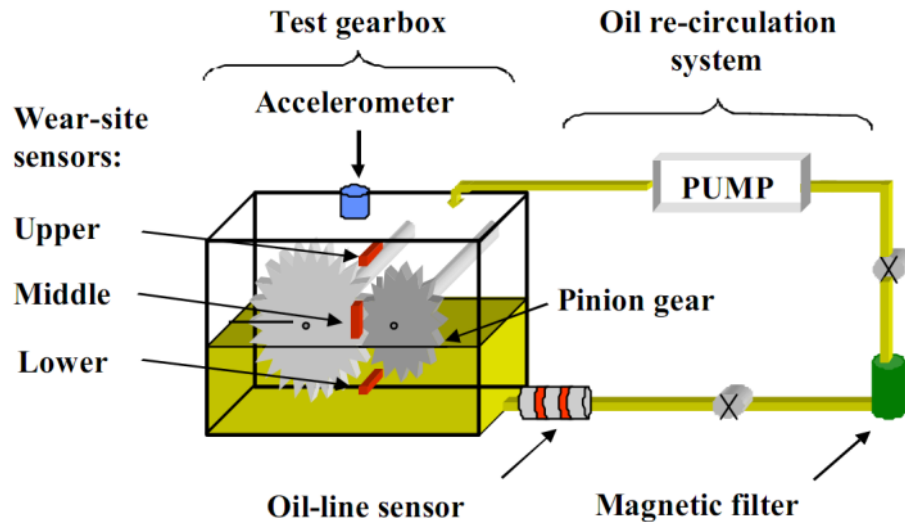


Figure 3.7: The layout of the FZG gear testing rig with the corresponding ES sensor locations [146]

WSS signals were analysed using frequency analysis and could subsequently identify a peak at corresponding to the frequency of the teeth passing the sensor faces. Analysing in the time domain, high charge signals were measured at the locations of two teeth at the onset of pitting wear.

As pitted surfaces extended to neighbouring teeth, the OLSs detected an increase in charge magnitude as a result of a prominent increase in debris production. Transient signal profiles with significant amplitudes were observed as wear debris passed the sensing faces.

ii. RE Bearing Monitoring

Electrostatic monitoring has been applied to the monitoring of steel rolling element bearings by Harvey et al [147], and Craig et al [14, 148] using a bearing test machine. The test machine incorporated four bearings (two support bearings and two test bearings) and monitored using three WSSs. Figure 3.8 below illustrates this set up.

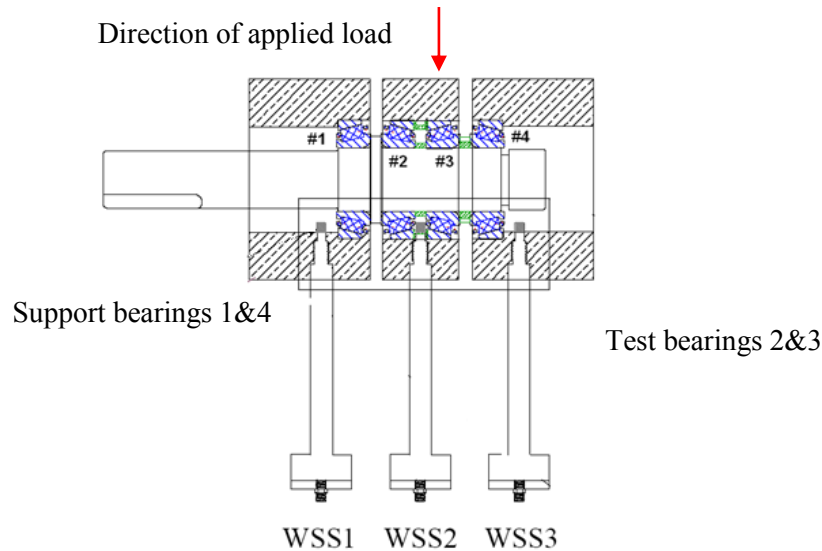


Figure 3.8: The locations of the WSSs in respect to the bearings. Harvey et al [147]

Furthermore, two OLS sensors were incorporated in the oil line downstream of the bearings and an accelerometer was fixed to the bearing test cell.

Tests were run until one of the bearings failed. This was accelerated by the incorporation of an indent seeded on the inner race of bearing 2, and by the application of relatively high test load producing a maximum contact pressure corresponding to 2.5 GPa.

Charge was found to demonstrate relatively high signal amplitudes (measured using RMS) during running in, and subsequently dropped off to a low and constant magnitude during steady state wear.

The results of a test performed by Harvey [147] is shown Figure 3.9 below.

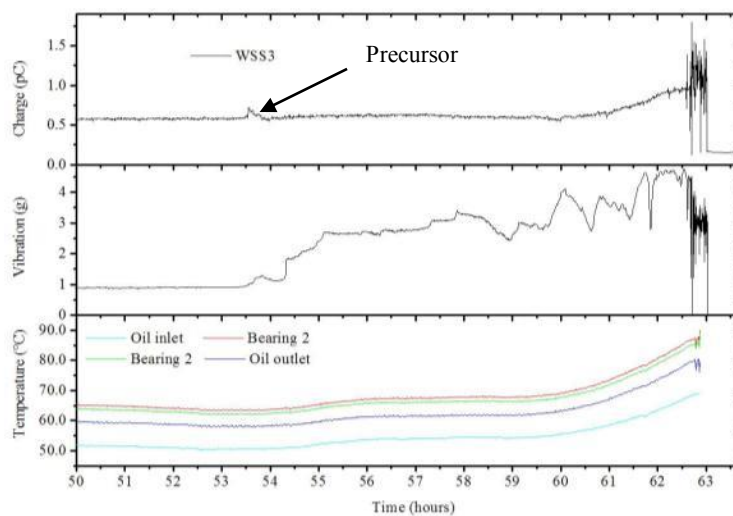


Figure 3.9: Sensor responses from a bearing rig test leading to spalling failure, WSS3 (upper), Vibration (middle) and Thermocouple (lower), during the final 13hrs of a bearing test [147]

High charge events were detected at 54 hours by WSS3 corresponding to bearing #4 and coincided with increased RMS vibration intensities. These events preceded increases in the temperature measurements and were attributed to the onset of a spall. High RMS charge signals were measured by the OLSs in the time following, as a result of debris production. Post-test inspection of the bearings confirmed that the spall did indeed initiate at bearing #4.

In a similar test, Harvey et al also detected increased RMS charge magnitudes corresponding to the WSS of the defected bearing, sometime before failure, corresponding to rises in both vibration and OLSs magnitudes.

3.2.4 Acoustic Emission

Acoustic emission (AE) can be produced by damages that can be attributed to microscopic material features, offering some attractive capabilities to condition monitoring systems - potentially providing earlier detection of RCF faults in bearings than that of vibration techniques [149, 150].

3.2.4.1 Principles of AE Monitoring

AE is the propagation of elastic waves generated by the rapid release of energy from localised sources within a material as a result of an applied stress field [151]. During the rapid release of localised strain energy, a wave is set-up around the source producing an impulse like signal - the character of which is dependent on the source process. Processes such as microscopic crack jumps are often completed from a few fractions of a microsecond to a few microseconds [152], producing a wave of short duration.

The waves propagate through the componentry of the system as oscillations of mechanical stresses in several different forms, classified based on the relationship between the direction of particle motion and propagation. These can be generally be separated into bulk and surface waves, and their properties are greatly dependent on the physical attributes of the medium [153, 154]. It is estimated that surface waves (Rayleigh type) carry two thirds of the total AE energy, whilst shear and longitudinal waves (bulk waves) consist of around 26% and 7% respectively [153] all of which attenuate with distance from the source⁵. This attenuation can be advantageous whilst also an obstacle, providing a natural filter of AE responses from distant components; but requiring sensors be placed in close proximity to the source of interest, where space can be limited.

⁵ Notably, surface waves attenuate at a comparatively slow rate that is inversely proportional to the square root of the distance from the source.

During material damage processes, notable AE sources include: plastic deformation [155], crack propagation [156], phase transformations [157], and asperity collisions [158]. In particular, for RE bearing analysis, AE can be produced during the over rolling of RCF defects as a result of localised material damages. These can be produced before the onset of material loss [150] [159] i.e. before they become evident in the vibration range [121, 122].

In order to measure and identify AE sources, an appropriate sensor and signal analysis method are required and are reviewed in the subsections following.

3.2.4.2 AE Sensors

An AE transducer consists of a piezoelectric material (commonly PZT), that produces an electrical output proportional to the magnitude of an applied mechanical stresses - consisting of high frequency acoustic waves. The frequency range of AE waves are broad, and can range from as low as 40 kHz to values exceeding 1 MHz; depending on their application the AE sensors are produced to detect waves anywhere within this range. For research investigations, wide-band sensors with a frequency range of between 0.1 and 1 MHz are often used (for example by [160]) to maximise the detection capability for capturing a wide range of AE events for subsequent analysis.

3.2.4.3 AE Signal Analysis Methods

For bearing monitoring, AE waveforms can be complicated by noise produced by various nearby components, resulting in signal profiles consisting of multiple sources superimposed. Boundaries of components and interfaces between them can further complicate a signal profile by producing reflections and attenuate waveforms [154]. Subsequently, the choice of signal processing method depends critically on these characteristics as well as the characteristic of the sources.

Because of the relatively high sampling rates required in AE monitoring, demands for signal computation and data storage are also high. Subsequently, researchers often employ three general analysis methods, namely: *continuous monitoring*, *hit-based* and *raw waveform analysis*. Details of these methods and their applications are reviewed in turn.

i. Continuous Monitoring

Continuous monitoring techniques rely on the implementation of signal extraction techniques performed in real-time. Subsequently, methods are generally limited to processing that requires low computational demands. In particular, RMS calculations have been applied to monitoring wear regimes - notably sliding wear. The method has been used to identify the onset of failure, detecting events such as asperity interactions [158], phase changes [144] and adhesive wear

[161]. These events occur rapidly and overlap in time considerably, producing a signal with complex features and high amplitudes and subsequently RMS techniques were able to provide a suitable statistic to identify these features.

Complex rotating machinery however can produce complicated waveforms with high background noise profiles, which can vary with operating parameters such as applied load and speed [116]. RCF fault signal components can therefore represent a small fraction of the signal energy and as a result RMS values can be dominated by the noise profile and subsequently are unable to identify the fault features concisely. A more appropriate technique therefore, is to analyse signal features that exceed the background profile using hit-based methods.

ii. Hit-Based Monitoring

Commonly, investigators are interested in the AE signal resulting from transient events or burst waveforms occurring over a background or baseline signal profile [154]. By pre-setting a threshold above this profile, signal components that exceed this value are extracted and analysed. A signal component that exceeds this threshold is referred to as a '*hit*', and parameters describing parameters of a hit can be recorded in real-time. The most commonly applied examples include the following:

- '*Counts*'- the number of threshold crossings
- '*Peak amplitude*'- the maximum signal magnitude
- '*Rise time*'- the duration between the initial threshold crossing and the peak amplitude.

An example of a hit is shown in Figure 3.10 below illustrating these parameters and comprehensive reviews on further techniques can be found in references [154, 162].

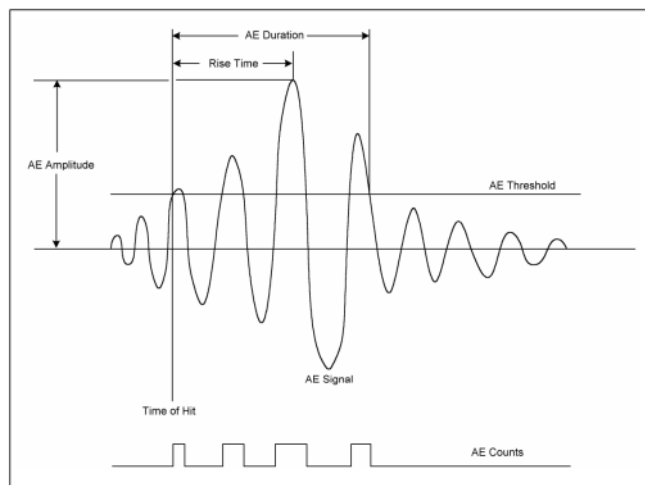


Figure 3.10: An illustration of a simplified signal arising from an AE event with the associated hit-based features.

The advantages of this analysis method are that some distinction is made between the background and fault components, and that both computational and data storage demands are significantly reduced.

Ichikawa, Takamatsu et al [163], were able to detect the initiation of ring cracks in silicon nitride using AE counts in compression tests. Rahman, Ohba et al [164] used an AE sensor of bandwidth 100-400 kHz to monitor an oil lubricated steel-steel wear test using a twin disc machine. A steel disc of low hardness (300 Hv) was rotated with a 9% slip against a hardened steel disc. AE counts were monitored for the duration of the test and were referenced to the positions of rotation on the unhardened disc using a tachometer. The disc was separated to angular segments of 3.6° to which recorded hits were binned. Transverse cracks were found on this disc (corresponding to segment 19) where the highest counts were recorded as illustrated in Figure 3.11. Rise time and duration were also investigated for the monitoring of incipient spall and its progression. It was found that the number of hits increased in response to the appearance of transverse cracks and the rise times increased from about 200 μ s to 600 μ s as a spall is formed on the disc.

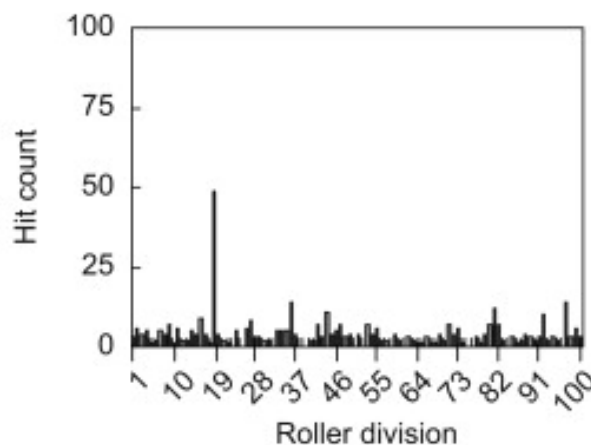


Figure 3.11: AE Hit count rate at the incipient of damage located at segment 19 where transverse cracks formed.[164]

Various researchers have used the hit-based monitoring methods to monitor bearings and some notable examples are discussed below.

Choudhury and Tandon [116] investigated steel RE bearings with various geometries and seeded defects on both the inner race and REs. They found that AE counts increased with the rotational speeds with little change with increasing load producing a substantial and changing background profile.

Al-Ghamd [159] found that peak amplitude and RMS were potential indicators for steel RE bearing defect sizes where point, line or large defects were seeded on the inner races. Both AE RMS and peak amplitude increased with defect size. However, Al-Ghamd found that defect-free bearings could not be distinguished from defects that exhibited a smooth surface profile using

these techniques. Subsequently, the AE hits were thought to be the result of surface protrusions around the defects and the resultant impacts, fractures and deformation associated with over-rolling.

Hit-based monitoring for complex machinery requires careful set up, where indirect wave paths and various AE sources can significantly attenuate fault signals [102]. Various parameters can be set to reduce the effects of reflected signals, for example by stopping the recording of AE signal some duration after a hit – known as *lockout time* [154]. However, the method is inherently reliant on AE events reaching the sensor with magnitudes greater than the background noise profile. Also, AE sensors exhibit resonant frequencies where sensitivities are highest, subsequently enhancing events corresponding to these frequencies and potentially masking fault components.

Raw waveform recording is a way of enabling different processing techniques to be explored, such that appropriate feature enhancement and analysis methods can be determined.

iii. Raw Waveform Analysis

Raw signal profiles are recorded and analysed using digital processing techniques. Because of the high sampling rates required for AE signal processing, raw signals are recorded in *bursts* lasting a short duration with an interval between each bursts (‘*dead time*’). By recording data in bursts capacity demands on both data storage and processing are reduced. However, some of the AE signal is not recorded. Nevertheless, the method has been applied in a number of investigations (for example [155, 165, 166]) and permits various analysis techniques to be investigated. Some notable techniques used are described in the subsections following which have been separated into time and frequency domain.

Time Domain

Mechefske et al, [165], analysed the AE waveforms by filtering the signal into respective frequency bands and calculating the RMS for each band. The method was applied to monitoring lubrication regimes for sliding wear using a PoD tribometer with a zirconia pin contacting a steel disc. A bandwidth corresponding to 100-500 kHz was found to demonstrate a strong dependence on the lubrication regime, i.e. high RMS values corresponded to severe wear regimes and a subsequent reduction indicated the activation of anti-wear additives.

Faisal, Steel et al [166], were able to separate the AE response during Vickers indentation of WC-Co coatings on 440c stainless steel into 3 distinct stages based raw waveform analysis (refer to Figure 3.12), these were:

Stage 1: Marked by low AE activity (magnitudes ~15% above background noise)

Stage 2: High AE response found to demonstrate highest intensity

Stage 3: Low amplitude discontinuous AE response with occasional high amplitude bursts

In a follow-on study, Faisal, Reuben and Ahmed [155], related the AE response to the load-displacement characteristics using a linear variable differential transformer (LVDT) and load cell incorporated in the indentation hardware. Faisal observed that for stage 1, load and displacement were essentially linear indicating elastic deformation. In stage 2, a steep gradient shows plastic deformation. However, for indentations made in softer materials (copper and aluminium), little or no distinction could be made to separating the 3 stages, indicating the dominance of fracture processes to generating AE in brittle materials. Furthermore, Faisal et al proposed a definition for separating the AE response into stages using a ‘zoning’ technique. Here, absolute raw waveform was smoothed using averaging, and a threshold was then made to form the boundaries of Stage 2 based on 60% of the peak magnitude as shown in Figure 3.12 below.

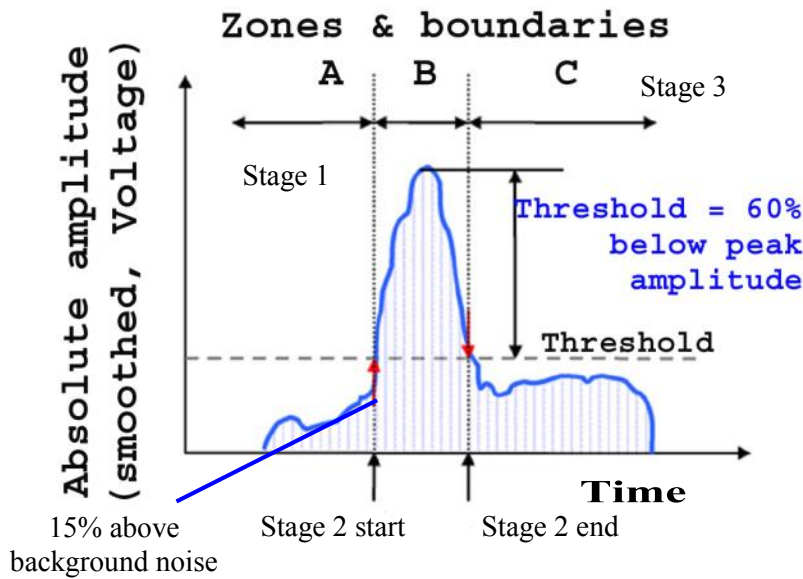


Figure 3.12: Illustration of the ‘zoning’ technique proposed by Faisal et al [155]

Al-Balushi [167] adapted a statistical processing technique for AE monitoring for noisy signal profiles, namely ‘*Energy Index*’ (EI). The technique is based on dividing the signal into segments and calculating the RMS of each (using equation 3.5). This produces a smoothed envelope (consisting of RMS values $RMS_{segment}$) which are then compared to the RMS value calculated for the entire signal length (RMS_{total}) using the equation below.

$$EI = \left(\frac{RMS_{segment}}{RMS_{total}} \right)^N \quad 3.9$$

The number of segments N can be changed depending on the fault signatures and noise profiles.

The technique was applied to steel RE bearing tests, where defects were seeded on the outer raceways using indentations [168]. The technique was able to successfully enhance fault signal components. Furthermore, to demonstrate the ability of the method, Gaussian noise was added to a recorded signal profile before applying the technique. The results are shown in Figure 3.13 below.

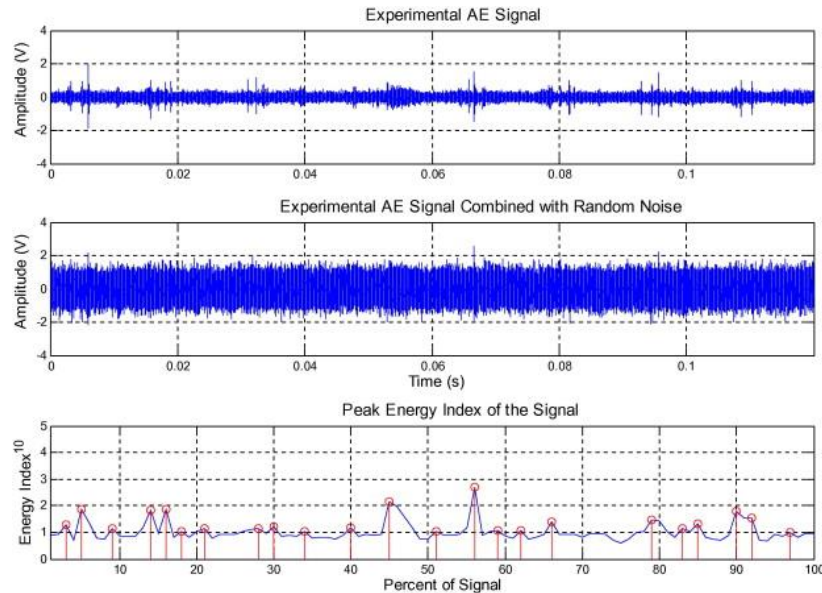


Figure 3.13: Analysis of a AE signal from a defective bearing [168]- the raw signal (upper), with added Gaussian noise (middle) and EI applied (the red circles correspond to peaks of the fault components).

The figure shows the amplitudes of the noise profile exceeds that of the fault components subsequently masking them from a hit-based analysis approach. Correspondingly, the EI technique was able to enhance many of them as the technique is able to expose transient signal features whilst attenuating a stationary background noise profile.

He et al [162] also applied the EI to AE signal profiles from defected bearings to produce an envelope profile. The periodicities of the envelope signal were analysed and compared to the calculated defect frequencies of the bearings (refer to section 3.2.2.1). The method was able to extract the periodicity of faults on the inner races where the wave propagation path to the sensor significantly attenuated the signal⁶.

⁶ The AE sensor was fixed to the outer structure of the test machine

Frequency Domain

Analysis in the frequency domain has been applied to AE raw waveforms in a number of investigations (some notable examples include [160, 169, 170]) and have been used to identify various AE events. This is because various material processes can produce signals with characteristic frequencies.

Price et al [169] found frequency shifts at the onset of pitting failures in oil lubricated steel-steel sliding wear tests from a four ball test machine. Increased signal energy was observed at 50 kHz, and in the region of 180 – 230 kHz as precursors to pitting failure. Wang et al [160] used spectral analysis in the monitoring of hybrid and Si_3N_4 - Si_3N_4 lubricated sliding contacts. High signal intensities with frequency in the range of between 500 and 550 kHz were a result of intergranular and transgranular cracking in the silicon nitride.

Navarro et al [170], applied Vickers indentations to alumina titanate coated steel, finding that plastic deformation produced a signal with low power and frequency at around 40 kHz. Cracking was found to produce signals with significantly higher amplitudes with frequencies in the range 220-280 kHz.

Restall and Wang [11] monitored the formation and progression of ring cracks in compression tests on silicon nitride balls. Using frequency analysis, the peak amplitude range for the formation of the ring crack was found to be between 0.225 to 0.275 MHz. During Vickers indentations made to alumina coated steel, Faisal and Ahmed [171] found that cracking events produced signals with dominant frequencies in the range 200-400 and 550-650 kHz.

3.3 Data Enhancement and Feature Extraction

The previous section reviewed each of the sensing techniques of interest to the project and introduced the signal analysis methods commonly used to identify various physical events associated with material wear. In particular, frequency based analyses have been identified as productive for vibration and AE signals. This section reviews several techniques that can be used to facilitate these analyses and presents an additional technique based on spatial separation.

Subsection 3.3.1 below reviews various techniques to analyse signals based on their frequency profile progression in time, with an associated filtering technique.

Subsection 3.3.2 below reviews a technique supporting the *Time Synchronous Analysis*, a method of signal analysis commonly used outside the RE bearing field, for example in gear monitoring, but which can potentially offer features suitable for RCF monitoring.

3.3.1 Time - Frequency Analysis

3.3.1.1 *Fourier Methods*

The Fourier Transform can transform a signal $x(t)$ into a set of sine and cosine functions with the general form:

$$x(f) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos 2\pi nft + b_n \sin 2\pi nft) \quad 3.10$$

Where a_n and b_n are Fourier coefficients.

A significant milestone in the development for signal processing was the establishment of the ‘*fast Fourier transform*’ (FFT) by Cooley and Tukey [172], and is subsequently able to provide fast computation of the Fourier transform for a digital signal.

A representation of the frequency content of a signal can be provided by estimating the signal power distribution in the frequency range using the *power spectral density* (PSD) estimate. This can be calculated using the nonparametric technique: the *Welch* method [173]. Here the signal is segmented using a window function translated across the signal and the Fourier transform is computed at each translation. An average is calculated across the signal using the squared modulus of the Fourier coefficients. The method is subsequently able to reduce the variance of noise [174].

To reduce border distortion during computation, a window function that provides a smooth profile with low weighted values at the edges is applied. The Hanning window exhibiting this feature is predominately used for analysis [174].

Notably, Wang and Wood [160] applied the Welch method to estimate the PSD of AE signal data recorded in bursts from a pin on disc test using silicon nitride components. A frequency-time profile was produced by estimating the PSD at each burst and arranged in time order; an example is shown in the Figure below.

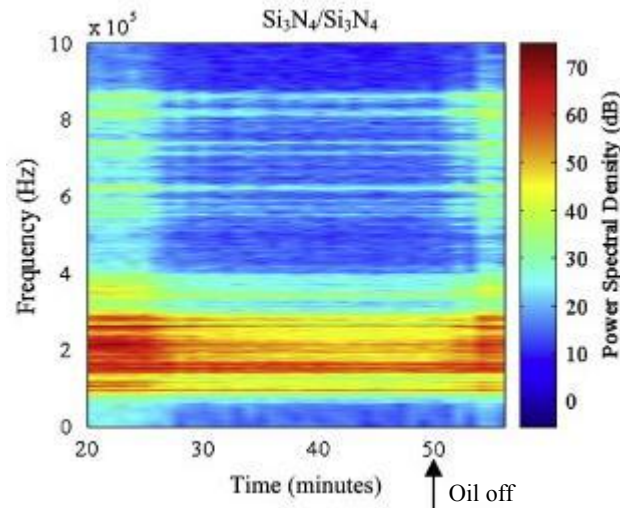


Figure 3.14: Frequency-time profile of AE signals recorded from a pin on disc machine produced using Welch PSD method [160].

The method was able to track the frequencies of the signal in the time domain, subsequently identifying new frequency bands at the onset of transgranular and intergranular cracking after the lubricant was switched off (at 50 mins).

In practice, signals are rarely stationary; instead, fault components are often expressed as short-lived transients consisting of various active frequency components representing only a small proportion of the overall signal energy in a noisy machine [175]. Subsequently, fault components can be significantly masked during the Fourier transformation.

A way to overcome this drawback is the *short-time Fourier Transform method* consisting of segmenting the signal in a manner similar to the Welch method and performing FFT for each segment separately. Therefore, the overall signal profile is represented by a series of segment profiles. However, an inevitable trade-off occurs in the choice of segment size resulting in the difficulty in capturing both slow and fast variations of a signal simultaneously.

To overcome these problems, alternative methods have been developed, namely:

- Wigner-Ville Distribution
- Wavelet Transforms
- Hilbert-Huang Transform

The following subsections reviews these techniques.

3.3.1.2 Wigner-Ville Distribution

The Wigner-Ville Distribution (WVD) $W_s(t, \omega)$ is essentially a correlation of a signal, x with a time-shifted version of itself, forming the instantaneous autocorrelation function. The Fourier transform is applied to the instantaneous autocorrelation function, and is repeated for all possible lags τ , as shown in the equation below:

$$W_s(t, \omega) = \int_{\tau} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i\omega\tau} d\tau, \quad 3.11$$

where t time and x^* is the complex conjugate of signal x .

An important aspect of the WVD is that it does not contain a windowing function as do the Fourier and wavelet frameworks. This feature frees the WVD from ‘smearing’ effects by averaging over a finite windowing function [176]. As a result, the WVD can provide high resolution in both time and frequency domains [177].

However, the quadratic nature of the WVD introduces cross terms for a multi-component signal that can have significant amplitudes and can corrupt the transform space considerably [176]. Various methods have been introduced to mitigate these effects including: pseudo-WVD function [178], *smoothed pseudo-WVD* [179], *Choi-Williams distribution* and *Cone-Kernel representations* [180]. Each is essentially based on the application of smoothing functions (the latter two combining weighted averages).

Other notable methods to reduce cross terms include filtering methods to decompose a signal into mono-components before applying the WVD [181, 182]. Liu [177] and Tang et al [183] combined the continuous wavelet transform, Cohen [184] adopted the wavelet packet technique, and Li [185] and Hui et al [186] applied the empirical mode decomposition method to signals before computing the WVD.

However, the adoption of these methods inevitably results in the loss of time-frequency concentration and significantly increases the latency in their computations [187].

In an investigation, Kim et al [188] compared the smoothed pseudo-WVD with a discrete wavelet transform to analyse the vibration signal of a cracked rotor shaft. The wavelet method was found to enhance resonant vibration modes of the machinery more concisely without corruption from the cross-terms. Subsequently, the wavelet method was able to track a cracked tooth fault in both fixed and variable rotating speed conditions.

3.3.1.3 Wavelet Transforms

The use of wavelet transform techniques has grown rapidly since their establishment by Morlet and Grossman [189] in the 1980s, as it encompasses many signal processing innovations - notably de-noising, image analysis and compression. Their use in the condition monitoring applications has also grown rapidly since its popularisation in the field in the mid-1990s.

Currently three wavelet transform methods are widely used: ‘*Continuous Wavelet Transform*’ (CWT), ‘*Discrete Wavelet Transform*’ (DWT) and ‘*Wavelet Packet Analysis*’ (WPA), and are discussed in turn in the following subsections.

i. Continuous Wavelet Transform

A wavelet is a smooth and quickly vanishing oscillating function, with localization in both frequency and time domains. A wavelet family $\psi_{a,b}$ is the set of elementary functions produced by dilations and translations of a ‘*mother wavelet*’ ψ expressed using the general equation:

$$\psi_{a,b} = a^{-1/2} \psi\left(\frac{t-b}{a}\right), \quad 3.12$$

where a is the scale parameter (representing the frequency of the wavelet), b the translation parameter, and t time.

Correspondingly, a series of coefficients $W_x(a, b; \psi)$ representing both time and frequency components of a signal $x(t)$ is calculated as follows:

$$W_x(a, b; \psi) = a^{-1/2} \int x(t) \psi\left(\frac{t-b}{a}\right) dt; \quad 3.13$$

it describes the similarity between a signal and the wavelet ($\psi_{a,b}$).

The resulting wavelet coefficients can be presented as a frequency-time graph known as a ‘*scalogram*’, illustrating the power intensity of the coefficients and calculated by squaring the modulus of each coefficient (an example is shown in Figure 3.16 below).

Tse and Peng et al [110] have used the scalogram to diagnose various defects of steel RE bearings and compared it to a conventional enveloping technique. The wavelet analysis method was able to extract defects located on ball, inner race and outer race compared to a conventional band-pass filtering technique. The latter was only able to enhance outer race faults due to the generation of large sidebands corrupting the output in the enveloping technique.

To improve the energy concentration of the scalogram in both time and frequency domains, Peng and Chu et al [105] developed a reassignment method based on generating local ‘gravity centres’. Subsequently, the method was applied to vibration data of a pumped storage power

generator unit, enabling various faults to be diagnosed including: stator rub, oil whirl and coupling misalignment. The technique was able to highlight faults at lower signal energy and at an earlier stage than that of the conventional scalogram.

The energy concentration of the wavelet transform in both time and frequency can also be improved by an appropriate choice of mother wavelet; which can reduce the effects of redundancy and noise [190], subsequently enhancing signal features. The mother wavelet should exhibit similarity in both shape and frequency bandwidth to fault signal features of interest [191], enhancing the correlation (inner product) between them.

A variety of mother wavelet bases are available for signal decomposition; some of the most commonly applied CWT wavelets to condition monitoring include the following: ‘*Morlet*’ [100, 103, 192-195], ‘*Gaussian*’ [191], ‘*Daubechies*’ [108, 196] and ‘*Laplace*’ wavelets [197-200]. An example signal-time profile of each is shown in Figure 3.15.

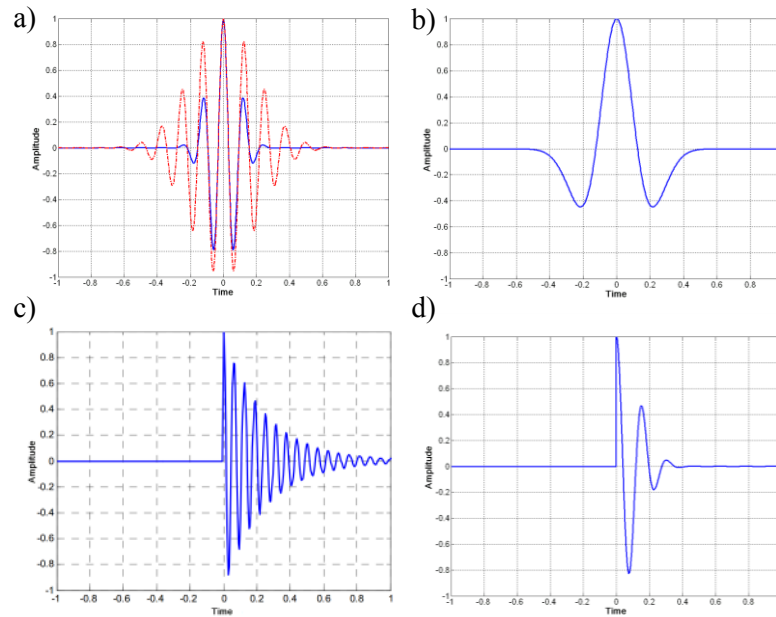


Figure 3.15: Examples of various wavelet functions a) Morlet (with two different β values), b) Gaussian (order 2 - *Mexican Hat*), c) Laplace and d) One sided Morlet wavelet.

Lokesha [194] compared the use of the Morlet and Laplace CWT bases on vibration signal data from a FZG gear test. Both mother wavelets show similarity in shape to that of a mechanical impulse (refer to Figure 3.15 c & d). In particular, the symmetrical nature of the Morlet wavelet was found to enhance the impulsive faults relating to a crack gear tooth, above that of the Laplace.

The Morlet wavelet is essentially a Gaussian window function applied⁷ to a cosine wave function and is expressed in the following equation:

$$\psi(t) = \cos(\pi t) e^{(-\beta^2 t^2)/2} \quad 3.14$$

The exponential decay characteristic of the function shares similarity to the vibration impulse response of a mechanical system [111, 201, 202]. The β parameter can be chosen by the user and adjusted to suit a particular signal.

Lin and Qu [100] investigated the β parameter for the production of an optimal Morlet in the analysis of vibration signal data from a faulty gearbox. Using various β values, signal data was decomposed and assessed by analysing the ‘*sparsity*’ of the coefficients produced at each decomposition. Subsequently, the Shannon entropy was applied to the coefficients using the equation following [203]:

$$Entropy_{CWT} = \sum_{i=1}^N \frac{|C(s, i)|^2}{\sum_{i=1}^N |C(s, i)|^2} \cdot \log_2 \left(\frac{|C(s, i)|^2}{\sum_{i=1}^N |C(s, i)|^2} \right) \quad 3.15$$

where $C(s, i)$ describes the coefficient value of scale s and datapoint i and N is the number of respective data points per scale.

The β value of the Morlet wavelet with coefficients corresponding to the minimum entropy was found to significantly enhance the fault features. This is because the technique can provide an assessment of the energy concentrations in both the time and frequency domains. This is demonstrated in the Figure below, showing a simulated signal decomposed separately using two mother wavelets. The first is an optimal decomposition (Figure 3.16a), with a corresponding low entropy score; the second is a suboptimal decomposition (Figure 3.16b), with a corresponding high entropy score, both are based on data from [201].

⁷ The Gaussian wavelet (unlike the Morlet wavelet), is the derivative of the n th order of a Gaussian function. A profile illustrating the 2nd order is shown in Figure 3.15a above.

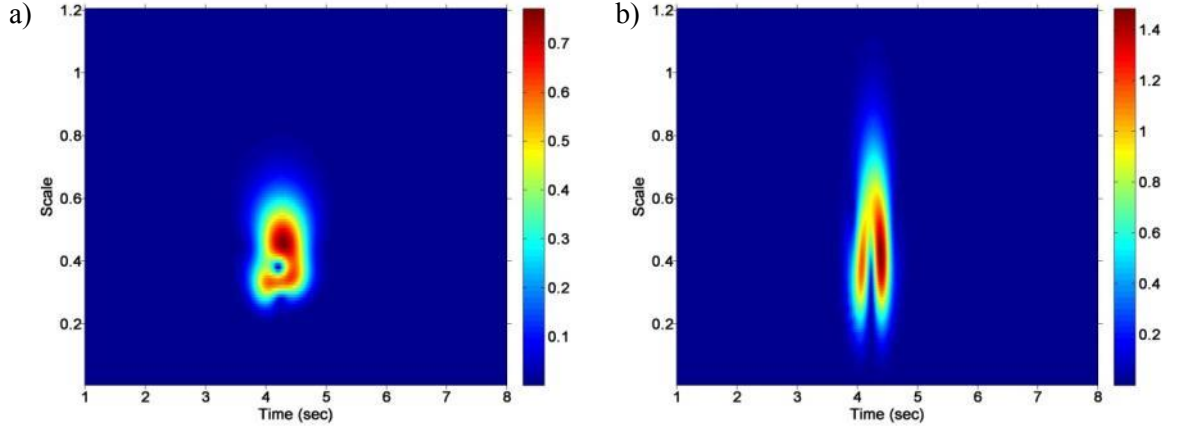


Figure 3.16: Frequency-time profile (scalogram) comparisons of a signal decomposed using CWT using a wavelet producing a) minimal entropy b) higher entropy values. Figures reproduced using the data provided in [201].

The optimal decomposition shows a frequency-time profile that separates the signal into distinct components (identified in Figure 3.16a as three prominent peaks i.e. dark red). In contrast, the profile in Figure 3.16b is much more extensive in the frequency domain and also varies prominently in time.

Because of these characteristics, entropy analysis has been used to select suitable mother wavelets to analyse the time-frequency characteristics of various signal data (notable examples include [195, 201, 202, 204-209]).

Lin and Zuo [205] used the Shannon entropy minimisation method to select an appropriate β value for the decomposition of vibration signal data recorded from a gear box with a broken tooth. A wavelet scale was then selected as a band-pass filter for envelope analysis. Because of its sensitivity to mechanical impulse components, kurtosis was used to select the appropriate scale.

Kurtosis is calculated for wavelet coefficients using equation 3.16 below [203] (adapted from equation 3.6, barred symbols indicate mean values).

$$Kurtosis_{WT} = \frac{(N) \sum_s \sum_i \left(C(s, i) - \overline{C(s, i)} \right)^4}{\left(\sum_s \sum_i \left(C(s, i) - \overline{C(s, i)} \right)^2 \right)^2} \quad 3.16$$

The selected scale was able to enhance the signal component relating to a broken tooth.

Because of its ability to enhance impulse-like faults, the method above has been applied to the analysis of vibration signal data in a variety of investigations, such as identifying cracked teeth in gears [201] and seeded defects in a RE bearing [206] from noisy backgrounds.

In a recent investigation, Hemmati [203] applied the kurtosis-to-entropy ratio (KER) to evaluate the ability of various mother wavelets to enhance RE bearing faults in AE signal data. KER was calculated using the following equation:

$$KER = \frac{Kurtosis_{WT}}{Entropy_{WT}} \quad 3.17$$

The resulting method is thus sensitive to both impulsive features and energy concentrations. The mother wavelet producing the highest KER value was subsequently able to significantly enhance the impulse signal from a noisy background profile. Rafiee et al [209] adopted the KER statistic to compare over 300 mother wavelets in their ability to enhance fault features in vibration signal data from RE bearings.

Complex Wavelets

A complex wavelet ψ^* is formed by taking the original wavelet and combining it with its phase as the imaginary component, thus forming a variable that provides information on both amplitude and phase simultaneously. A phase spectrum can be generated, similar to the aforementioned scalogram, describing local phase at various scale values. The method is able to highlight signal discontinuities. Furthermore, an envelope can be deduced from the modulus of the real and imaginary parts of the coefficients (refer to Equation 3.7 in Subsection 3.2.2.3).

Staszewski and Tomlinson [210] were able to detect a damaged tooth in a spur gear using the phase spectrum technique. Similarly, Boulahbal [211] adopted this method with the scalogram for the identification of cracked gears. Abhinav [193] applied the complex Morlet wavelet to monitoring planetary gear system in helicopters, using features extracted from healthy and faulty gear plates. Nikolaou and Antoniadis [103] used the complex Morlet wavelet to demodulate the vibration response from a tested steel ball bearing to extract an outer race fault. Rubini and Meneghetti [212] compared the conventional enveloping method and the complex CWT to detecting pitting faults in roller bearings and found that the former became insensitive to the defect as it smoothed.

Despite these successes, the CWT method can suffer from overlapping terms, providing difficulty in the interpretation of the results [213]. The discrete wavelet transform DWT has been developed to reduce these features.

ii. Discrete Wavelet Transform

To overcome the drawbacks CWT and reduce the latency in computation, Daubechies and Mallat [214] proposed a discretisation technique by setting:

$$a = a_0^m, b = na_0^m b_0, \quad 3.18$$

(where m and n are integers), and forming the discrete wavelet transform given by:

$$W_x(m, n; \psi) = a_0^{-m/2} \int x(t) \psi(a_0^{-m} t - nb_0) dt \quad 3.19$$

The signal is convolved with a low-pass filter and a high-pass filter resulting in two orthogonal vectors: the *approximate* (cA) and *detailed* (cD) coefficients respectively.

Warren and Liao [215] applied Daubechies (db) wavelet 1 (db1) to the AE response from a diamond grinding wheel against alumina, extracting features to distinguish between a ‘sharp’ and ‘dull’ grinding wheels. Purushotham [216] analysed a vibration signal from ball bearings using db2 wavelets to detect single and multiple defects with high accuracy. Chen et al [217] applied the db6 wavelet to extract seeded bearing faults from the machine’s background noise profile. The faults were found to be more prominent at particular coefficients.

However, the DWT in turn suffers from large shift sensitivity and low resolution in the high-frequency region, significantly limiting its usage [187].

iii. Wavelet Packet Analysis

Wavelet packet analysis (WPA) is an extension to the DWT to eliminate some of its drawbacks, particularly in the low-resolution of high-frequency components [104], whilst still providing relatively low latency in computation. The method is similar to DWT in its analysis, however both the approximation and detail coefficients are further split forming a decomposition tree composed of ‘*packets*’. An example is illustrated in Figure 3.17 below.

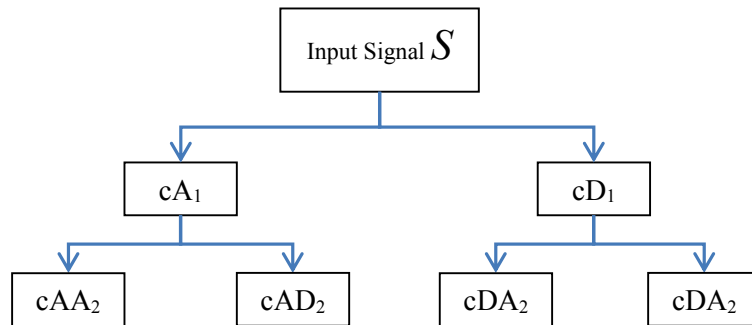


Figure 3.17: Illustration of WPA decomposition tree to level 2.

Decomposing a signal to increasingly deeper levels can reduce its time resolution [104]. The decomposition level is therefore an important aspect of the technique, and can be subsequently assessed using various methods as described below.

Hu et al. [196] assessed the most salient frequency band for a decomposition based on the wavelet packet demonstrating the highest signal power. The method was applied to analysing vibration signal data recorded from RE bearings with various faults. Similarly, Chendong [95] applied the method to extract gear faults from vibration data. Nikolaou [104] determined the optimum levels from the signal power of the coefficients above a threshold. The method was applied to enhance fault signal components in vibration signals recorded from a RE bearing with seeded faults.

Lei et al. [218] decomposed a vibration signal using the WPA. For each packet, the kurtosis value was calculated using the corresponding coefficients. The method was subsequently able to identify weak impulses corresponding to faults. This was attributed to the powerful filtering property of the WPA, demonstrating steep roll-off rates which were subsequently able to filter out noise effectively.

Wickerhauser and Coifman [219] proposed a fast algorithm to ‘prune’ a wavelet packet tree. The method was based on the assessment of the Shannon entropy at each adjacent packet compared to that of their expansion coefficients. Brechet et al [220] later adapted the method to compare the minimum entropy values in the computation of various wavelets to find an appropriate mother wavelet and corresponding packets for filtering.

Having chosen an appropriate mother wavelet base using the KER technique as described in Subsection i. above, Hemmati [203], further applied the KER measure to WPA nodes in the decomposition of an AE signal data from a bearing with a seeded fault. The packet corresponding to the highest KER values were subsequently used to enhance the impulse signal components attributed to fault over rolling.

3.3.1.4 Hilbert-Huang Transform (EMD method)

The Hilbert-Huang transform is one of the most recent of the time-frequency techniques considered. It is based on the well-established Hilbert transform, and is able to emphasise local properties of the analysed signal.

An important property of the Hilbert transform is that if the signal is of a mono-component character, then the time derivative of the instantaneous phase can deduce the frequency of the signal by the following:

$$\omega(t) = \frac{d\phi(t)}{dt} \quad 3.20$$

However, for signals produced from real systems, the Hilbert transform is rarely immediately applicable instead containing multiple components. In order to deal with this, Huang [221] proposed a sifting method to separate the signal into mono-components; subsequently a signal $s(t)$ could be expressed by the following:

$$s(t) = r_M(t) + \sum_{k=1}^M a_k(t) \cos \phi_k(t) = \text{residual}_M + \sum_{k=1}^M IMF_k \quad 3.21$$

Where $\sum_{k=1}^M IMF_k$ describe the mono-components referred to by Huang as ‘*intrinsic mode functions*’ (IMFs). To produce the IMFs of the signal, Huang proposed an iterative sifting process: the **Empirical Mode Decomposition** (EMD) forming the IMFs in a sequential manner as shown in Figure 3.18 below:

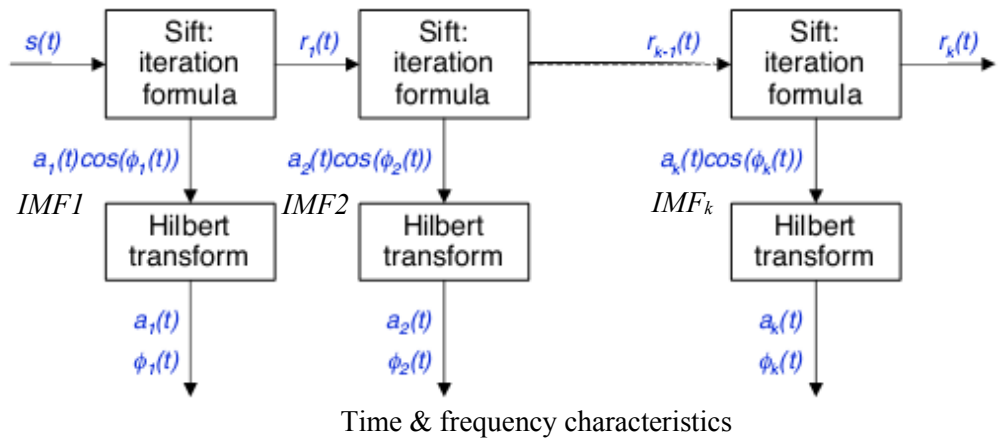


Figure 3.18: Illustration of the iterative sifting process proposed by Huang [222].

The signal is initially decomposed into the first IMF (representing the highest frequency components) of the signal and the residual $r_1(t)$. This residual is then itself decomposed producing the second IMF (representing the next highest frequency signal component) and the residual $r_2(t)$. This process is repeated until a final IMF is produced leaving a residual

component $r_M(t)$ whereby no further decompositions are made. The process results in a decomposition similar to dyadic filter banks resembling that of wavelet based processing [223].

To determine each of the IMFs and residual components, Huang developed a sifting procedure based on the following processing:

1. Identify the local extrema in the test data $x(t)$ (for the initial iteration $(t) = s(t)$) separating the maxima and minima.
2. Fit cubic splines to the maxima and separately to the minima producing upper and lower envelopes ($e_{upper}(t)$, $e_{lower}(t)$ respectively).
3. Calculate the mean of the two envelopes $m(t)$ using:

$$m(t) = \frac{e_{upper}(t) + e_{lower}(t)}{2} \quad 3.22$$

4. Extract a *proto-IMF* $c_n(t)$ using:

$$c(t) = x(t) - m(t) \quad 3.23$$

The steps are repeated on the proto-IMF (i.e. $c(t)$ becomes the test data $x(t)$) until the proto-IMF satisfies some conditions. Ideally, for the proto-IMF to be considered an IMF it should consist of an equal number of extrema and zero crossings, or differ at most by one, and the mean value of the envelopes defined by the local extrema should be zero [221].

The details of these processes are described in detail by Huang [221] and Meeson in [222].

To separate the amplitude $a_i(t)$ and phase $\phi_i(t)$, the Hilbert Transform $H(t)$ can be applied to the IMFs by:

$$H(t) * (a_i(t) \cos \phi_i(t)) = (a_i(t) \sin \phi_i(t)) \quad 3.24$$

And subsequently the envelope $a_i(t)$ and phase $\phi_i(t)$ can be calculated by the following equations:

$$a_i(t) = [(a_i(t) \cos \phi_i(t))^2 + (a_i(t) \sin \phi_i(t))^2]^{1/2} \quad 3.25$$

$$\phi_i(t) = \text{atan}[a_i(t) \sin \phi_i(t) / a_i(t) \cos \phi_i(t)] \quad 3.26$$

The calculations described in 3.24, 3.25 and 3.26 are identical to the common demodulation techniques used in vibration envelope analysis described in section 2.3.

The EMD method is described by Yang [224] as the best way to deal with the intermittent signals thus far, as it is able to present the local character better than the wavelet method, without the production of cross terms as the WVD method. Additionally, decomposition is based on the local characteristic time-scale of the data and subsequently described as adaptive [221].

Liu [225] compared the EMD method with the complex Morlet wavelet to analyse vibration signal from a gearbox. The EMD method was able to extract the low frequency transients resulting from tooth cracking. The CWT method was thought to have suffered from its inherent poor time ability in representing low frequency transients. Dong [199] employed the EMD method to deconstruct vibration data from cracked rotor beams, the Laplace wavelet was then applied to the first three IMFs to analyse the modal response.

Ruqiang and Gao [226] applied the technique to vibration data of RE bearings, producing a frequency-time graph for both healthy and defected bearings, as shown in Figure 3.19 below. Impulses at 6 ms apart corresponded to an outer race defect and were clearly identifiable as illustrated in Figure 3.19b. Furthermore, the method was able to track the fault as it progressed, resonating different frequencies and was attributed to the adaptive capability of the method.

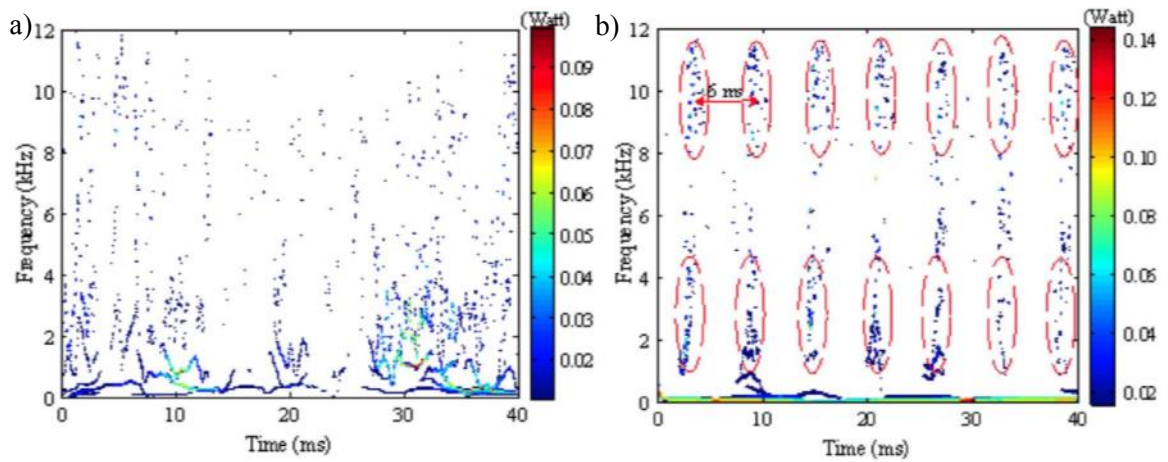


Figure 3.19: Frequency-time profiles of the Hilbert-Huang technique applied to vibration data from a RE bearing a) healthy, b) with an outer race defect [226].

Li [227] highlighted the suitability of the HHT technique for acoustic emission signal analysis in bearing monitoring, extracting the transient signal components relating to various faults. He et al [228] applied the HHT method to the acoustic emission signals for monitoring full ceramic bearings. The first two IMFs were computed and various statistics including kurtosis and RMS were calculated for both. Signal data produced during faulty bearings were compared to data from a healthy test using a 'k-nearest neighbour condition based indicator'. The method was able to identify seeded faults⁸ including outer race, inner race, ball and cage, based on the magnitude of the computed values.

The EMD method, because of its iterative nature, produces increasing computation demands with the increasing data length, potentially leading to significant latency during computation.

⁸ The size of the defects were not detailed

3.3.2 Time Synchronous Analysis

Time synchronous analysis (TSA) is a technique for analysing periodic signal waveforms and can therefore be particularly useful in the analysis of signals related to rotating machinery.

The technique requires the use of a trigger signal that is phase-locked with the angular position of a rotating shaft in the system being monitored. This is often in the form of a once-per-rev tachometer, or an optical encoder (with multiple separable pulses per rev) for increased angular precision⁹.

For a once-per-rev tachometer, the start and end points for each shaft revolution are correspondingly identified on the signal data. For each revolution, data length r , is interpolated into a vector of length m . This is iterated for a number of rotations n , producing a matrix of signal data of size $m * n$ synchronised to the rotations of a particular rotating component within a system. The procedure is illustrated in the Figure 3.20 below.

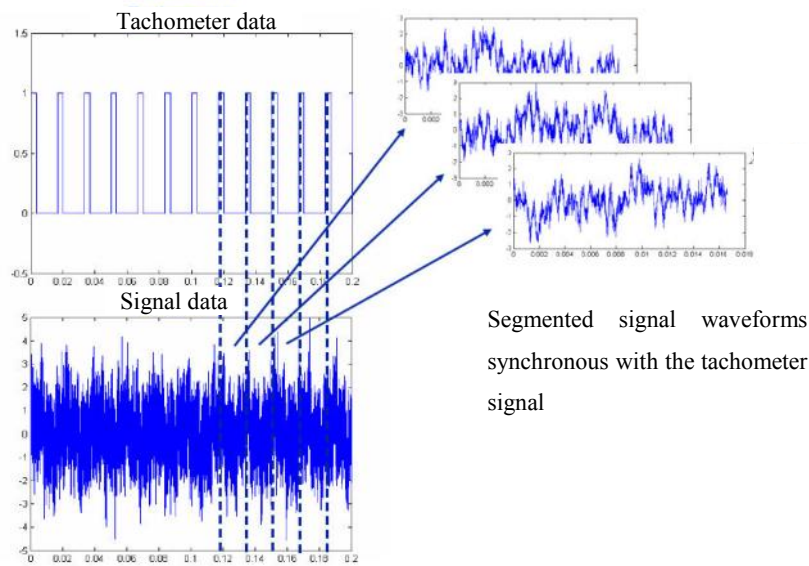


Figure 3.20: Illustration of the TSA technique.

Several interpolation functions are available, and commonly applied methods are listed below in order of computational demand [231] (starting at the lowest) with a brief description:

- *Linear*, a straight line is assumed between data points.
- *Cubic*, a curve is fitted to the data, in the form of a cubic equation, which is used to generate an interpolated data point.

⁹ Techniques are available to monitoring rotations of various systems without a trigger signal and instead are based on the identification of characteristic and periodic components within a signal. Because of the characteristic vibration signal they produce during meshing, they have been applied to gear vibration monitoring [229, 230].

- *Cubic Spline*, a series of separate cubic functions is formed between sets of data points. An interpolated data point is found which has a smooth first derivative and continuous second derivative of the proceeding and following functions.
- *Fourier Based*, the discrete Fourier Transform is calculated and zero-padding is applied to increase the signal length accordingly, before transforming back to the time domain.

Decker [232] reviewed the first three techniques for experimental vibration signals from damaged gears. The cubic spline method produced significant latency during computation with little improved accuracy in the diagnosis. More recently, however, Bechhoefer and Kingsley [233] compared all the above methods on gear vibration data with chipped gear teeth. Using various statistics applied to the averaged data, the authors assessed each method in their ability in enhancing fault features. The Fourier based technique was found to outperform that of the other interpolation methods considerably, as it was able to interpolate signals with greater accuracy.

TSA has been applied to wear monitoring using electrostatic charge sensing in various investigations [125, 126, 141, 146]. Harvey et al, [126] applied the TSA method to investigate the effects of debris in the monitoring of lubricated steel contacts using a pin on disc machine. Two tests were performed, in the first, the applied load was gradually increased from 10 N to 90 N; in the second steel debris was seeded on the disc at particular times. A background average signal recorded from a 'fault free' disc was subtracted during the TSA computation and subsequently angular-time profiles, showing the intensity of the signal (indicated in colour) in both the angle of rotation and time were produced. The results are shown in Figure 3.21 below.

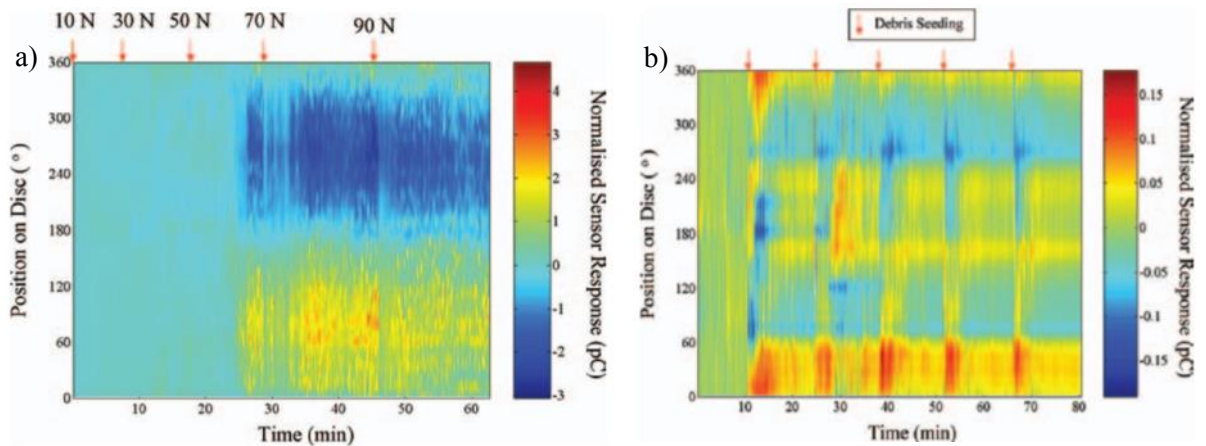


Figure 3.21: Angle-time profiles, for the two test types a) gradual increasing load, b) debris seeding.

The technique was able to show the dynamic charge features that were generated during debris production at a threshold load value <70 N (in Figure 3.21a), and during the debris seeding (Figure 3.21b).

By subtracting a background average signal, Harvey et al [126], reduced the amplitude variation from attitude changes of the disc and benign wear modes, subsequently enhancing the dynamic charge features attributed to debris. Booth [125] adapted the technique to electrostatic monitoring of charge at a cam follower, where the once-per-rev variation in distance from the surface to the sensor was considerable.

Powrie [141, 146] applied the TSA technique to electrostatic WSS data from FZG gear scuffing tests. The technique exposed the teeth meshing response and an increasing level of charge at the onset of scuffing [141]. In a reanalysis, Powrie [146] subtracted a background average profile, and subsequently was able identify adhesion wear on various teeth using in the angular-time profiles. This was indicated by increased signal intensities at the locations corresponding to the adhesive wear

3.3.2.1 Time Synchronous averaging

For a signal matrix of size $m * n$ synchronised to the rotations of a shaft or component, a mean can be calculated, providing an approximation of the cyclo-stationary periodic signal response described for a single rotation. As the number of shaft rotations in the calculation increases, the ensemble mean value approaches an ergodic synchronous waveform, and the standard deviation of the noise decreases by the following equation [233]:

$$\text{Standard deviation}_{\text{Noise}} = 1/\sqrt{n} \quad 3.27$$

The technique can provide a method to decouple signal features of various rotating components in a machine. This feature has made it particularly useful, for example, in the vibration monitoring of gear systems where signal components corresponding to individual gears can be analysed [233-242].

However, the computed average can still be dominated by non-fault related signal features which are synchronous to the rotation of the component being monitored and therefore unable to remove them. This can mask fault features exhibiting low signal power and amplitude. This is particularly evident in the vibration monitoring of gears where ergodic waveforms are dominated by meshing and shaft harmonics. Subsequently frequency filtering methods are used to remove them. These include filtering both shaft and meshing harmonics using a digital filter-forming a ‘*residual*’ TSaA signal. A ‘*difference*’ signal can also be produced by further filtering of the 1st order meshing sidebands from the residual. These techniques have been shown to enhance fault features exhibiting low signal energy [233-239].

Yip and Markis [240] using vibration data from in-service planetary gearbox failures, compared the use of TSaA applied directly to raw waveform and data analysed using the CWT method.

The CWT methods were able to decouple and enhance fault signatures, otherwise masked by synchronous noise components (e.g. meshing signals), providing higher accuracy in their diagnoses.

Li et al [241] and in a closely related study by Boulahbal et al. [211], applied the complex Morlet wavelet to vibration data of gears during start-ups with large speed differences. The phase spectrum was computed for vibration signal data recorded from a gear with a cracked tooth. Cracked teeth were identified by prominent changes in the outputs at their angular locations. This was attributed to the increased deflection of the cracked tooth during meshing. This technique has also adopted by Baydar [242] in the analysis of vibration data from helical gears. Selected scales above the frequencies of meshing signal components were extracted, significantly enhancing faults features relating to both cracked and worn teeth.

McFadden and Toozhy [243] developed a method to apply TSaA to vibration analysis of RE bearings. The principle was to measure the relative speed of the cage to the shaft, using separate tachometers to monitor both shaft and cage rotation to extract vibration data relating to multiple damages on inner races. Komgon et al. [244] used an approximation of the cage speed, in relation to the measured shaft speed for monitoring in-service RE bearing data. Here raw vibration data was convolved with periodic functions corresponding to each of the defect frequencies and was able to enhance signal intensities of faults.

3.4 Summary

This chapter reviewed in two parts the various signal processing strategies relevant to the project.

The first part (Section 3.2) reviewed how each of the three media of interest (vibration, electrostatic and AE) could be disturbed by various physical processes associated with wear, such as developing RCF faults and, in particular, the type of signal profiles they might generate. The general findings presented there are now, as a way of summary, translated into a specific situation involving a hybrid RE-raceway with a developing RCF fault (as reviewed in detail in the previous chapter) and presented in Figure 3.22 below.

The figure presents a life cycle of a RE in a hybrid bearing comprising of several stages of wear development. For each stage it illustrates the state of the contact and the physical processes that contribute to each of the media of interest.

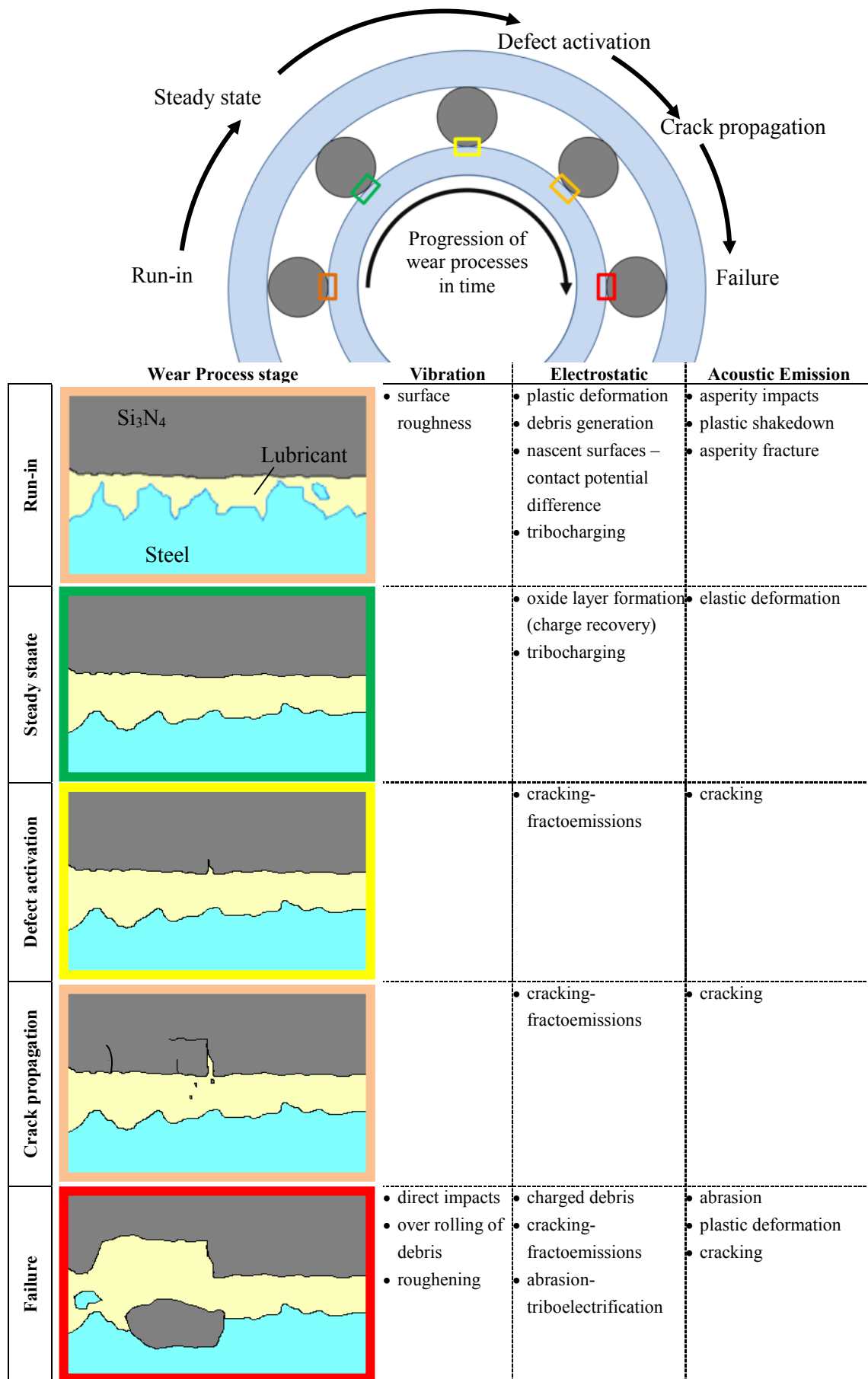


Figure 3.22: A life cycle model of a RE in a hybrid bearing indicating dominant physical and their signal-generation features.

The second part (Section 3.3) reviewed the techniques that can be employed to enhance particular signal profiles based on frequency and spatial distinctions.

The general conclusions from that can be drawn from both parts are as follows.

Vibration presents relatively low signal levels until the onset of material loss leading to surface deformation and a direct contact between the RE and the steel race. The contact then excites the resonant frequencies of the surrounding structure in a way amenable to frequency based analyses. The review highlighted that the WVD technique was poor at resolving fault signatures, but the wavelet (particularly CWT based) and EMD based techniques were found to be suitable filtering methods.

In *electrostatic* media, tribocharging can produce significant profiles, distributed over the entire ceramic surface and therefore presenting a prominent background. However, crack generation, being the dominant process during the fault development, can produce significant charge effects through fractoemissions. The resulting fault signature should remain localised due to the non-conductivity of silicon nitride, but the high levels of charge created through fractoemissions can ionise the surrounding matter. The essence of the data enhancement/ filtering is to reduce the background noise, using its known profile such as in TSA. In addition, filtering techniques based on spatial separation could distinguish the fault profile from the background due to its localised character.

Crack propagation should produce prominent signatures in *AE*; the transmission path of the signal is also important. The running processes of the device operation may produce a significant background, therefore the essence of the data enhancement/ filtering will be to reduce noise. For this a potential technique was found to be based on the EI technique, possibly aided by filtering methods; of these the WP presents themselves as the most suitable because the WVD and CWT have been found to be inefficient, and EMD possesses high levels of computational latency for large data sets. Various techniques have been identified for the wavelet optimisation, in particular using the KER assessment.

Chapter 4. Hardware Preparation

This chapter details the hardware (both mechanical and electronic) and its configuration for the execution of the (mainstream) tests. The first two sections cover a test machine (Indentation and Twin Disc respectively), its sensors and the data collection systems. The subsequent section presents the details of the test samples. Finally, the tools used for separate inspections of the samples are detailed in the last section.

4.1 Indentation Machine

The machine was used in two modes:

- a) to produce controlled defects in the silicon nitride specimens for the Production Runs
- b) to generate ‘calibration’ AE profiles to be used to determine the parameters for the AE sensor processing method.

For mode a) only the standard Vickers machine was used; for b) it was configured with an AE sensor and connected to a data acquisition system.

4.1.1 The Standard Machine

A standard macro Vickers hardness machine, model HTM 356 manufactured by Vickers Ltd, was used to generate indentations on the surfaces of the silicon nitride specimens. The tip of the indenter was a standard diamond square-based pyramid with 136° between opposing side faces. Indenter loads were set using selected weights in the range of 1 to 30 Kg which the machine applied through a system of levers to the indentation tip. The indentation process, from when the tip of the indenter first contacted the surface of the specimen to its eventual withdrawal from the specimen, was timed to be approximately 11.9 seconds at 10 kgF load.

4.1.2 Configuration and Instrumentation

The additional reconfiguration of the machine for the generation of the AE profiles was as follows.

The resulting set-up is shown in Figure 4.1 in a steel sample configuration. A cylindrical base plate was machined from mild steel with dimensions 100 mm length and 203.2 mm (8”) in diameter to provide an attachment for the indentation specimen and also for the AE sensor. The flat surfaces of the plate were machined smooth and parallel to ensure direct transmission of the AE signals, with minimum distortion. Also, in order to ensure that the AE signal path was similar to that in the Production Runs, the specimen was positioned at the plate centre where

PAO4 oil was applied as a couplant between the two interfaces, thereby producing the path specimen→ oil→ steel→ sensor (i.e. similar to the Twin Disc machine).

The AE sensor was the same model as used on the Twin Disc machine (PAC WS α - see description in the relevant subsection of 5.2.3 below). The sensor was glued flush with the outer edge of the plate using epoxy based glue (also used to attach the sensor to the Twin Disc).

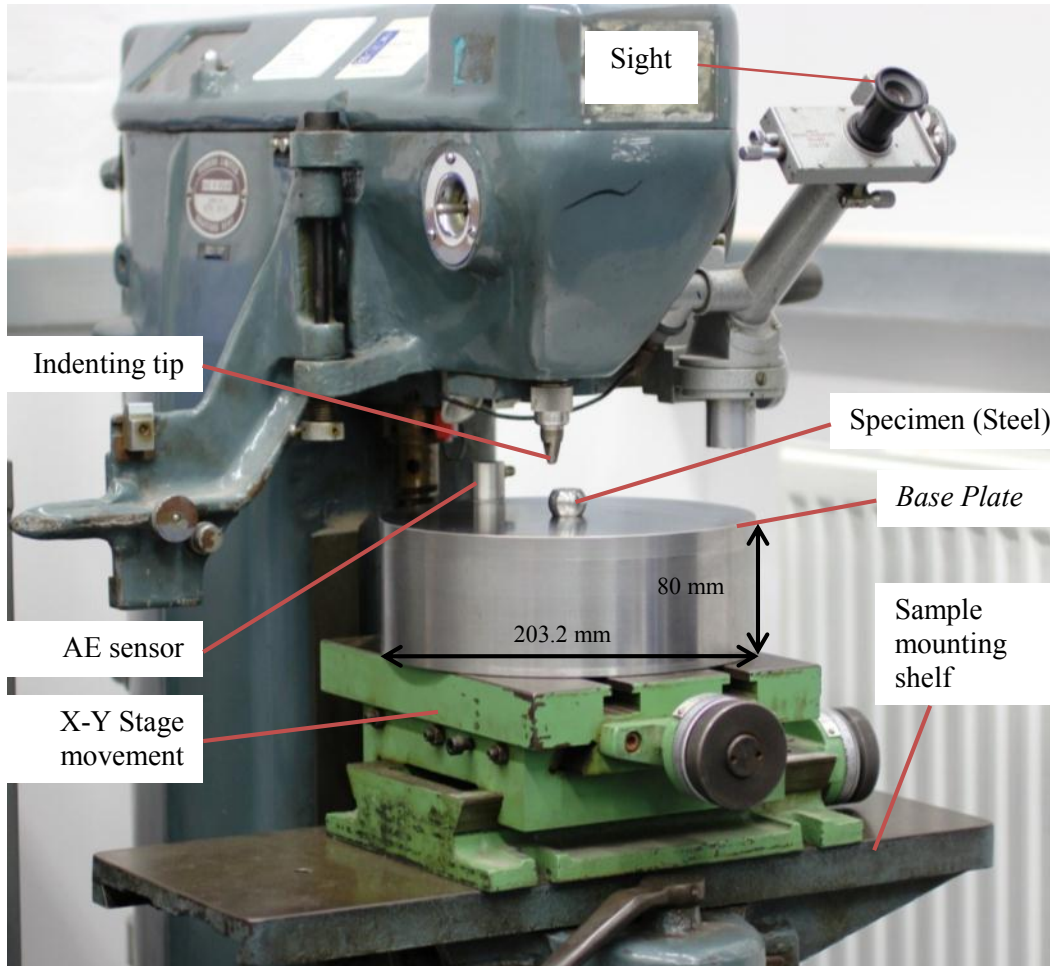


Figure 4.1: A photograph illustrating the experimental hardware for Vickers indentation tests with a steel ball sample.

4.1.3 Data Acquisition System

The sensor was connected up in a standard fashion as recommended by the AE equipment manufacturers (Physical Acoustics Corp. - PAC). The data collection device was a dedicated PC with a PAC 2/4/8 pre-amplifier set to a gain of 66 dB. An on-board digital-to-analogue converter on the PC (a PAC PCI-2 acquisition card) acquired data at a resolution of 18 bits and a sampling rate of 5 MHz. On-board analogue high-pass and low-pass filters were set at 0.1 and 1 MHz respectively. Acquisition was controlled by the PAC *AEwin version 3* software and raw AE waveform signals were recorded for each of the indentation tests performed¹⁰.

¹⁰ The setup is the same as for the AE sensor in the Twin Disc machine as described in subsection 5.2.3.3.

4.2 Twin Disc Machine

As the Twin Disc machine was new to nCATS and had to be additionally instrumented with sensors, its actual suitability had to be confirmed in a feasibility test. This identified the need for further modifications to both the machine and the instrumentation.

The subsection below presents the details of the original ('standard') machine at the start of this research. The subsection following describes the required modifications to the machine and instrumentation; these resulted from the feasibility test performed with the original instrumentation the details of which are described in Appendix B. The resulting modified machine and instrumentation are described in subsection 5.2.3.

4.2.1 The Standard Machine

The Twin Disc machine, a Phoenix Tribology TE74 machine shown in Figure 4.2 below, incorporates two disc specimens mounted on parallel shafts with 40 mm between their centres. Each shaft can be driven independently, each by a 5.5 kW three-phase motor controlled by vector drives. In this configuration, sliding was induced to the contact by reducing or increasing the rotational speed of the lower specimen relative to the upper. Motor speeds were restricted to 0– 3,000 RPM; however, the upper specimen shaft could be driven up to 6,000 RPM by the incorporation of a 2:1 pulley ratio between the drive shaft and its motor. The coupling between the lower specimen shaft and its motor incorporated an in-line torque transducer, measuring the input torque to the system.

The machine was capable of applying up to 12 kN to the contact via a servo controlled pneumatic actuator incorporating a force transducer feedback. The actuator applied load to the contact through the action of a pivot mounting the load arm (see Figure 4.3). Lubricant was pumped directly to the specimen contact and the support bearings (see Figure 4.3) by a recirculation line incorporating a 40 µm filter downstream of the test cell. A heating element in the oil sump heated oil to a pre-set temperature from ambient up to 180°C. To avoid catastrophic failures to the machine and specimens, the machine is equipped with a vibration sensor mounted on the supporting bench near the test cell. The sensor provides an adjustable threshold setting for an automatic shutoff (different settings were used for different tests).

The machine was connected to a dedicated PC installed with Phoenix Tribology's software *Compend v2000* which controlled and monitored the testing parameters such as torque, oil temperatures, load and the shaft speeds could be recorded by the program up to a sampling frequency of 10 Hz. Using this software, a sequence of test conditions could be pre-programmed for automated control during tests.

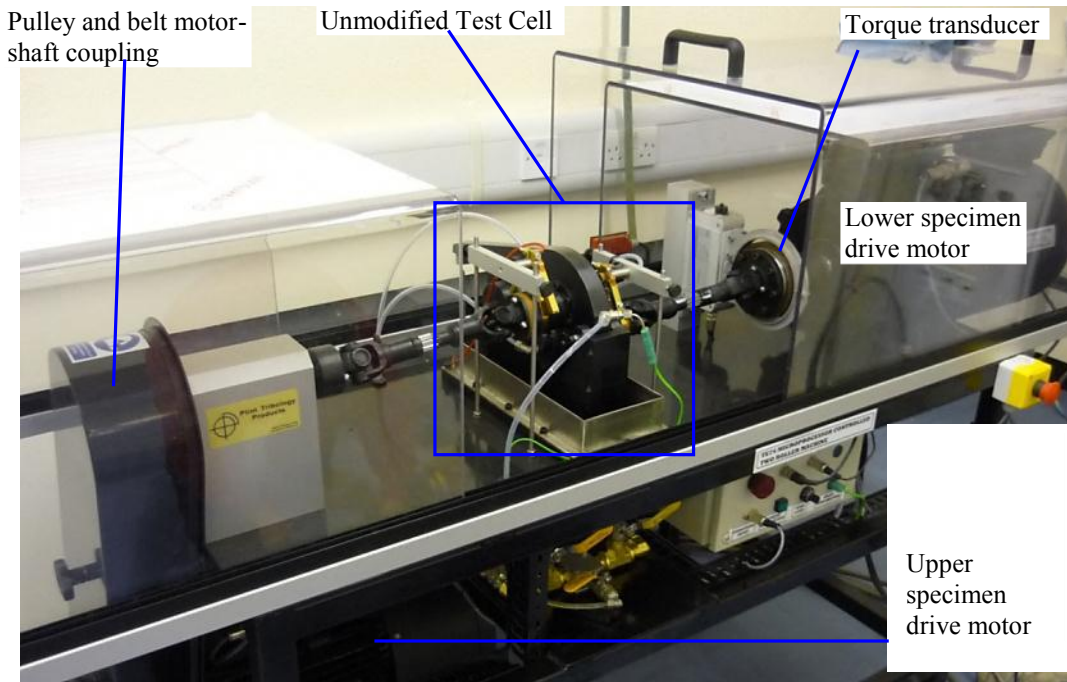


Figure 4.2: A photograph of the Twin Disc machine with the original unmodified test cell

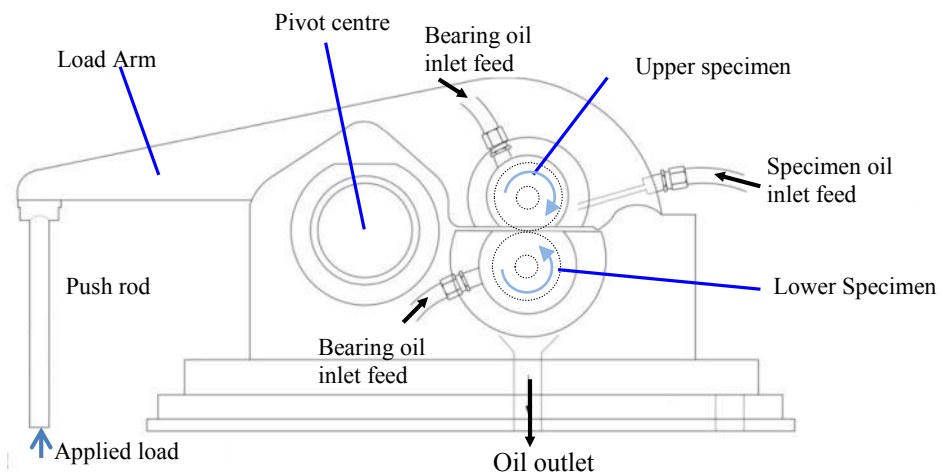


Figure 4.3: An illustration of the side on view of the original test cell and the specimen locations.

The standard machine was equipped with sensors to enable the feasibility test to be performed. The sensor arrangements are shown in the figure below.

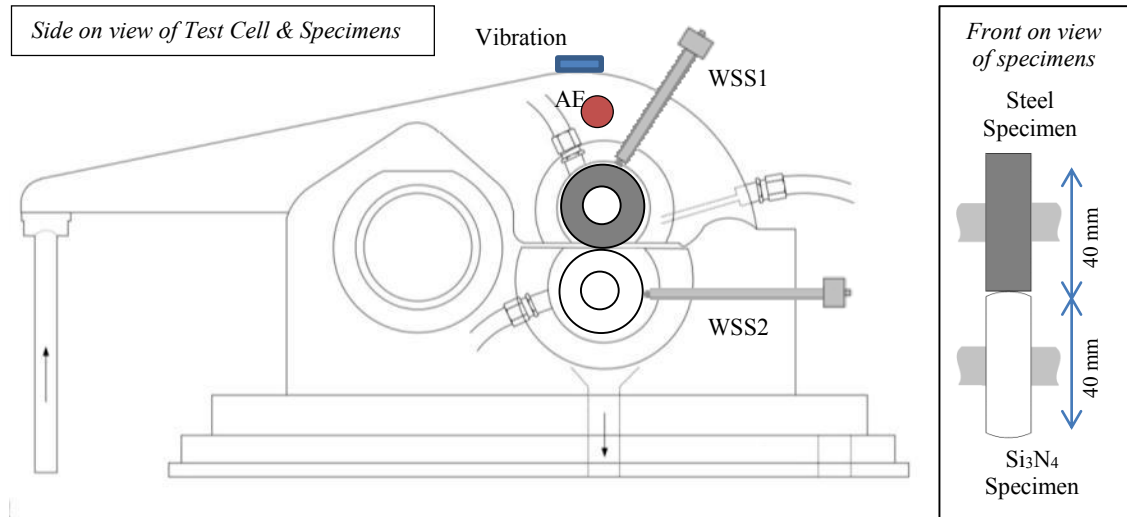


Figure 4.4: Schematic of the instrumented test cell in the original machine configuration.

WSS1 was used to monitor the charge on the surface of the upper specimen, whilst WSS2 monitored charge on the lower specimen.

4.2.2 Modification Details

The standard machine as described above was instrumented with sensors and subjected to an initial feasibility test. The details of this are presented in Appendix B together with the analysis of the results and the modifications identified from the test. The latter were as follows.

- Improved shaft and specimen mounting arrangements for more accurate and consistent/repeatable positioning and avoid multiple silicon nitride specimen breakages.
- Better positioning of the AE sensor to shorten and simplify the AE wave path through the machine.
- A new WSS head for the ceramic specimen to reduce the background noise level
- The silicon nitride specimen to be made from bearing-relevant material with a surface finish of adequate smoothness.
- A new WSSs and vibration signal acquisition system to avoid system freezing and data loss, and enable more advanced signal processing methods.

The solutions to these requirements were implemented by several modifications to the test cell, the sensors, the specimens and the data acquisition system. These are now described below.

4.2.2.1 Test Cell Changes

A key design decision was to avoid altogether the taper based specimen mounting for the ceramic specimen; instead, a straight-hole mounting on a redesigned shaft was used. In order to drive the specimen, it was equipped with anchoring keyways that locked into a new driving face attached to the new shaft (see Figure C.1 in Appendix C). The redesigning of the specimen shape also provided an opportunity to reduce its overall diameter to make it similar to the actual size used in industrial bearings (other researchers also used specimens with similar size as reported in the literature) and achieving a higher number of revolutions for each steel specimen cycle. The latter was beneficial for accelerating the fatigue in the ceramic specimen in the tests. As a result the steel specimen size had to be increased correspondingly, to maintain the 40 mm distance between the specimens' axes. In contrast, as each WSS was held in place by a grub screw it could be easily adjusted to the changed sample dimensions.

The changes in the shaft required a repositioning of its supporting bearings. Changing one of the bearings to a deep-groove bearing also significantly reduced the axial play of the shaft, thus increasing the accuracy of the specimen positioning. Finally, the changes to the bearing housing also allowed a new base for the AE sensor to be provided, much closer to the contact.

The detailed specifications of these modifications are presented in Appendix C part 1. The actual modifications were made by nCATS workshop based on similar drawings. They were implemented in a way that allowed their subsequent removal and replacement with the original components. Therefore it was possible to restore the original machine configuration if required (for example the calibration test for the new WSS).

4.2.2.2 Specimen Changes

The modifications for the silicon nitride specimen, identified in the feasibility tests, were to improve the material quality and surface finishing, and to remove the tapered mounting arrangement. Additional requirements arose from the changed test cell as described above, namely the reduction of the diameter and the modifications to the anchoring keyways. In addition, it was realised that by changing the overall shape of the specimen from a disc to a ball would achieve the following benefits:

- provide a shape matching that of a rolling element used in real bearings
- remove the need for additional costly surface finishing processes, as balls can be polished in the course of their mass-oriented production.

The resulting specimen could therefore be adapted from standard industrial bearing elements.

The diameter of the steel specimen was increased to accommodate the smaller silicon nitride specimen, in order to maintain the necessary 40 (± 3) mm distance between their axes.

The feasibility test incorporated a steel disc of standard geometry, which was machined from an EN31 steel bar in the nCATS workshop before heat-treating and then polishing by an external company; each step adding an associated cost which would increase with a larger specimen diameter. Therefore, instead of continuing with the rigid set up, the steel specimen was redesigned into a two component set up: a replaceable cylindrical sleeve that provided the contact counterface, fitted over a mounting disc attached to the original shaft.

The resulting specimen would now consist of the cylindrical sleeve only, a shape that could be easily machined from a standard industrial bearing inner race. This would also avoid the need for post-machine heat-treatments and surface finishing processes.

The detailed specifications of the modified specimens are presented in Appendix C parts 2&3.

4.2.2.3 WSSI Sensor Changes

A new sensor head was required in order to reduce the background noise in monitoring the silicon nitride specimen. The existing sensor heads deployed for the test were both of circular face with 4.5 mm diameter, a standard sensor design used in nCATS and specifically adopted directly from a steel bearing test machine that was used by Harvey and Craig [148].

Their use with silicon nitride has however uncovered a number of issues, notably:

- They were designed to monitor severe wear modes dispersed across a significant part of the wear track and were tailored to monitor conductive materials (notably steel) exhibiting low charge levels. Therefore, they had to be of large size which consequently provided low spatial resolution. In contrast, silicon nitride produces wear with localised and high levels of charge; therefore the new head should be more location sensitive.
- As the feasibility test indicated, an eccentricity of the specimen mounting made a major contribution to the signal pattern for the ceramic specimen, as the associated data acquisition system has a finite capacity (range and resolution) for signal recording, a large portion of this capacity was taken up by the eccentricity signals leaving only a limited range for the crack fault signatures. A head with high location sensitivity should be able to distinguish better between the eccentricity and fault signatures, therefore providing more selective data to the signal acquisition system.

To overcome these problems, the following approach was adopted.

To increase the spatial resolution of the sensor for ceramic monitoring, different sensing faces were investigated to enable the detection of the highly localised charge produced by faults in non-conductive ceramic material. The signal components due to the eccentricity of the machine were to be then separated by a suitable choice of filtering method.

In order to find a suitable sensor design, a model was run that emulated four separate sensor face sizes with four different defect sizes. The results of the modelling then led to the selection of an appropriate size that was subsequently made for the sensor.

The model-based analysis was adapted from the work performed by Zhe et al [140], describing an impulse-like moving charge source as it passes a sensing face. The details of the modelling are presented in Appendix D. The signal characteristics in both spatial and frequency domains were investigated for various sensor face geometries that could be implemented on the machine and various charge source sizes that corresponded to RCF failures expected to be found in practice (as guided by references [51, 84, 245]).

The results are as follows.

The smallest sensor size (at 0.3 mm in length) gave the highest spatial resolution, the poorest frequency resolution and the lowest signal strength; whereas the opposite was found for the largest sensor. The selection was based on the rationale that spatial resolution was the determining factor, that the drawbacks in the weak signal strength could be overcome by higher amplification, and that the poor frequency resolution could be overcome by a suitable choice of filtering method. Correspondingly the smallest head size was chosen.

4.2.2.4 Data Acquisition System Changes

The VEE system was replaced by a NI 6212 BNC USB card controlled by a MATLAB software system. This system is well supported by the manufacturers and is easily updateable and provides a wide range of functions. The hosting PC was retained.

4.2.3 The Modified Machine

A photograph showing the resulting modified machine is shown in Figure 4.5 and a detailed sketch showing the new arrangement of the specimens and the sensors within the test cell is in Figure 4.6. Further details of the various components are provided in Appendix C.

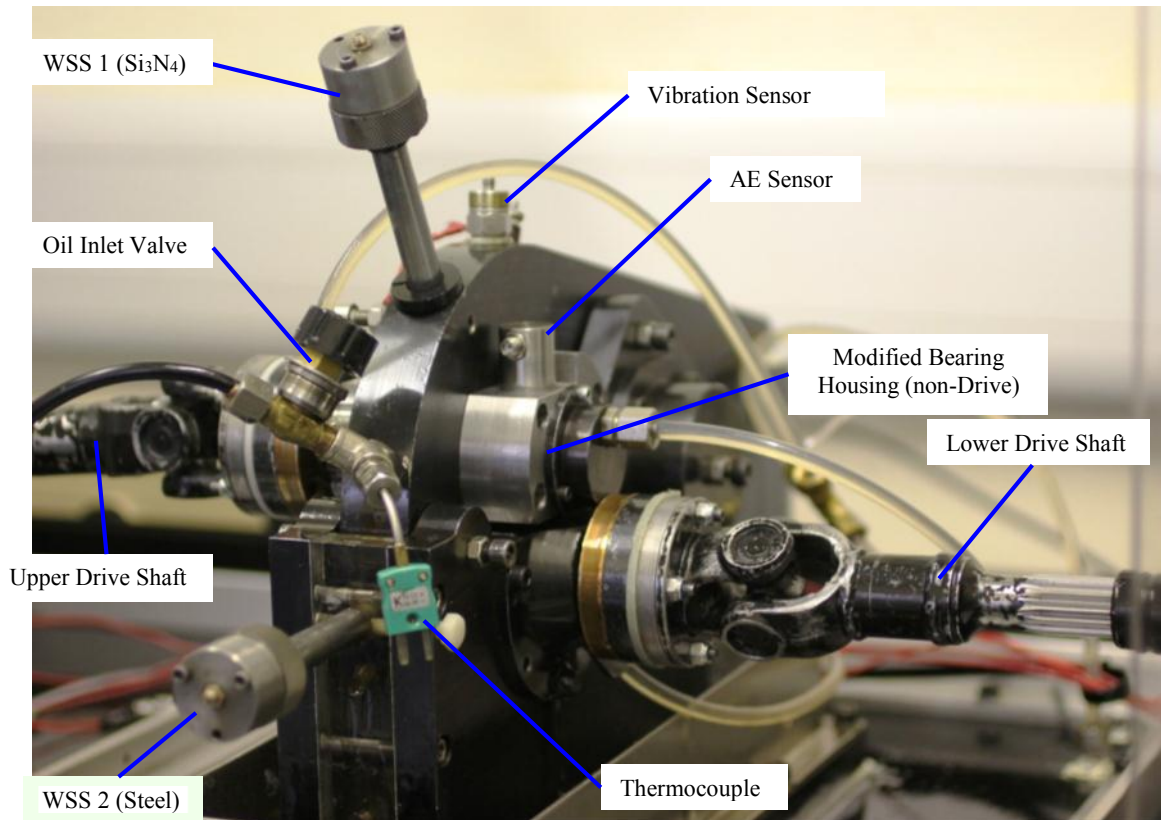


Figure 4.5: A photograph of the test cell of the modified machine

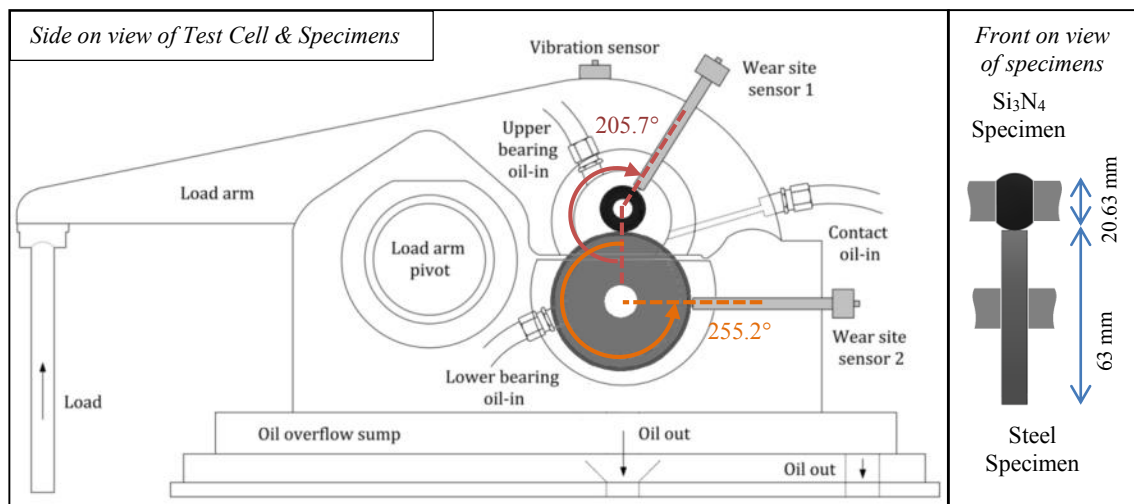


Figure 4.6: Schematic of the test cell of the modified machine. The rotational angle between the contact and WSS1 for the upper specimen is illustrated and indicated in red, and similarly in orange for WSS2.

4.2.3.1 Specimens

Figure 4.7 below shows the two-component steel disc and the new ball shaped silicon nitride specimen.

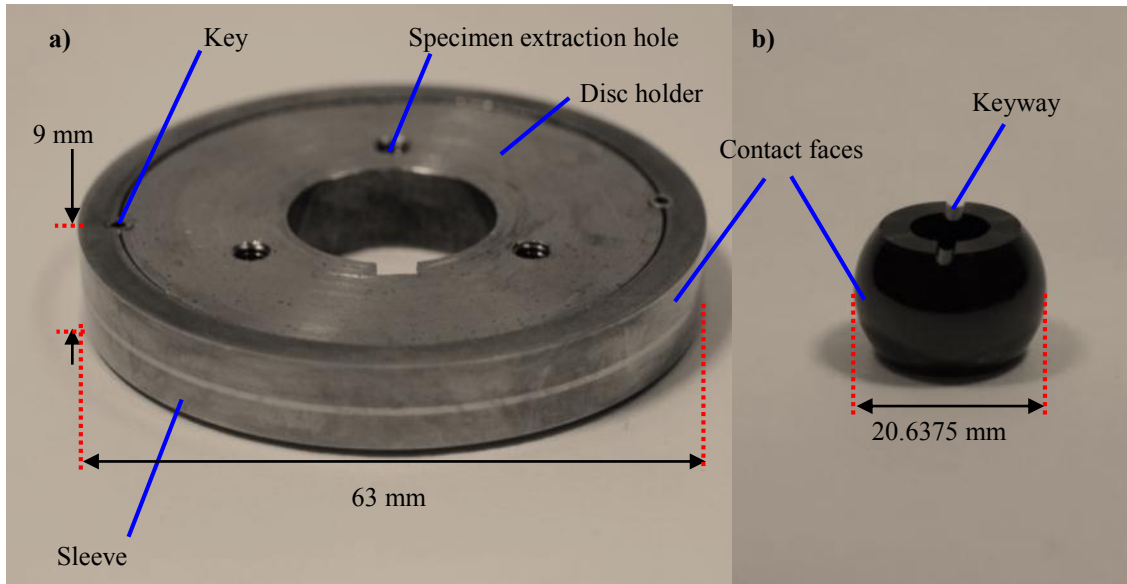


Figure 4.7: Sample pair geometry for modified machine showing a) steel, b) silicon nitride

The sleeve of the steel specimen (providing the counterface) was made from a through-hardened AISI 52100 steel bearing inner race (product name IR 55x63x45) with a diameter of 63 mm. The component was machined at nCATS to provide a race width of 9 mm and to incorporate the keyways. The keys fixing the component to the mounting disc were hardened-steel spring dowels.

The silicon nitride balls were provided by SKF Ltd. in The Netherlands, with 20.6375 mm (13/16") diameter and made from HIPed Toshiba TSN-03NH material. These standard ball bearing components were sent to Ceratec Technical Ceramics Ltd. in The Netherlands for subsequent machining, to incorporate an inner hole of 9 mm diameter and the keyways.

Table 4.1 below summarises the material properties of the two specimens.

Property	Si ₃ N ₄ (TSN-03NH)	Steel (AISI 52100)
Density (g.cm ⁻³)	3.22	7.81
Elastic Modulus (MPa)	309	210
Poisson's Ratio	0.28	0.29
Vickers Hardness (Hv)	1550	830
Tensile Strength (MPa)	900	2,240
Fracture Toughness k_{1c} (MPa.m ^{1/2}) ¹¹	6.7	14.7
Additives	Al ₂ O ₃ (6.6%), Y ₂ O ₃ (4.4%), TiO ₂ (1.3%)	C (0.95-1.1%), Cr (1.30 – 1.6%), Mn (0.25%)

Table 4.1: Material properties of the specimens (information provided by the manufacturers i.e. Toshiba for silicon nitride and SKF for steel).

Average roughness measurements of the specimens using the Talysurf Profilometer indicated Ra values of 0.01 and 0.09 for silicon nitride and steel respectively.

4.2.3.2 Sensors

i. WSSs

The resulting shape and dimensions of the new head are shown in the figure below, alongside the original.

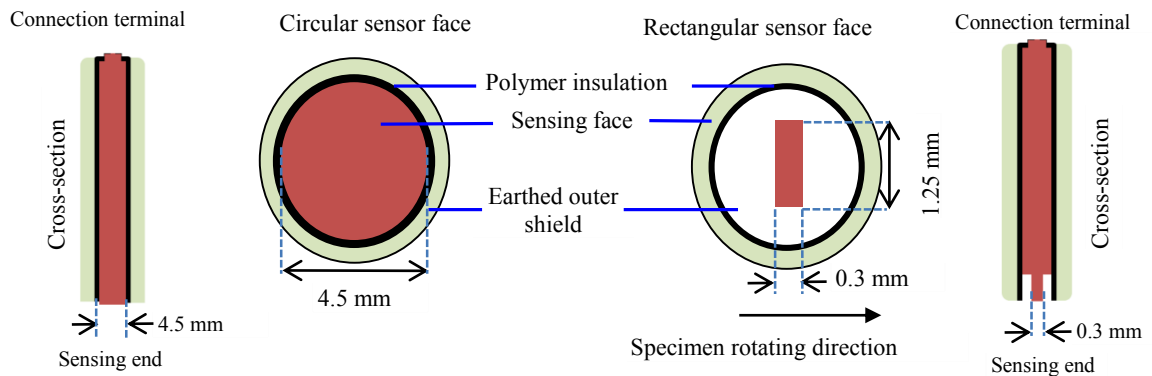


Figure 4.8: An illustration of the two WSS sensor head geometries

The new component was made from mild steel (the same material as of the circular sensor), measuring 1.25 mm x 0.3 mm in the transverse and rolling directions respectively. Because its outer dimensions were the same as the existing sensor, the original arrangements for its placement and adjustment could be used without modification. The sensors were connected to the data acquisition hardware by microdot cables.

ii. Vibration

An accelerometer, model A/20 sensor manufactured by DJ Birchell Company was used for the duration of the research. The sensor was positioned at the top of the load arm vertically above

¹¹ For silicon nitride fracture measured using the indentation method (Japanese Industrial Standard No. JIS R 1607-1990)

the contact, attached by a screw. The sensor has an operating frequency range of 0.01 - 12.5 kHz with a resonant frequency of 28 kHz and linear sensitivity of 28.9 pCg⁻¹ (manufacturer's data). The sensor was connected to the data acquisition hardware by a microdot cable.

iii. Tachometers

Two once-per-rev tachometers were incorporated on the modified machine to monitor the rotation speeds of the two specimen shafts. Tachometer 1 was an optical based sensor generating a step voltage at the passing of a reflective tape on the shaft of the upper (silicon nitride) specimen. Tachometer 2 was a Hall Effect based sensor, generating a step voltage at the passing of a protruding key on the shaft of the lower (steel) specimen. The tachometers were connected to the data acquisition system by standard BNC connectors.

iv. Acoustic Emission

The sensor used was the PAC WSc sensor which was a single-ended wideband sensor, with an operating frequency range from 0.1 to 1 MHz, and a peak sensitivity of - 62 dB (with a reference of 1 V.μbar⁻¹) at its resonant frequency 0.125 MHz.

The sensor was fixed using epoxy based glue to the non-drive side of the upper modified bearing housing as illustrated in Figure 4.5 above and connected to the data acquisition hardware by a R15 cable.

4.2.3.3 *Data Acquisition Systems*

Two PCs were used to record and further process the signal data: one PC was used solely for the AE data, whilst another PC handled both WSS, vibration and tachometer data. The details of these systems are given below.

i. WSSs and Vibration Acquisition Systems

The acquisition system for the vibration and WSS data included a separate analogue amplifier (Endevco model 133) connected to the data acquisition PC via BNC connectors. The amplifier incorporated switchable (on/off) Butterworth low-pass and high-pass filters with cut-off frequencies of 1 Hz and 10 kHz respectively. The amplification and filter settings are listed in Table 4.2 below.

Sensor	Scaling (mV/pC)	High-pass filter	Low-pass filter
Vibration	50	On	Off
WSS1	250	Off	On
WSS2	500	Off	On

Table 4.2: Charge amplifier settings for the vibration sensor and WSSs.

The data acquisition system for the tachometers included a dedicated integrated circuit card connected to the data acquisition PC via BNC connectors. The card was made in-house and provided Butterworth low-pass filtering with cut-off frequencies of 15 kHz to improve the signal-to-noise ratios with minimal phase distortion.

The data acquisition PC system had installed a dedicated National Instruments DAQ USB card (model 6212-BNC) to which the amplifiers and tachometer filters were connected. The PC card was used to provide signal acquisition with sampling rates of 32 kHz for each acquiring channel with a resolution of 16 bits.

The PC software system was based on MATLAB which also incorporated a plugin for compatibility with the card. Burst acquisition was adopted to provide high speed data acquisition whilst reducing data size and simplifying analysis. Bursts were set for the duration of 20 seconds with intervals of 5 minutes, to operate simultaneously for all channels in parallel.

i. AE Acquisition System

The data acquisition system for the AE sensor included a standalone pre-amplifier (PAC model 2/4/6) which was connected to the PC by a BNC cable to a port on a PCI card. The pre-amplifier provided a gain of 60 dB.

The PCI card (PAC model PCI2v.3) was providing further processing as follows:

- additional 6 dB amplification,
- high-pass and low-pass analogue filtering set to 0.1 and 1 MHz respectively,
- data sampling of 5 MHz,
- resolution of 18 bits.

The PC system was based on the PAC AEwin v.3, set to record bursts at 5 minute intervals lasting 1 second. Further processing was performed using MATLAB on another PC and described in the chapters following. Even with this setting, the data generated was close to the capacity of the machine. Because of this, it was not possible to record any additional data such as the tachometer output.

4.3 Physical Inspection Tools

4.3.1 Talysurf Profilometer

A Form Talysurf 120 L (Talyor-Hobson) laser profilometer was used to provide standard roughness measurements for samples and generate 2D wear track surface profiles. The stylus was diamond tipped with radius of 2 μm and a vertical resolution of 0.25 nm.

4.3.2 Alicona Infinite Focus

The Alicona Infinite Focus (AIF) profiler provided 3D optical imaging of up to 100 x magnification with a maximum vertical resolution of up to 10 nm. The machine was equipped with a polarizing light filter that could improve the clarity of the images.

The machine was particularly useful in generating profiles of failure regions with steep vertical side faces, unobtainable by the Talysurf Profilometer (- as the latter used sensing based on physical contact via a stylus with finite dimensions).

4.1.1 Digital Anglemeter

The digital anglemeter incorporated a fixed base with a rotatable shaft for specimen mounting. A digital readout indicated the angle of rotation of the shaft from 0 to 360° with a measuring accuracy of 1°.

4.3.3 Scanning Election Microscope

A JSM 6500F scanning electron microscope (SEM) was able to provide up to 3 nm resolution and equipped to perform energy-dispersive X-ray analysis (EDX). To analyse silicon nitride, specimens were first gold coated using a Hummer 6.2 sputtering machine. Sputtering was applied for a duration of 1.5 minutes, resulting in gold coating thicknesses of approximately 10 nm.

4.3.4 Computer Tomography Analysis

A custom 224 kV Nikon/Metris HMX ST at the μ -VIS centre at the University of Southampton was used in the rotating target configuration to scan a RCF failure region of one selected test sample to determine its potential usefulness for this type of research. The resolution achieved with the silicon nitride specimen was 6 μm .

Chapter 5. Signal Processing Development

With the modified hardware in place and ready for productive testing, the next step was to address the issue of how the generated signal data should be processed in order to achieve the required project aims. This chapter describes the results of this part of the project.

5.1 Overview of Process

The approach adopted was based on generating, for each sensor type and under controlled conditions, separate reference signal patterns for both a generated fault and a background noise. These would then be used, together with recommendations published by other researchers, to determine the appropriate methods for extracting the fault signatures under normal test conditions. This was done in three stages, first a test was done to emulate the generation of the fault and the signal characteristics, typically in the frequency range, were analysed. This was then followed by a test that emulated the background. An important part of the background signature is that generated by the test machine in the absence of any specimen faults. For this, a separate run (*'Healthy Run'*) of the Twin Disc machine with an un-indented sample was performed. Having obtained data from both the fault and background, suitable feature extraction methods were set up for each sensor type.

The procedure adopted is sketched in the diagram below and detailed in the subsections following.

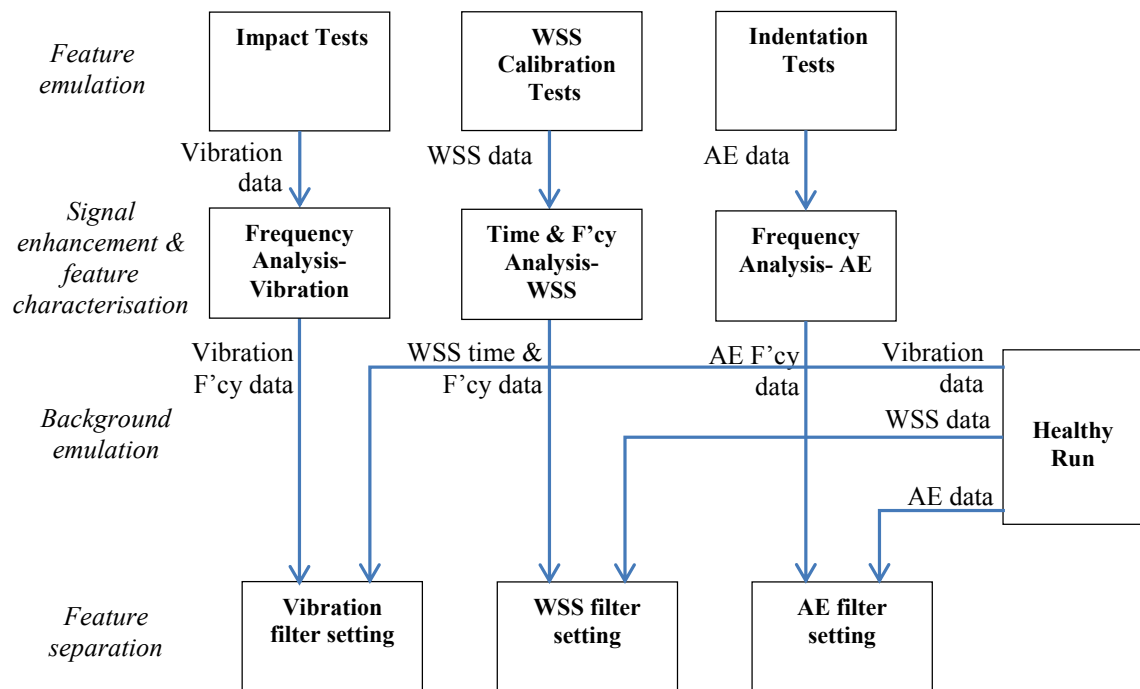


Figure 5.1: Organisation of the main tasks in the signal analysis development

5.2 Signal Generation Tests

In order to generate suitable signals, different style of tests were employed for each sensor type; these and their results are described below.

5.2.1 Vibration Sensing

Vibration sensing works on the principle that faults create mechanical impacts against the counter specimen therefore creating vibration that propagates through the Test Cell assembly. The Test Cell itself exhibits characteristic vibration frequencies of its own and, despite having a degree of separation from the rest of the machine, can also pick up additional vibration from the operation of the components within the machine and a background or environmental noise. The vibration sensor located on the Load Arm of the Test Cell (refer to Figure 4.5) directly above the contact will therefore detect multiple sources in a profile dominated by resonant frequencies. In a way, the contributory sources modulate the resonant frequencies of the Test Cell. All of the contributory sources (except perhaps the background) can have cycle-synchronous characteristics; therefore the separation of the fault signal from the remainder is much more difficult than, for example, WSS (see below). Nevertheless, researches analysing bearing vibrations have identified that fault signals can be separated by focusing on high frequency part of the vibration spectrum (see for example [101, 246]).

Following this advice, the approach adopted was based on first identifying the high resonant frequencies and then examining the signal profiles at these frequencies to separate the features that could be attributed to faults. The data for the initial step was obtained in the tests following.

5.2.1.1 Tests Details

To excite the resonant frequencies of the machine, simple impact tests were performed by hitting it gently with a hammer and the resulting data was collected. The sampling frequency of the acquisition card was set to 32 kHz, making it able to capture vibrations with frequencies of up to 16 kHz (by the Nyquist criterion). This was beyond the nominal maximum response of the sensor at 12.5 kHz, therefore ensuring that any fine features within the sensor-generated spectrum were preserved.

These tests were performed on the stationary Test Cell with different loads applied to the specimens to investigate the influence of loading on the resonant modes. The loads tested are summarised in Table 5.1.

Test No.	Load applied to specimens (kN)
1	0.5
2	0.75
3	1
4	1.25

Table 5.1: Loads applied between the specimens for vibration impulse response measurements

5.2.1.2 Results

The raw time profiles for each of the four impact tests are shown in Figure 5.2. Their vibration frequencies were analysed using the power spectral density (PSD), estimated using the Welch method^{12,13} – a common method of frequency analysis (see for example [247]). The resulting spectra, representing the frequency distribution of the vibrational energy, are shown in Figure 5.3 below.

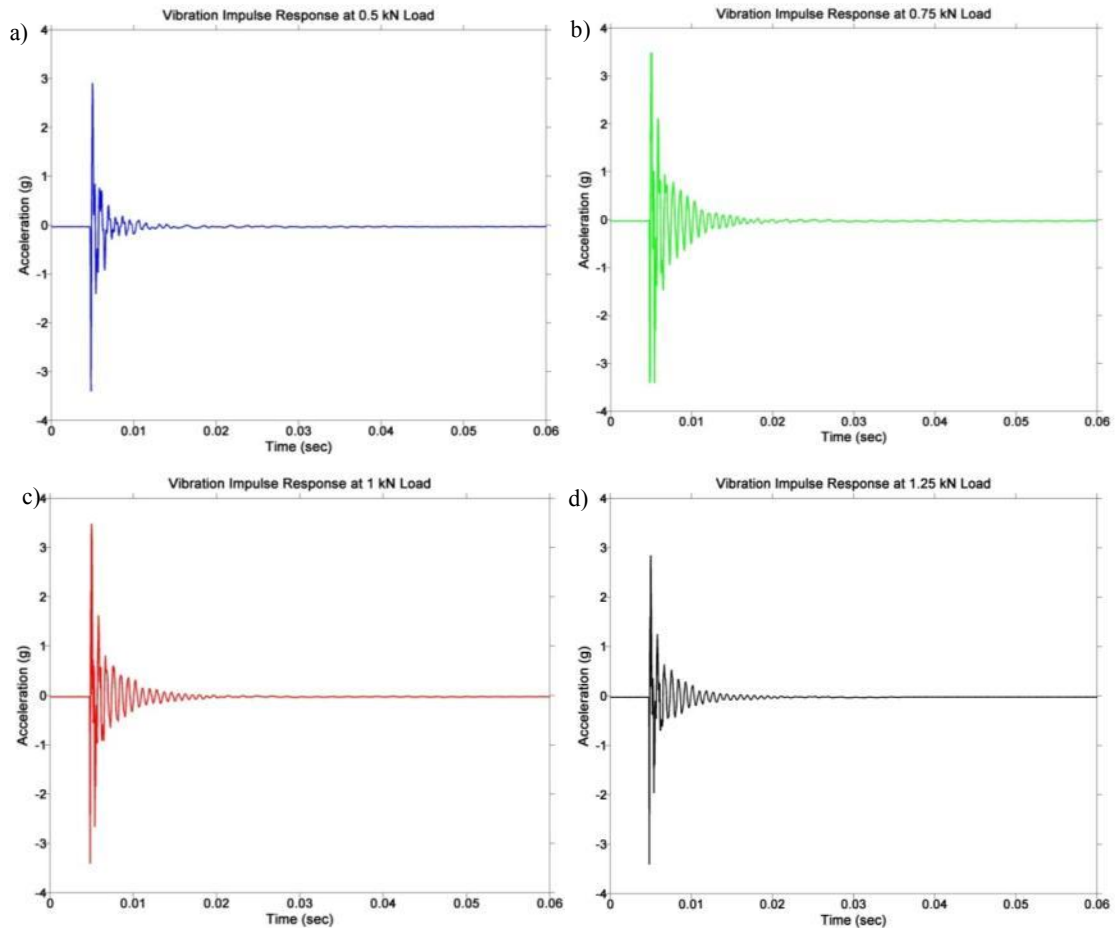


Figure 5.2: Recorded vibration responses of the four impact tests impulse measurements at various loads (indicated in the figure labels)

¹² As described in Subsection 3.3.1.3, the Welch method applies a window function during signal segmentation, reducing the effects of border distortion in the computation of the Fourier coefficients. For a transient signal which begins and ends at zero, such as those produced by the vibration impulse response tests, the signal is self-windowing. Despite this, the method is applied in order to produce outputs that could be directly compared to the vibration analyses of the Healthy and Production Runs – where border distortion is an issue.

¹³ A window size of 2^8 and a 50 % overlap was used in the calculation and is the same as the parameters used in the Healthy and Production Runs.

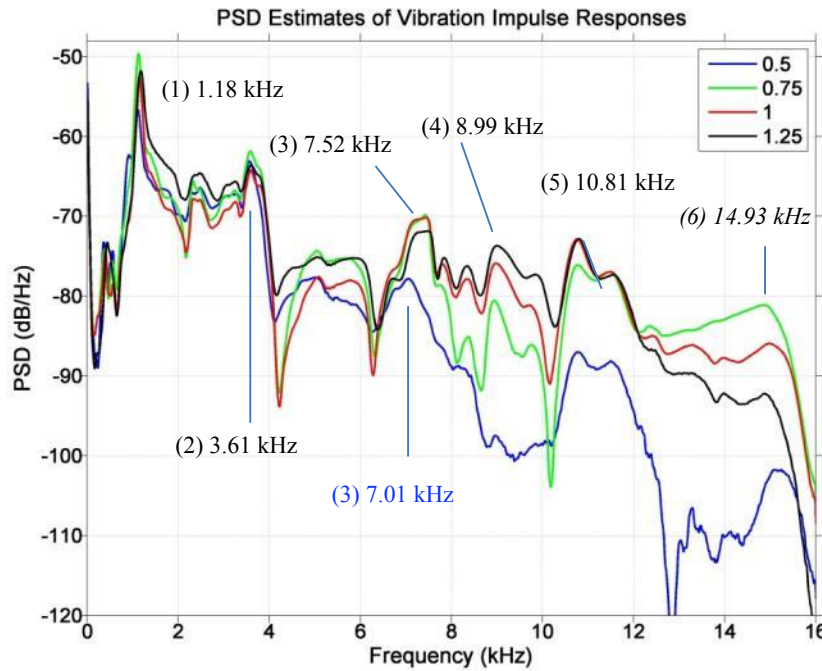


Figure 5.3: PSD Estimates of the vibration impulse responses at various loads (highlighted in the legend entries (in kN). The main resonant vibration modes are labelled and their respective frequencies are given.

They demonstrate a general common pattern of which the most prominent parts are multiple peaks representing resonant modes and labelled (1) to (6). They all appear for loads 0.75 to 1.25 kN. For the 0.5 kN test, mode (4) was not present and the resonant frequency mode (3) has shifted from ~ 7.5 to ~ 7 kHz. These results therefore indicate that modes can be load sensitive (something to bear in mind when planning and interpreting the Production Runs), and that the observed mode (6) was beyond the nominal range of the vibration sensor (1-12.5 kHz), thus showing that the sensor was fully operational within a wider range.

With these findings, both mode 5 and 6 (highest resonant components) were considered potentially suitable candidates for vibration monitoring. The final choice was made after the completion of the Healthy Run.

5.2.2 WSS Sensing

WSS method works on the principle that creation and progression of faults generate distinct charges. Therefore, a sensor that is sensitive to electrostatic charge can produce a signal as the specimen rotates underneath.

For this project, a new sensor head was made to extracting RCF faults in silicon nitride as detailed in Appendix D. In particular, the sensor was designed to be focussed on a small area, reducing a background dominated by high amplitude, low frequency charge signal. It was therefore important to get an indication that the results of the model were in at least a broad agreement with practice and that the planned experiments were feasible with the technology

available. Of particular interest was the sensitivity of the new sensor head, made specifically for this research and described in the section on hardware (5.2.3). The test was run on a calibration disc that was purpose-made for this work. Following this, another comparison test was run with the signal data taken by the other, unmodified, sensor.

5.2.2.1 Test Details

i. Calibration Disc

A disc was designed to generate a contact potential difference between two dissimilar metals at a precise location. This approach has been used in nCATS to calibrate WSSs by inserting metal discs into pre-drilled standard specimens. It was adapted for this machine by making a groove in a steel disc which was subsequently filled with silver (chosen because it could adhere to a steel casing, generating a potential difference as a result of the differences in the work functions of the two materials: Iron 4.4 eV and Silver 4.26 eV). A standard EN31 steel disc of 40 mm diameter and 10 mm width was manufactured to fit the machine in its original configuration shown in Figure 5.4a. A ‘V’ groove was machined on the outer diameter transverse to the rolling direction with a width and depth of 500 and 750 μm respectively (this was the finest groove that could be machined in the University workshop and deemed to be adequate for this test¹⁴). The groove was then filled with silver solder and the disc polished on the outer diameter removing any protruding solder from the testing face, see Figure 5.4.

Before its placement at the testing position, the disc was cleaned in an ultrasonic bath with ethyl acetate for 15 minutes, rinsed in deionised water and dried with high pressure air.

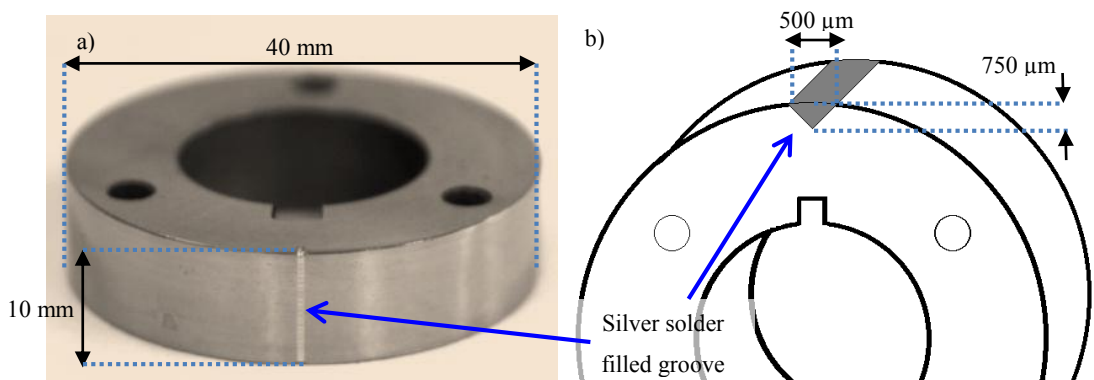


Figure 5.4: Illustrations of the Impulse Disc specimen a) photograph b) schematic showing the weld geometry

¹⁴ It turned out that this was of comparable size to the actual faults produced in the Production Runs.

ii. Test Procedure

The calibration disc was mounted on the upper specimen shaft and rotated using the upper drive motor whilst contact with a counter disc specimen was avoided to prevent wear charging which was unwanted in this case. To avoid contact between the disc and the sensor resulting from the absence of the damping effect from a counterface contact, the disc was rotated at a relatively low speed of 500 RPM. However a basic eccentricity should still be present even at this low speed.

Both tests (with the modified and unmodified sensors) were run with the WSSs positioned at the WSS1 location (see Figure 4.5) with 0.3 mm between the disc and the sensor face. Signals were recorded using the Twin Disc data acquisition system developed for the Production Runs and described in section 4.2.3.3 above, over test runs lasting 15 seconds. Both WSS and Tachometer signals were acquired with sampling rates at 32 kHz and the amplification scaling value for the WSS charge amplifier was set to 500 mV.pC⁻¹.

The rotational angle between the upper tachometer trigger point and the silver defect passing the WSS1 sensor face was measured using the digital anglemeter connected to an exposed end of the specimen shaft. The angle from the Tachometer trigger (indicated by the real-time output of the Tachometer signal) to the point at which the silver defect aligned with the WSS sensing face was measured at 233°.

5.2.2.2 Results

The raw data acquired from the WSSs showed complicated shapes with no discernible patterns. To interpret them, a key fact was used that the signals from both the emulated defect and the background eccentricity were synchronous with the disc rotations. To extract these signatures from any other background (non-synchronous) noise, TSA was applied. Subsequently, the emulated fault signal could then be discerned by its angular location using tachometer data. To do this, the peaks of the tachometer which identified the starts of each cycle, were digitally extracted and used to segment the WSS signal data into 100 consecutive rotation cycles. The background noise was then removed by averaging the amplitudes taken at the same points within each of the 100 cycles.

To overcome the problem that the raw data sampling did not produce values corresponding to the same points within each rotation cycle, the Fourier based interpolation technique discussed in section 3.2, was employed (3,840 appx data points per-rev were interpolated to 4,096 Fourier points). The resulting attenuation of the asynchronous noise was up to 10 dB.

The TSA outputs for both of the sensor tests are presented in Figure 5.5 below. To ensure comparability, they have been each normalised with their respective maximum values.

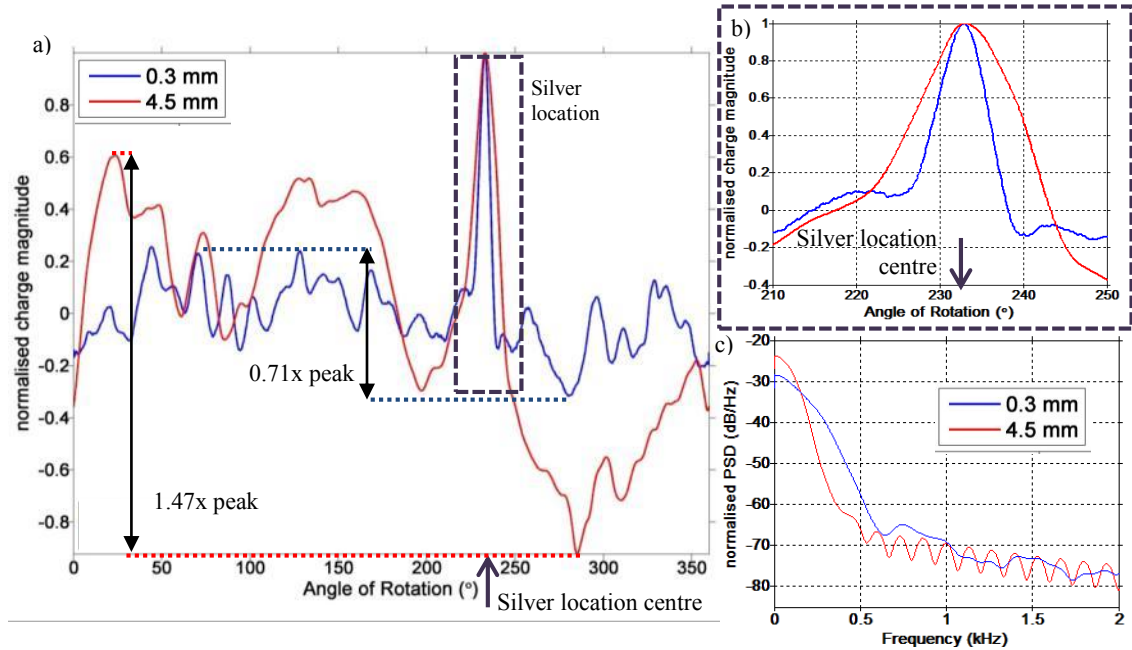


Figure 5.5: Normalised TSA responses of the calibration disc for both sensor sizes (indicated in the legend), a) in the angular domain b) a close-up of the weld location and c) in the frequency domain (PSD).

In the angular profiles Figure 5.5a&b the location of the silver defect is also indicated. As expected, the relative magnitude of the response from the defect and the response from the eccentricity was lower for the new, modified (0.3 mm) sensor as a result of the increased spatial resolution (8.2 times lower than that of the 4.5 mm sensor). However the relative amplitudes between the defect and eccentric signals were higher in the new, modified sensor (approximately 2x, refer to their measurements in Figure 5.5a).

Figure 5.5c shows the comparative frequency profiles. In particular, a reduction in the frequency resolution for the modified sensor is visible. This was expected as per the modelling performed in its design (refer to section 4.2.2.3) and this drawback could be overcome by a suitable choice of filtering method (to be made below).

Therefore, the new smaller head was appropriate choice for the subsequent runs.

The other important conclusion from this test was that any analysis of the WSS data must first enhance it by focussing on the cyclical characteristics of the source of interest (the fault) i.e. by performing the TSA transformation.

5.2.3 AE Sensing

To generate AE sensor data to be used for setting up the AE signal processing system, the indentation machine was utilised.

The indentation process produces substantial cracks in ceramics as discussed in Chapters 2&3. Given that RCF mechanisms are also dominated by crack propagation (see e.g. [50]), the data generated by the indentation process can therefore provide a suitable basis for the development of the processing method. There are several advantages of using an indentation machine rather than the Twin Disc machine: the tests are significantly simpler to set up, it has a relatively quiet operation (i.e. it can produce high signal-to-noise ratios) and it allows variable indentation load settings thus providing a degree of control over the crack generation. Importantly also, it allows collecting separate signal profiles from the silicon nitride and steel specimens¹⁵.

Therefore the test bed used consisted of the indentation machine equipped with an AE sensor and processing system as described in Chapter 5.

Their AE signatures can be isolated from other indentation-related processes using the smoothing technique developed by Faisal et al [166].

To provide a comparison and further aid in the identification of the AE crack propagation events, monitored indentations were made to both steel and silicon nitride specimens at the same load conditions. Furthermore, the AE characteristics were then compared.

5.2.3.1 Test Details

i. Test Conditions

Separate tests were run with silicon nitride and steel samples, both of the same geometry.

The silicon nitride sample was the modified type as described in Chapter 5. The steel specimen was made from through-hardened AISI 52100 steel¹⁶ machined to the same dimensions as the silicon nitride. The indentations were made at the specimens' flat surfaces as they provided an optimum seating arrangement. In order to achieve a uniform smooth surface there, both specimens were polished using a polishing solution with a diamond size of 1 μm , resulting in average roughness (Ra) values of 0.02 and 0.01 μm for the steel and ceramic respectively. Before placing the specimens to the testing position, they were first cleaned in an ultrasonic

¹⁵ Unfortunately the same set up could not be used to emulate the vibration and WSS signals as the rolling processes are essential for their generation.

¹⁶ This was the same steel specification as used in the Production Runs.

bath consisting of ethyl acetate for 15 minutes before rinsing in deionised water and drying with high pressure air.

Indentations were made using three different indentation loads summarised in Table 5.2 below. To assess consistency of results, two indentations were made at each load (referred to below as *A* and *B*), altogether therefore 12 indentation tests were run in total.

Test Name	Load (kgF)	Load (N)	# Indentations
15	15	147.1	2
20	20	196.1	2
30	30	294.2	2

Table 5.2: Indentation loads for Indentation Tests¹⁷

To distinguish the noise generated by the machine itself from the persistent environment noise, two tests were performed. *Background 1* was a measurement of the signals produced when the machine was not in operation. *Background 2* was a measurement of the signals generated by the machine during the indenter tip initiation, motion and retraction. For this the machine was set up with a 30 kgF load and the indenting process was done without any contact being made by the indenter tip with a test specimen. For all tests performed, raw AE waveforms were recorded by the AE data acquisition system.

ii. Test procedure

For the indenting tests (i.e. non background), the initial indenter stage height (refer to Figure 5.1) was set such that the distance between the indenter tip and specimen was 0.2 mm; a spacer was used to calibrate the distance.

The AE events were expected to be detectable sometime after the contact between the indenter and the specimen surface was made. As only the time of the initiation of the indentation test run (*Indentation Start*) can be directly known, it was necessary to estimate the time interval it takes for the indenter to reach the surface of the specimen and commence the indentation (*First Contact*). These times are illustrated in Figure 5.6 below. The First Contact point was determined by using a mounted magnifying glass focused at the indentation location and the time interval was measured by a stopwatch. Intervals were recorded for all three testing loads listed above in Table 5.2. To reduce the inaccuracy of using a manual stopwatch, 5 indentations were performed per load and an average was used to determine the length of time that was used subsequently.

¹⁷ kgF are used for the machine settings and translated in the table into N

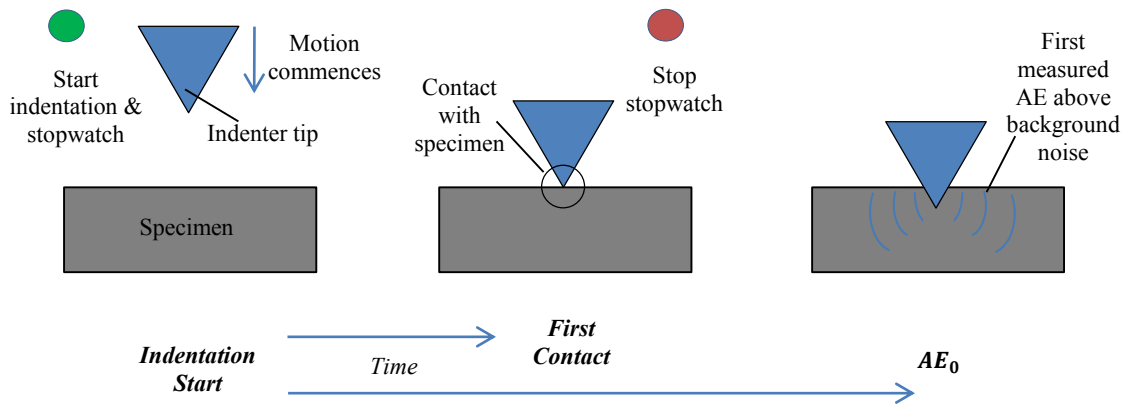


Figure 5.6: An illustration of Indentation Start, First Contact and AE_0 (introduced in the following subsection). The start button represents the starts of the indentation tip, the stopwatch and AE recording

To check the attenuation of the AE signal path between the specimen and the sensor, a standard pencil lead break test was performed. For tests involving indenter tip motion (i.e. all tests except *Background 1*), the commencement of AE recording was synchronised to coincide with Indentation Start, manually performed by the operator. The indentation took 11.9 seconds, so to ensure that AE signals were recorded from the entire run, the recording was programmed to run as follows:

- Background and steel tests: 13 seconds
- Silicon nitride tests: 14 seconds.

In the latter case, the extended recording time was set to capture possible AE events post indentation. As each sample was reused several times, indentations were made with a minimum distance of 500 μm from the edges of one another to avoid interactions between them. Alignment of the indenter tip and the indentation location was made through movement of the machine's x-y stage and the use of the machine's viewing and alignment sight (refer to Figure 5.1).

5.2.3.2 Results

i. Sample Inspection

The indentations produced in the tests were measured to allow comparisons between their sizes and the signals collected during the tests. The measurements were performed on images taken on the AIF machine that were also used to provide some indication of the mechanisms that led to their development for example by their exact shape, presence of additional cracks etc (in addition to their general form, sketched in Figure 5.7 below). Additional information for this was provided by SEM images.

The overall size was characterised by two measurements:

- average diagonal lengths, $2a$,
- average radial crack length l_a , for indentations exhibiting radial cracks,

derived as shown in Figure 5.7.

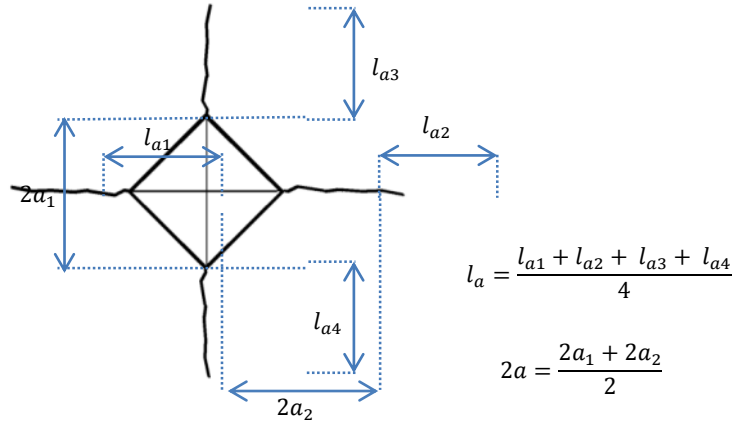


Figure 5.7: An indent with radial cracks and its measurements and calculations.

Steel

Figure 5.8a presents a SEM image of indent 20-A, representing a typical form of indents produced in the tests (others are presented in Appendix E). No cracks were observed at the surface around the indents, suggesting that the dominant mode for material accommodation was through plastic deformation. Also, little evidence for piling up mechanisms was found, indicated by the straightness of the edges of the indent and suggests limited work hardening has undergone in the material. Surface indent geometry increased linearly with load as illustrated in Figure 5.8b. Measurements of the indentations indicated a hardness value of 825 ± 5 Hv. Notably, defects were observed in the indentations, which corresponded to material chipped from the indenter tip and are indicated in Figure 5.8a.

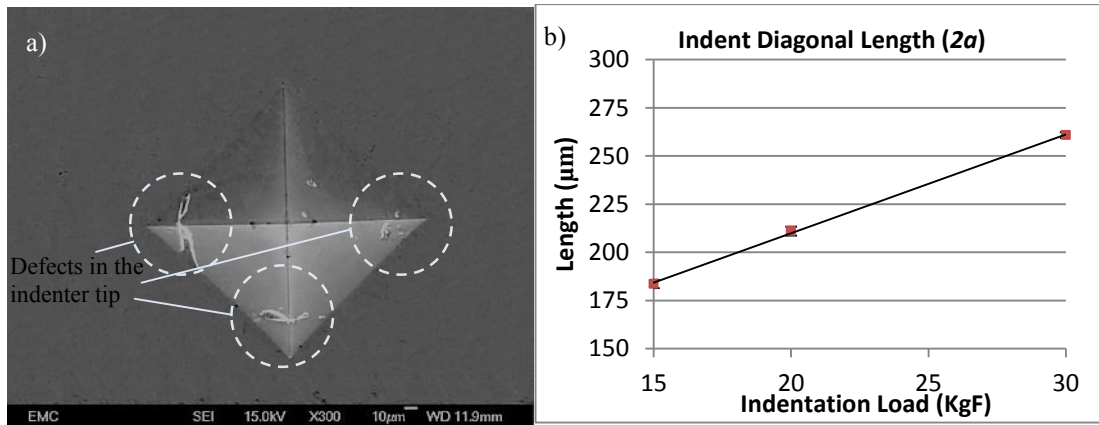


Figure 5.8: a) A SEM image of indent 20-A, b) average diagonal surface lengths ($2a$) with load for steel indentations.

Silicon Nitride

Figure 5.9a presents an AIF image of the indent 30-A, whilst all AIF images of the silicon nitride tests are presented in Appendix E. Radial cracks were observed in all tests using the AIF machine (indent , and their respective surface lengths are plotted in Figure 5.9b. The indent diagonal length $2a$ demonstrated an apparent linear increase with the load, whilst the radial average crack length l_a increased such that the ratio $l_a/2a$ reached unity at 30 kgF indentation load indicating a deepening of the radial cracks forming median crack geometry. Lateral cracks observed through polarised light indicated that the majority of indents demonstrated some asymmetry. Measurements of the indentation diagonals indicated an average hardness of 1598 ± 15 Hv for the silicon nitride balls.

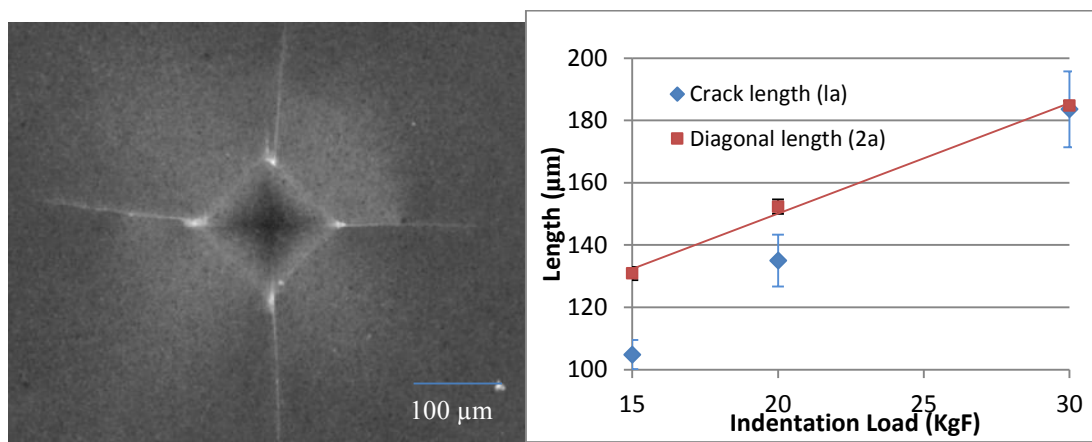


Figure 5.9: a) An AIF image of indent 20-A, b) Average indent geometries with load for silicon nitride indentations.

ii. AE Signal Characteristics-General

Raw AE data were collected for all indentation and background measurements (these are presented in Appendix E). The first step in their analysis was to remove any characteristics associated with background. For this the background data was analysed.

Background Levels

Background waveforms showed a persistent noise with a maximum amplitude of 26 mV. Their frequency analysis showed a maximum frequency at 250 kHz¹⁸. The machine movement in the *Background 2* test did not produce any distinguishable AE signals above the noise baseline; also, no variation was observed in the frequency domain.

With the background levels analysed, the next stage was to determine the first point at which a significant AE signal becomes visible. For this, a threshold was set at 15 % above the maximum noise found above i.e. 29.9 mV and the time at which the AE amplitude first exceeded this

¹⁸ The Welch PSD method was used in frequency analysis using a window size of 2^{10} (1024) data points with an overlap of 50%.

threshold (AE_0) was recorded. The results for both steel and silicon nitride tests are summarised in Figure 5.10 below.

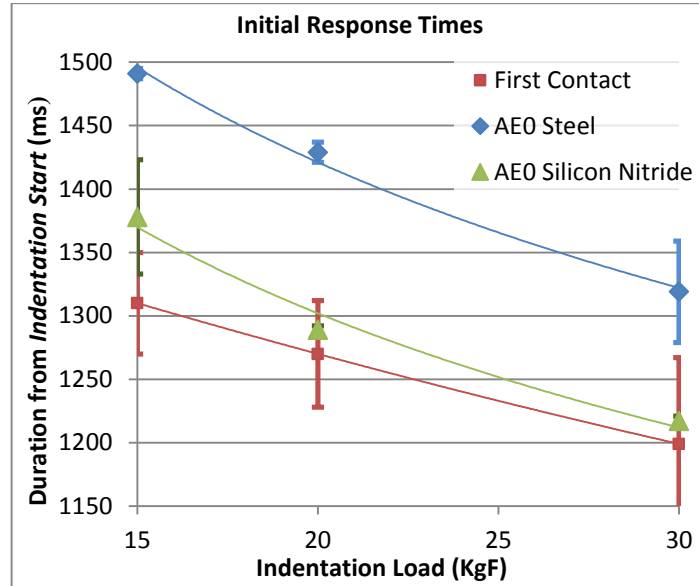


Figure 5.10: Measured first contact times and first AE responses (AE_0) for both steel and silicon nitride indentations

The time interval from Indentation Start to First Contact was observed to decrease linearly with load, the difference over the load range tested being approximately 100 milliseconds. Of particular interest was that the silicon nitride exhibited an AE response almost immediately after First Contact; in contrast, steel demonstrated a delay of approximately 120 milliseconds (for tests at 30 kgf load), before the onset of a measurable AE response. This delay also decreased with load.

Having determined the point from which the signal bears the signatures of interest, the focus of analysis was on the period commencing from this point, with the objective to relate the signal patterns to the material accommodation processes.

A technique for this has been developed by Faisal et al [166]. This is based on a two stage process consisting of an initial signal enhancement followed by a division of the enhanced signal into time zones, using a threshold criterion, each with a distinct signal pattern.

The technique was adapted for use in this project by selecting a particular signal enhancement approach. This was the Energy Index (EI) technique developed by Al-Balushi [168] (described in Chapter 3), chosen in preference to the traditional smoothing function to reduce the large signal volume ($>3 \times 10^7$ data points) and also to significantly reduce latency in computation without the loss of significant features as demonstrated by the author [168].

A rectangular window function with a size of 2^{12} (4096) data points was applied with a 50% overlap and subsequently the output was normalised relative to the respective maximum value.

The following discussion presents the resulting profiles.

iii. AE Signal Characteristics -Steel

Figure 5.11 presents the results of the smoothed waveforms truncated at 0.5 seconds (from AE_0) showing both A & B repeats for each load level. Also shown is a derived frequency pattern discussed below.

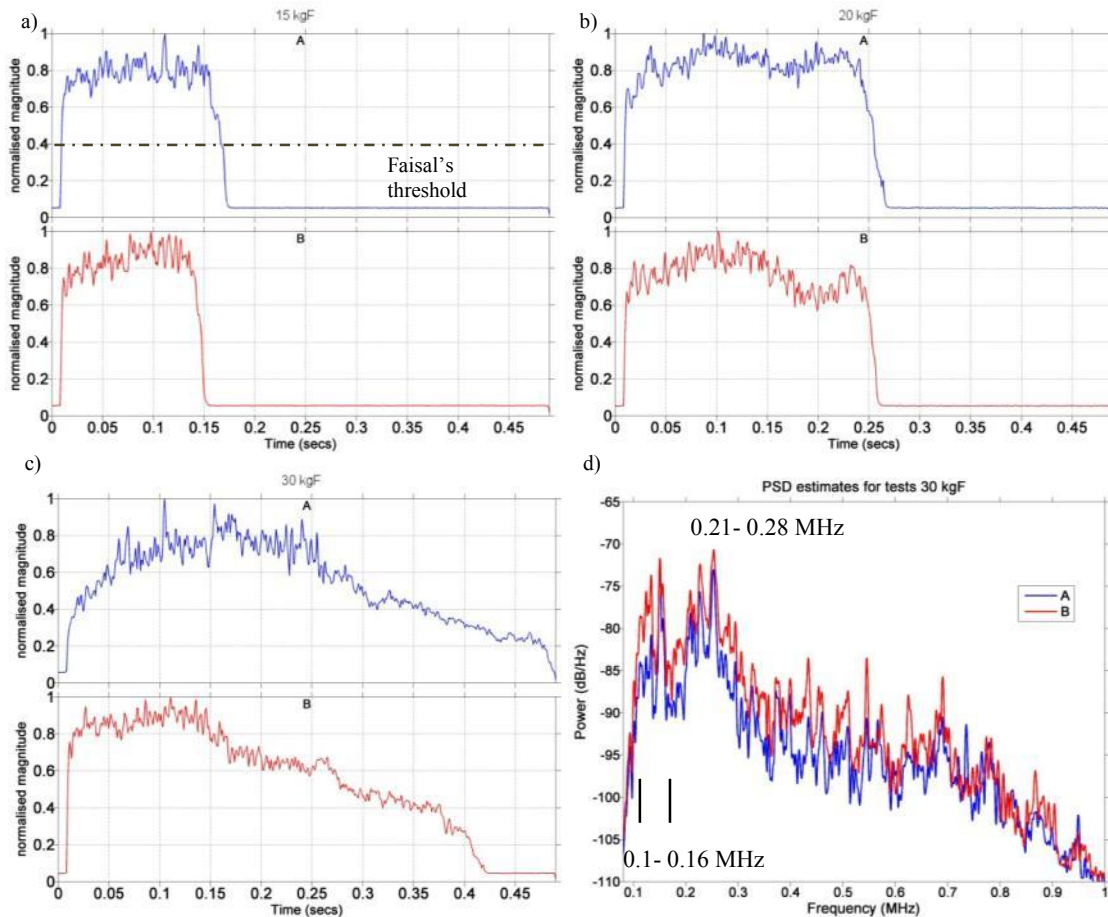


Figure 5.11: The smoothed AE profiles for steel indentation tests, at a) 15, b) 20 and c) 30 kgF loads. The PSDs of the tests at 30 kgF loads are illustrated in d).

The profiles for all tests increase rapidly from the AE_0 baseline to near maximal amplitude and maintains high values for a minimum duration of 150 milliseconds (for test 15-A) and then decreases back to the baseline. Because of this simple trend, there is only one zone (also indicated by the threshold shown).

Figure 5.12 illustrates the duration and mean amplitude variations between loads.

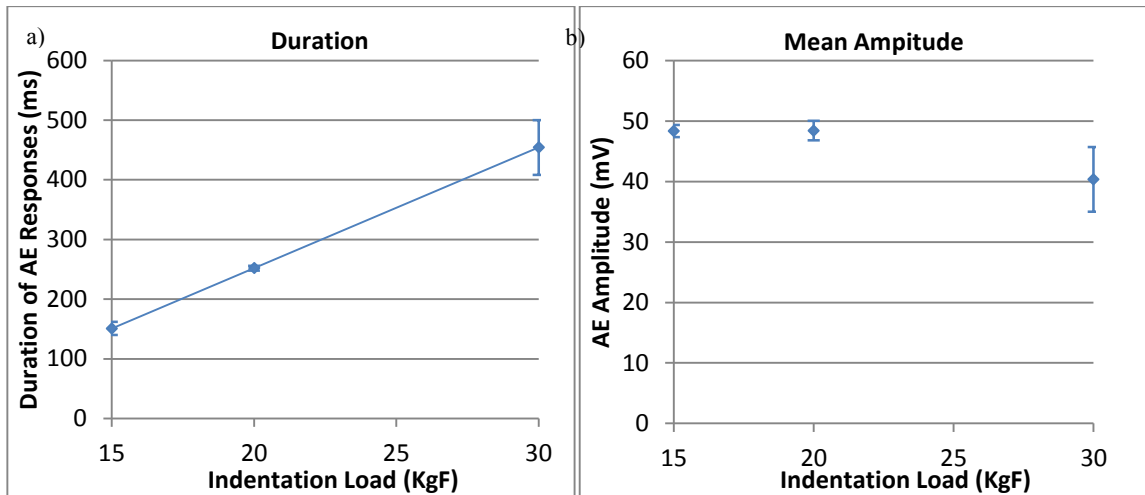


Figure 5.12: Graphs of mean duration and mean amplitude responses for the steel Indentation Tests

The duration of the AE response increased linearly with load, correlating with the indentation geometry parameter $2a$ (see Figure 5.8). In contrast, the magnitude was relatively insensitive to load and no distinct trend was apparent.

Responses in the frequency domain (refer to Figure 5.11d) did not demonstrate any prominent peaks; however, the frequency bands 0.1 – 0.16 and 0.21- 0.28 MHz appear to dominate, these were consistent across all steel tests (refer to Figure E.9 to Figure E.11 in Appendix E).

iv. AE Signal Characteristics -Silicon Nitride

Figure 5.13 presents the results of the smoothed waveforms truncated at 4 seconds (encompassing the most relevant part- for details refer to Appendix E) showing both A & B repeats for each load level. Also indicated are Faisal's threshold at 40% of peak, and a derived frequency pattern which is discussed separately below.

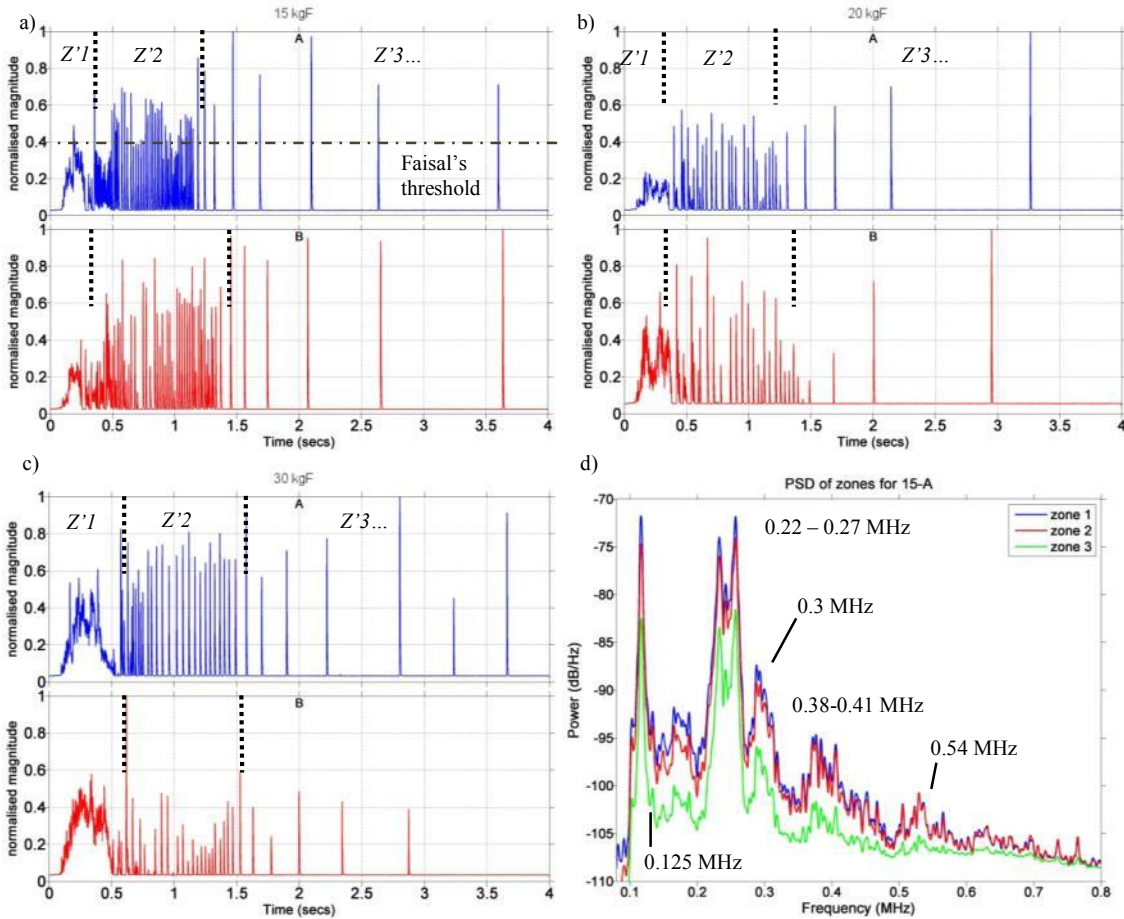


Figure 5.13: a) to c) The truncated and smoothed AE waveforms for all silicon nitride indentation tests for each of the three test loads from a) to c) 15, 20, 30 kgF. The subsequent zones have been indicated for each test and their respective boundaries highlighted d) PSD estimates of the three zones for 15-A.

Unlike in steel, the AE signal profile in silicon nitride indentations is characterised by a varied and complex signal pattern from the start and pertaining for a period (<1.5 seconds in all tests) after which the pattern settles into a series of simple sharp peaks with increasing intervals in between.

The responses did not demonstrate the three classical zones described by Faisal [155] and comprising a central zone of high signal intensity values flanked by zones of lower intensity values, separated by crossings of the threshold.

However, by using the criterion of distinct pattern (instead of threshold crossing), three zones can be identified for all the tests and their boundaries are shown by dashed vertical lines in Figure 5.13a-c.

Figure 5.14a&b show the mean durations and amplitudes (calculated on their rectified waveforms¹⁹) of each of the three zones respectively.

¹⁹ For zones 2 and 3, a mean was calculated on the extracted signal peaks above Faisal's threshold.

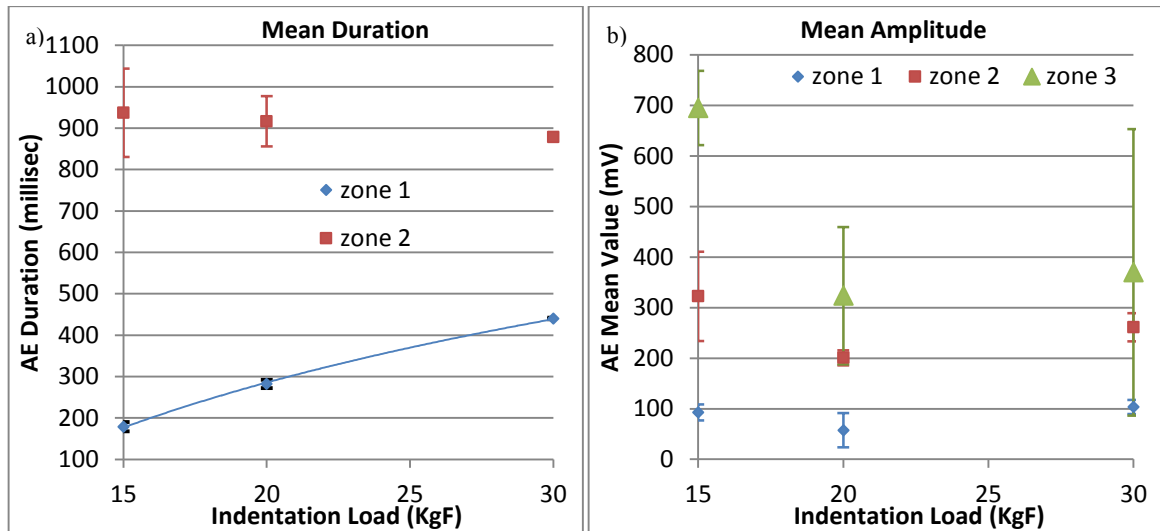


Figure 5.14: Graphs of a) mean duration and b) mean amplitude responses of each of the three zones of the AE waveforms in silicon nitride indentation tests.

Zone 1 commenced at approximately First Contact (see Figure 5.10) and its completion was marked by the signal response returning to the baseline level. The duration of Zone 1 was observed to increase with load, whilst no discernible trend was found for the mean amplitude between loads (refer to error bars of Figure 5.14b).

Zone 2 exhibited a dense pattern of sharp peaks of approximately the same high amplitude and a maximum of 40 milliseconds in between. After approximately 1 second, the duration between peaks increased above 40 milliseconds, marking the end of zone 2. The duration of this zone appeared to reduce with increasing load, the difference being approximately 40 milliseconds.

Zone 3 also exhibited a sparse pattern of high amplitude peaks. The duration between the peaks were greater than 40 milliseconds and increased with time. The duration of Zone 3 continued until the full removal of the indenter tip (and is therefore out of scale in Figure 5.14a see Figure E.15 in Appendix E). The mean amplitude of peaks was considerably higher than for Zone 2 (refer to Figure 5.14b).

For all zones, AE amplitudes were found to be highest at the tests performed at a load applied at 15 kgF, and lowest for tests performed at 20 kgF. There was no obvious explanation for this.

Frequency analysis was performed by calculating the PSD estimates for each of the zones and the results are presented in Figure 5.13d. The resulting profiles shows 5 characteristic peaks with similar magnitudes across the zones.

5.2.3.3 Discussions

The AE responses from indentations made to silicon nitride demonstrated significant difference to that observed in steel.

i. Steel

For steel indentations, a single zone of AE activity was distinguishable, commencing with a substantial delay from the First Contact (refer to Figure 5.10). This feature is similar to the findings of Faisal [155] in steel, where a delayed signal with consistently low amplitude was observed. The delay is attributed to the elastic accommodation response of the material as the indenter progressed into the specimen: at a critical load the stresses below the indenter tip surpassed the material yield, resulting in substantial plastic deformation and the onset of an AE response. The on-going continuous movement of the indenter is then accommodated by the material in a smooth fashion therefore producing a relatively featureless AE signal, with a duration that is proportional to load (compare Figure 5.8 with Figure 5.12a) and the resulting indent size as shown in the visual inspections.

ii. Silicon Nitride

Unlike in steel, the accommodation mechanisms in silicon nitride includes not only elastic and plastic deformations (as reported for example by Akbari [248]), but also cracking (see [249]). Of these, cracking is much more sudden, representing discontinuities in the deformation processes and therefore expected to produce more pronounced AE signatures (as found for example by Bouras [250]). In particular, research by Restall and Wang [11] and Zhe [251] found cracking in silicon nitride producing a signature signal with a characteristic frequency of 0.54 MHz.

Therefore the presence of three zones with distinctly different signal patterns indicates that these processes occur in varying proportions in the course of the indentation.

The AE profile, whose onset followed immediately the First Contact indicates that, in a brittle material such as silicon nitride, the initial accommodation by deformations is severely limited leading almost immediately to cracking. The relatively smooth profile of Zone 1 is explained by the continuing accommodation by elastic and plastic deformations due to an indentation load which is still increasing in this zone. However, due to the brittleness of the material, the deformations are producing dispersed micro-cracks at a rapid pace thus maintaining a smooth overall profile. A further evidence of cracking is the presence of a frequency peak at 0.54 MHz in Figure 5.13d.

The regular pattern consisting of dense peaks in Zone 2 suggests the micro-cracks have coalesced into the full cracks resulting in incremental crack growth as the stress intensities at the crack fronts periodically reach the critical level (K_{1c}) due to the continued descent of the indenter into the material. Correspondingly the signal intensities also become larger.

Zone 3 shows a significant increase in the intervals between the peaks, signifying a reduction in crack propagation. This is due to the reduction of the descent speed of the indenter and a possible toughening of the material following a rising R-curve behaviour (as described for example by Okada [252]). The withdrawal of the tip at the end of this period at approximately 11.9 seconds also produces a crack signal characteristic as a release of the residual stress built up in the material previously (as identified by Lawn [48]).

The quality of the AE signals obtained in both silicon nitride and steel tests was sufficient to perform this detailed analysis above. It therefore demonstrated that the sensor had adequate sensitivity for both materials and that there was no significant attenuation over the wave propagation path.

The high magnitude of the silicon nitride response, (at ~700 mV average for indent load 20 kgF during zone 3, refer to Figure 5.14b) compared to steel (at ~50 mV refer to Figure 5.12b), was an indication of the suitability of the sensing technique for monitoring crack propagation in hybrid systems as it appears to be able to distinguish between the two different materials.

The analysis shows that the prominent frequency peaks do not vary significantly over the range of loads tested and therefore, that any of the loads settings is suitable for the subsequent signal processing development. Also, because the frequency peaks are of similar magnitude across the frequency range, several frequency bands could be chosen for subsequent filtering such as 0.125, 0.22 – 0.27 MHz etc.

5.2.4 Healthy Run

The objective of the Healthy Run was to provide background data under situations where no pre-crack propagation was generated. In particular, to determine:

- the signature of the background noise (for AE data processing)
- the dominant background resonant frequencies of the machine (for vibration data processing)
- the signature of the ‘natural’ eccentricity of the machine (for WSS data processing).

The subsections below describe the organisation of the test and the data collected.

5.2.4.1 Run Description

The test was run on the modified Twin Disc machine instrumented with all sensor types and data acquisition systems. The specimens were the standard Production Run samples as described in section 4.3. The run lasted 75 hrs, anticipating the typical length of the Production Runs. The significant run parameters are detailed in the following table:

Parameter	Value
Load (kN)	1
Max. contact pressure (GPa)	3.5
Slide-Roll Ratio (%)	2
Upper Specimen (Si ₃ N ₄) Speed (RPM)	3,300
Lower Specimen (Steel) Speed (RPM)	1,125
Calculated Lambda Ratio	3.2

Table 5.3: Healthy Run settings.

The load was chosen at the ‘moderate’ setting (see next Chapter for details) in order to provide a baseline signal data²⁰.

The data were acquired as follows:

AE Sensor data was recorded in burst samples of 1 second length with a sampling rate of 5 MHz taken at 5 minute intervals.

Vibration Sensor and WSSs data was recorded in burst samples of 20 seconds length with sampling rates of 32 kHz taken at 5 minute intervals.

5.2.4.2 Specimen Inspections

Following the run the specimens were visually inspected to identify the presence of any faults and to detect any significant wear effects. The following instruments were used:

- AIF machine to visually inspect the entire wear track of the specimens for their surface status
- Talysurf Profilometer to quantify the generated wear (for the steel)

The results of the Talysurf Profilometer inspections showed that no measurable wear was observed in the silicon nitride specimen (similar to the findings in [61]), whilst the steel surface showed a thin and shallow wear track shown below:

²⁰ The other possible setting, ‘severe’, was likely to induce significant wear.

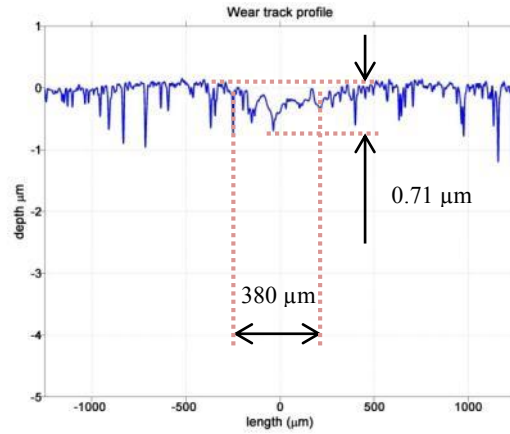


Figure 5.15: Talysurf wear track profile of the steel specimen from the Healthy Run.

A standard assessment of the wear is the wear rate:

$$\text{Wear Rate} = \frac{\text{Volume loss (mm}^3\text{)}}{\text{Load (N)} \times \text{sliding distance (m)}} \quad 5.1$$

for which the volume loss was calculated by the equation below, where the area was calculated using the wear track profile within the section bounded by the vertical markings indicated in the figure above.

$$\text{Volume loss (mm}^3\text{)} = \text{area} \times \text{specimen diameter} \quad 5.2$$

The *sliding distance* was calculated using the test time (75 hrs) and sliding speed (150 mm.s^{-1}) of the run giving a wear rate of $7.4 \times 10^{-7} \text{ mm}^3.\text{Nm}^{-1}$. This result indicates a mild wear regime [57], indicating that the lubrication regime was sufficient to separate the specimens over the duration of the run, as intended.

The mean coefficient of friction value was calculated from the torque and load measurements taken throughout the run as a reference for the Production Runs. This value was 0.07.

5.2.4.3 Signal (Quality) Analysis

The data obtained are discussed in the course of the signal processing development in the section following, before that a preliminary assessment was made broadly reflecting on the quality of the run (i.e. whether some important assumptions and objectives were satisfied).

For this, the raw data collected were processed in a similar way to the above (feature emulation) tests i.e. their frequency profiles were produced by PSD estimates.

i. Vibration Signal Data

The frequency profiles are presented in the following figure:

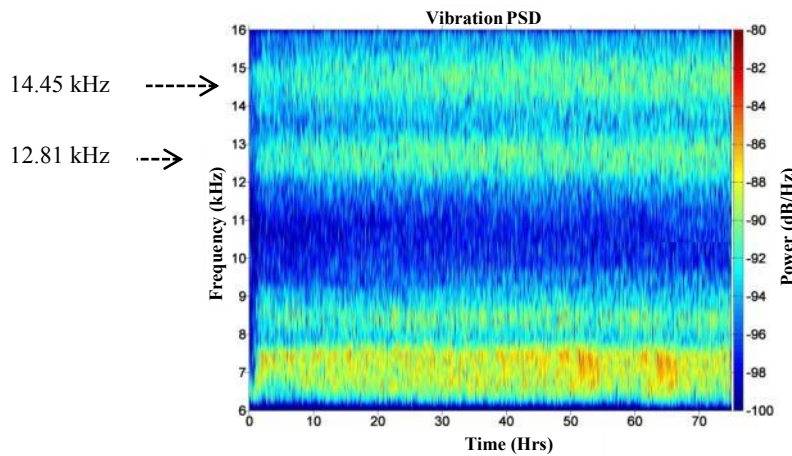


Figure 5.16: Vibration PSD profile for the Healthy Run

The profile shows the presence of bands. These are explained as resonant frequencies of the assembly; they remained relatively constant throughout the run. A closer comparison with the results from the impact tests shows that one of the frequencies, 10.81 kHz (5), see Figure 5.3, is missing whereas a new peak appears at around 12.81 kHz, below the next impact peak at 14.45 (6)²¹. These indicate that there are differences in the resonant characteristics between a stationary and a running machine, thus making the choice of the optimum peak frequency for feature extraction more difficult.

ii. WSS Signal Data

The following figure represents the RMS time profiles of the two sensors:

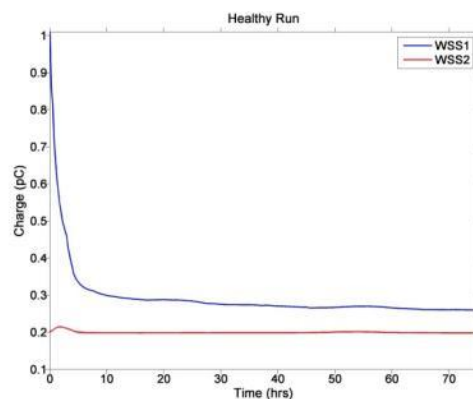


Figure 5.17: WSS RMS profiles for the Healthy Run

The diagram shows very high charge at the beginning on WSS1 (ceramic) with a sharp drop off within 5 hrs and a continued decay thereafter at a significantly reduced rate. This behaviour is as

²¹ This appears in both impact and Healthy Run profiles.

expected from the tribocharging effects of the oil (and has been observed in hybrid sliding contacts [145]). The smooth curves (on both sensors) are further confirmation that no RCF faults were being generated.

TSaAs were computed for both sensors taking averages from each burst period. The purpose was to see how the charge develops in time and is distributed on the surfaces of the specimens. The following figures present the results:

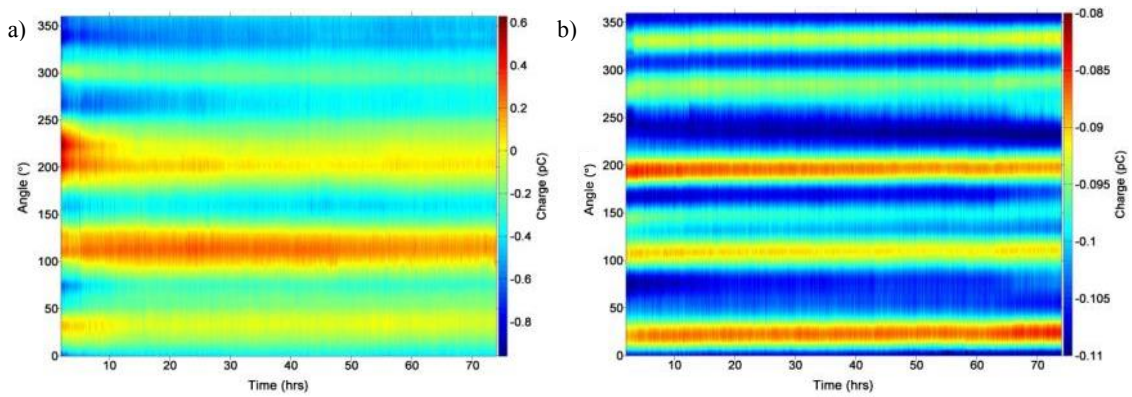


Figure 5.18: WSS signal profiles distributed across the rotation of the specimens a) silicon nitride and b) steel

The figures show the presence of bands of varying charge across the surfaces. As expected for a conductive material, the steel charge distribution is much finer range than that of silicon nitride (refer to scales). Also, it shows that only negative values are seen, thus indicating that the positive ions are being removed by oil only, i.e. that a separation of the specimens was maintained throughout the run. The silicon nitride showed more pronounced bands; both positive and negative charges are seen, explained by the non-conductivity of the material i.e. that charges remain in situ.

Figure 5.19 below provides an illustration that charge is also being exchanged between the specimens at these bands.

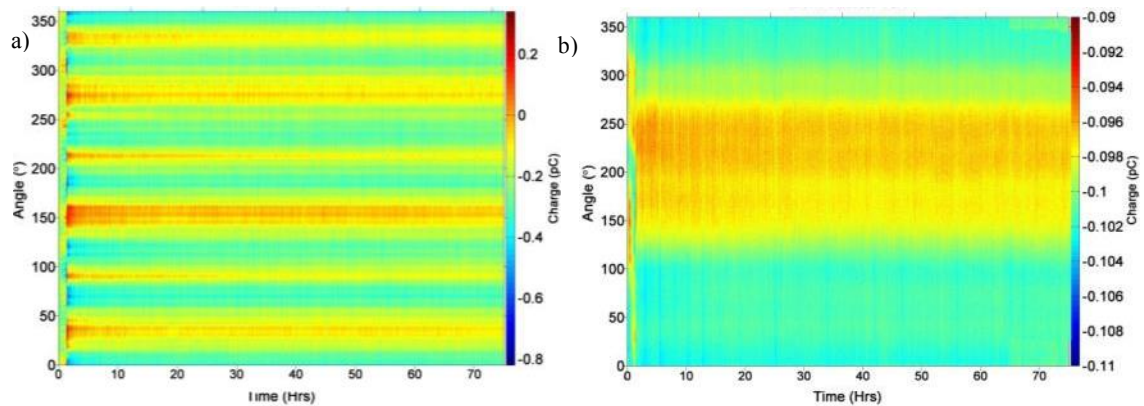


Figure 5.19: WSS signal profiles distributed across the rotation of the specimens a) silicon nitride, b) steel, cross-timed (lower and upper tachometer signals respectively)

It presents the profiles timed against the counter specimen tachometer. The resulting bands reflect therefore the ‘natural’ bands of the counter specimen. For example WSS1 (monitoring silicon nitride refer to Figure 5.19a) has been timed against the steel tachometer and therefore reveals the bands that correspond to the steel bands in Figure 5.18b - with much higher charge levels as explained above. Similarly, Figure 5.19b shows that the corresponding effects of the silicon nitride specimen on steel produce very low charges.

Notably, when timed against the steel specimen, WSS1 shows more prominent negative charge values. The triboelectric series indicates that steel strips the electrons from silicon nitride when rubbed together, thus silicon nitride becomes positively charged (as observed by Morris [132]). With the presence of the oil in the Healthy Run, negative charges are attracted to the triboelectrified positively charged region of the silicon nitride surface and through tribocharging effects (described in Subsection 3.2.3.1), produces a double-layer that is negatively charged at this region. The consistent pattern throughout the run indicates that the natural eccentricity bands of the steel face are responsible for inducing charge. This effect is further discussed in Chapter 8 section 3.

The following figure presents frequency spectra for silicon nitride taken at the start and end of the run.

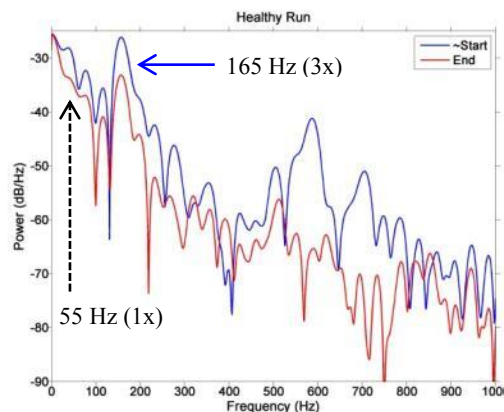


Figure 5.20: WSS1 frequency profile of the silicon nitride specimen

The higher profile at the start of the run confirms again the tribo-charging effect at the start of the run. The peaks at 165 Hz are due to an eccentricity of the specimen corresponding to 3x the rotational frequency.

iii. AE Signal Data

Presented below are the frequency profiles:

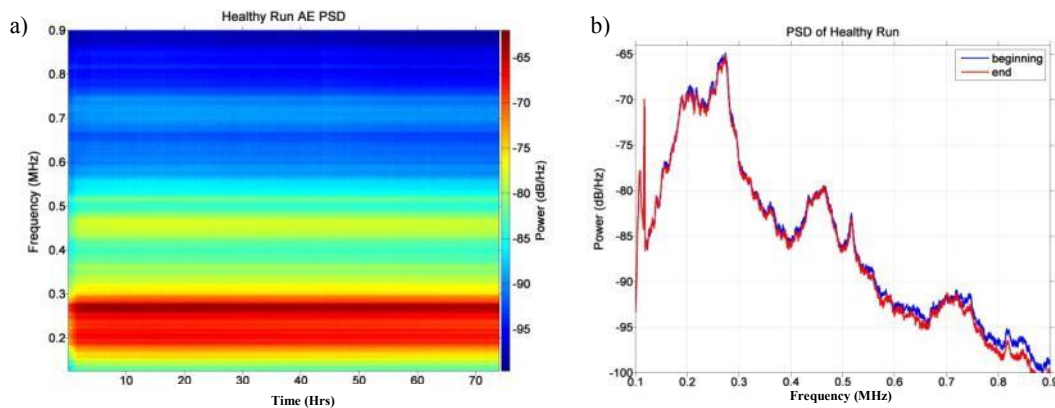


Figure 5.21: Frequency profiles: a) time profile, b) the frequency spectra at the beginning (5 hrs) and the end (70 hrs) of the run.

The frequency-time profile (a) shows that the frequency spectrum was constant throughout the run- therefore, there were no RCF faults being generated, as required from the run. This is also confirmed in detail by the frequency spectra in b).

5.2.4.4 Run Evaluation

Based on the data presented above, it was felt that Healthy Run fulfilled the basic objectives set for it by the measuring of:

- the resonant vibration frequencies of the test cell under running conditions
- the dominating eccentricity components of the upper specimen.
- the AE background noise profile.

The work following utilised the data to set up the signal processing methods.

5.3 Signal Analysis Development

The following section describes how the signal processing systems, notably signal enhancement, were developed.

5.3.1 Vibration

The results from the Impact tests and Healthy Run produced vibration spectra with complex patterns (refer to Figure 5.2 & Figure 5.16) with portions that varied significantly between the two tests (notably the absence of impact peak (5) in Healthy Run and a new peak occurring in the latter). Therefore, it was difficult to select a suitable high resonant frequency (as was the intention following the discussion in 5.2.1) for further signal analysis and feature extraction.

Other researchers (see for example [253]) have found that similar complex vibration spectra are produced in machines in general, and a technique has been developed recently that obtains positive results from their analyses. The technique, empirical mode decomposition (EMD), is an alternative way of decomposing the raw (TSA) profile into elementary frequency components (resonant oscillatory modes - ‘intrinsic mode functions’, IMFs), effectively representing the ‘carriers’ of the fault signals.

It therefore presented itself as a suitable candidate to be used in this project and by this perform its assessment for use in hybrid bearings. The data from the Healthy Run test was used to learn in detail about the technique, to assess its capabilities and to identify any parameters that should be carried forward to Production Runs.

The technique has been incorporated in several operational algorithms; the one created by M.D. Ortigueira [254] was used as it is one of the most recent ones and was readily available from the internet as a MATLAB add-on. The method also claims to improve on several previous algorithms, removing some of their deficiencies that have been identified in practice detailed in [254].

The method was set up as follows:

Its fundamental parameters, the sifting parameters $q_{residual}$ and $q_{resolution}$, were set to 60 dB, representing a most rigorous sifting.

The TSAa vibration data from the middle of the Healthy Run²² was chosen for analysis and the periodic padding technique [255] was applied in order to reduce the effects from border distortion. The results were then processed by the EMD algorithm producing a decomposition of the spectrum into elementary modes. Two elementary modes (IMF1 & IMF2) were selected because they emphasised the highest portions of the frequency spectrum and the resulting EMD profiles are presented together with the original raw (TSAa) in the figure below.

²² the spectrum did not vary significantly over the run

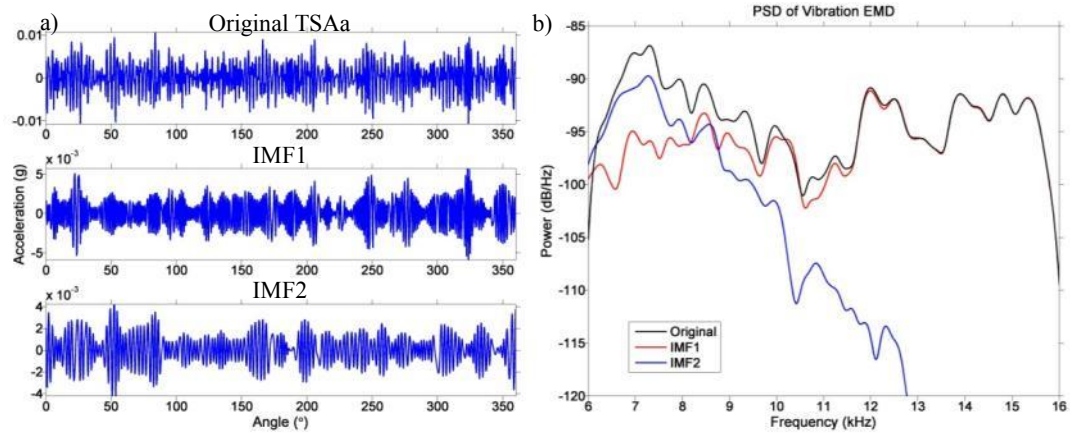


Figure 5.22: The raw (TSAa) and the decomposition mode profiles IMF1 & IMF2 showing a) angular profiles and b) frequency profiles (PSD)

The most prominent features of the figures are that the two profiles (IMF1 & IMF2) are significantly different from each other and that IMF1 (this was the highest frequency mode) appears quite close to the original TSAa profile in Figure 5.22b, especially in the high frequency range. The reason for this was not apparent at this time, it was hoped that the Production Runs would illuminate the matter more. As the IMF2 profile appears to cover only the lower end of the frequency spectrum, the IMF1 setting was carried forward into the Production Runs.

5.3.2 Electrostatic Charge

The objectives of the signal processing system were to remove the eccentricity effects in the signal profiles. As the Healthy Run results show (Figure 5.18a) these represented a significant proportion of the profiles, masking potential fault contributions. An additional difficulty was that the amplitude of the signal varied in the course of the test (continually dropping), therefore traditional eccentricity profile removal methods (based on fixed profiles, e.g. [125]) were not applicable.

The solution was to generate an adaptive method based on continuously updated profiles. Such a method was provided by a band-pass filtering technique, followed by a subtraction of the filtered signal from the original raw profile.

A digital band-pass filter works on the principle that only a specified band of frequencies is retained in the signal. To set up such a system for this application is straightforward as the frequency of the eccentricity is known from the Healthy Run (see Figure 5.20). A spline is generated representing the eccentricity profile using a Hanning digital-filter with centre pass-band at the frequency of the eccentricity profile – found to be 3 x the specimen rotating speed in the Healthy Run.

This is then subtracted from the original profile, therefore producing a signal profile describing the contribution of the remaining charging effects. For the Healthy Run these represent only the

effects inherent in the set up (the additional fault-induced effects will be present in the Production Runs).

The diagrams below show the corresponding profiles at the beginning and the end of the Healthy Run:

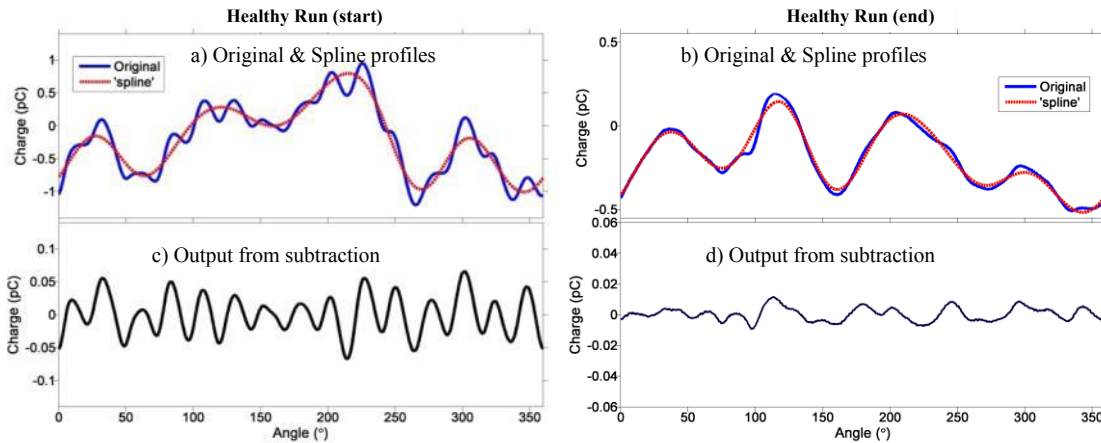


Figure 5.23: WSS1 profiles taken at various stages of the data processing development.

A significant aspect of the result is that the remaining charge signals are significantly reduced, below the expected fault-induced effects. Therefore the resulting technique was considered sufficiently sensitive to be carried forward to the Production Runs.

5.3.3 AE

The diagrams in Figure 5.24 below show the frequency and time patterns taken from the silicon nitride indentation test (test 30-A start of zone 2) and the Healthy Run based on a 1 second burst recording taken towards the end of the test²³. The first represents a ‘pure fault’ profile and the latter is a ‘pure noise’ profile; a key assumption was made that the results obtained from the two tests were comparable (ran on different machines therefore with different signal propagation paths but the specimen, the sensors and the acquisition system were the same).

²³ The Healthy Run discussion above shows that the precise location of the selected period within the run not critical as the frequency profiles remained almost constant throughout

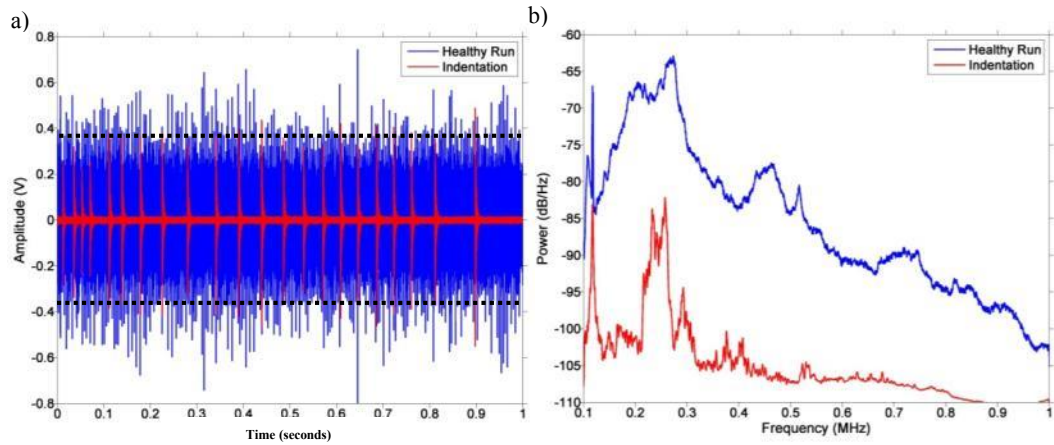


Figure 5.24: Time (a) and frequency (b) patterns of Indentation 30-A (Zone 2) and Healthy Run (end of test).

Figure 5.24a shows that the amplitude levels from the faults are surpassed by the noise levels throughout the duration of the burst. This potentially renders Hit-based analysis as inappropriate for RCF failure detection, as any threshold would be swamped by the noise data (the horizontal lines in the diagram indicate a potentially suitable threshold level for a Hit-based analysis). However, as the two profiles were obtained directly (without any normalisation applied) from two different machines a Hit-based analysis method was still considered worthwhile to be performed in the Production Runs.

Considering the frequency profiles, Figure 5.24b shows that the noise level is significantly higher than the fault level but, more importantly, that both show prominent peaks in the same (low) range of frequencies. In contrast, at the higher end of the spectrum (>0.55 MHz) the noise shows patterns with more features than the fault signal. Because of these characteristics, any frequency-based filtering method could be difficult.

As the frequency profiles (Figure 5.24b) nevertheless did show perceptible variation between the fault and the noise, a more detailed frequency analysis was performed. In order to do this, it was necessary to adopt a filtering method that was computationally efficient (to cope with large volumes of Production Run data) and could provide steep roll-off rates (i.e. high separation between the pass-band and unwanted frequencies). As the first criterion ruled out the EMD method²⁴, a wavelet packet (WP) filtering method was subsequently selected following the review of time-frequency signal enhancement methods in Section 3.3.

²⁴ The method used in the vibration analysis was appropriate as the amount of data was significantly smaller.

In order to set up the WP method, two parameters defining the characteristic frequency pattern had to be specified:

- A mother wavelet that represented the feature (fault) shape
- A wavelet packet representing the frequency characteristics (bandwidth and centre-frequency)

In-order to select the mother wavelet, 69 candidate wavelet patterns were chosen and each used in turn in the continuous wavelet transform (CWT) process applied to the indentation signal over the full frequency range. The results were compared with one another and the one demonstrating the best selectivity²⁵ was chosen (for details see Appendix F).

The indentation signal was combined with a Gaussian noise profile and then transformed using the discrete wavelet transform (DWT) with the selected mother wavelet into a set of wavelet packets ('nodes'). The wavelet packet with the best selectivity was then chosen as input to the next iterative DWT process. This procedure was terminated when the selectivity of the resulting wavelet packet was not greater than that of the input packet. The outcome of this procedure was therefore a transformed profile that has been optimally enhanced to extract the fault from a random background, and a set of indices designating the nodes selected in the iterative transformation process producing that profile (referred to subsequently as '*WP filter*'). For details of the above see Appendix F.

To test this transformation process, a simulated RCF profile was created by adding the indentation signal to the Healthy Run signal (i.e. those shown in Figure 5.24a above)²⁶.

A wavelet analysis of the combined data was then performed with the WP filter above (the details of this are also presented in Appendix F).

²⁵ Selectivity was measured by the kurtosis-to-entropy ratio (KER)- higher values indicating better selectivity. Essentially it measures how well the profile distinguishes between a source signal and a random background.

²⁶ This test was an exploratory i.e. the possible different weights of these profiles were ignored.

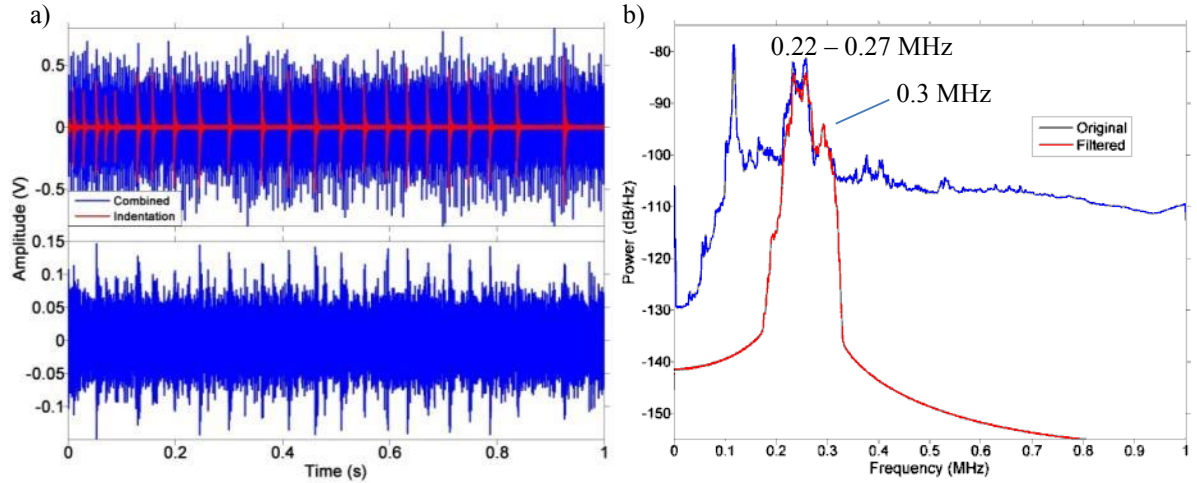


Figure 5.25: a) combined signal profiles: pre- (top) and post- (bottom) WP filter, b) Indentation frequency profiles: pre- and post- WP filter.

As seen in the figure, the results of the WP filter did produce a more pronounced distinction between the fault signatures (seen as peaks) and those of the background. Interestingly it has selected two of the most prominent frequency peaks.

Nevertheless a decision was made to investigate whether the use of an additional processing analysis, namely the EI method, (being appropriate for the extraction of signals from high-noise data in AE profiles, reported by Al-Balushi [168]) would further enhance the profile to make the fault features more prominent. In order to do this, the method was applied to both the unfiltered and filtered (WP) profiles and the results were compared and also related to the original fault signal from the indentation machine.

A window of 4096 was chosen (same as that applied in the indentation tests) and the RMS_{total} was set to incorporate the whole signal waveform (refer to equation 2.20 in section 3.1.2).

The results are shown in Figure 5.26 below.

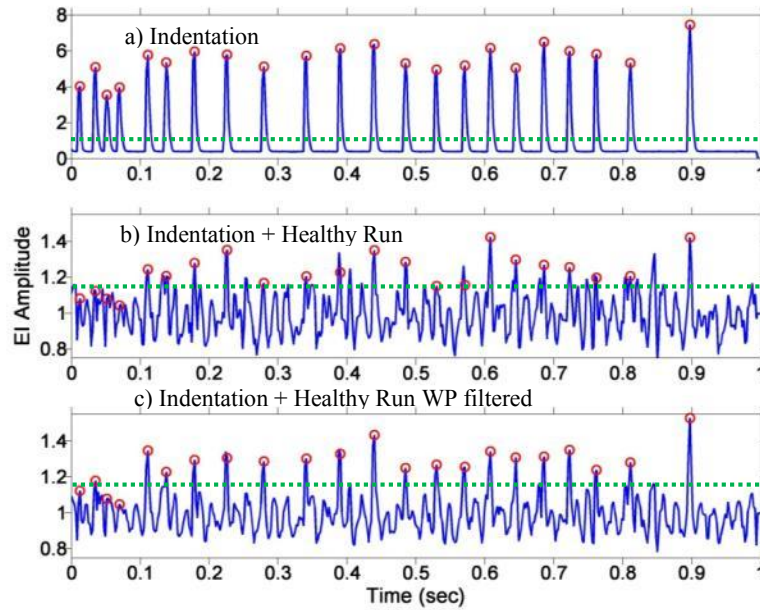


Figure 5.26: EI of AE waveforms with the fault-related peaks indicated in red

The horizontal lines in the figures represent a value 15% above the normalised mean signal (i.e. 1) simulating a sensitivity threshold for the selection of peaks representing significant events. With this setting, all of the peaks in a) are included (as expected), the peaks in b) include both fault and background (and moreover some faults are lost) and c) shows that most of the faults are recovered and the background is attenuated. This is illustrated in Table 5.4 below.

Waveform	Fault-related peaks (F)	Noise peaks (N)	Efficiency $F/(F+N)$ (%)
Indentation	22	0	100
Indentation+ Noise	15	11	42
Indentation+ Noise filtered	19	3	59

Table 5.4: Prominent fault and noise peaks and the filtering efficiency

The results shows that to make AE sensing effective, an advanced filtering method (consisting of wavelet filtering and EI method) has significant advantages as it provides a better distinction of the fault signals from those of the background than conventional analyses (i.e. standard frequency filtering and possibly also Hit-based); however, even this is not perfect. Nevertheless, the work performed in the analysis of the AE data resulted in a sensor and processing set up that was deemed appropriate for the Production Runs.

5.4 Summary

Two distinct series of tests were performed for each sensor to generate reference data for the development of the signal processing techniques for use in the Production Runs. The first series emulated the generation of RCF faults and the second series (Healthy Run) was designed to emulate the background. The resulting data were analysed using several methods to determine their characteristics. Based on the two analyses for each sensor, an appropriate signal processing method was sought that could enhance the fault features.

Vibration produces a spectrum characterised by resonant frequencies. These get excited by various sources, notably progressing faults in specimens, the basic operation of the machine and also potential external influences. Based on the findings of other researches, the focus of the signal processing was on high resonant frequencies as they were reported to possess a better fault signal-to-background ratio than lower ones.

To excite the resonant frequencies, mechanical impacts were made to the stationary Test Cell of the Twin Disc machine. The resulting profile was analysed using PSD estimation, resulting in the identification of 5 prominent resonant frequencies. These were compared against the resonant frequencies generated in the Healthy Run resulting in some differences. In order to resolve them, the EMD method was chosen with a profile that had the best high frequency emphasis.

The **WSS** tests had an additional objective, to verify the suitability of the new sensor head designed specifically for use on silicon nitride. The first series tests were performed on the Twin Disc machine running with a calibration disc created to emulate a fault and equipped with both the old and new sensor heads in turn. The resulting signals from both heads were noisy and much of the noise was removed by the application of TSA; the transformed profiles confirmed the suitability of the new head as it provided a higher spatial resolution and a reduced eccentricity profile.

Nevertheless, eccentricity bands dominated the profiles potentially masking any fault signatures. The Healthy Run revealed further details about the eccentricity profiles, namely that their amplitude varied significantly over the course of the run, but the frequency bands remained constant. Therefore, the main objective for the signal processing was the attenuation of these eccentricity bands. This was achieved by applying a band-pass filter centred at the eccentric frequencies generating an eccentricity signal profile which was subsequently subtracted from the original.

Fault generation in **AE** tests was performed on the indentation machine as it could emulate the mechanisms responsible for RCF fault generation (specifically cracks), and provided a number

of advantages over the Twin Disc machine; notably the production of specific profiles of silicon nitride and steel, and a quiescent operation generating less background noise. The ‘correct’ background noise was subsequently supplied by the Healthy Run.

The generated AE signals were complex profiles in both time and frequency, and in addition the Healthy Run produced a very high noise profile over the entire frequency spectrum, thus presenting a serious problem for any fault signature identification. To overcome this, a 2-stage frequency filtering method was developed, consisting of a wavelet packet filter followed by the EI method.

The method was developed a series of calculations involving the processing of noisy data, and using several standard profiles to choose one that gave the best selectivity. The 2-stage process was shown superior to a single EI method, by demonstrating that is produced, from a noisy input, a filtered signal better matching the pure fault signal created in the indentation tests.

The following table summarises the signal processing methods developed that were carried forward to the Production Runs.

Profile	Technique
Vibration	TSA + EMD (IMF1)
WSS	TSA + Eccentric Profile Subtraction
AE	Wavelet filter + EI method

Table 5.5: Summary of the advanced signal processing techniques

Chapter 6. Production Runs Planning and Execution

6.1 Overview of Process

Having completed the setup of the machines and the acquisition and processing systems as described in the previous chapters, the equipment was ready for Production Runs. These were at the heart of the project, performed to emulate RCF fault propagation and provide signals for the collection and enhancement and for their subsequent analysis and feature extraction.

To provide a wide variety of signal profiles, several Production Runs were executed with different pre-crack configurations and with different load and slide settings.

6.2 Runs Planning

The Twin Disc machine allows variable settings of several parameters. In order to maintain the same charge characteristics for all runs, it was decided to keep both the oil temperature and type unchanged. This left the main parameters to be varied the load and slide-roll ratios, combined with different pre-crack sizes, selected as follows.

Two **load settings** were adopted²⁷

- 2.5 kN resulting in a maximum Hertzian contact pressure of 5 GPa- comparable to the conditions used by many researchers in a similar work on hybrid rolling contacts (e.g. Hadfield et al [55, 83]); this setting represented a '*Severe Load*'
- 1 kN resulting in a maximum Hertzian contact pressure of 3.5 GPa (the same value as applied in the Healthy Run to ensure minimal wear), representing a '*Moderate Load*' and selected to emulate conditions near the high end of real operations of hybrid bearings.

Both load settings represent higher contact pressures than those operated under service conditions (typically < 2 GPa); they were chosen to accelerate the development of the faults thus to reduce run times and the amount of signal data to be analysed.

Three **slide-roll ratio** (SRR) values were considered:

- 0 %, representing pure rolling condition
- 2%, representing bearing realistic conditions [14]
- 4%, providing severe slippage.

²⁷ The pressures were calculated using the equations located in Appendix A.

The pre-cracks generated on the silicon nitride specimens for each run were of three **pre-crack sizes**:

- 5 kgF, replicating the indent adopted by Hadfield et al [55, 83], representing ‘small defect size’ (S)
- 10 kgF, representing the indent adopted by Kida et al [51], representing ‘medium defect size’ (M)
- 12.5 kgF, set above the moderate level and representing a ‘large defect size’ (L)

Each specimen was indented with three pre-cracks. The advantage of this was that multiple fault data could be generated in each run therefore making the testing more efficient (given that the lengths of the runs were in some cases exceeding 100 hrs).

Runs were lubricated with a synthetic base oil available in-house: PAO4, pumped to the contact at a rate of approximately 60 ml.min⁻¹ at 45°C with the resulting viscosity of 14.6 cSt.

Each run was terminated automatically by the in-built vibration cut-off sensor in the Twin Disc machine, set to prevent damage to the machine due to excessive vibration levels. The setting was done after the start of each run based on the vibration levels being reported by the in-built sensor; the value set was just above that level.

The actual runs performed were as follows (indicating also their rationale):

Test No.	Load Setting	SRR	Pre-crack Configuration	Comments
1-1	Severe	0	L, M, S	Scoping run 1
1-2	Moderate	4	L, M, S	Scoping run 2
2-1	Moderate	2	L, M, S	Full pre-crack range- shaft breakage
2-2	Moderate	2	L, M, S	Full pre-crack range- re-run of 2-1
2-3	Moderate	2	M, M, S	Investigating consistency of failures
2-4	Moderate	2	M, M, S	Repeat
2-5	Moderate	2	M, M, S	Repeat

Table 6.1: Production Run parameter configurations.

The objective of the scoping runs was to determine the general characteristics of the test runs under the two load settings, starting with the more demanding.

6.3 Execution

Each run was performed in a series of the following steps.

6.3.1 Preparation of Specimens

The silicon nitride specimen was first cleaned in an ultrasonic bath consisting of ethyl acetate for 15 minutes then rinsed in deionised water and dried with high pressure air.

Samples were gripped on their side faces using a mechanical vice, and the crest of the sample was found using a microscope with a ring light. When observed using the microscope, a ring of reflected light formed with the crest of the sample at its centre. This region was then marked for indentation using a fine marker pen at the intended position, on either side of the contact path. The marked sample and vice were then transferred to the indentation machine where an indent was made to the region indicated by the pen marking. Alignment was achieved using the machine's sight and movement of the x-y stage (see Figure 3.1). The sample was rotated in the vice, and the procedure was repeated to generate the remaining indentations.

Some trial and error was required before it was possible to make the indentation machine produce the cracks at precisely their intended locations.

6.3.2 Pre-Run Inspection

The resulting indentations on each specimen were inspected using the AIF machine to broadly check that they were as intended i.e. no significant loss of material and positioned within the contact path. Snapshots were taken under the polarizing light filter to indicate the shape of the subsurface lateral cracks such that any asymmetry could be identified that may have relevance to the subsequent fault development. The angles between indentations and the keyway were measured using the digital anglemeter with the rotational alignment made using digital cross-hairs against the perceived centre of the crack. The resulting values were recorded in the test log.

6.3.3 Setting Up the Twin Disc Machine

Before placing in the machine, both specimens (the silicon nitride and the steel counterface ring- refer to section 3.3) were cleaned in an ultrasonic bath consisting of ethyl acetate for 15 minutes then rinsed in deionised water and dried with high pressure air.

The rotational angle between the upper tachometer trigger point and the keyway passing the WSS1 sensor face was measured using the digital anglemeter connected to an exposed end of the specimen shaft. This measurement was required so that the precise time points at which the faults passed WSS1 could be determined, and was recorded in the test log.

A run was programmed into *Compendx* for automated test control and was configured to execute in 5 general steps summarised in Table 6.2 below. **Step 1:** the upper specimen (silicon nitride) speed was linearly increased from zero reaching 3,300 RPM in 2 minutes, whilst the lower (steel) was correspondingly increased such that no sliding was induced to the contact during the run up. A pre-load of 0.1 kN was applied to the contact. **Step 2:** the load was ramped up with a constant rate such that the intended full load was reached within 44 minutes after the test start. **Step 3:** appropriate slip was induced to the contact (as planned for the run) by varying the rotating speed of the lower specimen (see details below). At 45 minutes after the run start, the testing conditions were reached commencing the main run step. **Step 4:** the settings reached were held during the step which ran until the automatic shutdown sequence, Step 5, was initiated. **Step 5:** a fast removal of the load (typically achieved within 1 second) and the rotations halted.

Step Number	Specimen Speeds (RPM)		Load (kN)	Duration (min)
	Upper (Si ₃ N ₄)	Lower (Steel)		
1	0→3,300	0→1,081	0.1	2
2	3,300	1,081	0.1→Run condition	42
3	3,300	1,081→Run condition	Run condition	1
4	3,300	Run condition	Run condition	To shutdown
5	3,300	Run condition →0	0	0.06

←Main run step

Table 6.2: Summary of the test sequence programme implemented by Compendx for Production Runs

To achieve the desired levels of SRR in Step 3, the lower specimen rotating speed was altered according to equation:

$$SRR = \frac{v_l - v_u}{(v_l + v_u)} \quad 6.1$$

Where v_u and v_l were the surface velocities of the upper and lower specimens respectively.

The resulting specimen speeds are summarised in Table 6.3 below.

Run	SRR (%)	Upper Specimen (Si ₃ N ₄) Speed		Lower Specimen (steel) Speed		Calculated Lambda Values (Applied Hertzian pressure indicated)
		RPM	mm.s ⁻¹	RPM	mm.s ⁻¹	
1-1	0	3,300	3,566	1,081	3,566	3.1 (5 GPa)
1-2	2	3,300	3,566	1,125	3,711	3.2 (3.5 GPa)
2-1 to 2-5	4	3,300	3,566	1,193	3,935	3.1 (3.5 GPa)

Table 6.3: SRR values and their corresponding applied specimen rotating speeds

The remaining parameters have been calculated for subsequent post-run analyses to be as follows:

Run	Max. Principal shear stress		Max. Tensile Stress			Calculated λ Values ²⁸	Load (kN)	Applied Max. Pressure (GPa)
	Magnitude (MP) Si ₃ N ₄ /Steel	Depth (μm) Si ₃ N ₄ /Steel	Magnitude (MP) Si ₃ N ₄ /Steel	Contact widths (μm)				
				- Roll	- Trans			
1-1	1,641/1,630	250/252	752/718	960	1160	3.1	3	5.13
1-2	1,138/1,130	173/174	521/498	666	804	3.1	1	3.55
2-1 to 2-5	1,138/1,130	173/174	521/498	666	804	3.2	1	3.55

Table 6.4: Calculated values for various contact parameters corresponding to the run conditions²⁹. Unless otherwise stated, calculations were performed using the equations in Appendix A.

The shape of the contact made between the specimens is illustrated in the diagram below.

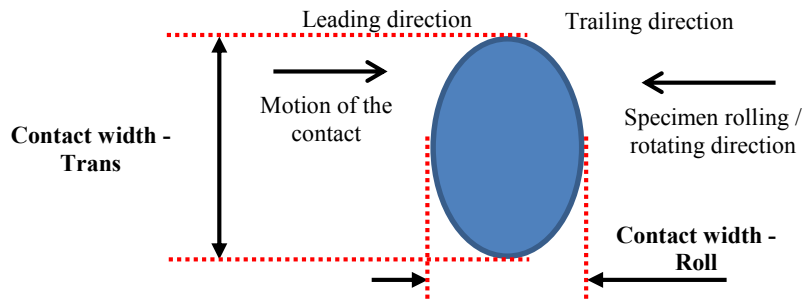


Figure 6.1: Sketch of the contact shape on a specimen, illustrating also the motion of the contact and specimen rotating direction. Leading and trailing directions refer to the rolling contact

6.3.4 Monitoring of the Run and the Data Collected

Because of the length of the runs (in some cases exceeding 90 hrs), the machine was left unattended with occasional checks performed at convenient intervals. The actual run durations achieved were as follows.

Run	1-1	1-2	2-1	2-2	2-3	2-4	2-5
Time (hrs)	2.1	21.5	47	94	76	98	106

Table 6.5: Actual run durations.

The data collected were the raw signal waveforms as described in the hardware description in section 4.2.3.3 and used in the Healthy Run (Section 5.2.4), and repeated below for convenience:

AE Sensor data was recorded in burst samples of 1 second length with a sampling rate of 5 MHz taken at 5 minute intervals.

Vibration Sensor and WSSs data was recorded in burst samples of 20 seconds length with sampling rates of 32 kHz taken at 5 minute intervals.

²⁸ Lambda values for each of the run conditions calculated using equation 2.7.

²⁹ friction not included in the calculations

6.3.5 Post-Run Inspection

Each sample was analysed post-run and the results were recorded in a similar fashion as for the pre-run. Residual oil was removed from the specimens by submerging them in a beaker of ethyl acetate before drying in air.

The Talysurf Profilometer was used to generate a profile of the steel wear track in the transverse direction at a random point along the circumference, indicating the wear track geometry. Specimens showing significant wear were further analysed using the AIF and inspected using SEM where snapshots were taken for further analysis.

The test results consisting of the signals acquired during the runs and the specimen snapshots taken before and after the runs formed the basis for detailed analyses, the outline of which is presented in the next subsection.

6.4 Overview of the Analysis Process

The programme of the results analysis was divided into the following stages:

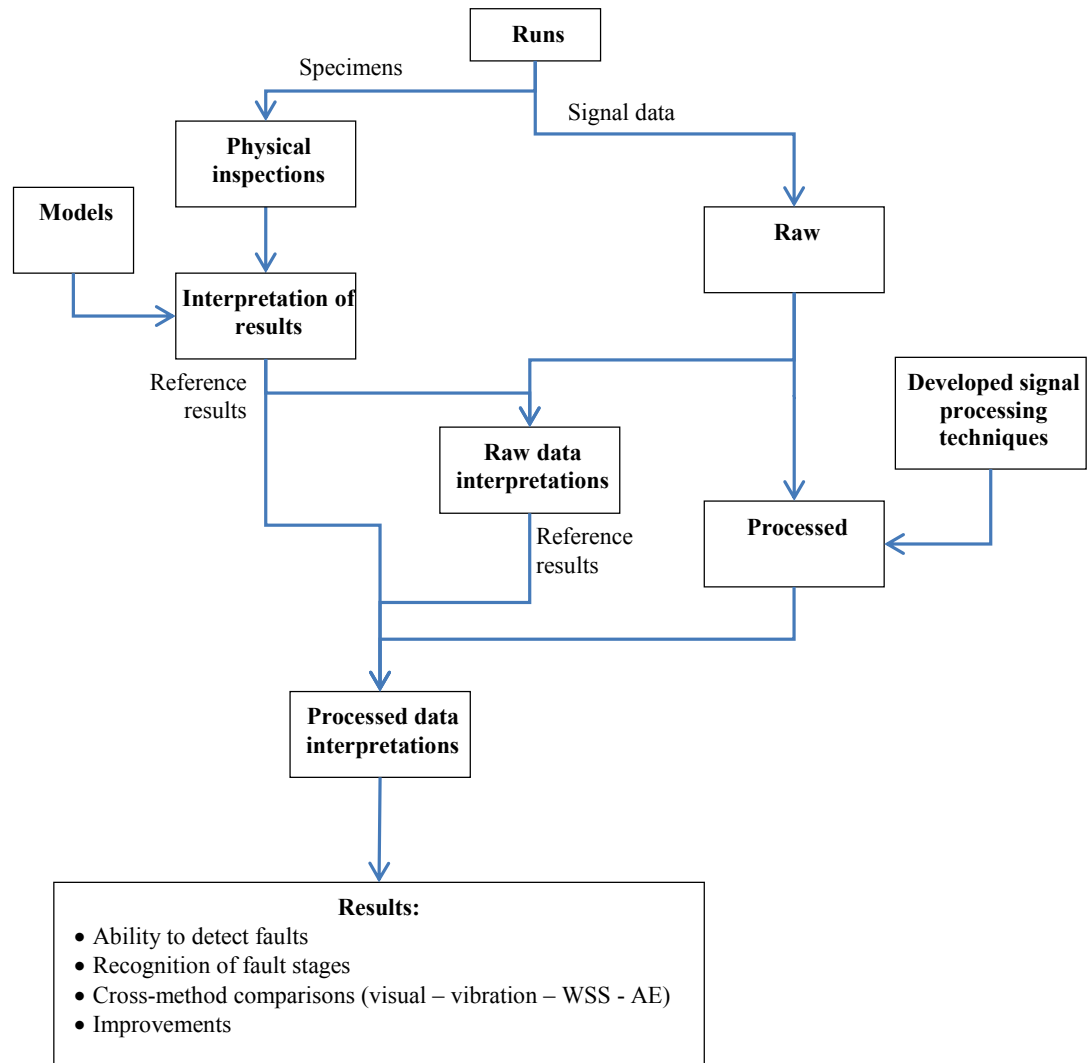


Figure 6.2: Flow chart illustrating the stages for the analysis of results

The stages are organised as follows.

Analysis of the physical inspections of the faults on the specimens provides a reference basis for the subsequent evaluation of the signal outputs. This is because it enables direct classification of the faults produced in the runs and, with the help of published and developed models, to deduce the basic mechanisms of their progressions. This analysis is detailed in Chapter 7.

The assessment of the signal data needs to consider both their raw and processed forms as the processing techniques themselves are the subjects of the evaluation. The main steps of these techniques have already been developed before the Production Runs, as described in the previous chapter. These techniques were applied to the raw data collected and the resulting profiles were assessed for potential fault signature identification. These analyses are detailed in Chapter 8.

Chapter 7. Visual Inspection Analysis

7.1 Introduction and Overview

Visual inspection of the faults provides a fundamental reference point against which the sensor data will be compared.

The dominant focus is on the ceramic specimens, however, (for large faults at least), the contact with the steel counterface becomes also relevant; therefore the steel surfaces are also considered. The situation is, however, asymmetrical as far as the two specimens are concerned as explained below.

Because of the characteristics of the ceramic material, the three indentations can be reliably considered as independent of each other. However, each of these makes several contacts with the steel interface over each steel cycle and due to the difference in the hardness of the two materials, the ceramics makes a significantly greater impression on the steel than vice versa. Consequently, the steel surface features at the end of a run are the results of the combined effects of all three ceramic indents / faults.

In practice, significant wear in the steel was produced only by large faults which were also the dominant faults on the specimen – therefore the contributions of the remaining faults could be safely ignored.

This chapter is organised as follows:

Section 7.2 presents the post-run fault images and their basic classifications.

Section 7.3 presents the images of the pre-cracks and their classification in terms of their visual characteristics.

Section 7.4 introduces a basic model of fault propagation that could link the pre-cracks to the resulting faults.

Section 7.5 develops the dynamics of the CRC fault propagation.

Section 7.6 develops the dynamics of the Flaking fault propagation.

Section 7.7 presents the summary of the process and the findings.

7.2 Post-Run Fault Snapshots

The AIF images of the faults at the end of the 7 runs are shown below in Table 7.1. In addition Table 7.2 shows the AIF images taken at the point of interruption of the Run 2-1 due to a shaft breakage.

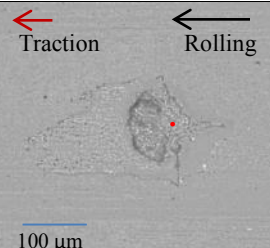
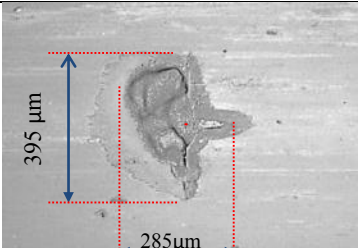
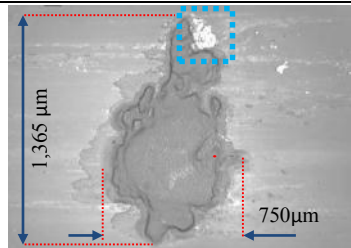
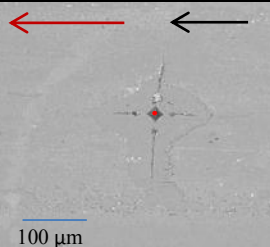
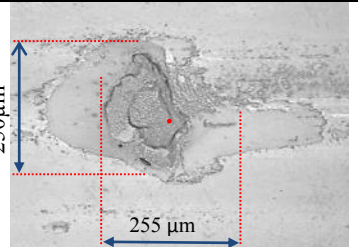
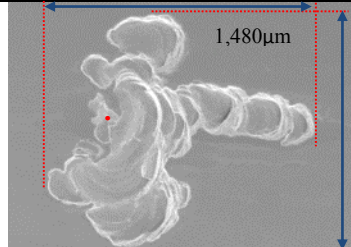
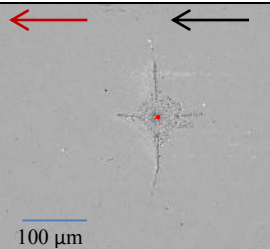
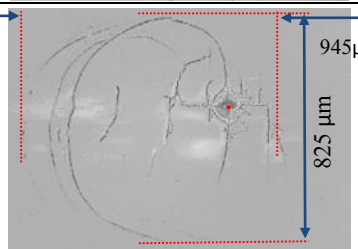
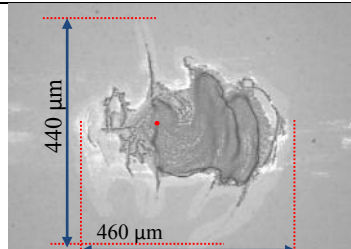
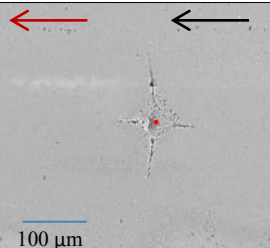
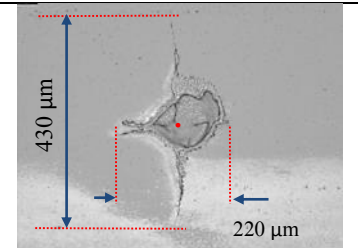
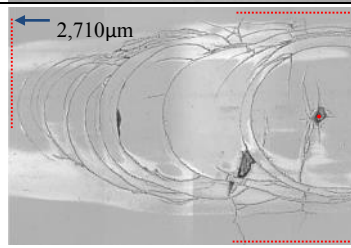
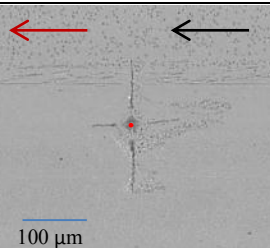
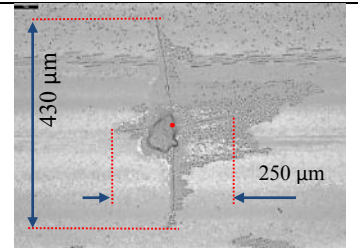
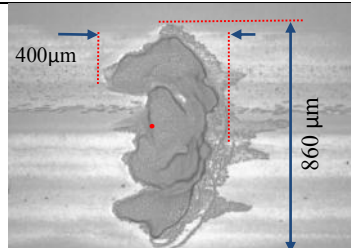
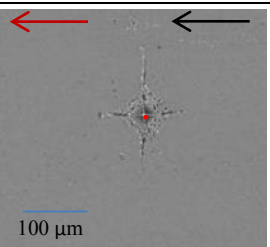
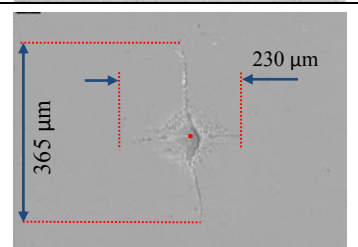
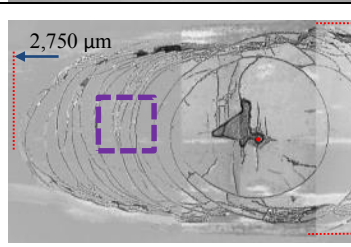
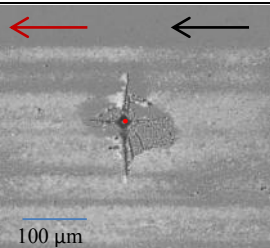
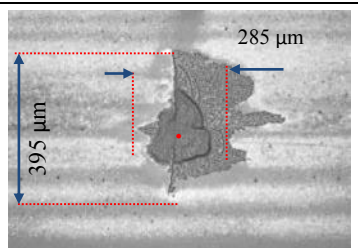
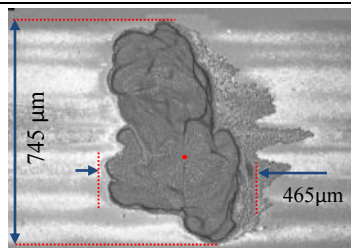
	Fault 1	Fault 2	Fault 3
1-1			
1-2			
2-1			
2-2			
2-3			
2-4			
2-5			

Table 7.1: Post-run fault snapshots (taken using the AIF). Fault 1 images taken at 20x magnification, remaining faults taken at various magnifications and fault dimensions are instead indicated. The rolling and traction directions are indicated for each run in column 2. The centres of the seeded indents are marked with a red dot. Coloured boxes are referred to in the sections following.

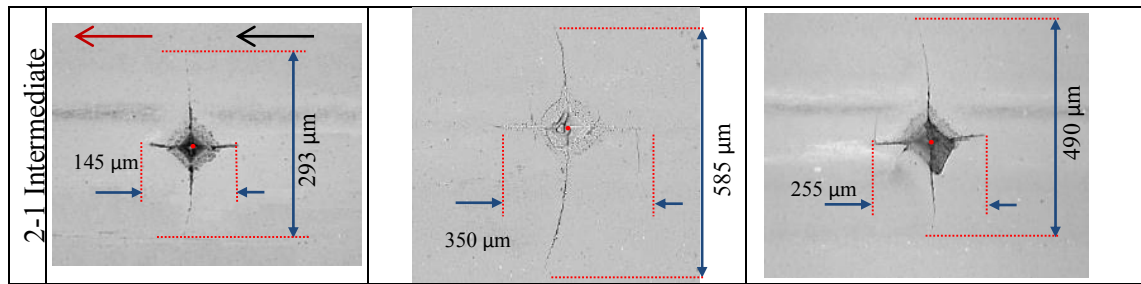


Table 7.2: Intermediate fault snapshots (taken using the AIF) at run interruption for Run 2-1. The rolling and traction directions are indicated in column 2.

These faults exhibit a wide variety of shapes and sizes, however, can be classified using the types reported in the literature (notably [51, 256]) as occurring in silicon nitride rolling tests. Based on their dominant features they are termed here as follows:

Flaking (spall): where the dominant visible feature is a pronounced area pattern with clearly identifiable boundaries. These demonstrate a substantial material loss resulting in relatively deep spalls with irregular contours forming a single complex, and exhibiting a wide range of sizes (from 40 to 750 μm in length with depths of between 15 to 135 μm). Based on the published results some of these are comparable to RCF faults occurring in steel. An example of this fault type is Fault 3 of Run 1-1 shown in the Table above.

C-/Ring crack complex- (hereafter ‘CRC’): where the dominant visible features are curved line patterns. In particular, the complex is a set of several co-located crack features in regular patterns such as part circles (c-cracks) or full circles (ring-cracks). The diameters of these are comparable to the wear track width (i.e. in the range of $>800 \mu\text{m}$). These appear similar to the faults observed by Kida et al [51] in ceramic-ceramic runs. An example of this fault type is Fault 3 of Run 2-4.

Delamination: where the dominant visible feature is a small and symmetrical area pattern without prominent boundaries of around 50 μm in length. These are identified with the faults observed by Hadfield et al [256], in which a removal of surface material bounded by the lateral crack occurred without substantial lateral progression³⁰. Therefore they were not seen as significant cause of failure. An example of this fault is Fault 1 of Run 1-2.

In addition, these faults in some circumstances triggered secondary wear mechanisms which involved wear away from the fault locations. Two different types were evident from the visual inspections: 3-body abrasion and adhesion. Their presence was investigated because they could both have substantial influence on the sensor signals, however, not all specimens showed a visual evidence of these mechanisms. They were identified using their following characteristics:

³⁰ Hadfield et al [256] referred to these faults as ‘*lateral crack spalls*’; to avoid confusion between flaking failures, they have been referred to in this thesis as *delaminations* - emphasising their characteristic material loss above cracks which lies almost parallel to the specimen surface.

3-body Abrasion is characterised by a presence of dots distributed around the sample wear track as a result of debris entrainment impinging on the surface. Having observed evidence of these, a confirmation was obtained using SEM with higher magnifications. An example of this effect is seen in all of the faults of Run 2-3.

Adhesion is characterised by a presence of shiny/reflective parts due to the adherence of steel from the component counterface as debris. Having seen evidence of these, a confirmation was obtained by performing chemical analysis using SEM-EDX. An example of this effect is seen in Fault 3 of Run 1-1.

Using these visual characteristics enabled a classification of the faults presented in the snapshots above as follows:

Run No.	Fault Type			Secondary Wear Mechanisms	
	Flaking	CRC	Delamination	3-body Abrasion	Adhesion
1-1	1, 2, 3	-			3
1-2	2, 3	-	1	1	
2-1 int.	3	-			
2-1	3	2			
2-2	2	3			
2-3	2,3	-	1	1, 2, 3	
2-4	-	3			3
2-5	2,3	-	1		

Table 7.3: Specimen faults and secondary wear mechanisms identified from post-run fault snapshots and the intermediate snapshots from Run 2-1. The numbers indicate fault numbers.

Having classified the faults at the end of the runs, the next issue refers to similar situations at the start of the runs and to the examinations of possible connections between them.

7.3 Fault and Indentation Correlations

The pre-cracks have also been examined using the AIF with a polarising filter and the corresponding snapshots are presented in Table 7.4 below. These show a greater uniformity in shapes and sizes than the final faults above, however, they cannot all be easily classified using the types identified there. Instead, their classification was subsequently based purely on their visual characteristics as revealed by the polarised pictures.

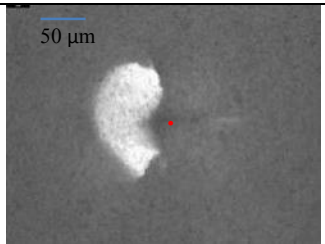
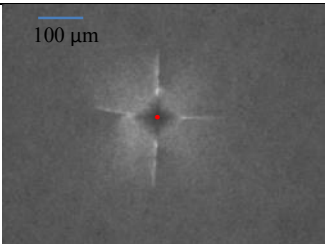
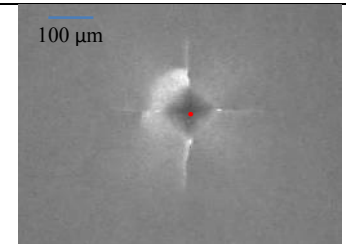
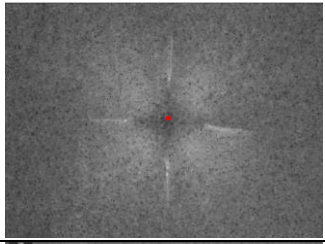

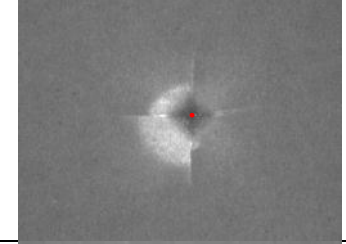
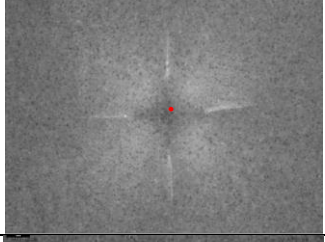
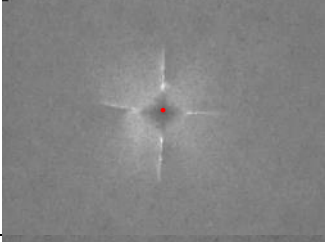
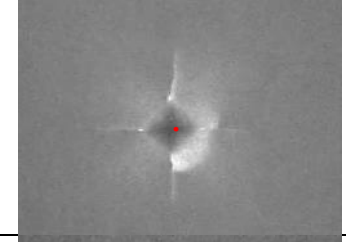
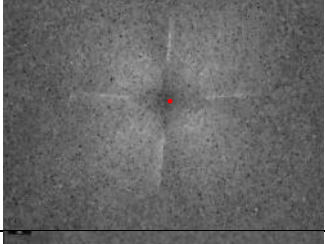
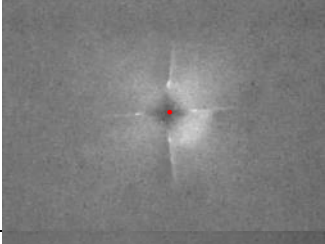
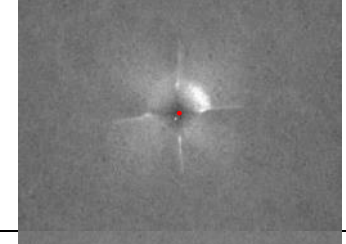


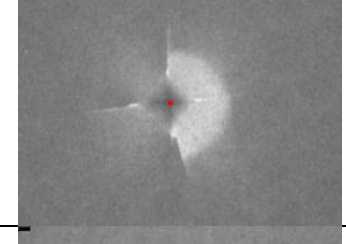
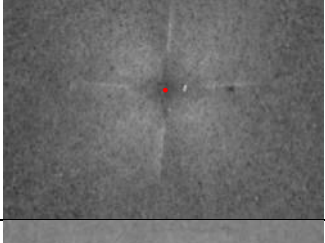
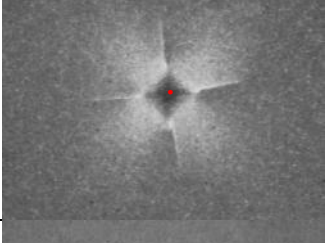
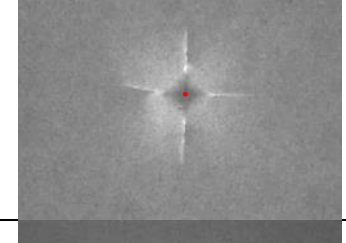
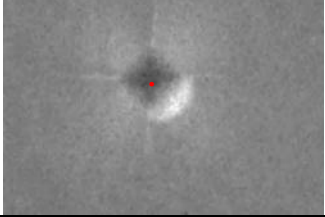
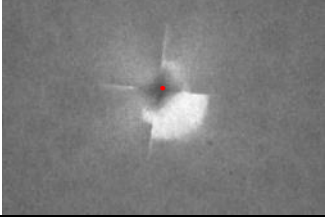
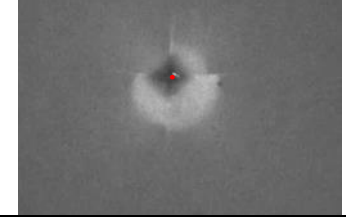
	Fault 1	Fault 2	Fault 3
1-1			
1-2			
2-1			
2-2			
2-3			
2-4			
2-5			

Table 7.4: Pre-run fault snapshots (taken using the AIF under polarised light filter). Fault 1 images taken at 20x magnification, remaining faults taken at 10 x magnifications. The rolling and traction directions are indicated for each run in column 2. The centres of the seeded indents are marked with a red dot

Comparing the pre-crack shapes and the failure pictures hinted that certain factors in the pre-cracks were influential in the subsequent fault developments. These were: the indentation load (as expected) and the symmetry or asymmetry of the subsurface lateral pattern, made visible in the AIF by the polarised light. The symmetry or the skewness of the lateral crack with respect to the rolling direction was classified as follows:

- 's' = symmetrical, exhibiting no prominent skewness (but still a presence of a lateral crack)
- 'l' = indicates a dominant crack in the leading direction
- 't' = dominant in the trailing direction.

The figure below illustrates an example of each of the three pre-crack types.

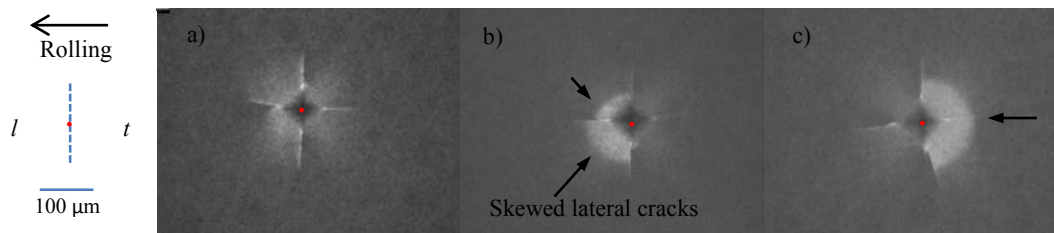


Figure 7.1: Example pre-crack snapshots, illustrating the symmetry (or asymmetry) characteristics a) symmetrical, b) leading and c) trailing edge skewness. The red dot indicates the centre of the indentation (i.e. the tip of the indenter). Snapshots taken from Runs: 1-1 (Fault 2), 1-2 (Fault 3), 2-3 (Fault 3) respectively.

The pre-crack fault categories for each of the runs are presented in Table 7.5 below. They have been matched with the corresponding end-run faults with an additional *broad* indication of their size (S, M or L)³¹:

Test No.	Fault #	Pre-crack Size	Pre-crack Type	Fault Category	Secondary Wear
1-1	1	S	<i>l</i>	F(S)	
	2	M	<i>s</i>	F(M)	
	3	L	<i>l</i>	F(L)	Adhesion
1-2	1	S	<i>s</i>	Delamination	
	2	M	<i>l</i>	F(M)	3B
	3	L	<i>l</i>	F(L)	
2-1	1	S	<i>s</i>	npf	
	2	M	<i>s</i>	CRC(M)	
	3	L	<i>t</i>	F(L)	
2-2	1	S	<i>s</i>	npf	
	2	L	<i>t</i>	F(M)	
	3	M	<i>t</i>	CRC(L)	
2-3	1	S	<i>s</i>	npf	3B
	2	M	<i>t</i>	F(M)	3B
	3	M	<i>t</i>	F(L)	3B
2-4	1	S	<i>s</i>	npf	
	2	M	<i>s</i>	npf	
	3	M	<i>s</i>	CRC(L)	Adhesion

³¹ Meaning: Small, Medium and Large. The classifications are based on the fault length. As discussed in the previous subsection, different scales apply to Flaking and CRC faults.

Test No.	Fault #	Pre-crack Size	Pre-crack Type	Fault Category	Secondary Wear
2-5	1	S	t	Delamination	
	2	M	t	F(M)	
	3	M	$t (+l)$	F(L)	

Table 7.5: Pre-cracks and corresponding fault categories (F= Flaking, CRC, npf=no prominent features) and wear categories (3B= 3body abrasion and A= adhesion).

The table indicates the following correlations between the pre-crack size and the resulting faults:

- Large magnitude indents lead to flaking faults
- Medium magnitudes produce either flaking or CRC faults
- Small magnitude indents either produce delamination or lead to no significant fault.

The correlations between the symmetrical or skewed nature of the pre-cracks and the resulting faults are not so clear-cut; the following tendencies appear:

- A skewed crack leads to Flaking type faults
- Trailing skewness produces significant (L) Flaking failures
- A symmetrical pre-crack develops into CRC or a fault with no prominent features (npf)

The first exception to this rule (skewed pre-cracks leading to CRC) in Run 2-2 is explained by the skewness being significantly less pronounced than in the other cases and is shown in the figure below:

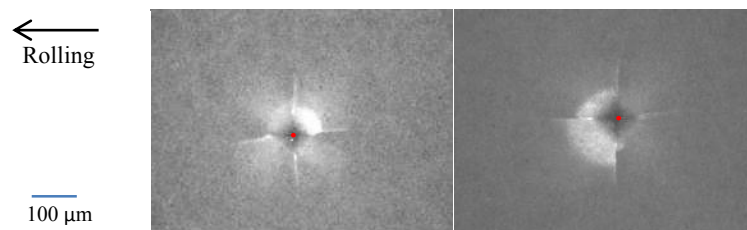


Figure 7.2: left snapshot: pre-crack Fault 3 snapshot before Run 2-2 illustrating the lateral crack skewness was relatively small and subsequently led to a CRC failure. The right snapshot shows a more typical skewed crack for comparison (Fault 3 of Run 1-2).

It therefore appears that size is the more prominent factor in determining the fault category than symmetry/skewness.

The second exception is the M-sized Flaking (rather than CRC or npf) produced from a symmetrical pre-crack during Run 1-1. This is explained by the exceptional severity of the loading applied in that particular run, indicating that fault development is not determined solely by the starting conditions but also by the detailed dynamics of the runs.

7.4 Basic Model of Fault Propagation

In order to interpret these observations in more detail a basic model was constructed based on the models and mechanisms reported in the literature.

Existing models describe situations that are similar to parts of the results presented above, notably:

Kida et al [51] describe semi-circular cracks generated around pre-cracks and eventually leading to substantial material loss. They describe the dynamics of this process in a concise model (see Section 2.2.3 for details). Because of the shape of the cracks, this process is identified here with the development of CRC faults.

Wang et al [256] and Zhao et al [257] describe flaking failures initiated from seeded c-cracks involving progressive material loss and provide models of the associated mechanisms (see Section 2.2.3 for details). These processes are identified here with the development of Flaking faults.

The model constructed to explain the empirical findings above was based on a fundamental idea that a fault develops in a systematic manner (in a way indicated by the researchers' models above) from a certain critical point on. The latter determines precisely which of the two fault modes (Flaking or CRC) are in play from there on. This leads therefore to a 2-phase model comprising:

Early development- the formation of a defect from the pristine material before the two mode mechanisms (Flaking/CRC) set in. For example the formation of a structure with non-prominent radial and lateral cracks (either of which can subsequently become dominant).

Subsequent development- one of the two mode mechanisms becomes dominant and determines entirely the subsequent progression of the fault (along the lines indicated by the researchers above).

The model is presented below with some indications of dominant mechanisms associated with each stage.

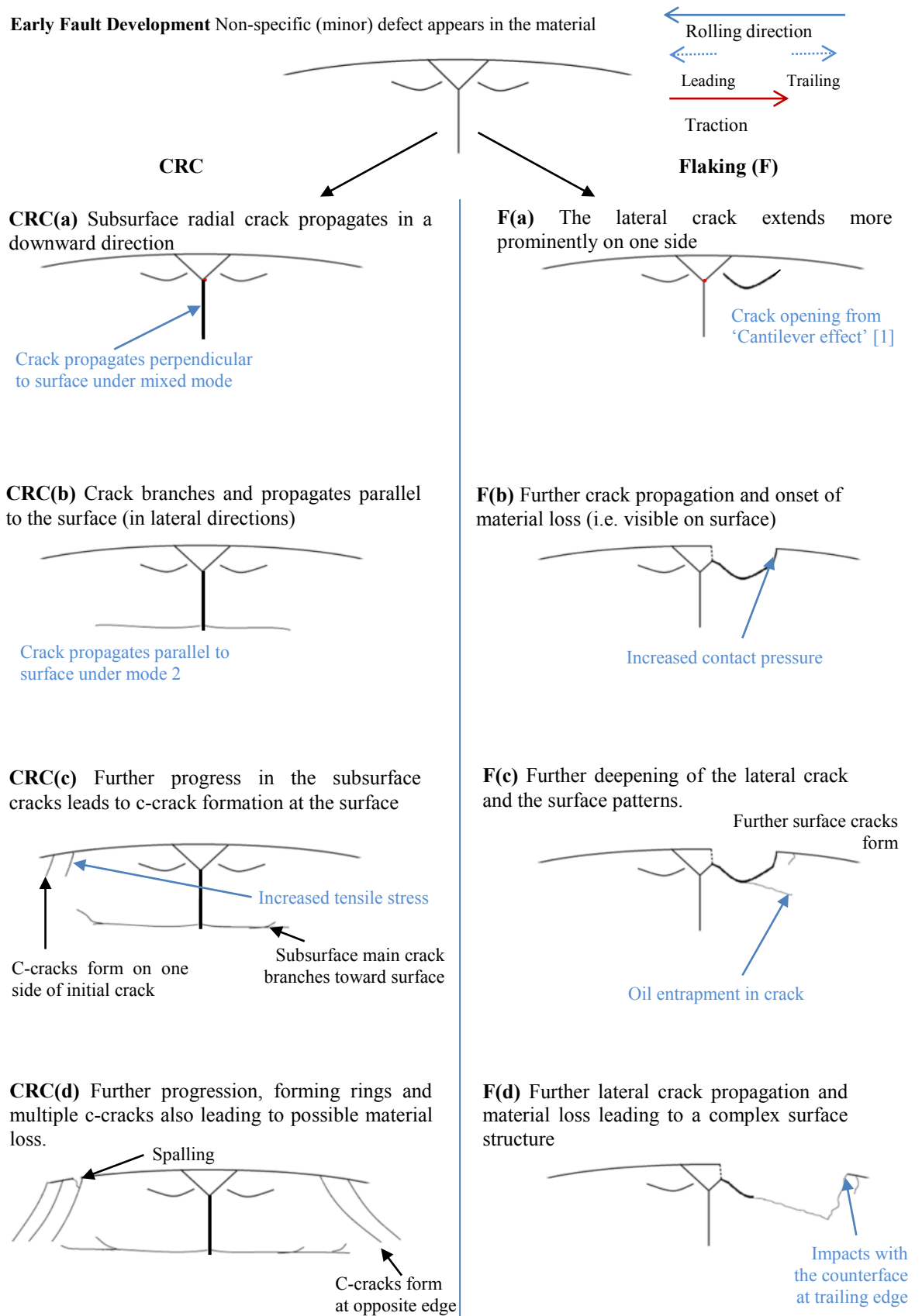


Figure 7.3: Stages of fault development

Stages (a)-(c) are those described in the literature mentioned above ([51] for CRC); corresponding stages for Flaking were derived from [256] & [257]. Additional stage (d) was added for both CRC and Flaking based on the evidence from the runs which is detailed later below.

Delaminations fit into the model as follows: by their definition they are closely associated with Flaking faults as they involve material removal initiated by the lateral crack. They differ from the Flaking faults because of their small size and the shallowness of the lateral crack. Therefore, they can either be produced from EFD or F(a).

Based on the descriptions of the fault progressions it is now possible to reinterpret the results of the runs (using their visual characteristic i.e. Table 7.1 & Table 7.2) in terms of the various stages of the model. An important assumption is being made therefore, that each of the faults across the runs represents a stage of progression within this single general model.

In the attempt to similarly assign the pre-cracks to the model fault stages (using their visual characteristics i.e. Table 7.3), only a partial match is achievable. This is essentially because the visual snapshots represent a plan view of the faults i.e. it is not possible to discern the relative weights of the subsurface lateral and radial cracks. Because of this it is not possible to directly distinguish between the stages EFD (i.e. undetermined), CRC(a) and F(a) (i.e. not yet prominent). However, using additional information some simple rules can be identified:

A significantly skewed pre-crack = F(a): because the skewness of the lateral crack suggests that it is more pronounced than the radial crack (in addition, all the corresponding runs without exception resulted in a Flaking fault).

Small pre-crack = EFD: because generally it did not develop into a recognised primary type.

The reinterpreted results are presented in the Table below:

Test No.	Fault #	Pre-crack Stage	Pre-crack Type	Pre-crack Category	Fault Category
1-1	1	S	<i>l</i>	F(a)	F(b)
	2	M	<i>s</i>	?	F(c)
	3	L	<i>l</i>	F(a)	F(d)
1-2	1	S	<i>s</i>	EFD	D
	2	M	<i>l</i>	F(a)	F(c)
	3	L	<i>l</i>	F(a)	F(d)
2-1	1	S	<i>s</i>	EFD	npf
	2	M	<i>s</i>	?	CRC(c)
	3	L	<i>t</i>	F(a)	F(d)
2-2	1	S	<i>s</i>	EFD	npf
	2	L	<i>t</i>	?	F(c)
	3	M	<i>t</i>	?	CRC(d)

Test No.	Fault #	Pre-crack Stage	Pre-crack Type	Pre-crack Category	Fault Category
2-3	1	S	s	EFD	npf
	2	M	t	F(a)	F(b)
	3	M	t	F(a)	F(d)
2-4	1	S	s	EFD	npf
	2	M	s	?	npf
	3	M	s	?	CRC(d)
2-5	1	S	t	EFD	D
	2	M	t	F(a)	F(c)
	3	M	$t (+l)$	F(a)	F(d)

Table 7.6: Identification of pre-cracks and faults with the model stages.

Further information can be obtained by examining the mechanisms responsible for the generation of the pre-cracks and in operation during the runs that develop the pre-crack stage to the end-fault.

The skewness of the initial pre-crack results from the positioning of the specimen's centre being off the centre-axis of the indenter. The latter therefore meets a sloping rather than a perpendicular surface as illustrated in Figure 7.4a below.

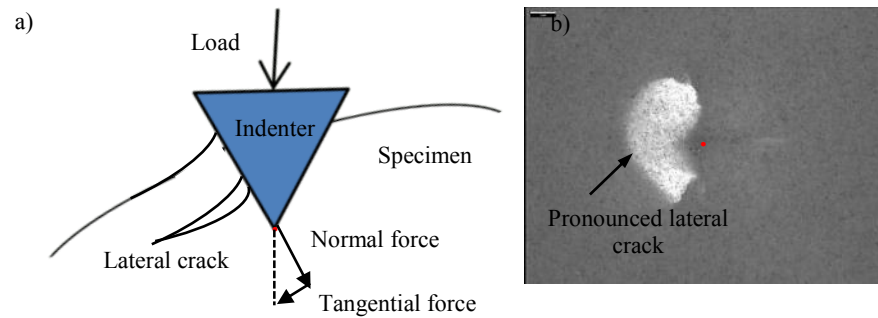


Figure 7.4: a) indentation process made at an oblique angle b) AIF snapshot from Run 1-1 (Fault 1) showing a prominent skewness in the pre-crack.

This generates an additional tangential force parallel to the surface. Subsequently, the lateral crack as shown above is more pronounced and closer to the surface than in a symmetrical indent making it more prominent under a polarised light source as shown in Figure 7.4b.

The indentation machine offers a high degree of precision in the alignment of the specimen with the indenter tip - hence many of the generated pre-cracks appear symmetrical. However, lateral cracks generated in highly brittle materials are generally larger than in materials with lower brittleness (as reported in reference [48]). Consequently, a highly brittle material such as silicon nitride is more sensitive to the precise alignment i.e. the load angle. Because of this, indentation-generated skewed cracks tend to be more numerous than symmetrical cracks. Based on the above rules therefore, they are more likely to be in the stage F(a) rather than CRC(a) or EFD.

The following sections describe in detail the mechanisms in play during the runs as evidenced by the various snapshots; they are separated into the two main branches of the model, i.e. CRC and Flaking.

7.5 CRC Propagation Dynamics

As explained earlier, the snapshots do not provide enough information to distinguish between the stages EFD and CRC(a). Because of this, the development of the fault from EFD is presented in one step directly to CRC(b).

7.5.1 EFD → CRC(a) → CRC(b)

The mechanism that takes a non-specific weakness of EFD to the beginning of the CRC stage has been described by Kida [51] as a deepening of a surface crack which lies perpendicular to the rolling direction (this crack must therefore be present at the end of the EFD stage). This propagates almost vertically into the material driven by mixed stress modes (notably mode 2).

At a critical depth (near the depth corresponding to maximum orthogonal shear stress) the crack branches, propagating in directions parallel to the surface of the specimen, also described by Kida [51]. Evidence of this process is provided by the intermediate Fault 2 of Run 2-1 (reproduced in Figure 7.5a below) where the presence of a new surface crack is explained according to Kida [50] as a result of subsurface crack propagation from the initial indentation crack. In this snapshot one can also see a lengthening and curving of the original radial crack transverse to the rolling direction as a precursor to the formation of a c-crack (observed in the end snapshot and described also by Kida [46]).

7.5.2 CRC(b) → CRC(c)

This stage is characterised by the formation of semi-circular cracks in front of the indent (i.e. leading direction) as a result of a relaxation of the contact, combined with the traction from the induced sliding, leading to an increase in the tensile stress at the leading edge of the contact at the leading edge of the fault as found by Wang and Hadfield in [256].

An example of this is provided by the end-run snapshot of Fault 2 from Run 2-1, reproduced in Figure 7.5b below:

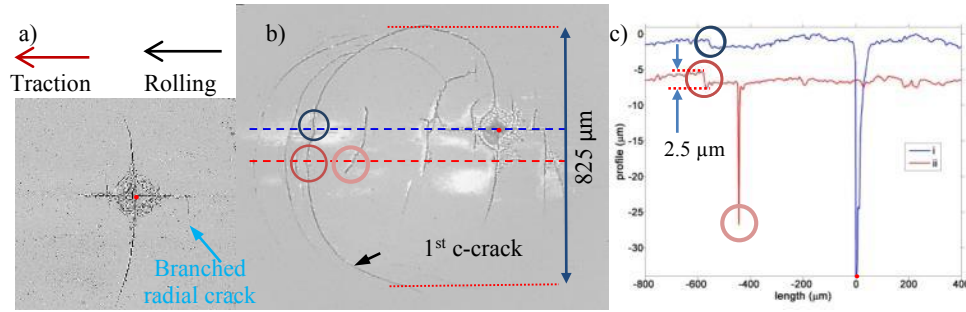


Figure 7.5: AIF Snapshots of Fault 2 from Run 2-1: a) intermediate b) end-run fault surface profile measurements c) corresponding depth profile measurements.

Here, a curving of the upper radial crack forms an arc and the first c-crack. Continued over-rolling generates further c-cracks in front of the initial, corresponding to the maximum tensile region of the contact. This was further evidenced by the c-crack diameters (approximately 825 μm) being comparable to the calculated contact width of the run (800 μm) using equation 2.8.

A profile measurement obtained using the AIF 3D analysis tool is shown in Figure 7.5c above. This indicates an irregular surface profile with steep jumps at the locations of the cracks, the most substantial being at the 1st c-crack³² with a magnitude of approximately 2 μm protrusion. The film thickness was calculated to be 0.48 μm and subsequent abrasion with the softer steel counterface was expected³³.

7.5.3 CRC(c) \rightarrow CRC(d)

CRC(d) stage is characterised by the formation of c-cracks at the *trailing* edge of the fault accompanied by a substantial material loss.

The mechanism responsible is based on a substantial relaxation in the contact at the fault location; material is removed around the fault where cracks coalesce. Fault 3 from Run 2-2 demonstrates these features and is shown in the figure below.

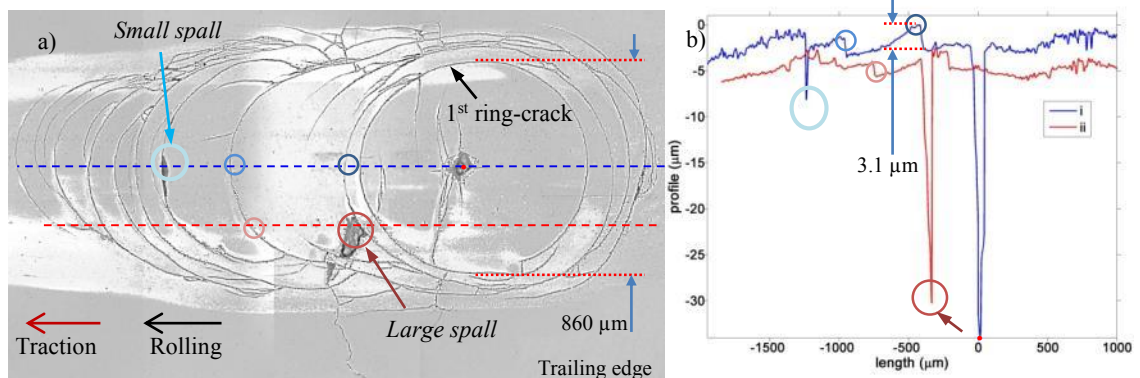


Figure 7.6: a) Post-run snapshot of Fault 3 from Run 2-2 b) corresponding surface profile measurements.

³² Wang and Hadfield [82] observed that the c-cracks forms initially closest to the pre-crack whilst additional c-cracks form progressively in a direction away from the pre-crack.

³³ A wear track profile deeper than that of the Healthy Run was observed for Run 2-1, however a substantial Flaking fault was observed in the run, and decoupling the steel wear between the two faults was not possible.

The figure illustrates also other features:

The 1st crack surrounding the pre-crack centre (generated at CRC(c)) now becomes a ring crack, and additional ring cracks develop around it (with larger diameters), illustrating the effect of the contact relaxation and subsequent increases in the contact radius.

Further, additional leading c-cracks are generated joining those already present at the previous stage CRC(c).

These features are also present in the other example of CRC(d) fault, Fault 3 of Run 2-4, thus indicating the possibility that they are an intrinsic characteristic of this stage.

In order to describe the above fault in more detail, further measurements were performed and are described in the following.

The fault region was scanned using x-ray CT and analysed to identify subsurface cracks around the fault which were subsequently sub-selected from the rest of the scan volume before image rendering³⁴. The resulting analysis was superimposed with the surface cracks from the image in Figure 7.6a above and is shown below:

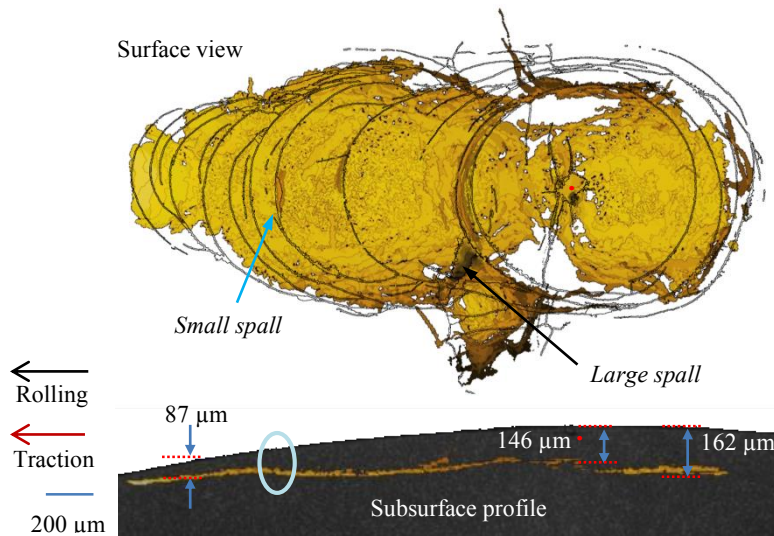


Figure 7.7: Rendered CT images of Fault 3 Run 2-2 combined with the surface AIF image in the surface view (the subsurface cracks shown in variants of yellow colouring with darker colours indicating shallower crack regions, surface cracks are shown in black).

The image shows the extent of the subsurface parallel cracks under which the profile and depth of the cracks are shown. The branching of the main crack at the location of the indent occurred at 146 μm deepening to 162 μm, comparable to the calculated depth of maximum principal shear stress for the run at 173 μm.

³⁴ Unfortunately the resolution was not high enough to expose the vertical cracks from the ring/c-cracks.

The subsurface parallel crack becomes significantly shallower at the region in front of an area marked as '*small spall*'. Consequently, this indicated that oil was entrapped in the Small Spall lubricating the opposing crack faces of the c-crack located here. Also, the Small Spall represents a widening of the crack gap size at its location³⁵. Subsequently, spikes in the tensile stresses at the leading edge occur as the contact centre passes over the corresponding c-crack, generating additional c-cracks with higher density than in the earlier stage.

7.5.4 The Contribution of the Steel Counterface

The steel counterface begins to play a new prominent role in the later stages of the fault development when direct contact between it and the ceramic specimen occurs.

This is seen in Figure 7.6b as a protrusion of the c-cracks above the surface level by a maximum distance of more than 3 μm , this being also above the minimum oil thickness. The resulting direct contact produces an increase in the traction forces at these locations. This reduces the depth of the maximum shear stress location and increases the tensile stress at the leading edge of the contact. A positive feedback mechanism results as further protrusions cause increasing wear in the steel face and subsequent increasing traction coefficients.

The resulting effect on the steel surface is evident in the AIF snapshot of the steel wear track at the end of the corresponding run, shown below.

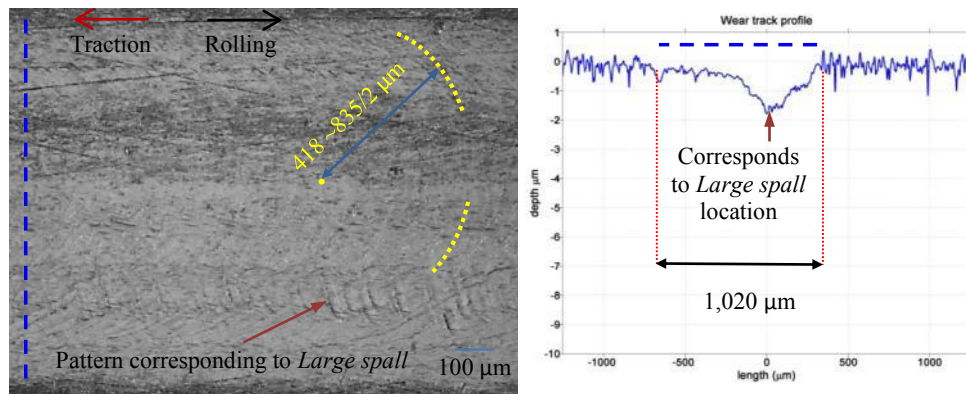


Figure 7.8: Analyses of the steel wear track from Run 2-2 a) AIF image b) profile measurement.

A mild abrasive wear is evident in the softer steel counterface by the presence of repeating semi-circular patterns partially visible matching the c-crack contours (with a diameter comparable to the diameter of the 1st crack location shown in Figure 7.6a). In addition, an imprinted repeating pattern covering the entire circumference is seen in the steel at the transverse location corresponding to the Large Spall (refer to Figure 7.6a).

³⁵ Wang and Hadfield [82] describe the influence of an increased crack gap size to promoting c-crack formation at the leading edge of the fault.

The extent of the increase in the tensile stress at the leading edge through dry sliding between the protruding crack regions and the steel face could be estimated from the difference in the CoF under dry and lubricated conditions. The CoF of silicon nitride and steel has been reported by Wang et al [25] to be 0.15 compared to the baseline 0.07 observed in the Healthy Run. The tensile stresses are calculated using equation 2.8 to be 783 MPa for a dry sliding contact compared to 521 MPa observed for the Healthy Run i.e. a 50 % increase (the tensile strength of the silicon nitride specimen is 900 MPa); i.e. a substantial factor in promoting further c-cracks.

Further evidence of direct steel interaction is provided by the CRC fault in Run 2-4 (Fault 3) which exhibits even more pronounced spalls and more numerous, denser c-cracks. SEM images of parts of the fault are shown in the figure below.

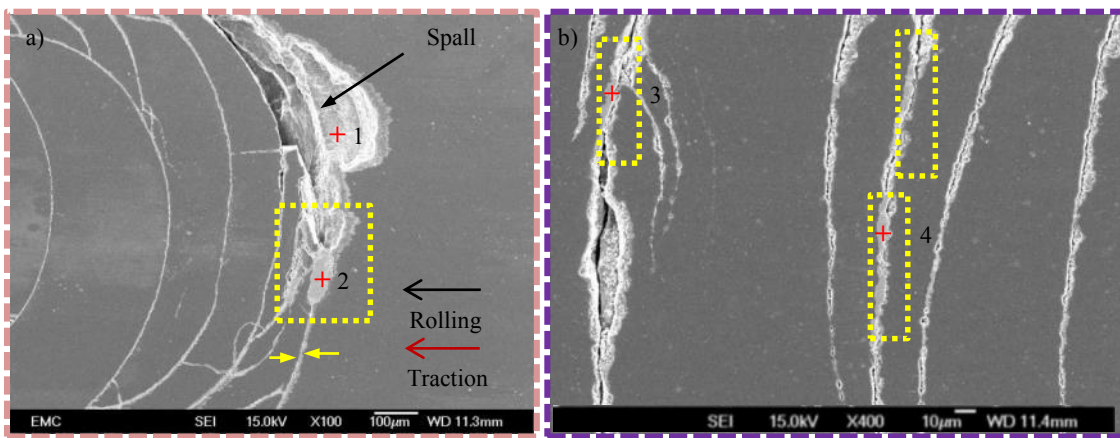


Figure 7.9: Post-run SEM analyses of Fault 3 from Run 2-4 indicating (with yellow markers) the locations where steel has adhered (confirmed by EDX analysis). Image locations are indicated in the relevant image of Table 7.1 with the respective box colours. a) Cracks and material loss at the trailing edge of the fault, b) Cracks located at the leading side of the fault.

These figures show the presence of steel debris adhered to the spalled area (Figure 7.9a) and to the cracks located at the leading side of the pre-crack (Figure 7.9b). The presence of steel was confirmed by EDX analysis, the results of which are shown in Table 7.7 below.

1		2		3		4	
Element	Weight%	Element	Element	Weight%	Weight%	Element	Weight%
N	19.36	C	5.39	C	7.40	C	8.59
Al	1.01	O	2.01	O	-	O	-
Si	70.11	Si	2.92	Si	2.92	Si	2.34
		Cr	2.96	Cr	2.96	Cr	0.1
N	19.36	Fe	78.32	Fe	78.32	Fe	80.26
Au	9.52	Au	8.40	Au	8.40	Au	8.70
		Totals	Totals	100.00	100.00	Totals	100.00

Table 7.7: EDX measurements of the corresponding locations from Fault 3 in Run 2-4 shown in Figure 7.9.

Locations 2 to 4 above reveal the presence of iron, carbon and chromium (the main constituent elements in the AISI 52100 steel specimen).

The pronounced imprints on the steel wear track resulting from the protrusions at the surface cracks and at the spalled region, and illustrating significant abrasive wear, are shown below.

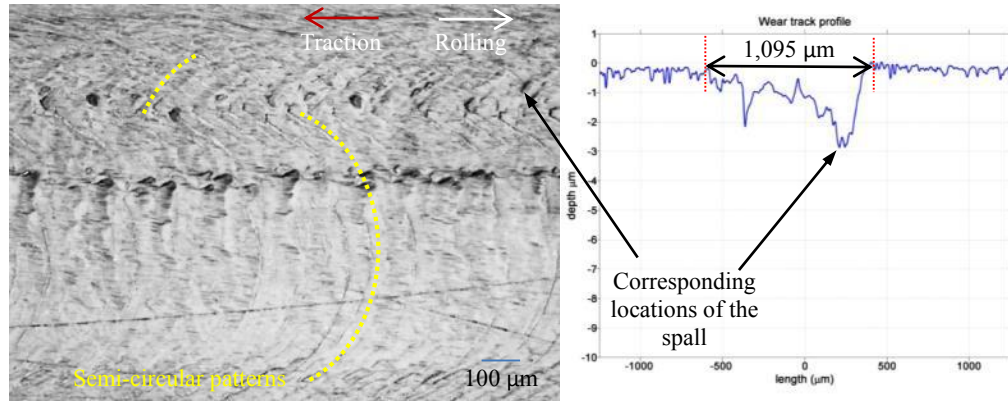


Figure 7.10: a) AIF image of the steel wear track from Run 2-4 b) corresponding profile measurement

In particular, the semi-circular patterns, partially visible in the steel wear track in Run 2-2, are significantly more pronounced in this run with arcs visible in both the leading and trailing directions. Significant wear is found at the location of the spall at the upper trailing edge (indicated in Figure 7.9a) and is shown in Figure 7.10b.

7.6 Flaking Propagation Dynamics

As explained earlier, the snapshots do not provide enough information to distinguish between the stages EFD and F(a). Because of this, the development of the fault from EFD is presented in one step directly to F(b).

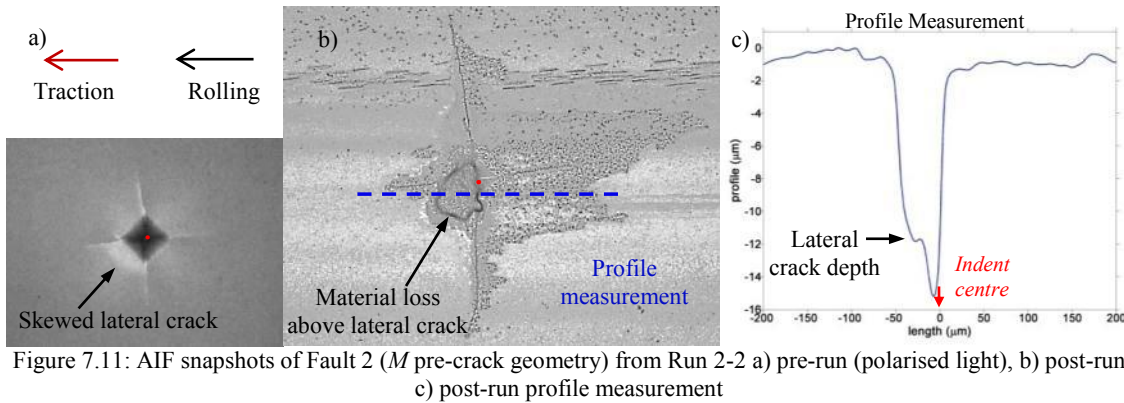
7.6.1 EFD → F(a) → F(b)

Subsection 7.4 explained how stage F(a) is generated in the indentation process as a result of a tangential force producing a pronounced lateral crack. The lateral crack exhibits a shallow incline from the material surface; Kadin et al [1], demonstrated through modelling that a crack with a shallow incline propagates differently to a near vertical crack through the cantilever effect described in section 2.2.3 and the effects of lubricant entrapment.

It is proposed that a similar process occurs during the course of the over-rolling cycles: a crack with a shallow incline (similar to the one sketched in Figure 7.4a for indentation) develops in the EFD stage. This then provides a seed that leads to a pronounced lateral crack thus producing stage F(a). For example, the presence of hard, round debris in the contact can form c-cracks with sloping subsurface features in the ceramic.

Stage F(b), distinguished by a detachment of the surface material, forms an area bounded by a prominent lateral crack. This is a straightforward consequence of the continuous over-rolling process over the location weakened in stage F(a).

The figure below shows an example of a fault exhibiting this Flaking stage and its corresponding pre-crack (being at F(a)). This fault is Fault 2 of Run 2-3.



The flaking wear was initiated by a *leading* edge skewness of the pre-crack (Figure 7.11a) whose location corresponds closely to the observed flaking shown in Figure 7.11b. A profile measurement shown in Figure 7.11c indicates the depth of the lateral crack at 12 μm. Furthermore, it shows a deepening of the fault as the crack propagated into the material to 14.6 μm before flaking off.

This example also illustrates the effect of the direction of the skewness. As identified in subsection 7.3 trailing skewness produces significant flaking faults; in contrast, leading skewness is more benign: over the 100+ hrs duration of this run, the fault developed only to the next stage, F(a) → F(b). The effect of the location of the spall in respect to the pre-crack centre is discussed in the following stage.

7.6.2 F(b) → F(c)

F(c) is characterised by the formation of secondary surface cracks and a deepening of the subsurface lateral crack arising from the basin of the flaking area.

Following F(b), the removal of material reduces the contact area to the edges surrounding the flaking basin leading to increasing contact pressures. This is more prominent at the trailing edge (contact is leaving the flaking area) than on the leading edge. Increasing tensile stresses lead to the formation of secondary cracks at the surface close to the flake as observed in Fault 2 of Run 2-5 and shown in the figure below.

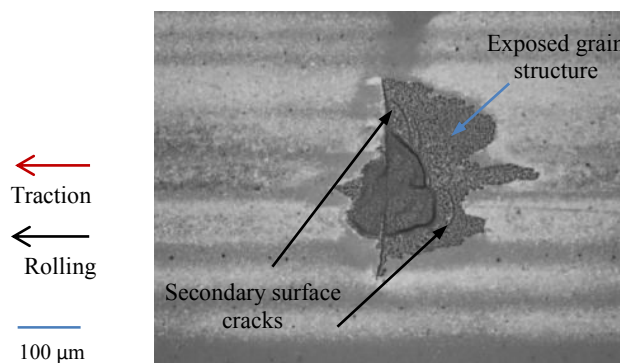


Figure 7.12: Post-run AIF snapshot of Fault 2 (M pre-crack geometry) from Run 2-5.

These appear similar to the failures observed by Zhao, Hadfield et al [53] initiated from c-cracks. Here a progressive material loss was observed as the initial pre-crack propagated subsurface at a shallow angle away from the surface driven by increasing subsurface shear stresses. In addition, secondary cracks formed at the surface above the subsurface cracks. Eventually the two sets of cracks met producing material loss and a progressive flaking wear (thus producing F(d) of the model).

Figure 7.12 also illustrates the prominence of the trailing edge over leading as the increased tensile stresses there led to the removal of the surface oxide layer, exposing the grain structure below. This was also found in similar tests with steel on steel rolling-sliding contacts by Ueda et al. [258].

The preference of the trailing direction over leading indicates the substantial role the oil played in propagating these faults also. Oil is pushed into the fault as the contact approaches the fault. As the contact centre passes directly over the fault, oil becomes trapped in the spall generating high hydrostatic pressures at the crack tips, promoting subsurface crack propagation. Furthermore, oil pushed into the cracks lubricates the opposing crack faces reducing the CoF between them³⁶. This, and the incorporation of traction from the induced sliding³⁷, further promotes crack growth through mode 2 (shear) stresses and assists in opening surface cracks as the contact approaches the fault. For cracks orientated in the leading edge direction, the rolling direction, in contrast to the above, squeezes oil out of the crack during the approach of the contact and subsequently the effects of oil entrapment are not evident.

An apparent exception to the above rule appeared in Run 1-1, where a *substantial* flaking failure occurred in Fault 3 at the *leading* edge. This is explained simply by the combination of the following factors:

- The failure developed from a pre-crack (F(a)) with a skewness in the leading edge
- The severity of the applied contact pressures at 5 GPa.

7.6.3 F(c) → F(d)

F(d) is produced when the flaking off of material produces a fault area comparable (in width and length) to the curvature of the counterface, thus allowing the specimen to come closer to the centre of the counter steel disc at the contact point (effectively the steel disc ‘drops’ into the fault). This has an effect of significantly increasing the rate of material removal as a result of impacts on the trailing edge from that point on. This mechanism has been observed in steel by Arakere et al [71].

³⁶ The influence of the reduction of CoF was modelled in [81]

³⁷ The influence of traction was modelled in [54] and the effect of direction is discussed in [258]

This process is illustrated in Fault 3 from Run 2-1 shown in the figure below together with a detailed topography of the F(d) fault, as well as the intermediate fault F(b) leading to it.

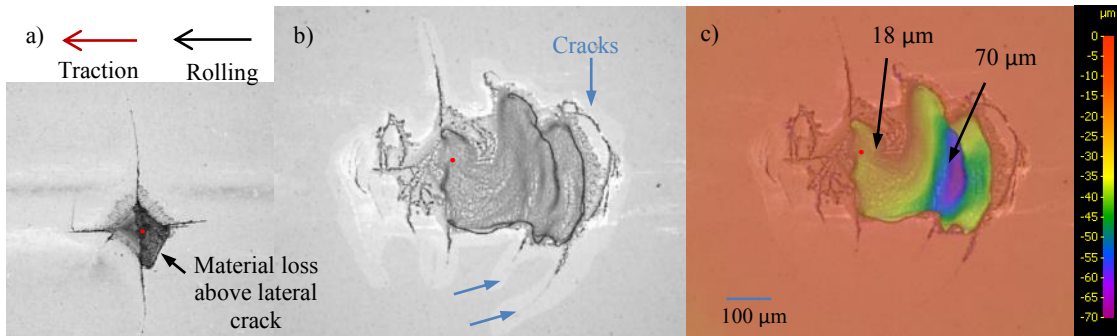


Figure 7.13: AIF snapshots of Fault 3 from Run 2-1 a) intermediate stage, b) post-run c) post-run colour map showing the surface topography obtained by the 3D analysis tool of the AIF.

The figure shows that the fault area consists of multiple flaking zones, much more pronounced than in F(c), widening and deepening in the trailing direction- thus confirming that the fault has reached the F(d) stage. The widening and deepening is the result of the increased subsurface shear stresses at the trailing and side edges, which increase with fault size (increasing the magnitude of the load imparted during impacts with the counterface), thus producing a positive feedback mechanism also reported in [71]. This effect was further aided by sliding, set at 2% SRR for this run.

The effect of high level sliding (4%) in the production of a F(d) fault is illustrated in the following figure, showing Fault 3 from Run 1-2, its detailed topography as well as the initiating pre-crack.

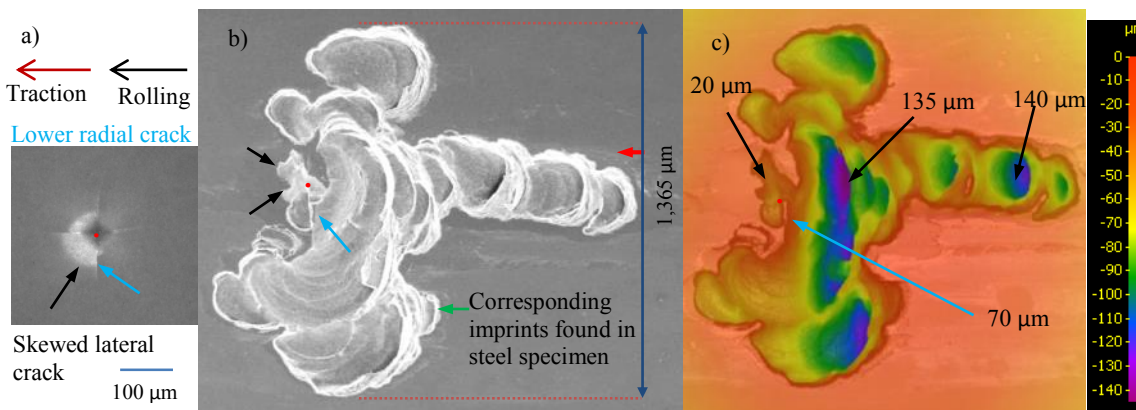


Figure 7.14: AIF snapshots of Fault 3 from Run 1-2 a) pre-run indentation, b) post-run c) colour map showing the surface topography. The markers are positioned for reference of the corresponding steel wear track shown in Figure 7.17 and discussed in Subsection 7.6.5.

In this run, the failure was initiated by a pre-crack (F(a)) with a *leading* edge skewness yet it produced a significant *trailing* edge F(d). This is explained by the following mechanisms:

The pre-crack F(a) progressed in the normal manner to produce a leading edge F(b). The evidence of this is seen in Figure 7.14b where the F(b) area is marked by black arrows. As a

result, the associated increase in contact pressure combined with the high traction promoted lubricant entrapment in the lower radial crack (running through the indent centre and already present in the F(a) stage), therefore reducing the CoF within it. The associated weakening and the high traction applied during the run caused subsequent cracks propagating in the trailing direction (evidenced by the prominent vertical edge at the lower radial crack location with a drop in the trailing direction as shown in Figure 7.14c), until the stage F(d) was reached.

The run also highlights the effect of the impact energy induced from the high traction at the trailing edge of the fault. This is seen by the substantial area of the fault, measuring 1,480 and 1,365 μm in length and width respectively and with prominent material loss in the trailing edge of the fault producing deep contours with maximum depths of 140 μm .

The effect of the reduced traction magnitude is shown in Run 1-1, in which all flaking faults exhibit material loss on the leading edge. This is most pronounced at the stage F(d) (Fault 3), as would be expected.

7.6.4 3-body Abrasion

Evidence of 3-body abrasion on the silicon nitride specimen was detected solely in Run 2-3. Figure 7.15 below shows the post-run snapshot around the prominent F(d) fault (Fault 3) and in addition two bands of surface patterns which have been shown to cover the entire circumference of the silicon nitride specimen (for example, they are also seen in the snapshot of the other faults). For comparison, no such *obvious* bands were found on the steel counterface (Figure 7.15b&c).

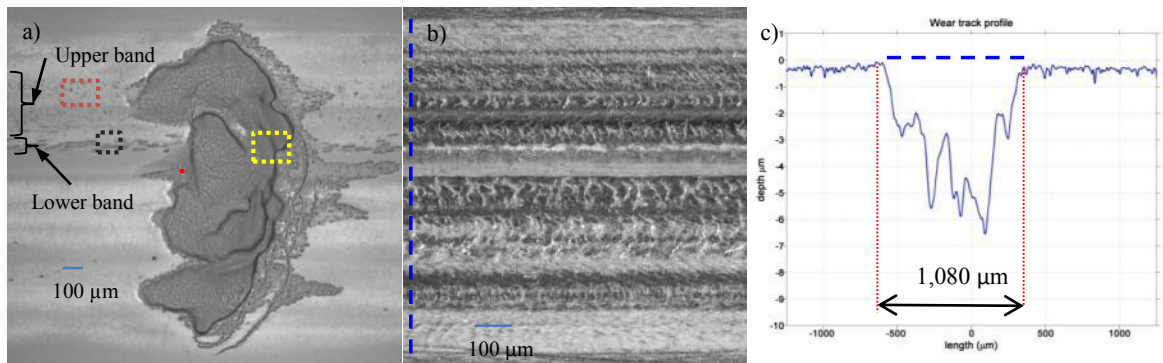


Figure 7.15: a) AIF image of Fault 3 and surrounding areas from Run 2-4 b) AIF image of the steel wear track and c) its profile measurement.

The details of sections of the bands are shown in greater magnification below. They show that the upper band consists of a series of pits and the lower band consists of multiple scratch marks.

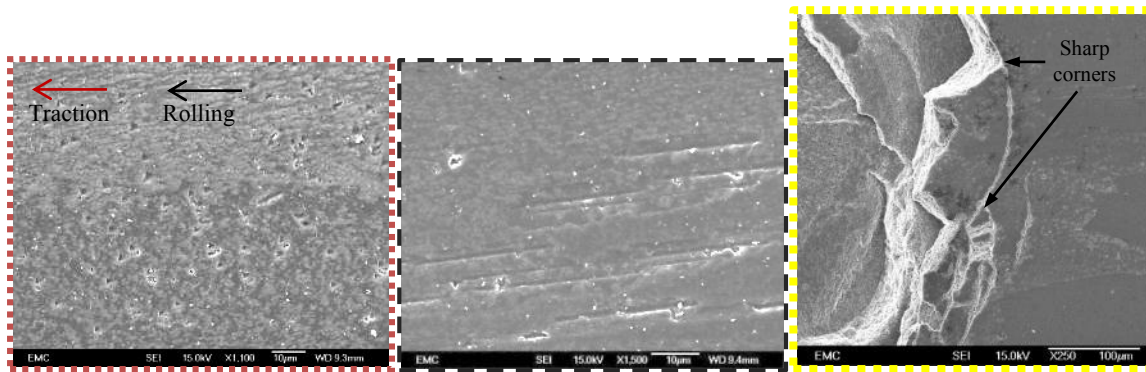


Figure 7.16: SEM images of the silicon nitride sample detailing the areas within the boxes shown in Figure 7.15 a) portion of the upper band b) lower band c) trailing edge of the fault area.

Similar scratches and pits were observed by Mitchell and Mercholsky et al [68] in hybrid rolling as a result of oil contaminated with Alumina debris. As a continuous band of markings in a hard material such as silicon nitride cannot be explained by any of the fault mechanisms above, it is taken as evidence of 3-body abrasion and the following establishes further supporting arguments for this. For debris to produce such markings it would have to be from a sharp and hard material. As the steel is unlikely to produce such material³⁸, the obvious source remains the silicon nitride specimen itself and the fault F(d) in particular. The detailed area around it, presented in greater magnification in Figure 7.16c above, shows the presence of cracks that could further eventually produce further flaking with very sharp boundaries/edges. Once dislodged, the presence of soft steel, possibly aided by the induced sliding, would have helped to keep the debris within the contact area over at least one rotation (for example by partially impinging it), sufficient to cause cutting and ploughing wear [68] over the entire silicon nitride circumference, before being further dislodged and removed by oil or centrifugal forces.

Rather than two separate pieces of debris, both originating in the flaking area being responsible for the two bands, it is likely that the upper band was produced by multiple small debris generated in the above process that was responsible for the lower band. This process, referred to as comminution, has also been described in [68]. Further evidence for this is that the resulting upper band consists of pits (from smaller debris) rather than scratch marks and that it lies in a peripheral direction away from the bottom band as debris was pushed from the wear track centre before being removed from the specimen.

The fact that no obvious bands were found on the steel counterface is explained by the smoothing action of the large flaking fault (F(d)) over subsequent cycles (this mechanism is detailed below in section 7.6.5).

³⁸ Work hardening of the steel is possible, however Mitchell and Mercholsky et al [68] found that only the hard alumina ceramic was able to generate significant abrasive wear on the silicon nitride specimen.

It is not entirely clear why similar 3-body wear effects were not detected in the other F(d) faults. It appears that the sharpness of the debris plays a key role and that the circumstances of this run helped to generate such debris. Comparison of the F(d) faults (see Table 7.1) shows that this flaking fault consisted of multiple cracks, more extensive than the other F(d) faults, with sharp corners and spanning the entire trailing edge of the fault.

7.6.5 The Contribution of the Steel Counterface

With a flaking fault of large magnitude (F(d)), substantial interaction with the steel counterface occurs as a result of direct (oil free) contact between it and the ceramic specimen. This happens due to the large depression of the ceramic fault and subsequent protrusions of the surrounding contours above the oil film. The consequences of the direct contact on the ceramic were mentioned above (as increased rate of material removal); the consequences on the (softer) steel surface are various impression and abrasion marks produced by the surface features of the ceramic specimen. The following analysis attempts to identify the main mechanisms that produce these features.

These features were observed most prominently in the Severe Runs (i.e. Runs 1-1 & 1-2) and were reflected by significant wear in the steel specimens.

An image of part of the steel wear track from Run 1-2 is shown in the Figure 7.17 below.

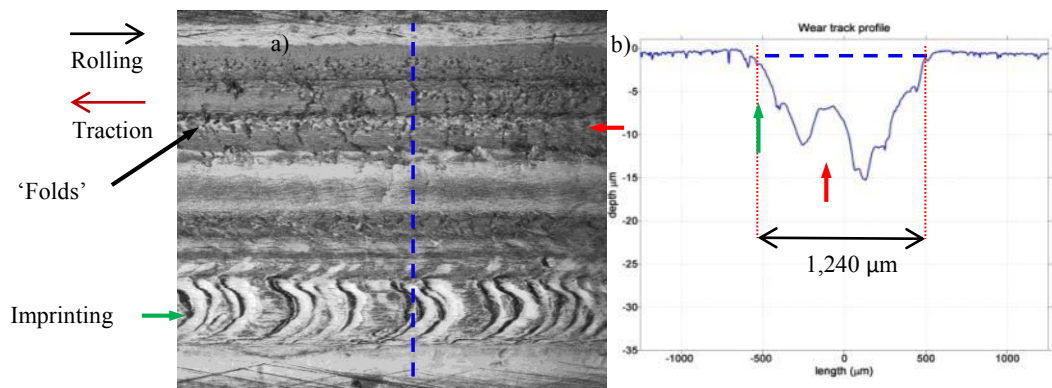


Figure 7.17: a) AIF post-run snapshot of the steel sample from Run 1-2 b) profile measurement. Corresponding markings of the ceramic fault are shown in Figure 7.14

The snapshot shows two main bands along the rolling track: a smooth band at the upper part and a series of repeated imprints at the bottom band. These bands were matched by their relative transverse positions on the track with two prominent zones of the F(d) fault indicated in Figure 7.17a: by a red arrow for the upper track and green arrow for the bottom. Despite its smooth appearance, the upper band represents severe plastic deformation of the wear track by the presence of 'folds'; these were reported also by Lakshminarayanan [67] in a hybrid rolling contact under high contact pressures >5 GPa. The bottom band is multiple direct imprints of the

corresponding section of the F(d) fault. Both bands are the results of a two-part process initiated by the F(d) fault:

First, a direct contact made between the steel and the F(d) fault generates impressions on the steel surface (as illustrated in Figure 7.18a) with possible associated abrasive wear (due to sliding) and material loss. A single ceramic fault produces multiple imprints on the steel surface as a result of the rotations of the two specimens, eventually covering the entire steel surface as the two rotations are generally not in resonance. The latter is due to not only the basic geometry of the two specimens but also due to differences in their surface speeds (i.e. sliding – reducing or extending the relative cycle track lengths).

Second, smoothing of the irregular steel surface occurs (in parallel with the above) through over-rolling by the ceramic counterface in the contact cycles as the imprint comes into contact with a smooth ceramic specimen surface, away from the fault location. Here, protruding steel features are folded over one another by the high contact pressures producing a smoother surface by severe plastic deformation (as illustrated in Figure 7.18b). This smoothing process has been described by Wan [259] as ‘healing’, associated with hybrid rolling contacts and the relative hardness of the ceramic material.

For a single (or dominant) imprint zone the second mechanism comes into effect immediately after the first. After a sufficient number of rolling cycles and repetitions of the above processes, the steel is worn sufficiently (forming a channel track) to accommodate the most protruding parts of the ceramic fault and, therefore, obtaining a smooth appearance as evidenced by the upper band above. With this zone accommodated, the process gradually shifts to other protruding parts of the fault, already existing or newly generated.

The bottom band is produced by a zone that is in a peripheral location on the track relative to the upper one. Due to the curvature of the ceramic specimen in the transverse direction the interaction with the steel of the former (peripheral) zone is less pronounced than the latter. Nevertheless, at the fault depression point it is sufficient to produce an imprint on the steel surface. At the subsequent smooth surface contact, however, the interaction is again reduced and therefore the smoothing effect is only minor. Therefore the initial imprints (repeated due to rotations of the specimen) remain.

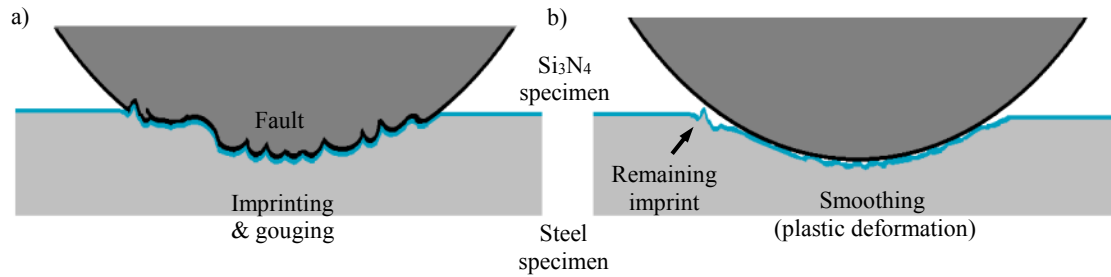


Figure 7.18: Cross-sectional view illustrating a) the imprinting and gouging of the steel by central and peripheral zones of a fault b) the smooth surface emphasises the central area (the specimen rotate in direction perpendicular to the page).

Similar mechanisms are responsible for a smooth wear track in the Run 1-1, produced in a straightforward way by Fault 3 (F(d)) as shown in Figure 7.19 below.

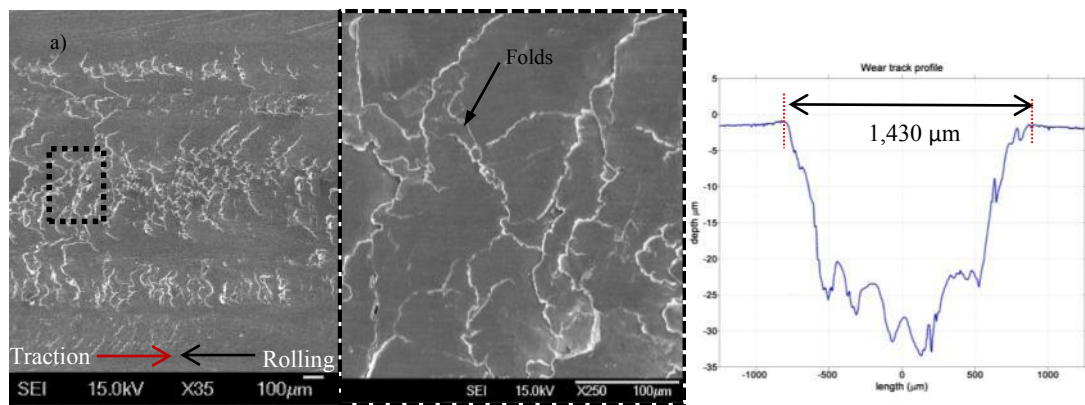
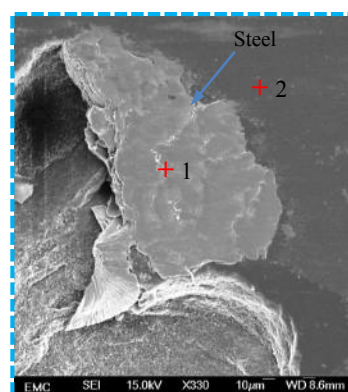


Figure 7.19: Run 1-1 steel wear rack analyses, SEM images a) 35x and b) 250x at the corresponding box colour in (a); c) a Talysurf profile measurement across the wear track.

Evidence of the initial mechanism is substantiated by the presence of steel residue in the fault as shown in the following figure.



1		2	
Element	Weight%	Element	Weight%
C	3.50	C	4.71
O	3.47	N	26.14
Si	0.58	O	4.31
Cr	1.48	Al	0.97
Fe	81.67	Si	54.02
Au	9.30	Au	9.86
Totals	100.00	Totals	100.00

Figure 7.20: SEM image of the upper part of Fault 3 from Run 1-1 (refer to Table 7.1 for its position), and in the table, the corresponding EDX measurements at the locations indicated.

Steel was confirmed by EDX analysis by the presence of high levels of iron and chromium at the location (1) of the residue, compared to the levels outside the fault area (2).

A close inspection of the entire steel counterface revealed two spalls of considerable size shown in the Figures below.

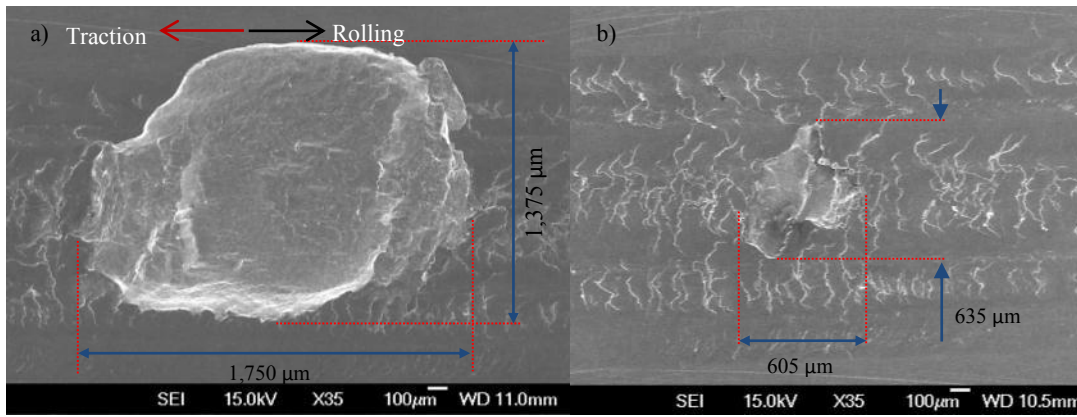


Figure 7.21: SEM images of the two spalls found in the steel of Run 1-1 a) *Large Spall*, b) *Small Spall*.

These spalls could not be produced by either of the two mechanisms above (because they would produce a smooth surface or a repeated pattern), but can be explained as a result of a debris impingement.

Because the traction in the run was minimal i.e. no sliding induced (unlike all the other runs), the effects of leading and trailing edges in Flaking are reversed and most importantly, the resulting debris tends to flake off on the leading edge i.e. in the direction of roll. Therefore, the debris is ejected into the path of the contact and subsequently pushed into the soft steel by the sloping trailing part of the fault basin. On the steel however, the traction is opposite to rotation (the steel was the driving interface); correspondingly loose debris gets effectively dragged over the steel surface by the leading edge of the ceramic. This results in a prolonged shape as in Figure 7.21a above. Examining the depth of the steel fault ($250\text{ }\mu\text{m}$ – see figure below), it is over several times the maximum depth of the ceramic fault ($97\text{ }\mu\text{m}$). Therefore, a second mechanism must have been present: RCF producing spalling. Spalling in steel has been observed by many researchers in the depth closer to the maximum principal shear stress; this is calculated as $252\text{ }\mu\text{m}$ in this Run, i.e. almost exactly the depth observed.

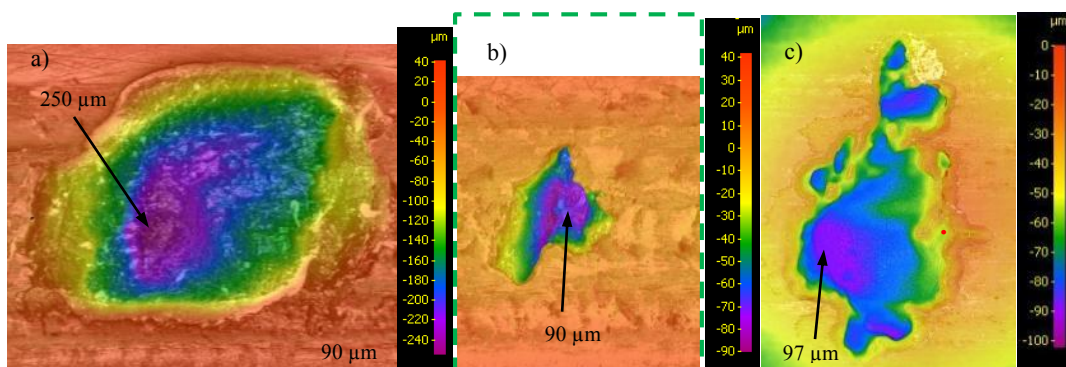


Figure 7.22: Topography images from Run 1-1 of a) the Large Spall, b) Small Spall and c) Fault 3 of the ceramic specimen.

Thus the above steel fault is caused by a 2-stage process: a weakening of the steel surface by debris dragged over a distance, followed by a spalling induced by subsequent continuous roll-

over at the weakened area. Because of the significant depth of the steel fault, subsequent roll-overs do not have a significant smoothing effect.

The second steel fault has been produced by similar mechanisms, however the ceramic debris was of much smaller size and therefore produced no spalling effect and occurred later in the run therefore was not subsequently smoothed out.

7.7 Summary

The Runs were performed on a machine emulating the rolling processes in a hybrid bearing which was also equipped with sensors of the three types of detection methods investigated (vibration, ES and AE). To accelerate the fault development processes, pre-cracks were seeded on the ceramic specimens using an indentation process performed on a separate machine. Visual snapshots and measurements of the pre-cracks and the final faults were made using inspection tools. The present chapter considers the interpretation of the results from the inspection tools in order to establish a basis from which the sensor data could be judged; the outcome of this is presented in the following chapter.

In order to interpret the results of the visual inspections and measurements of the faults generated in the Runs, a model of RCF propagation was constructed incorporating the models of the two major ways of fault propagation reported in the literature: near-vertical and shallow-inclined cracks, identified here as the dominant causes leading to CRC and Flaking type faults respectively. The model assumes that after an initial stage one of the types becomes dominant and governs the subsequent development of the fault.

The interpretation of the results was based on identifying each end-run snapshot with a stage in the model and reconstructing the mechanisms that led to its production from earlier stages, notably the initial pre-crack. Where the rolling process produced significant fault development, the end results were easily identifiable as either CRC or Flaking³⁹. The generated pre-cracks in most cases emulated realistically the near-vertical and shallow-inclined cracks as radial and lateral cracks respectively.

Similar matching of the pre-cracks to the model stages encountered a fundamental difficulty however, in that the relative prominence of the radial and lateral cracks could only be judged indirectly by their surface characteristics, notably their skewness and size. Prominent trailing edge skewness was interpreted as an assured sign of a dominant lateral crack, representing therefore an early Flaking stage, and small size indentations were interpreted as the initial stage

³⁹ Two runs have produced delamination type faults that were subsumed under the Flaking type.

that could potentially produce faults of either type. The remaining pre-cracks were left unclassified.

The faults encountered in the runs were then matched to the progressive stages of the model and various mechanisms responsible for the fault progression were described with the snapshots and associated measurements providing evidence for their presence/operation.

The results showed a general agreement with the overall structure of the model i.e. separate developments of the CRC and Flaking mode. The one case where a minor Flaking fault at the pre-crack stage developed subsequently into a major CRC fault was explained by the uncertainty of the initial pre-crack classification.

The results of the CRC faults investigation show that the fault patterns produced were in agreement with those found by other researchers and their production in successive stages could also be explained by the models they developed. Several of the runs produced an advanced stage of this fault (CRC(d)), not reported in the literature, interpreted as the result of the induced sliding which promotes the fault development into multiple and extensive c-cracks.

The effects of this advanced stage were also examined in the snapshots and measurements of the steel counterface where they produced abrasive and imprinting wear.

Similarly, the shapes of the Flaking faults were in agreement with those found in the literature and their production stage by stage by the detailed mechanisms described by various researchers. Supplementing these findings were additional factors identified as playing important roles in the flaking development.

An initial trailing edge skewness led to the development of a substantial fault (F(c)/ F(d)) due to oil incursion into the fault assisted by the roll-over process and the traction in its direction. The reduction of traction magnitude produced faults with opposite orientations (leading edge).

Having reached an advanced stage (F(d)), the process was accelerated by the onset of direct contact with the steel counterface, producing greater impacts on the fault. This led to the fault enlargement and debris production, and also abrasion and imprints on the steel counterface.

Three distinct types of abrasion and imprints were found.

The first led to 3-body abrasion, where ceramic debris retained in the contact over several rotations created multiple scratch marks and its comminution produced subsequent pit marks, both along the ceramic specimen.

The second was a direct impingement of a ceramic fault on the steel surface creating imprints and abrasion. These were smoothed out by the subsequent roll-over processes, however, these effects were found to depend critically on the transverse position of the impacting fault zone.

The third type was the impingement of debris into the steel counterface, during a leading edge flake, assisted by traction in the countervailing direction. The presence of large debris resulted in multiple deep imprints through dragging initiating a (RCF assisted) spall.

Additional comments on the run arrangements were as follows.

The indentation process employed for the runs produced a variety of pre-cracks shapes; this was an advantage because it created multiple testing conditions. However, the repeatability of this process is severely diminished through the lack of full control over the indentation results. Other researchers whose results were referred to here, were able to produce single types of pre-cracks through better control of crack generation processes; such ability may be required for our own future work.

The inability to assess crack features below the surface was a disadvantage in that many pre-cracks could not be reliably assessed for fault types (i.e. relative prominence of the radial and lateral crack). Such ability would be required e.g. to further investigate the conditions that lead to crack development from an initial indefinite fault stage to a CRC or Flaking type.

In interpreting some of the results, notably those that involved single or double imprints, an uncertainty related to the shut-down procedure of the twin disc machine. Namely, it was possible that the (gradual) shut-down process could induce additional effects, which would therefore appear as one-off. In the reasoning above these were dismissed on the statistical grounds (as the shut-down process was of negligible duration compared to the entire run), however future investigation could be helped with the shut-down process being instantaneous, for example involving instant disengagement of the rolling surfaces.

The use of three indents per specimen was advantageous in that it was an easy way of producing multiple results from a limited number of runs and the interpretation of the steel surface effects was aided by the fact that mostly a single fault became dominant. Having established the major conclusions from these runs, further work will however require the use of single pre-cracks per specimen to avoid cross-fault interference and obtain statistics of fault life-times.

Chapter 8. Sensor Data Analysis

8.1 Introduction and Overview

The objective of the sensor data analysis is to examine the signal profiles in order to extract the features that can be attributed to fault development. As explained in Section 3.1 and Chapter 5, this process is based on the combination of (i) signal data enhancement (essentially removing or attenuating the features known not to be associated with faults) and (ii) feature extraction/amplification (emphasising the parts known to be associated with faults). In Chapter 5 techniques were developed for vibration and WSS to attenuate the low frequency resonance and eccentricity parts of the profiles and by this effectively also achieving (ii) as the fault effects in the resulting signal profiles should be more prominent. For AE, an advanced filtering technique was developed that showed, on the test data, enhanced fault features over the raw profile.

Further steps to associate specific aspects of the profiles with particular faults are based on additional considerations, notably:

Time distinction, referring to distinct stages in the fault development, for example as identified in the general model presented in Chapter 7.

Location distinction, referring to a known position of the fault on the test specimen, for example by its angular measurement.

Frequency distinction, when it is known that the generated fault dominates particular frequencies; thus, for example, faults are more prominent in the high resonant frequencies of the vibration spectrum (as already used in the development of the vibration signal processing technique above).

Intensity distinction, when a fault produces an effect of a higher magnitude than other sources notably background, for example in the AE profiles after the enhancement developed in Chapter 5.

8.1.1 Analysis Tools

In order to analyse the possible distinctions, a number of graphical tools were implemented as follows.

Signal-time profiles (for example see Figure 5.17): these are line graphs showing the progression in time (x-axis) of an indicator of the signal intensity (y-axis) taken at sample intervals throughout the run - this could be e.g. the average value or peak value over that

interval (see Table 8.2 below for descriptions of calculations). Such graphs should reveal trends in time and thus providing indications of fault developments.

Angular-time profiles (for example see Figure 5.18): these are colour maps showing signal progressions in time (x-axis) distinguished by the angular location on the specimen (y-axis) and signal intensity (indicated by colour). These values are obtained by matching the signal data against tachometer signals. Typically a fault would be visible as a horizontal band against its angular position value; however additional non-fault related features are also revealed.

Frequency-time profiles (for example see Figure 5.16): these are colour maps showing the progression in time (x-axis) of the frequency values (y-axis) and the power intensity (indicated by colour). These PSD estimates computed using the Welch method as described in Section 5.2. Typically, prominent frequencies would appear as horizontal patterns developing in time.

Additional graphical tools were employed to illustrate the details identified in the above analyses.

8.1.2 Analysis Approach

The approach to the signal analysis builds on the findings based on the visual analyses presented in the previous chapter.

Initially, the signal-time profiles were produced for each of the sensor types and runs in order to identify any common patterns, relating also back to the fault types and sizes generated in the run, as assessed in the visual analysis (see Chapter 7). Based on this initial analysis, three runs were selected for subsequent detailed analyses as they represented a wide variety of signal patterns, fault types and sizes. The subsequent sections of the document include the results of the selected runs only.

The detailed analysis consisted of these steps:

The initial assessment of a run was based on a detailed examination of the combined signal-time profiles in order to discern the overall progression of the faults in the course of the run and how they were reflected in the various signal media. The combined profiles show the comparative signal strength values detected across all sensors and are presented in both raw and processed forms. To enable comparisons, these values have been normalised based on their respective maximum run values. Because of the summary nature of these profiles, any patterns/trends identified are of a global nature (i.e. relating to all faults); however, could be traced to individual, dominant, faults (or even their individual stages) if these developed distinctly over the course of the run.

In order to discern the individual contributions of the faults, the angular-time profiles were analysed starting with vibration profiles as they are produced by mechanical processes only, i.e. are closest to the visual data obtained earlier. Prominent features in the profiles were interpreted using the fault development model presented in the previous chapter. Also, frequency analysis was performed in order to make a comparison with the assumptions that have been made in the course of the signal processing development described in Chapter 5.

This was followed by a similar analysis of the electrostatic profiles, with any prominent features identified related to the findings of the vibration and visual analyses. As for vibration, a frequency analysis was also performed.

A similar analysis of the AE signal could not be performed as it was not possible to record tachometer data with AE, due to device capacity limitations – possessing a high-pass filter of a minimum of 40 kHz (above the maximum frequency of the filtered tachometer signal). The potential limitation to the research is discussed in section 8.6.

The runs selected for detailed analyses are presented with the rationale for their selection in Table 8.1 below.

Run	Rationale
1-2	Severe run (high slip) producing a large Flaking fault.
2-1	Moderate conditions, producing early CRC and Flaking faults
2-2	Moderate conditions, producing a large CRC and a minor Flaking faults.

Table 8.1: Runs selected for detailed analyses.

The sections of this chapter present:

Section 8.2: analysis of the run histories based on the signal-time profiles.

Section 8.3: analysis of the fault progression in the vibration medium based on both angular-time and frequency analyses.

Section 8.4: analysis of the fault progression in the electrostatic medium based on both angular-time and frequency analyses.

Section 8.5: a reconstruction of the fault development histories based on the visual, vibration and electrostatic analyses.

Section 8.6: analysis of the fault progression in the acoustic emission medium based on both angular-time and frequency analyses.

Section 8.7: a summary of the above findings.

8.2 Runs at a Glance: Signal-Time Profiles

To get a summary assessment of the runs in terms of their signal characteristics, the signal-time profiles are presented in the Table 8.3 below from each of the sensors. Their meaning is explained in the following Table 8.2.

Name	Profile description	Specimen	Tachometer	Statistic
Vibration	Raw (TSA) vibration	Si ₃ N ₄	T1 (upper)	RMS
Vib-EMD (IMF1)	Filtered (TSA) vibration	Si ₃ N ₄	T1 (upper)	RMS
WSS1	Raw (TSA) WSS	Si ₃ N ₄	T1 (upper)	RMS
WSS1 filtered	Filtered (TSA) WSS	Si ₃ N ₄	T1 (upper)	RMS
WSS2 wrt Steel	Raw (TSA) WSS	Steel	T2 (lower)	RMS
WSS1 wrt Steel	Raw (TSA) WSS	Si ₃ N ₄	T2 (lower)	RMS
CoF	Coefficient of friction	Both	-	RMSx10
AE-EI	AE raw Energy Index	Both	-	Peak counts
AE-EI	AE filtered Energy Index	Both	-	Peak counts
AE-Hit-based	AE Hits	Both	-	counts

Table 8.2: Signal-time profile descriptions

The CoF profile was calculated from the load and torque recorded by the machine's in-built sensor and multiplied by a factor of 10 for comparative purposes. The AE profiles were separated to make the graphs more readable.

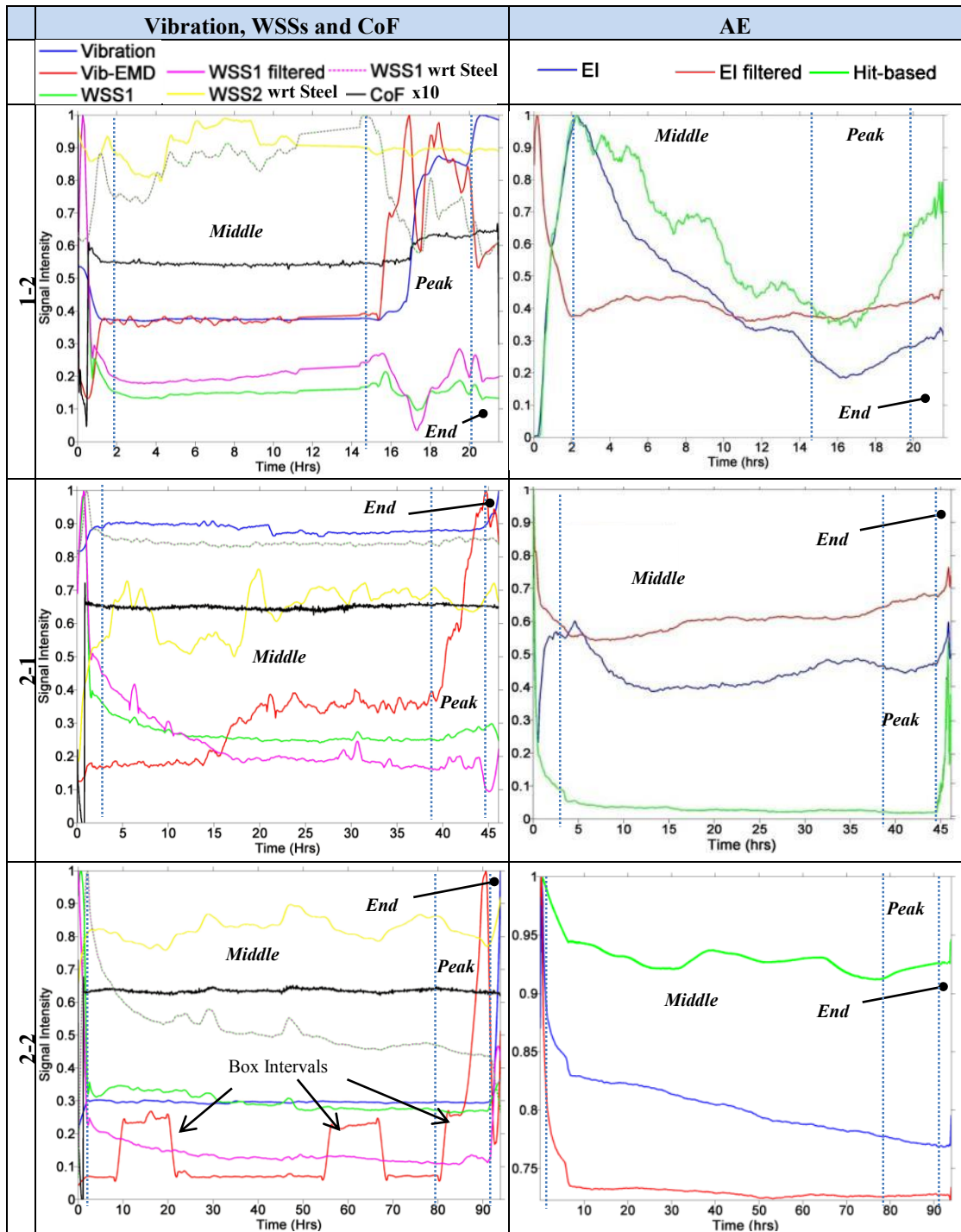


Table 8.3: Signal-time profiles for the runs selected for detailed analyses.

The *vibration* and *WSSs* profiles show four distinct periods indicating potentially different mechanisms at play (with possible transitional periods in-between - therefore times are approximate):

Initial: characterised by sharp changes in signal strengths and explained by the running-in processes including tribocharging.

Middle: relatively quiescent period indicating gradual fault progression.

Peak: rapid changes progressing through one or more peaks or troughs.

End: rapid gradient changes in the direction opposite to the previous period.

The values of these periods for the three selected runs are shown in Table 8.4 below.

Run	Initial (hrs)	Middle (hrs)	Peak (hrs)	End (hrs)
1-2	0-2	2-14.5	14.5-20	20-EoR
2-1	0-2.5	2.5-38	38.44.5	44.5-EoR
2-2	0-2.5	2.5-80	80-92	92-EoR

Table 8.4: Run periods of the three selected runs (EoR- end of run)

Taking the assumption that, once the steady state is reached, most profile variations originate from fault progressions (see Figure 3.22 in Section 3.4), the profiles indicate that most fault-related activity happens in the Peak and End periods. Broadly, they also demonstrate mutually corresponding patterns (for example peak/trough, inflection points etc.), reflecting intervals of increased activities.

Comparisons between the raw and filtered profiles show that these patterns appear in filtered ahead of those in raw, indicating increased sensitivity of the filtering methods. This applies to the vibration and WSS1 profiles. For example, the prominent rise in the vibration signal in the Peak period of Run 1-2 commences at 17 hrs in raw, but around 15.5 hrs in the EMD profile; the corresponding WSS1 profiles peak in the filtered before raw.

The box-like patterns in the vibration EMD profile of Run 2-2 are specific to this run and are discussed in detail below (see Section 8.3).

The CoF follows the vibration and WSS pattern only in the extreme run (1-2) whereas it remains flat in both of the Moderate runs. This indicates that the severe slip had a prominent role in the progression of the faults, however, that under normal load conditions the CoF is not a sensitive indicator of faults.

The AE profiles do not fit these periods well. They show rapid variations in the initial period of the runs followed by indistinct trends in the Middle and Peak periods and sharp variations again at the end of the runs. In particular they do not reflect any of the activities presumed to occur in the specimens from the vibration and WSS profiles.

At this *summary* level, the vibration and WSS1 profiles show a broad agreement with the RE life cycle model presented in the literature review above (see Figure 3.22 in Section 3.4), however, detailed analyses are required to discern the contributions of the individual faults.

The AE profiles do not fit the life cycle model in a clear way.

The detailed analyses now follow.

8.3 Vibration Sensing

This section presents a detailed analysis of the vibration profiles. Firstly, the angular-time profiles are investigated to discern the contributions of individual faults and deduce the various events in their progression as guided by the fault progression model and the results of the visual inspection analysis (Chapter 7). Selected burst data are examined for more detail. Finally, frequency-time profiles are analysed to obtain additional information on the findings.

The angular-time profiles based on the TSaA calculation and are presented below in Figure 8.1, both raw (left) and filtered (right); the final 20 hrs of Run 2-2 are presented in Figure 8.2 for improved resolution. The fault locations are indicated in each of the figures.

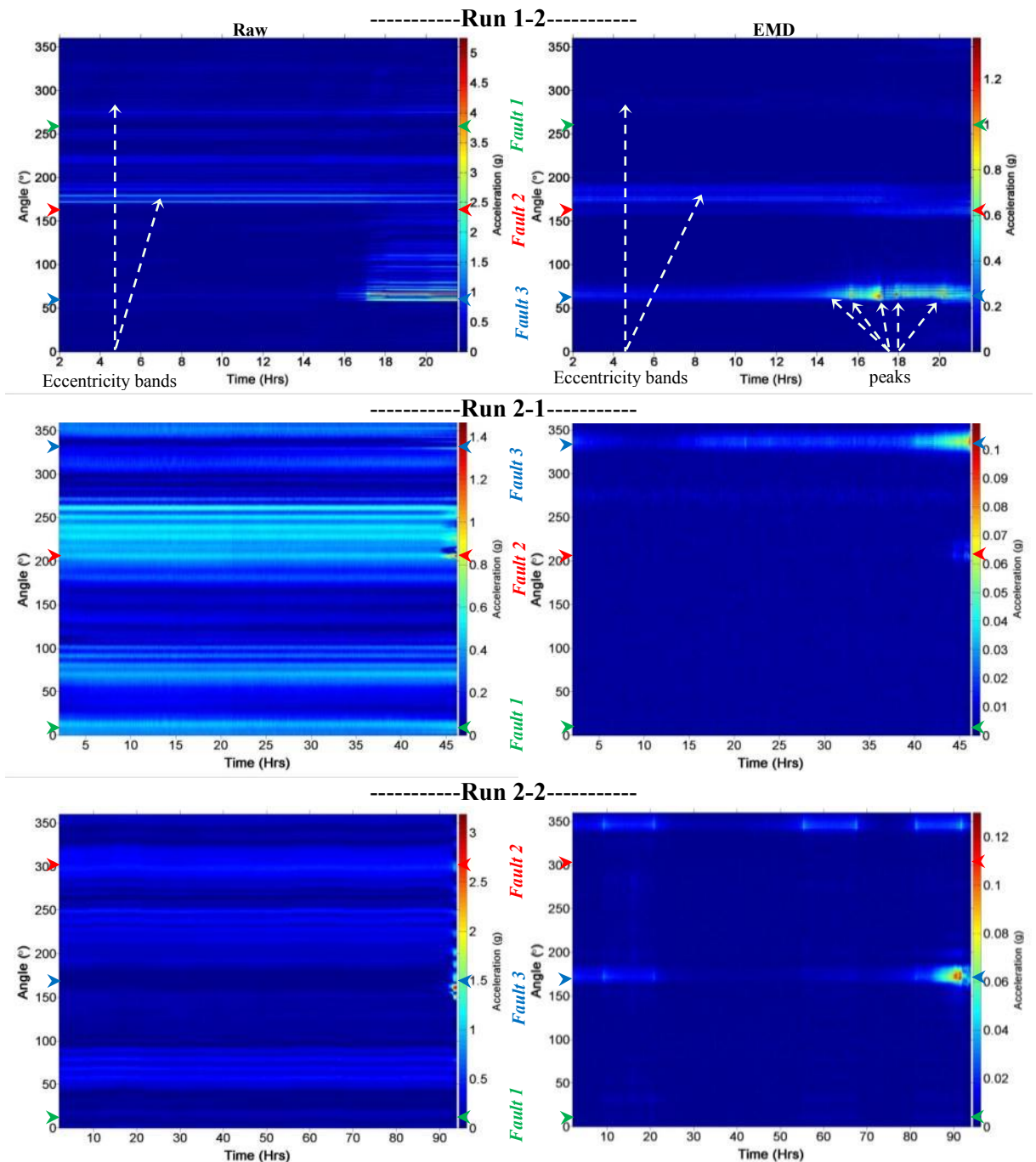


Figure 8.1: Vibration angular-time profiles for TSaA-raw (left column), TSaA-EMD (right column).

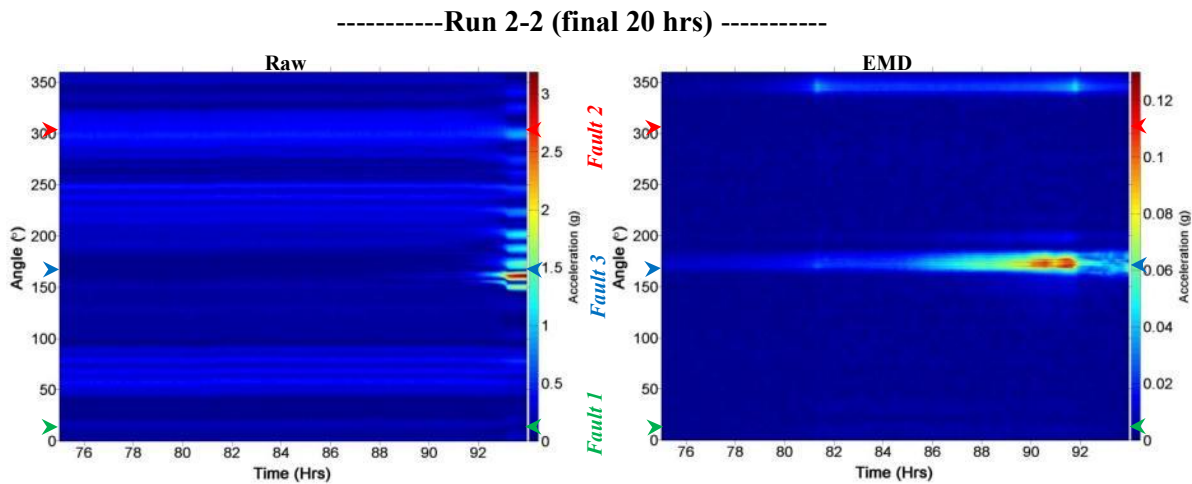


Figure 8.2: Vibration angular-time profiles for final 20 hrs Run 2-2 for TSaA-raw (left column), TSaA-EMD (right column).

The prominent features on these figures are the horizontal bands and peaks which are now discussed in detail below.

8.3.1 Horizontal Bands

Two types of bands are visible, differing in constant and varying signal levels. Each is discussed in turn.

8.3.1.1 Eccentricity

Both raw and EMD profiles are dominated by horizontal bands at fixed angles and constant signal levels throughout the run; these are the eccentricity bands. They are not fault-related as their angular locations do not match the faults. Similarly, they are not related to the general wear state of the specimens as they are unchanging over the runs. Their locations, identified in the *raw* profiles, are as follows:

Run	Angle (°)	Prominent/Faint
1-2	100	(faint)
	175-200	√
	220	√
	275	√
2-1	17	√
	70-100	√
	200-260	√
2-2	50-90	√
	110-150	(faint)
	170	(faint)
	180-250	√
	290-320	(faint)
	350	(faint)

Table 8.5: Eccentricity band locations as identified in the raw vibration angular-time profiles

All appear constant throughout the runs. The EMD filtering removes all with the exception of two:

Run 1-2, band 175° - 200° is attenuated and still visible in the EMD-TSaA, its centre coincides with Fault 2, so it is possible that it is a result of an interaction between eccentricity and fault.

Run 2-2, band 350° still appears prominent in the EMD profile, in addition three periods of constant high-level signal become visible at this angular location (therefore indicating a high frequency source >6 kHz). They cover the following intervals ('box' intervals⁴⁰):

- 9-21 hrs (12 hr duration)
- 55-67 hrs (12 hr duration)
- 81.5-92.5 hrs (11 hrs duration)

They appear consistently at 170 and 350° i.e. precisely 180° apart, twice per rotation. This synchronicity indicates therefore an internal, machine based source. They have a clean shape - the vibration profiles always revert to the pre-interval values (except the last period when there are other events happening), and they are not detected by either WSS (noted in the WSS analysis below, Section 8.4) or AE sensors (compare the signal-time graphs above in Table 8.3). This indicates that they are unlikely to be related to the actual specimens. Their presence, only discovered during the analysis of the results and not seen in any other runs, is therefore puzzling. Because of their anomalous nature and no apparent connection to the specimens, they have been ignored in the analyses following. They present a caution that sensor based detection requires a careful attention to all machine (and environmental) parts.

8.3.1.2 Bands Related to Defects and Faults

Contribution of an individual fault produces a horizontal line at its corresponding angular location, with the magnitude generally reflecting the intensity of its progression at a particular time. Many start only appearing at a point (from prior levels that were considered to be background) during the run; these represent their '*first sight*'. Also distinguishable are rapid, local changes in their intensities (referred to as '*peaks*'), representing sudden changes in the fault activity.

Table 8.6 below summarises the faults, their end-run snapshots, the first sights and any prominent peaks.

⁴⁰ Because of their box-like shapes in the signal-time profile in Table 8.3.

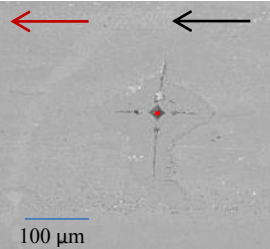
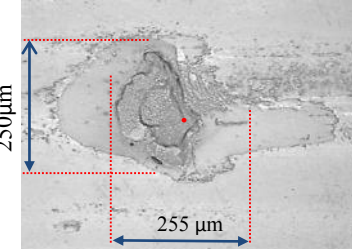
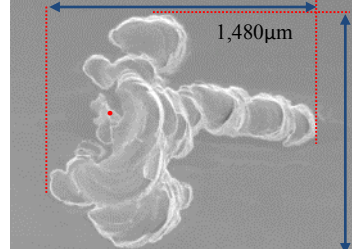
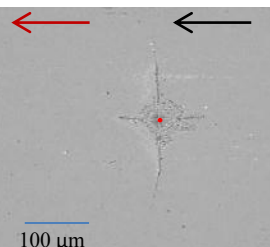
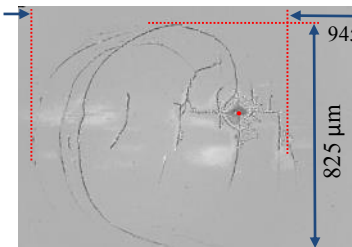
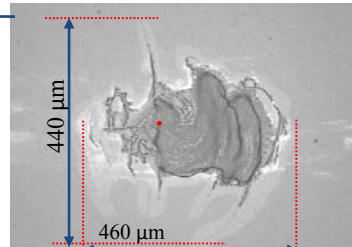
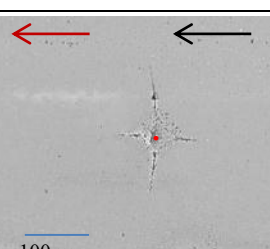
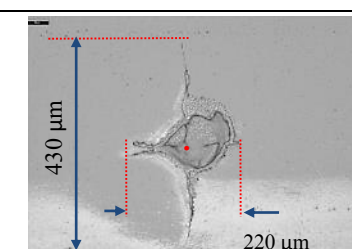
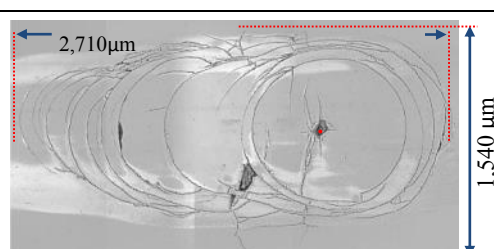
Run	Faults		
	1	2	3
1-2			
	268°	161°	60°
FS	-	14 hrs (EMD)	17 hrs (raw), 2 hrs (EMD)
Peaks	-	-	(a) 14.4 hrs, (b) 15.8 hrs, (c) 16.8 – 17.2 hrs, (d) 17.8-18 hrs, (e) 20- 20.5 hrs (EMD only)
2-1			
	10°	208°	328°
FS	-	44 hrs (both raw and EMD)	40 hrs (raw), 15 hrs (EMD)
Peaks	-	-	(a) 21 hrs, (b) 41 hrs, (c) end of run (EMD only)
2-2			
	18°	301°	162°
FS	-	93 hrs (raw), - (EMD)	90 hrs (raw), 80 hrs (EMD)
Peaks	-	-	(a) 85 hrs, (b) 90 hrs, (c) 91.5 hrs (EMD only) (d) 93.25 hrs (raw)

Table 8.6: Fault snapshots and prominent vibration events ('FS' = first sight, subsequent event peaks labelled by letters).

The analysis of the angular-time profiles (Figure 8.1&Figure 8.2) is presented in the following sequence. First the relative of the detectability of the faults is assessed, in both raw and EMD profiles. This is then followed by a detailed investigation of prominent events of selected faults.

The profiles indicate that all the small faults (Fault 1) were not detectable. However, both Flaking and CRC faults (due to Faults 2 & 3) are detectable.

The profiles indicate (by the relative intensities) that Runs 1-2 and 2-2 were dominated by a single fault (Flaking and CRC respectively), whilst both Faults 2 & 3 actively contributed in Run 2-1 (CRC and Flaking respectively).

Flaking faults produce more prominent vibration signatures than CRC. This is directly evident in Run 2-1 where a Flaking fault signal is detected much earlier in the run than that of the CRC, despite the latter having a considerably larger size. Further evidence is provided by the comparison of the signal levels between the runs 1-2 (approx. 5g) and 2-2 (approx. 3g). The dominance of Flaking over CRC is easily explainable in that the former produces more prominent surface features.

EMD enhances the prominence of the fault features, for example most faults appear earlier and none of the prominent peaks appear in raw.

An exception to this is that Fault 2 in Run 2-2 is not visible in EMD. The raw profile of the run also shows the presence of a series of horizontal bands from 93 hrs to the end of run – all removed by EMD. These are explained by a sudden increase in Fault 3 (CRC) to a large size that induces, through subsequent/trailing contact points (i.e. at higher angles), low-frequency, dampened oscillations. One of these coincides with Fault 2 which thus becomes ‘visible’ in the profile; however this is only an apparent effect due to a coincidence between a resonant and a fault location angles.

The prominent events in the dominant fault developments in all runs are explained by a direct contact with the steel, resulting from sudden increases in the fault size. The evidence for this is found in the steel vibration profiles, showing variations at these time points. These profiles are presented in Figure 8.3 below.

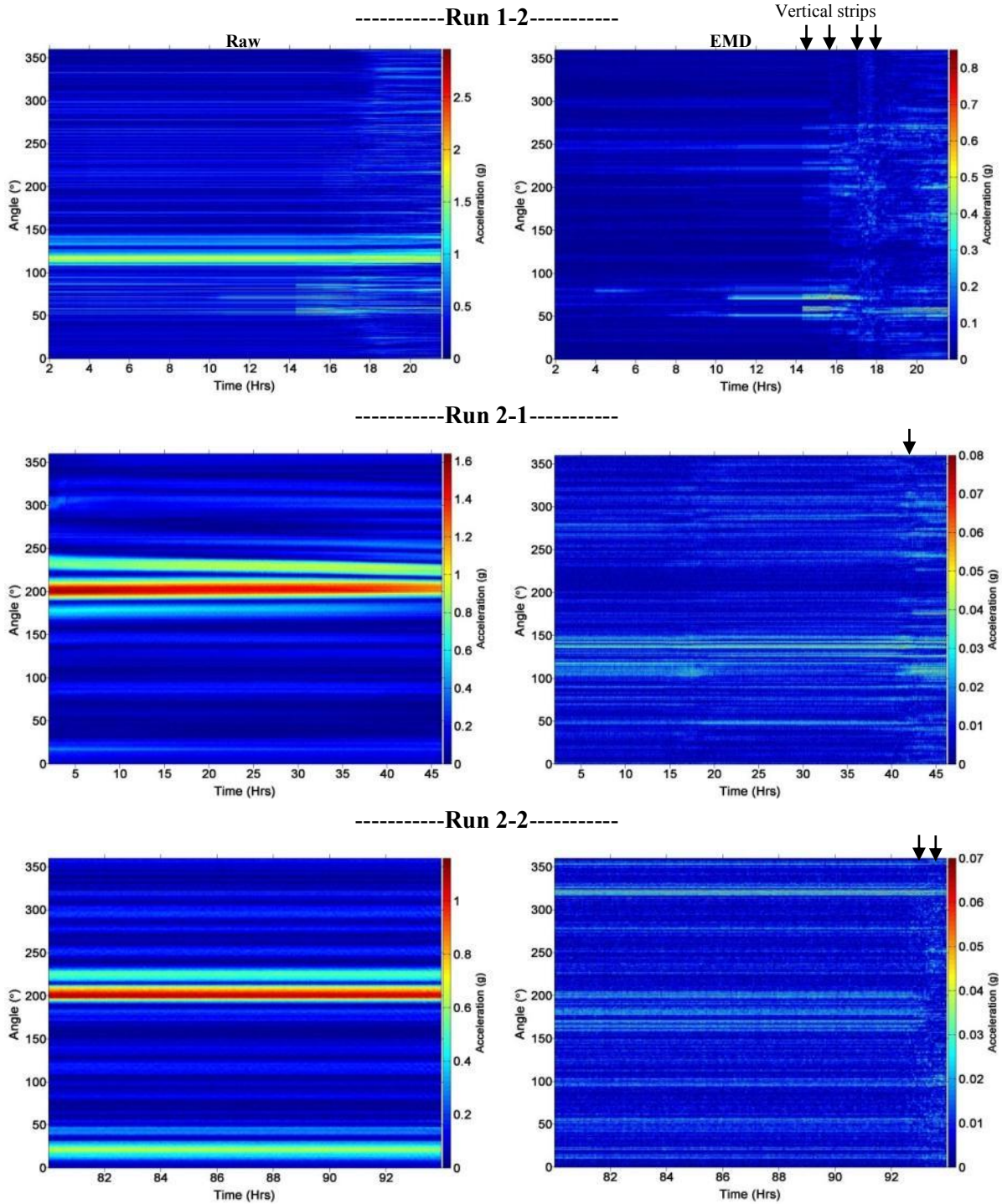


Figure 8.3: Vibration angular-time profiles with respect to the steel specimen for TSaA-raw (left column), TSaA-EMD (right column).

These profiles are constructed from the vibration signals by TSaA over multiple rotations of the steel specimen, obtained from the lower tachometer (T2), therefore reducing the contributions of the silicon nitride specimen. They are plotted against the angular positions along the steel specimen surface and display prominent eccentricity bands similar to those in silicon nitride and are also reduced by the EMD filtering.

The direct contact points induced by the faults are visible as vertical strips as the contacts are made rapidly over successive rotations covering the entire circumference. The following correspondences were identified in the EMD profiles (Figure 8.1&Figure 8.2) between the vertical strips and the prominent vibration events ('peaks' in Table 8.6):

Run 1-2: all events visible from first sight and peaks (a)-(e).

Run 2-1: peak (b)

Run 2-2: peaks (a) and (b).

The visibility of all the fault-related events in Run 1-2 is not surprising as this was a severe run – causing substantial interaction with the steel specimen.

Further evidence of the direct contact is provided by an examination of the steel counterface. As discussed in the previous chapter, direct contact with the steel specimen can induce additional wear in the steel and was confirmed by the end-run visual inspections of the steel specimens, reproduced below in Table 8.7 and interpreted subsequently.

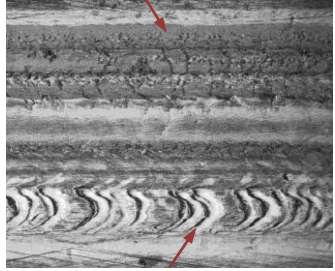
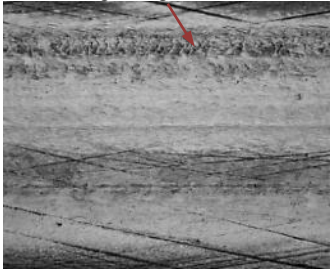
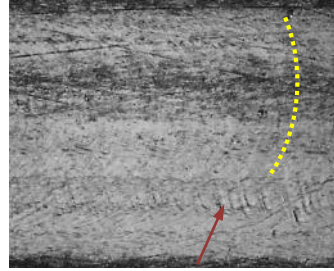
Run		
1-2	2-1	2-2
<p>'Folds' – plastic deformation</p>  <p>Imprinting</p>	<p>Roughening from Fault 3</p>  <p>Traction Rolling</p>	<p>Pattern corresponding to Fault 3 cracks</p>  <p>Pattern corresponding to <i>Large spall</i></p>

Table 8.7: Snapshots of the steel wear tracks.

For Run 1-2, as discussed in section 7.6.5, substantial wear in the steel was induced by roughening (through imprinting) and smoothing (through abrasion and plastic deformation) effects.

Runs 2-1 and 2-2 show the evidence of some direct contact with the fault, by a roughened track at the upper contact region from the Flaking fault (Fault 3), and imprints from the *Large spall* (refer to Figure 7.6) and semi-circular cracks of Fault 3 for Run 2-2.

8.3.2 Raw Burst Profiles - Contact Eccentricity Bands

In order to learn more about the details of the fault progression, raw burst data were analysed at selected points in the fault histories. These revealed an additional feature not seen in the diagrams above and described in the following.

Each profile represents data acquired over a selected burst and arranged against the time points as marked by the respective tachometer (T1 for silicon nitride and T2 for steel). In addition, only the signal magnitude was selected (i.e. taking absolute values).

The points of interest were selected to cover a ‘quiescent’ period during a normal fault progression (corresponding to an interval in the Middle period of the signal-time profiles in Table 8.3) and another (‘Active’) at a prominent peak point. Both silicon nitride and steel profiles were examined at each of these points.

The resulting profiles for each of the runs are shown in Figure 8.4 - Figure 8.6 below.

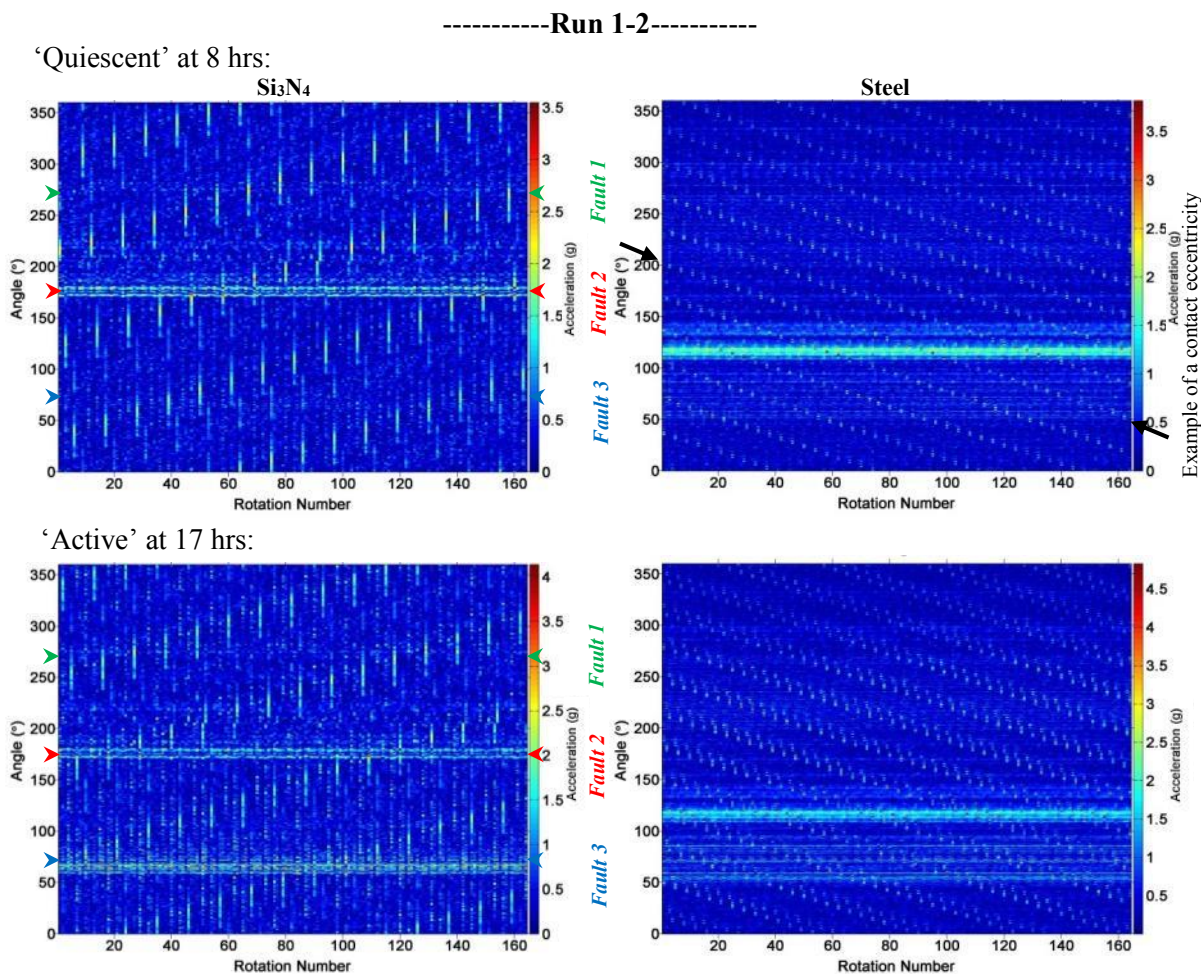
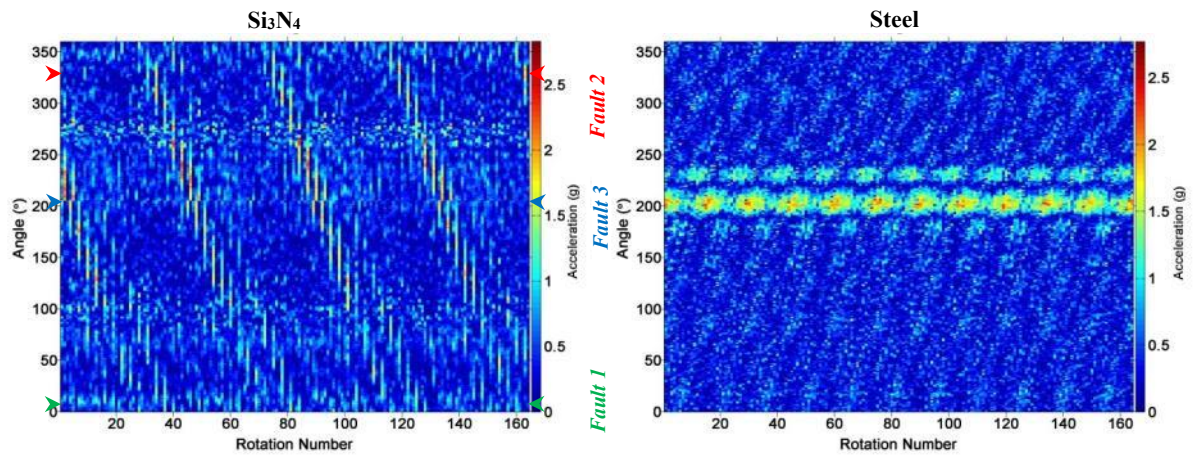


Figure 8.4: Raw burst vibration data for Run 1-2 with respect to silicon nitride (left column) and steel (right column).

-----Run 2-1-----

'Quiescent' at 16.5 hrs:



'Active' at 45 hrs:

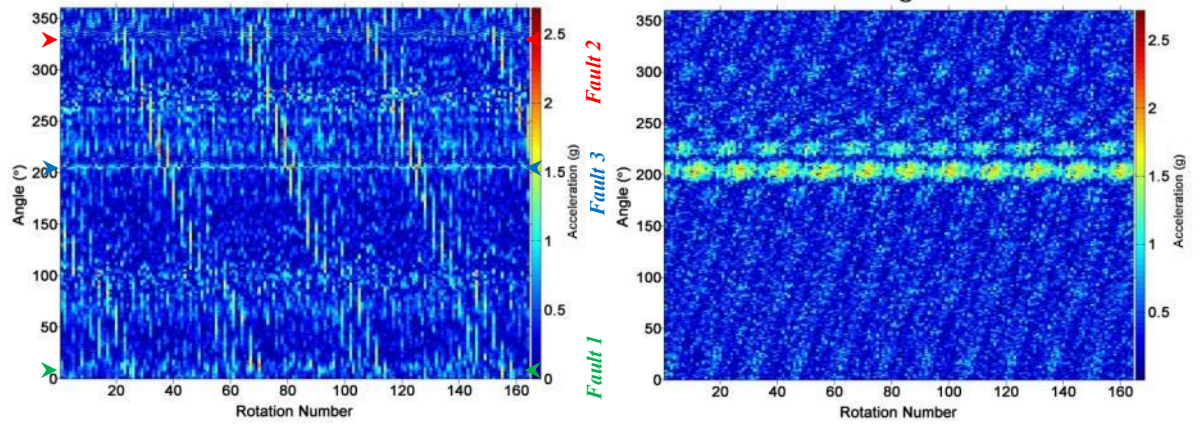
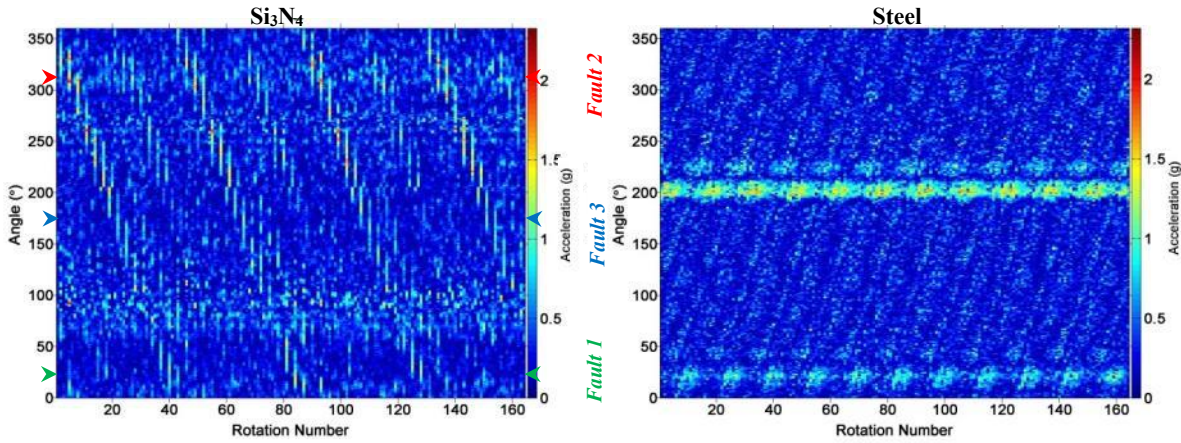


Figure 8.5: Raw burst vibration data for Run 2-1 with respect to silicon nitride (left column) and steel (right column).

-----Run 2-2-----

‘Quiescent’ at 84 hrs:



‘Active’ at 93 hrs:

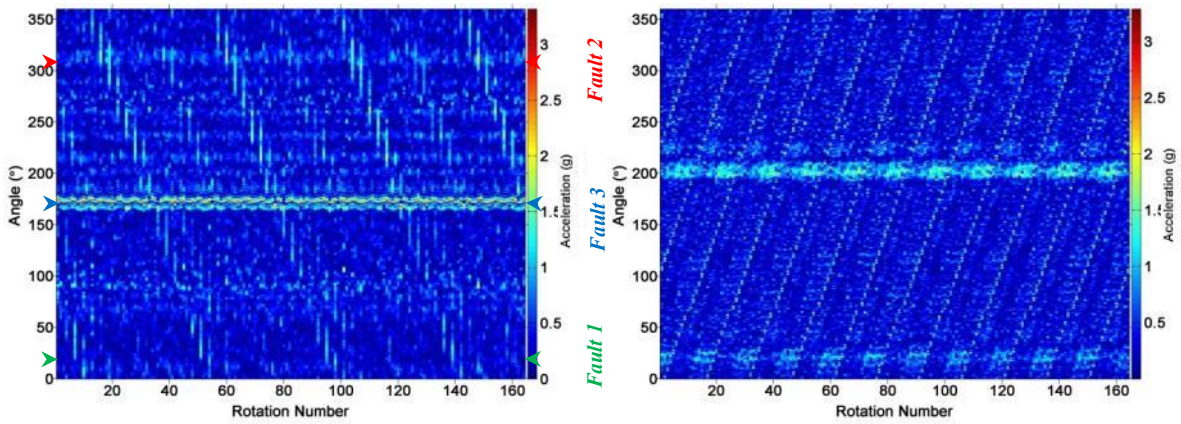


Figure 8.6: Raw burst vibration data for Run 2-2 with respect to silicon nitride (left column) and steel (right column).

All profiles show clearly the dominant fault (Fault 3) at the Peak periods. Run 1-2 also shows the second, non-dominant fault (Fault 2). In contrast, none of the faults are visible in the quiescent periods. This is because their signatures are masked by the high background noise, illustrating the efficiency of the TSaA and EMD techniques in removing it.

The prominent feature appearing in all of the above profiles are diagonal bands. These are the manifestations of the eccentricity points present in the counter specimen. For example, the diagonal bands in the left column (silicon nitride - related) are caused by the eccentricity points inherent in the steel specimen. To distinguish them, the former are referred to as ‘*contact eccentricity bands*’ (and therefore the latter are effectively ‘*natural eccentricity bands*’).

An eccentricity point on the steel specimen will come into contact with the silicon nitride specimen once per rotation, whilst the silicon nitride will perform several rotations depending on the relative speeds of the two specimens. This number is calculated as follows:

$$N_{Si_3N_4} = \frac{\text{Rotating speed}_{Si_3N_4}}{\text{Rotating speed}_{Steel}} \quad 8.1$$

For Run 1-2, this number is:

$$N_{Si_3N_4} = \frac{55}{19.9497} = 2.7570 \quad 8.2$$

The steel eccentricity point therefore makes subsequent contacts with the silicon nitride specimen at 2.757 rotations of the latter. The surplus of 0.757 translates into 272.52° of upper specimen rotation i.e. $\alpha = 87.48^\circ$ in the leading direction from the previous contact. For example, 6 consecutive contacts from a starting position at 0° will be as follows:

Steel Rotation number	0	1	2	3	4	5	6
Si ₃ N ₄ Angle (°)	0	272.5	185.04	97.56	10.08	282.6	195.12

These matches exactly the centres at consecutive peaks of the diagonal bands on silicon nitride (see Figure 8.4- left column).

For the steel specimen, the situation is reversed: each rotation of silicon nitride makes N_{Steel} contacts on the steel surface where $N_{Steel} = 1/N_{Si_3N_4} = 0.3627$, equating to 130.58° in Run 1-2. This is easy to confirm in Figure 8.4 -right column.

For Runs 2-1 and 2-2 (both of which were run with the same rotating speeds), the above effect is magnified by the resonant periods between the two specimen rotations:

$$N_{Si_3N_4} = \frac{55}{18.755} = 2.933 \quad 8.3$$

The surplus 0.933 translates into 24° in the leading direction. Therefore, the same contact point on the silicon nitride specimen surface will come into subsequent contact after 44 rotations, corresponding to 15 rotations of steel. These points are easily seen as prominent peaks at the intersections of the natural and contact eccentricity bands in either vibration profiles in both runs, refer to Figure 8.5 and Figure 8.6. Thus for example, moving along a natural eccentricity band, the regular changes in the signal intensity over cycles occur as the natural eccentricity bands of the two specimens reinforce or attenuate on contact.

In summary, the examination of burst profiles uncovered further effects of contact eccentricity; the use of TSaA data processing techniques (i.e. burst averaging) masks these effects. Further aspects of contact eccentricity are discussed in the electrostatic analysis (Subsection 8.4.3).

8.3.3 Frequency Analysis

Frequency analysis was performed in order to find how the various features identified above are reflected in the frequency-time profiles, and to what extent the EMD method is able to enhance or attenuate them.

The features identified above fall into three main categories:

- i. Eccentricity bands
- ii. Fault-related features (first sight, peaks)
- iii. Background

Features (i) and (iii) are generally constant over the course of the runs (an exception are the box intervals found in Run 2-2), so any pattern in the frequency-time profiles were matched against the fault progression features.

The frequency-time profiles are presented in Figure 8.7 below.

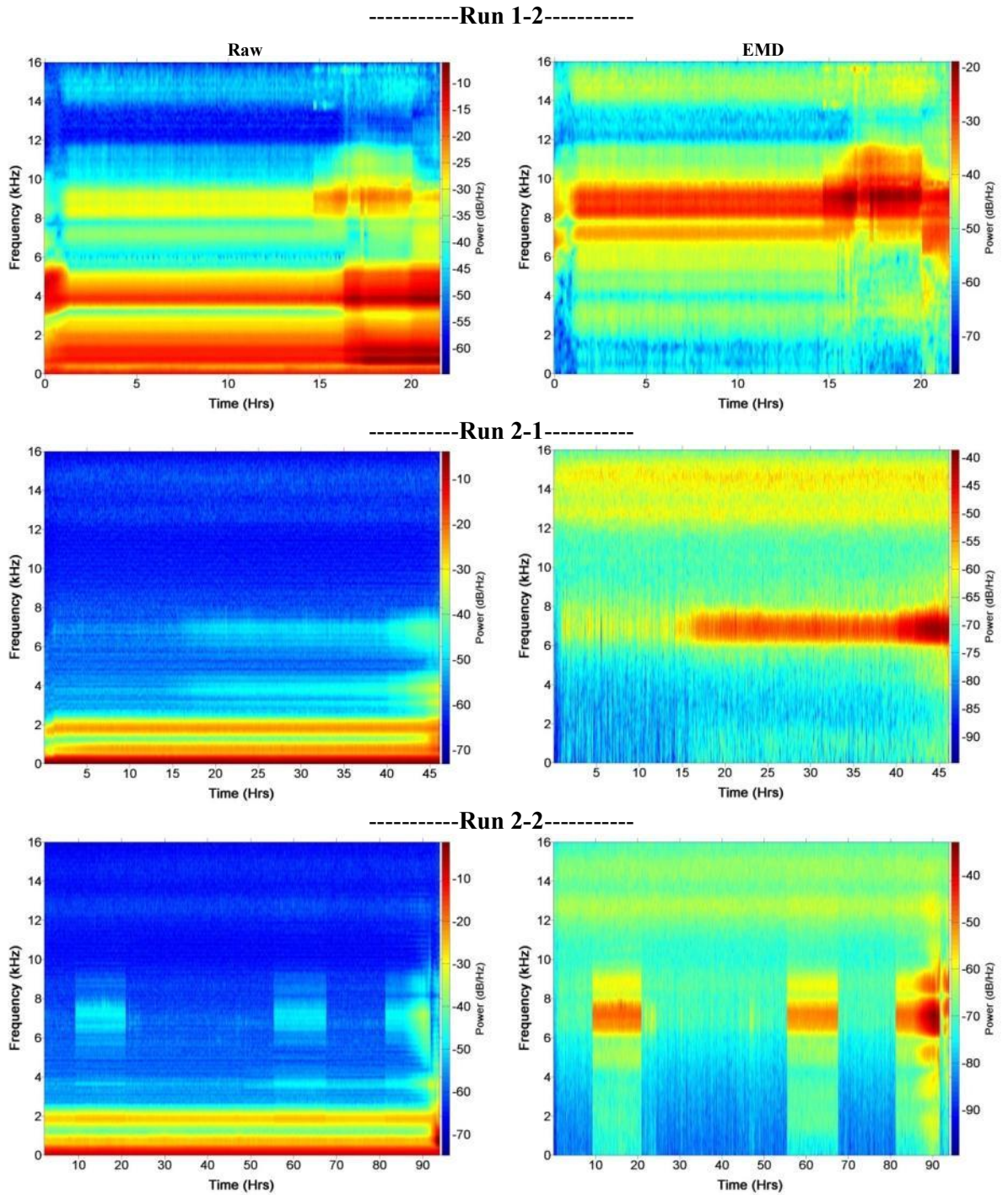


Figure 8.7: Vibration frequency-time profiles for raw-TSaaS (left column) and EMD-TSaaS (right column).

The profiles show distinct horizontal bands of prominent frequencies. Their presence and relative strengths differ from run to run; some additional frequencies were found above those in the Healthy Run. Table 8.8 below summarises the frequencies and the presence of the fault-related features in both raw and EMD.

Run	1-2			2-1			2-2		
Resonant Frequency (kHz)	Visible in raw	Fault features (Raw) (hrs)	Fault features (EMD) (hrs)	Visible in raw	Fault features (Raw) (hrs)	Fault features (EMD) (hrs)	Visible in raw	Fault features (Raw) (hrs)	Fault features (EMD) (hrs)
1*	√	16.8-EoR	-	√	45-EoR	-	√	92-EoR	90-EoR
2	x	-	-	√	45-EoR	-	√	92-EoR	88-EoR
4*	√	16.8-EoR	-	√	15-EoR	-	√	80-EoR	85-EoR
7	x	-	-	√	15-EoR	14-EoR	√	80-EoR	80-EoR
9*	√	14.5-20	14.5-20	x	-	-	√	80-EoR	80-EoR
10.5*	√	14.5-20	14.5-20	x	-	-	√	-	88-EoR
12.5*	x	-	-	√ (faint)	-	-	√ (faint)	-	85-EoR
14.5*	√	14.5-20 (faint)	14.5-20 (faint)	√ (faint)	-	-	√ (faint)	-	-

Table 8.8: Comparison between resonant frequency bands (*indicate the resonant frequencies found in the Healthy Run)

The table indicates that the middle-range frequencies (7 – 10.5 kHz) reflect the fault-related features more clearly and earlier than the extreme frequencies. Correspondingly, the EMD method (which emphasises frequencies above 6 kHz) shows the features more distinctly than the raw profiles over most of the runs. In the End periods (especially Runs 1-2 and 2-2), the middle-range frequencies show prominent drop offs, indicating that at a critical size, the dominant faults excite the lower resonant frequencies. Correspondingly, the EMD profile was unable to track this shift. This also explains the EMD drop offs at the end of the run that can be seen in both signal-time and angular-time profiles.

The box profiles are clearly identifiable in Run 2-2 as prominent vertical bands corresponding to the resonant frequencies mostly in the middle range. Even in the frequency profiles, they stand out as being different from normal background and fault-related features.

8.3.4 Vibration Results Summary

The analyses of the vibration profiles identified a number of features critical to the monitoring of fault developments via vibration sensing techniques:

Vibration profiles are dominated by eccentricity bands, caused by varying radial distances over the contact surfaces of both specimens. They are responsible for several distinct features.

- Prominent bands at or around a fixed angular locations ('natural eccentricity bands') caused by the specimens' own eccentricities.
- Regular and intermittent diagonal bands spanning the entire surfaces (0-360°) caused by contact with the counter specimen ('contact eccentricity bands'). Different patterns result from resonant and non-resonant rotational frequencies of the two specimens. As they are spread over the entire surface, they can contribute substantially to increased background signal levels.

Faults contribute to the vibration profiles by changing signal levels at or around their angular locations. Because of a generally noisy background, their first detection ('first sight') occurs well into the run, but after that they remain visible throughout the rest of the run. Their progression is not smooth, but exhibits sudden peaks. These are associated with direct contacts (without intervening oil) between the two specimens, evidenced by vibration increases in the profiles synchronised with the steel specimen rotations. The End period typically exhibits a final drop off from the mid-range to lower frequencies, signifying that, at a critical size, the dominant faults excite the lower resonant frequencies.

Both CRC and Flaking faults produce detectable vibration signals once they reach a particular stage in their development. Two runs (1-2 and 2-2) were dominated entirely by a single fault, Run 2-1 showed contributions of two faults of different type and size. Flaking produces significantly more prominent vibration signatures than CRC.

The EMD method was able to enhance all the features of the dominant faults and significantly suppress the eccentricity signatures, despite the dominant frequencies having been found to be different from those identified in the Healthy Run, and on which the method was designed. The only potential problem occurs with the final drop off of the fault-generated signal to lower frequencies and thus to the extreme of the range of the selected IMF profile.

An unusual feature was found in Run 2-2, showing sudden jumps to higher levels, constant over three long periods, each lasting 11-12 hrs. Their presence registered in all the resonant frequencies for the run, appearing consistently at precisely 180° apart, twice per rotation. This synchronicity indicates therefore an internal, machine based source. They present a caution that sensor based detection requires a careful attention to all machine (and environmental) parts⁴¹

⁴¹ A failure in the bearings, supporting the shaft which coupled the pulley and upper drive shaft, occurred in a test following Run 2-2 (refer to Figure 4.2). The 'box' profiles could have therefore been the first signs of the impending failure.

8.4 Electrostatic Sensing

This section presents a detailed analysis of the electrostatic profiles. Firstly, the angular-time profiles are investigated to discern the contributions of individual faults and deduce the various events in their progression as guided by the fault progression model and the results of the visual inspection analysis (Chapter 7). Selected burst data are examined for more detail. Finally, frequency-time profiles are analysed to obtain additional information on the findings.

The angular-time profiles are presented below for WSS1 (silicon nitride) in Figure 8.8; the final 20 hrs of Run 2-2 are presented in Figure 8.9 for improved resolution. The fault locations are indicated in each of the figures.

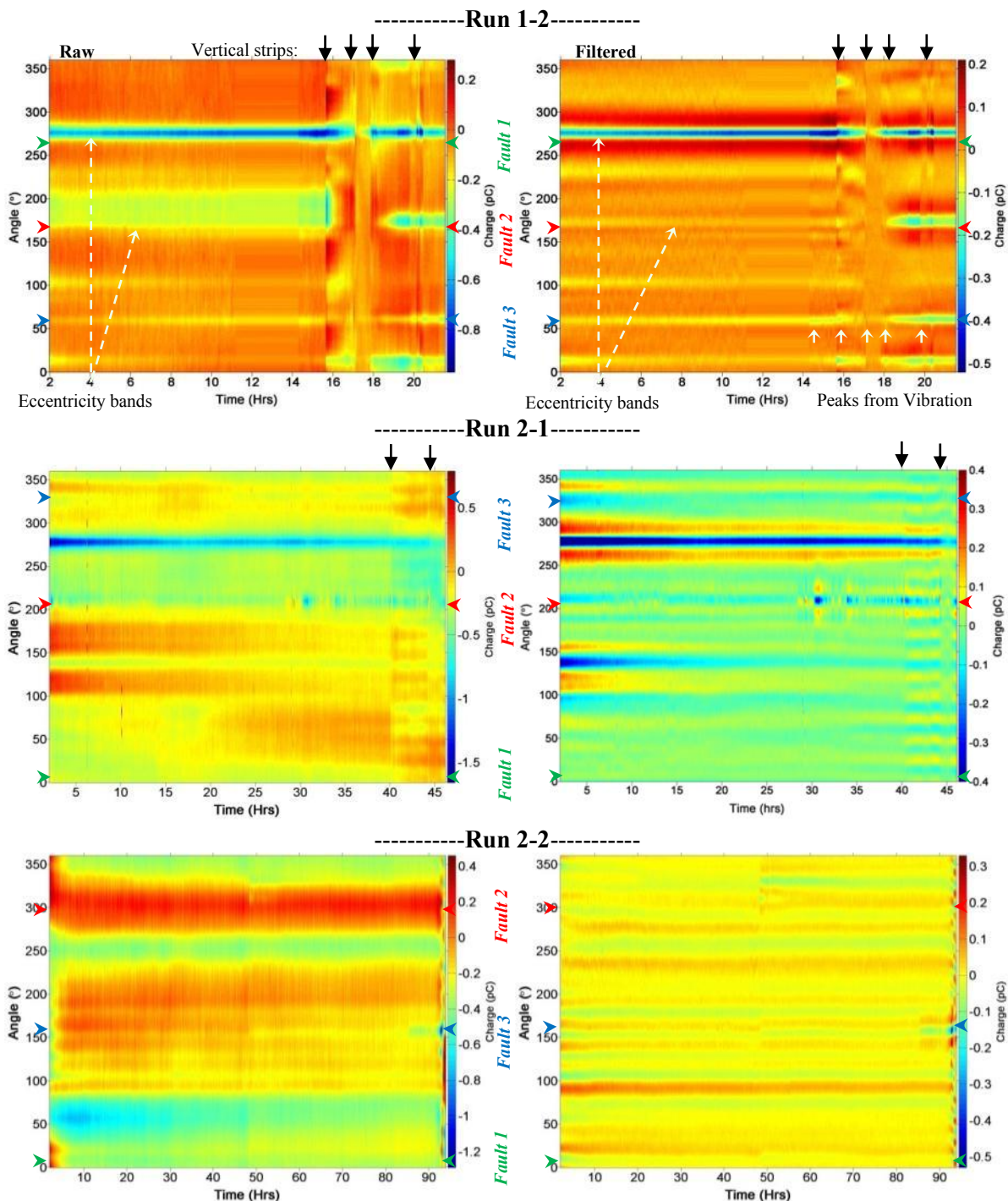


Figure 8.8: WSS1 angular-time profiles for TSaA-raw (left column) and TSaA-filtered (right column).

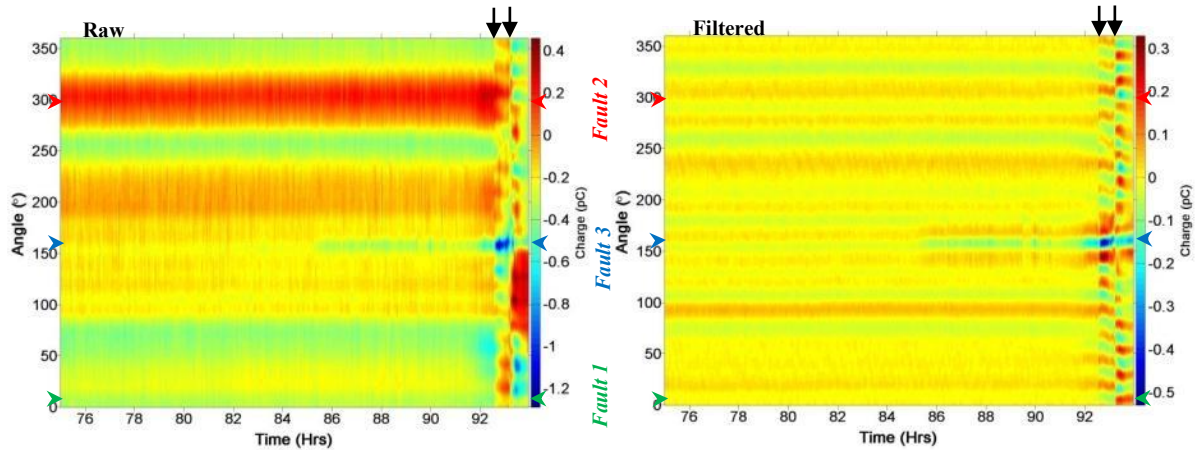


Figure 8.9: WSS1 angular-time profiles for the last 20 hrs of Run 2-2 for TSaA-raw (left column) and TSaA-filtered (right column).

8.4.1 Eccentricity Bands

Similar to vibration, the raw WSS1 profiles are dominated by horizontal bands representing both polarities. The following table notes their locations and whether they match those found in vibration.

Run	Angle (°)	Vibration detected	Raw WSS1 Prominent/Faint
1-2	15	-	√
	100	(faint)	(faint)
	175-200	√	√
	220	√	√
	275	√	√
2-1	17	√	(faint)
	70-100	√	(faint)
	140	-	√
	200-260	√	(faint)
	275	-	√
2-2	50-90	√	(faint)
	110	(faint)	√
	150		√
	180-250	√	√
	290-320	(faint)	√
	350	(faint)	√

Table 8.9: Eccentricity band locations as identified in the raw WSS1 angular-time profiles.

The table shows that all of the vibration eccentricity profiles are reflected in the WSS1 signals and that additional bands occur that have not been identified in vibration.

The bands can be produced from a uniform charge distribution on the surface by the variation in the distance between the specimen and the sensor head, pronounced at the eccentricity points⁴²⁴³. This is because electrostatic charge detected by the sensor is related to the distance between the surface of the specimen and the sensor surface by the relationship $\frac{1}{distance^2}$ [125]. For a

⁴² This effect was observed by Booth et al. [125]

⁴³ The good agreement between the vibration and the electrostatic bands points to the specimen eccentricity as the cause rather than, for example, surface waviness.

highly eccentric specimen, a significant distortion from the sinusoidal pattern occurs at the points of minimum distance, resulting in pronounced measured peaks.

In addition, charge changes can be induced at these points due varying pressure and sliding levels, activating the following mechanisms:

- varying levels of triboemissions, particularly enhanced in silicon nitride as demonstrated by Nakayama et al [134-137].
- triboelectrification, assisted by the variation in oil thickness (i.e. distance between the two materials), found to be more significant in insulator-conductive material combinations than in conductive-conductive pairs [125].

The majority of these bands appear less distinct after filtering (note the reduced charge scales). The fact that some of these cannot be eliminated altogether indicates that their actual shape differs from a (near-) perfect sinusoid (note especially 275° in Run 1-2 and 275° in Run 2-1).

8.4.1.1 *Bands Related to Defects and Faults*

It is expected that a portion of the charge created by a fault will remain at its location due to the non-conductivity of the silicon nitride (some charge will be removed e.g. by oil) and therefore will produce a signature at its corresponding angular location. Of particular interest is how the charge profile patterns reflect the various features identified in the vibration profiles (i.e. first sight, peaks and tail-offs).

Examination of the above profiles (Figure 8.8&Figure 8.9) confirms that there are indeed fault-related features such as those associated with first sight and peak points. Some of the peak points are accompanied by vertical strips covering part or most of the angular range (i.e. the specimen surface). In addition, there appear vertical strips which are difficult to associate with any particular fault. These features are summarised in Table 8.10 below.

The data in the table is based on the examination of the TSaA-filtered profiles (right columns of Figure 8.8&Figure 8.9). Those that are fault-related are matched against the vibration events as summarised in Table 8.6 and, in addition, non-fault related features (strips) are also noted; all these are assigned event labels for subsequent reference. One must bear in mind however, that these features are picked out from a noisy background and are therefore approximate.

Run	Time (hrs)	Fault-related				Strip	Event label
		Fault	Vibration	Event	Strip		
1-2	Start of Run	3 (60°)	-	First sight	-		3A
	14.4	3 (60°)	Peak (a)	Peak	-		3B
	15.8	3 (60°)	Peak (b)	Peak			3C
	15.8	2 (161°)	-	First sight	√		2A
	15.8-17.2					√	
	16.8-17.2	3 (60°)	Peak (c)	-	-		3D
	17.2-17.8		-	-		√	
	17.8	2 (161°)	-	Peak	-		3E
	17.8-18	3 (60°)	Peak (d)	Peak	-		
	20-20.5	3 (60°)	Peak (e)	Peak	√		3F
	20.5 – end of run					√	3G
2-1	15	3 (328°)	First sight	-	-		3A
	21	3 (328°)	Peak (a)	First sight	-		3B
	30	2 (208°)	-	First sight	-		2A
	34 - 40	2 (208°)	-	Peaks	-		2B
	40	2 (208°)	-	Peak	√		2C
	41	3 (328°)	Peak (b)	-	-		3C
	44	2 (208°)	First sight	visible	-		2D
	44.5- 45.5	-	-	-	-	√	2E, 3D
	end of run	3 (328°)	Peak (c)	-	-		3E
2-2	48	3 (162°)	-	First sight	√		3A
	85	3 (162°)	Peak (a)	Peak	-		3B
	90	3 (162°)	Peak (b)	Peak	-		3C
	91.5	3 (162°)	Peak (c)	-	-		3D
	92.25 – 93.25	3 (162°)	-	Peak		√	3E
	93.25	2 (301°)	First sight	First sight?	-	√	2A
	93.25- end of run	3 (162°)	Peak (d)	Peak	-		2B, 3F

Table 8.10: Summary of WSS1 fault-relevant features (event labels are referred to in section 8.5).

The smallest faults (Faults 1) are not detectable in any of the profiles. The remaining faults are discernible in all runs.

A further comparison between the fault feature detected in electrostatic sensing and those found in vibration reveals the following correspondence:

Two of the peaks that have been identified in the vibration profiles are absent in the WSS1 profiles: Peak (b) of Fault 3 in Run 2-1 and Peak (c) of Fault 3 in Run 2-2. Peaks (c) and (d) of Fault 3 in Run 1-2 are masked by a prominent strip extending more than an hour.

In contrast, several additional peaks appear in charge (right columns of Figure 8.8&Figure 8.9) that have not been identified in vibration, potentially providing more information on the nature of the fault and indicating that some faults, notably CRC, may be more visible in WSS than in vibration. This is illustrated in Table 8.11 below.

Run	Fault	WSS1 Time (hrs)	Vibration Time (hrs)	Start Fault Type	End Fault Type
1-2	3 (60°)	0	0	Flaking (a)	Flaking (d)
	2 (161°)	15.8	14.4	Flaking (a)	Flaking (c)
2-1	3 (328°)	15	15	Flaking (a)	Flaking (d)
	2 (208°)	21	44	Indefinite	CRC (c)
2-2	3 (162°)	48	80	Indefinite	CRC (d)
	2 (301°)	93	93.25	Indefinite	Flaking (c)

Table 8.11: Comparisons of the first sight points between WSS1 and vibration.

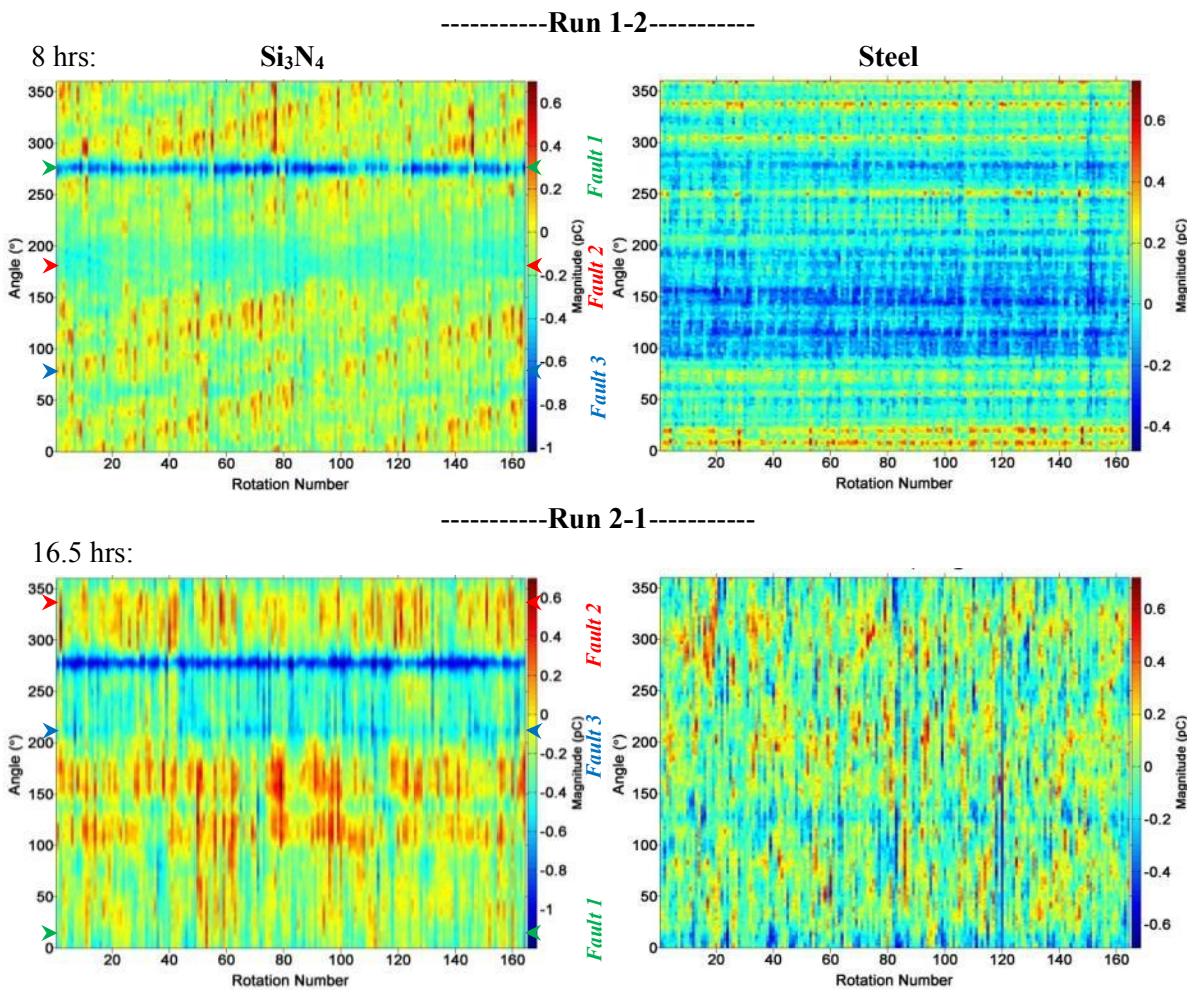
The higher visibility of CRC faults in charge profiles than in vibration is due to their extensive structures (producing charge), but near-smooth surface profiles. The dominant charge-generating mechanism is fractoemissions, found to be especially prominent in silicon nitride (see reference [138]).

Comparisons between the diagrams in the left and right columns of Figure 8.8 and Figure 8.9 show that the fault features and vertical strips are enhanced in the filtered profiles. The vertical strips are analysed in detail in subsection 8.4.4 below.

In order to find more details about the profiles, selected bursts are now considered as in the vibration analysis above.

8.4.2 Raw Burst Profiles- Contact Eccentricity Bands

In order to discern the effects of the contact eccentricity points, burst data were analysed at the ‘quiescent’ periods identified in the vibration analysis. These are presented below for the WSS1 profiles timed against both specimens’ tachometers.



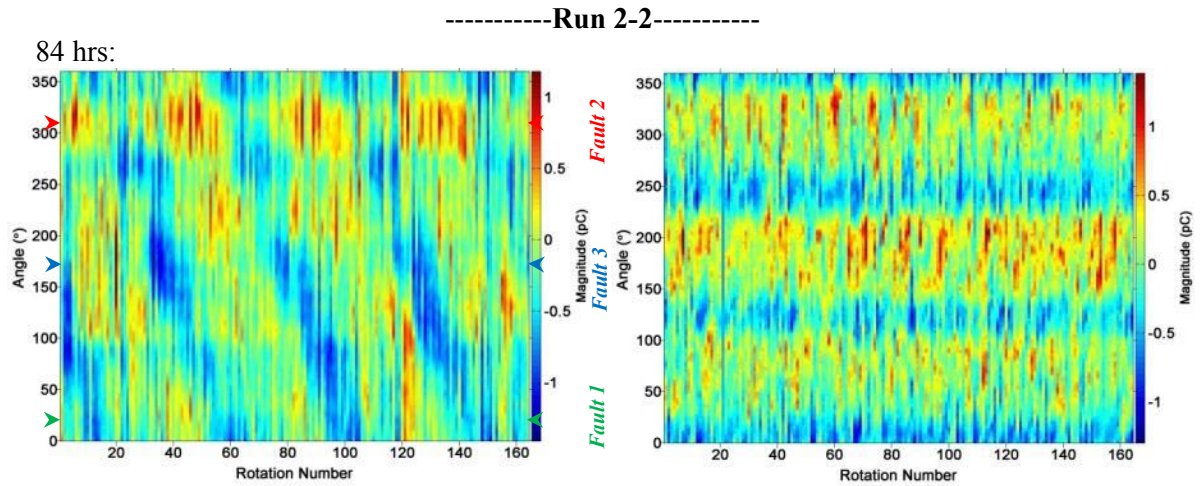


Figure 8.10: WSS1 angular-time profiles for raw burst data timed against the silicon nitride (left column) and steel (right column) tachometers.

The profiles are very noisy, nevertheless diagonal patterns can be seen in the left column with the slopes corresponding to those seen in the vibration burst data in Figure 8.4, Figure 8.5 and Figure 8.6. These diagonal patterns are straightened (into horizontal bands) when re-timed against the steel tachometer (right column), thus indicating their origin as steel surface patterns. Because there cannot be any enduring significant charge residues at the surface of the conducting (and earthed) steel specimen, this points to mechanical features i.e. eccentricity points, and thus provides evidence that these points invoke charge changes on the surface of the silicon nitride.

Unlike in the vibration profiles, a similar analysis of the steel surface charge does not yield any contact eccentricity patterns indicating that the charge induced there is minimal and/or is conducted away. This is illustrated in Figure 8.11 below taken from Run 2-2 (notice the very low charge values).

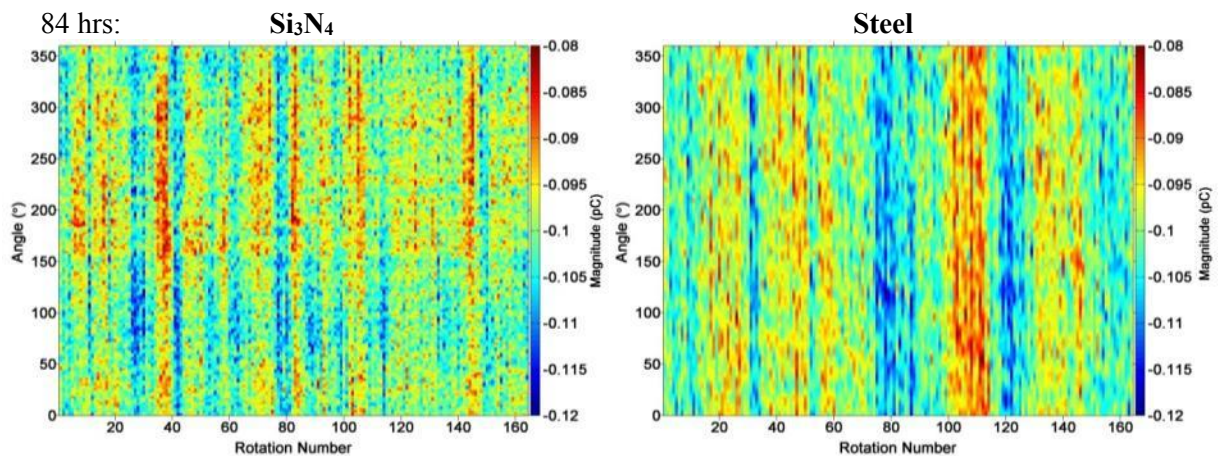
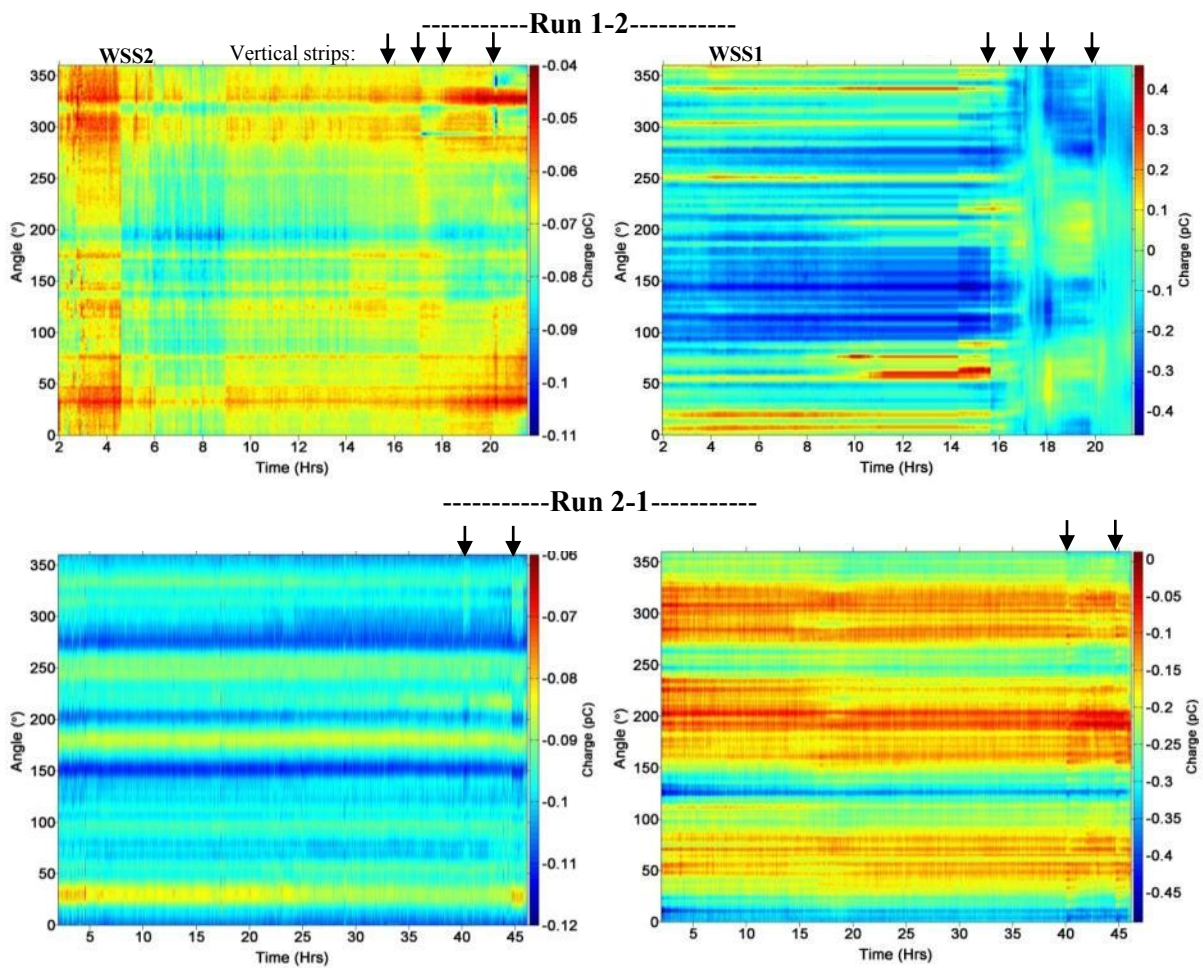


Figure 8.11: WSS2 angular-time profiles for raw burst data of Run 2-2 timed against the steel (left column) and silicon nitride (right column) tachometers.

8.4.3 Vertical Strips

Vertical strips indicate that there is a sudden change in charge intensity affecting a major part or entire surface of the specimens. Their detailed examination shows a variety of different patterns indicating that there are several mechanisms in play. Run 1-2 shows the most diverse vertical strip types⁴⁴.

All of the vertical strips are due to the interplay between the two specimens: to illustrate that the steel counterface has a role in the strips seen in the silicon nitride, both silicon nitride and WSS profiles are timed against the steel tachometer and presented in Figure 8.12 below.



⁴⁴ As the main focus of the project is on detection of fault features, the initial vertical strip appearing between 0 and 4.5 hrs in Run 1-2, attributable to run-in, is ignored.

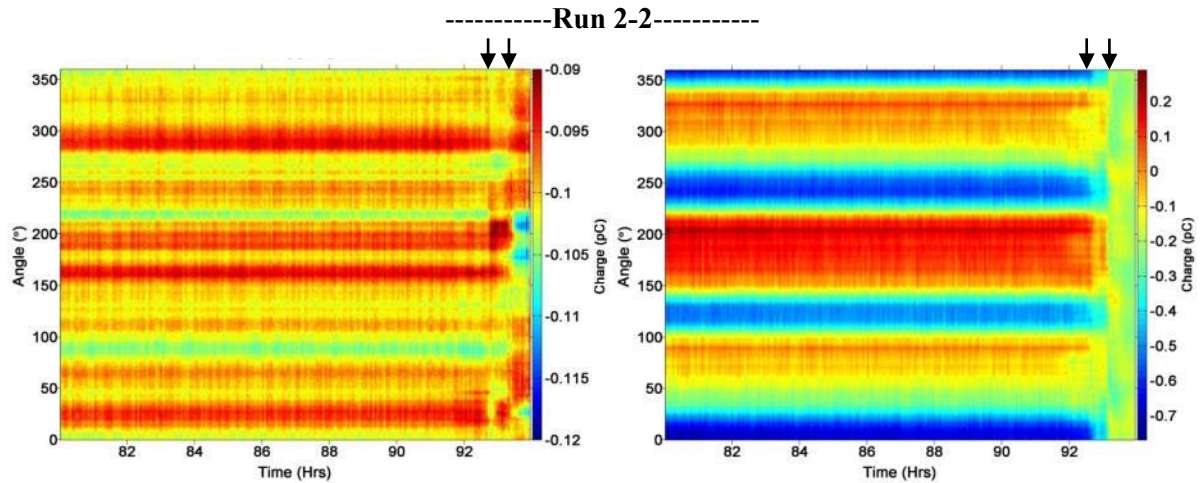


Figure 8.12: WSSs angular-time profiles for TSaA-raw timed against the steel tachometer showing WSS2 (left column) and WSS1 (right column).

All of the strips identified in the WSS1 profiles (see Table 8.10) appear in both of the above profiles. However, they are generally more prominent in the right column as both the charge magnitudes and duration over which it decays are significantly higher than on the steel specimen.

Steel can influence the charge levels in silicon nitride by at least two mechanisms: earthing and debris attachment. These are now considered in detail.

8.4.3.1 Earthing

This process is initiated when a fault makes a direct contact with the steel causing substantial impressions. Both CRC and Flaking faults were shown in subsection 8.3.1.2 above to be capable of achieving this. The roughened surface makes a direct contact with the silicon nitride specimen on subsequent rotations, repeating the earthing effect there (for an illustration of this process see Figure 8.18 below). If the rotational speeds of the two specimens are not in resonance, the direct contact rapidly covers the whole silicon nitride surface. The effect stops when the roughened surface is smoothed and the intervening oil layer is re-established, restoring also the charge to the previous levels.

The evidence for this is shown in the WSS1 profile of Run 1-2 between 17.2 – 17.8 hrs, seen in Figure 8.8 as a uniform band of near-zero charge spanning the entire silicon nitride surface. The band is also clearly visible when WSS1 is timed against the steel tachometer, and also faintly on WSS2, despite it being noisy with extremely low charge values- see Figure 8.12 above. Corresponding WSS1 burst data is shown below in Figure 8.13, demonstrating that most of the surface is already at a near zero charge level.

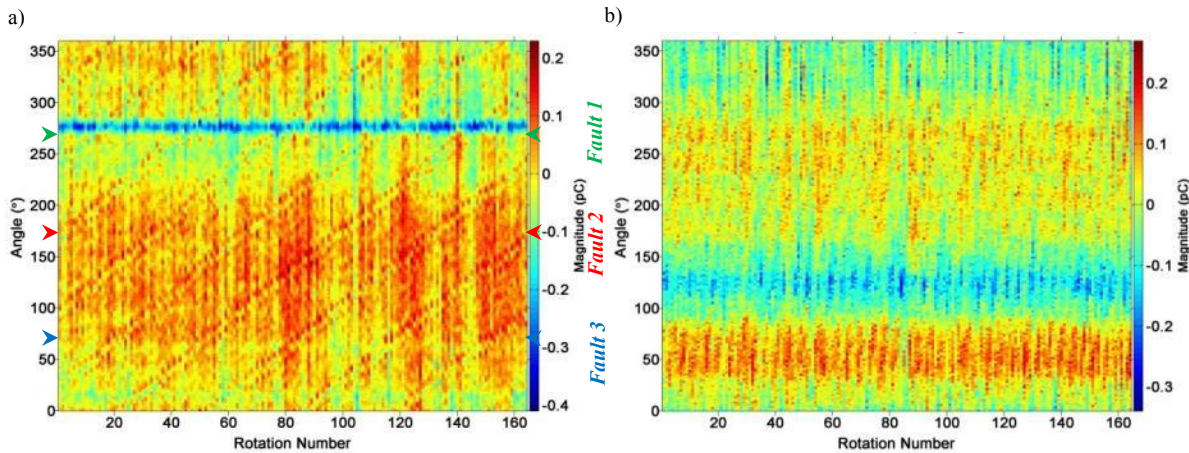


Figure 8.13: WSS1 burst profile at 17.7 hrs of Run 1-2 timed against a) silicon nitride, b) steel.

The initial direct contact occurred at the points coinciding with the vibration –EMD peaks (c) & (d) seen in Figure 8.1, and with the strips in vibration timed against the steel tachometer in Figure 8.3.

The direct contact between the two specimens can also explain the sudden increase in the charge intensity at Fault 2 location at 18 hrs, seen in both raw-TSaA and filtered-TSaA (see Figure 8.8), i.e. immediately following the cessation of the earthing effect. An explanation for this is that in the course of the direct contacts the fault becomes filled with steel debris. The presence of the two different materials, aided by the return of the oil into the contact, will form a contact charge between them thus adding to the charge level at the fault location. The evidence for the debris was found on the silicon nitride specimen in a visual inspection post-run (before cleaning the sample) and is shown in the figure below.

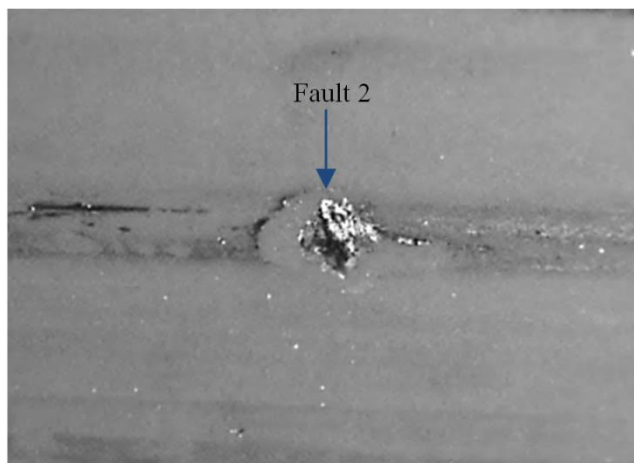


Figure 8.14: A photograph showing Fault 2 post-Run 1-2 before cleaning.

8.4.3.2 *Debris Attachment*

This effect is initiated when a fault produces debris that becomes attached to the steel specimen and induces charge changes on the silicon nitride specimen on subsequent rotations of the steel. As with the earthing contacts above, the precise effect will depend on the relative speeds of the two specimens. The effect ceases when the charged debris is removed by the oil and/ or centrifugal force. This could happen after only a few rotations of the steel, however the effect will remain on the non-conductive silicon nitride for much longer.

The evidence for this appears in the filtered WSS1 profile at 40 hrs in Run 2-1 (Figure 8.8 above, right column). Because the two rotational speeds are in resonance (44 rotations of silicon nitride = 15 rotations of steel, as discussed in Subsection 8.3.2 above), the effect is seen as a series of horizontal bands: 15 are seen in the ‘normal’ (i.e. timed against its own tachometer) WSS1 profile (Figure 8.8 above) and 44 are seen in the cross-timed WSS1 profile in Figure 8.12.

A more complex evidence for the operation of this mechanism is presented in Run 2-2 strips 92.25-93.25 hrs and 93.25- end of run (refer to Figure 8.9 right column). This run has the same resonance as Run 2-1 and correspondingly there are 15 bands seen in WSS1 profile, however, they appear to shift down gradually over the first strip and, at 93.25 hrs, drop and then begin to widen in the directions away from the feeding fault (Fault 3- 162°).

The fault responsible for this effect is a very large CRC fault (2,710 μm , i.e. 15° long) with a complex structure consisting of multiple rings indicating a staged progression and including a small spall.

The progressive shifts in the WSS horizontal bands occur when new crack substructures open and deposit charge at slightly different angular locations. As discussed in the previous chapter, for CRC faults this is predominantly in the leading direction. This growth occurs over the first vertical strip (92.25 -93.25 hrs).

At 93.25 hrs, coinciding with peak (d) in vibration, the spall forms ejecting debris which is subsequently smeared onto the steel by traction. Furthermore, this results in the slowdown of the relative rotations of the two specimens. Therefore, subsequent contacts will occur at longer time intervals, producing the widening of the horizontal bands away from the originating fault. This is substantiated in the figures below showing the charge smeared on the steel (WSS2, left hand side) and the slowdown of the steel rotating frequency (right hand side).

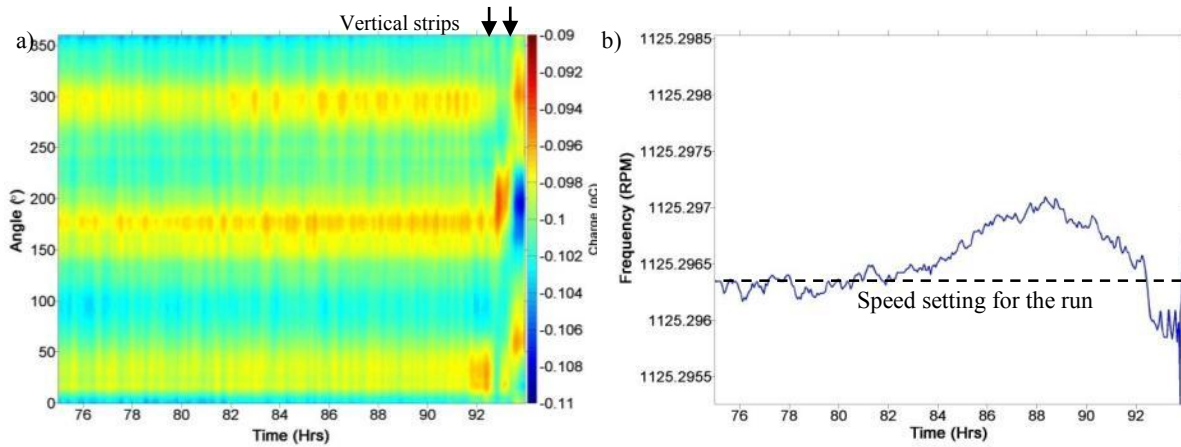


Figure 8.15: a) WSS2 angular-time profile for Run 2-2 from 75 hrs to end of run b) Steel specimen rotating frequency.

8.4.4 Frequency Analysis

As for the vibration analyses, frequency-time profiles were examined to see how the various WSS features identified above are reflected in them, and to what extent the filtering method is able to enhance or attenuate them. These were obtained by computing the PSD estimates for each burst using the Welch method. The profiles are presented in Figure 8.16 & Figure 8.17 below.

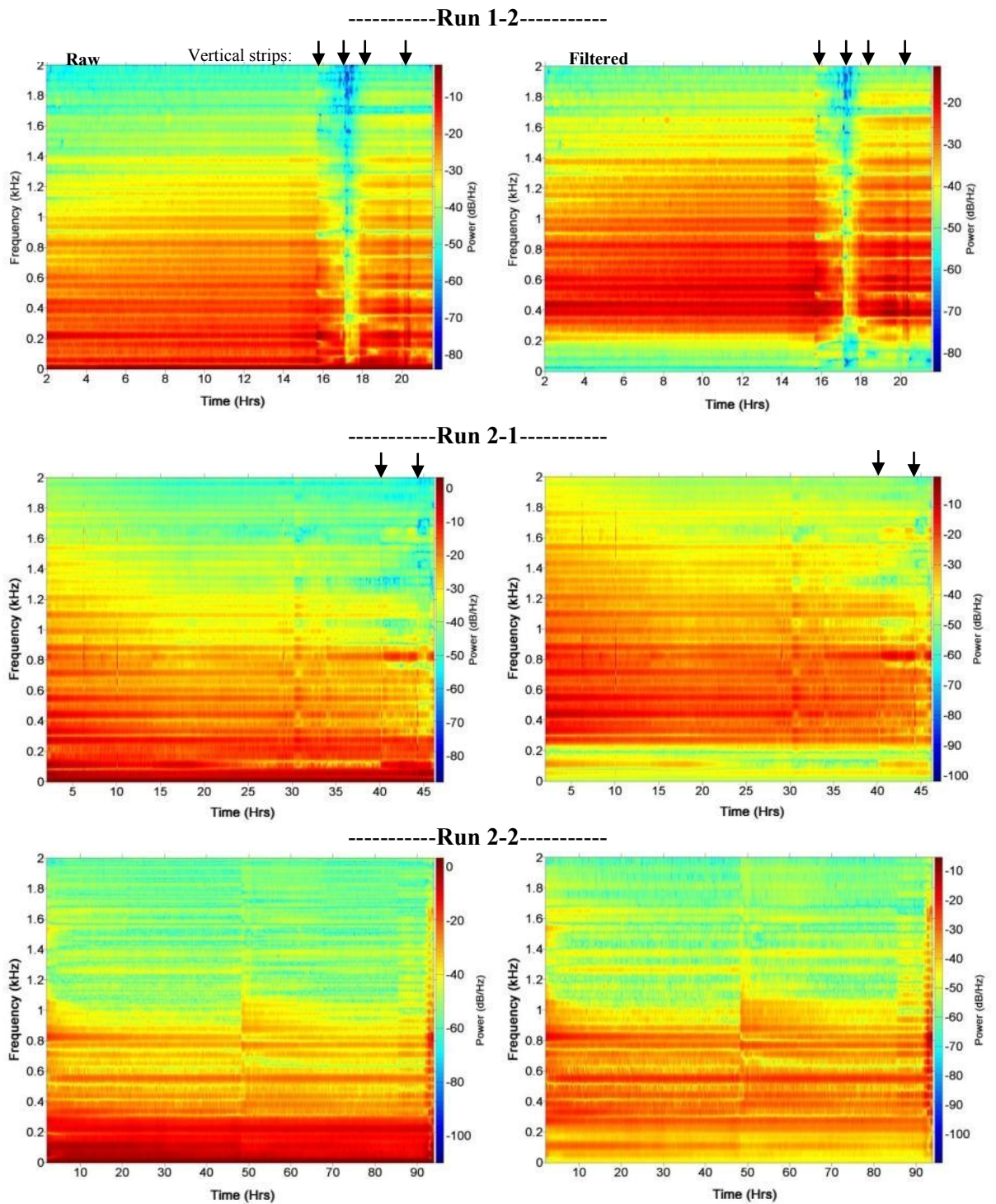


Figure 8.16: WSS1 frequency-time profiles pre- (left column) and post- filtered (right column).

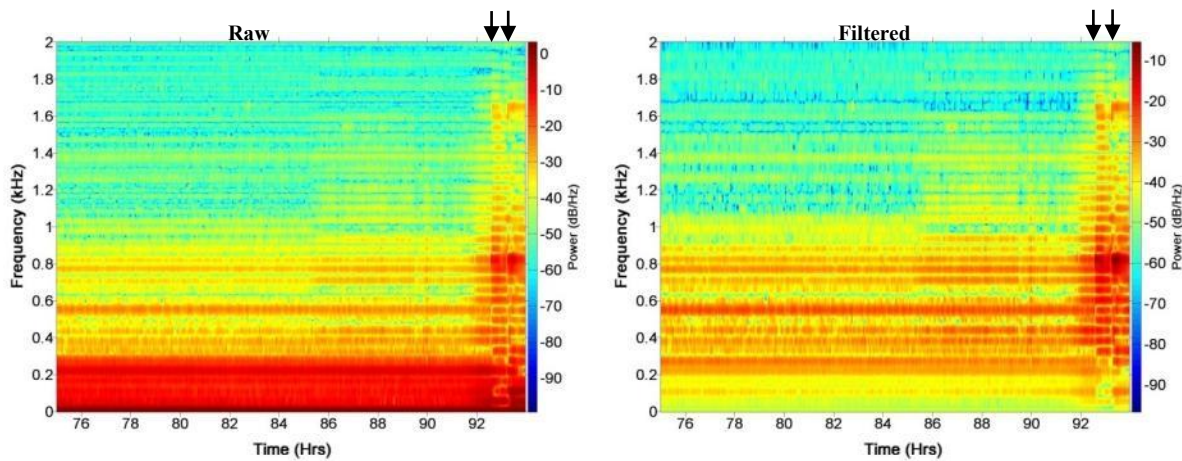


Figure 8.17: WSS1 frequency-time profiles for the last 20 hrs of Run 2-2 pre- (left column) and post- filtered (right column).

The above Figures show that the spectra are dominated by eccentricity bands therefore presenting a noisy background to any fault signatures. The situation is partly improved by filtering as seen by the increased sensitivity scale of approximately 10 dB.

A fault generates a broad profile covering a wide band of frequencies with a diminishing intensity at the higher end; increasing fault size reduces the broadness of the profile i.e. higher frequencies are reduced (see Appendix D). Superimposed on this is a series of horizontal bands generated at harmonics of the rotational frequency of the specimen combined with the influences of the steel such as contact eccentricity effects.

Significant fault events, such as those associated with rapid growths, therefore appear as decreases in the power levels across the higher range of frequencies. This effect is seen for example in the profile of Run 2-2 at 48 hrs, coinciding with the first sight of Fault 3 (refer to Figure 8.16, either column).

Those vertical bands in WSS1 that are associated with sudden increases in faults are reflected similarly. An example of this is a strip at 14.4 hrs in Run 1-2 corresponding to Peak (a) in vibration (peaks are identified e.g. in Table 8.6).

The fault events that involve a direct interaction between the two specimens and resulting in charge transfer are explained as follows:

As discussed in subsection 8.4.3.2, the fault produces, at a critical point, charged debris which attaches to the steel specimen (at a ‘primary transfer point’).

The primary transfer point comes into contact with the ceramic specimen at subsequent rotations (‘secondary transfer points’), producing a change in the charge level at that location of the ceramic. With rapid on-going rotations this effect will spread quickly over the entire

circumferences of both specimens. The actions of the mechanism are illustrated in the diagram below.

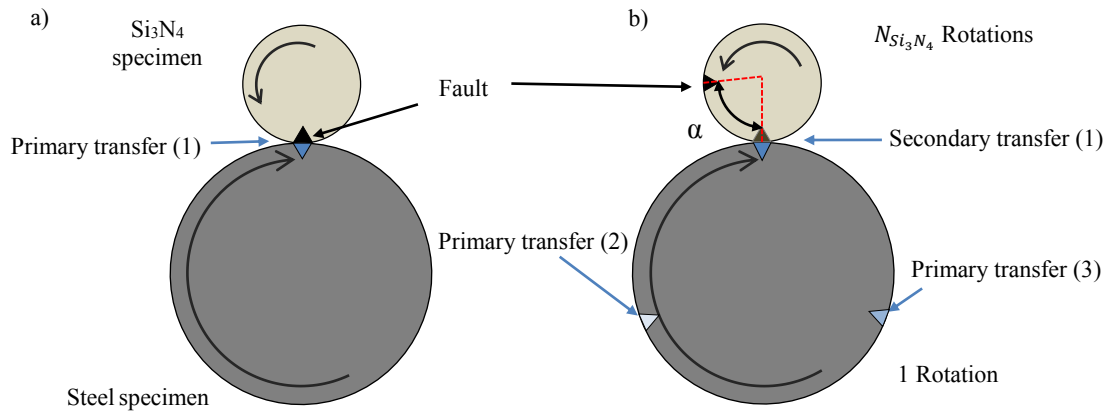


Figure 8.18: Illustration of the charge smearing effect on the silicon nitride via the steel specimen a) primary transfer and b) consecutive secondary transfer.

The charge changes at these points therefore present themselves as variations in WSS1 intensity at the rotational frequency of the primary and secondary transfers and its harmonics.

For example in the Runs 2-1 and 2-2, it was calculated above that 15 transfer points appear per silicon nitride rotation, producing a base rotational frequency of $15 \times 55 \text{ Hz} = 0.825 \text{ kHz}$. As discussed above, the charge transfers occur at 41 hrs in Run 2-1 and between 92.25 hrs and end of run in Run 2-2. The frequency-time profiles do indeed show significant increases at 0.825 kHz at these times. In fact, both of these are preceded by less significant increases (nevertheless still prominent) thus indicating that the charge transfer was starting earlier. These preceding increases are more prominent in the filtered profiles.

For Run 1-2 as calculated earlier, transfer points appear at 87.48° in the leading direction from the previous contact, producing $360/87.48 = 4.098$ transfer points per silicon nitride rotation and a corresponding base rotational frequency of $4.098 \times 55 \text{ Hz} = 0.2254 \text{ kHz}$. A prominent increase in the intensity at this frequency is seen, for example, at 15.8 hrs. The effect is reduced by the filtering method, as its focus is on the higher frequencies (it remains visible in the filtered angular-time profile in Figure 8.8 as a result of its effect in the harmonic frequencies).

The direct contact producing earthing is seen as a significant power reduction over all frequencies, therefore easily distinguishable from charge transfer. The mechanism was identified earlier as appearing at 17.2-17.8 hrs in Run 1-2 and is seen as a prominent band of low intensities in Figure 8.17 (in both unfiltered and filtered profiles).

8.4.5 Summary

The analyses of the charge profiles identified a number of features critical to the monitoring of fault developments via the electrostatic sensing techniques:

As for vibration, the profiles are dominated by eccentricity effects, however to a much greater extent. Therefore, the intensity of the non-fault related background is much higher in electrostatic than in vibration. The eccentricity bands are only partially attenuated by the filtering method as their profiles are not sinusoidal.

Nevertheless, fault events are detectable; the significant fault development events found in vibrations have been detected in WSS1 including additional peaks. Some faults, especially CRC, are detectable much earlier in electrostatic sensing than in vibration.

A new feature in the WSS angular-time profiles are vertical strips with different intrinsic patterns. They are caused by distinctive interactions between a fault on the silicon nitride and the steel counterface. Two types of such interactions have been indentified.

Earthing, when a direct (oil free) contact between the two specimens occurs resulting in fault impressions on the steel surface and the sinking of the charge from the silicon nitride. The resulting pattern in WSS is produced by the fault producing mulitple impressions, each of which also cause direct contacts at subsequent rotations. The effect ceases when the steel surface is sufficiently smoothed by the direct encounters and the direct contacts no longer occur. The associated WSS angular-time profile is a smooth vertical band of near-zero charge throughout. The drop off happens across the entire frequency spectrum as revealed in the frequency-time profiles.

The second type of the inter-specimen interaction is *Debris Attachment*, occuring when a debris produced by a fault becomes attached to the steel surface and, on subsequent contacts, produces changes in the charge level on the silicon nitride. These remain for some time even after the removal of the debris by oil or centrifugal force due to the non-conductive nature of silicon nitride. The pattern produced in WSS1 depend on the relative rotational speeds of the two specimens. In particular for resonant rotations, distinctive horizontal bands appear in the angular-time profiles. The resulting frequency-time profile is a combination of the density of the charge changes generated on the silicon nitride surface and the rotational frequency of the silicon nitride specimen.

The filtering process was able to enhance the fault-related features, making them better distinguishable from the background and detectable earlier. Some low-frequency dominant

features, for example debris attachment with low rotational frequency, are still detectable after filtering via their harmonic frequencies.

The electrostatic profiles are therefore able to discern the various stages in the development of individual faults to an extent exceeding the sensitivity of vibration. This also demonstrates that the new WSS1 sensor is able to provide good location and frequency resolutions, therefore confirming that its design and construction were successful.

8.5 Fault Histories Deconstruction

A deconstruction of the fault progressions was attempted using the information obtained in the above visual, vibration and electrostatic analyses, and the results are summarised in the tables below.

The tables show a timeline in the horizontal direction divided into separate columns, each presenting a significant event in the progression of a detectable fault. The fault events are presented separately in two horizontal sections: the upper indicates the type of event deduced from the signal profiles and the matching stage of the fault development model introduced in Section 7.4. The lower part presents the signal profile features that provided the basis for the deduction, using the arguments presented in the Sections 8.3 and 8.4 above.

The objective of signal monitoring is to detect the progression of faults through important stages, and is naturally superior to direct inspection of components at specified times, typically after a breakdown. The timelines below provide evidence that a combination of signal monitoring in the vibration and electrostatic spectra can provide sufficient information to deduce a fault progression history with some confidence, backed by suitable models.

Run 1-2		Time (hrs)	Start of Run	2	14.4	15.8	16.8	17.2	17.8	18	20	20.5	21.5	
Fault 3	Even	Label		3A	3B	3C	3D		3E		3F		3G	
	Deconstructed	Event		Flake off	Flake off	Continued flake offs							Direct contact and substantial wear in steel	
						Debris attachment		Imprinting and gouging steel			Debris attachment Imprinting and gouging steel			
	Fault stage	F(a)	F(b)	F(c)	F(d) → Large F(d)								Large F(d)	
	Evidence	Vibration	Visible	Peak (a)		Peaks (b) & (c)		Increased vibration in steel		Peaks (d)	Peak (e)		Increase in raw	
		Charge	Visible	Peak		Strip		Strip (drop off)		Peak	Strip		Drop off	
Fault 2	Even	Label			2A	2B								2C
	Deconstructed	Event				Flake	Flake	Filled with debris						Flake
		Fault stage	F(a)		F(b)	F(c)							F(c)	
	Evidence	Vibration			Visible	Peak								Peak
		Charge				Visible				Visible				

Table 8.12: Deconstruction of the fault developments in Run 1-2

Run 2-1		Time (hrs)	Start of Run	15	21	30	34	40	41	44	44.5	45.5
Fault 3	Event	Label		3A	3B				3C		3D	3E
	Deconstructed	Event		Flake off	Flake off				Flake off		Contact with steel	Flake off
		Fault stage	$F(a)$	$F(b)$	$F(c)$					$F(d)$		$F(d)$
	Evidence	Vibration			Peak (a)	Remains visible			Peak (b)		Increased vibration in steel	Peak (c)
		Charge										Strip
	Fault 2	Event	Label				2A	2B		2C	2D	2E
Deconstructed		Event	Subsurface crack growth			Surface cracks propagate intermittently					Partial contact with the steel	
		Fault stage	$Indeterminate$	$CRC(a) \rightarrow (CRC(b)?)$			$CRC(b) \rightarrow CRC(c)$					$CRC(c)$
Evidence		Vibration									FS	Remains visible
		Charge				Peak	Peaks				Strip	Strip

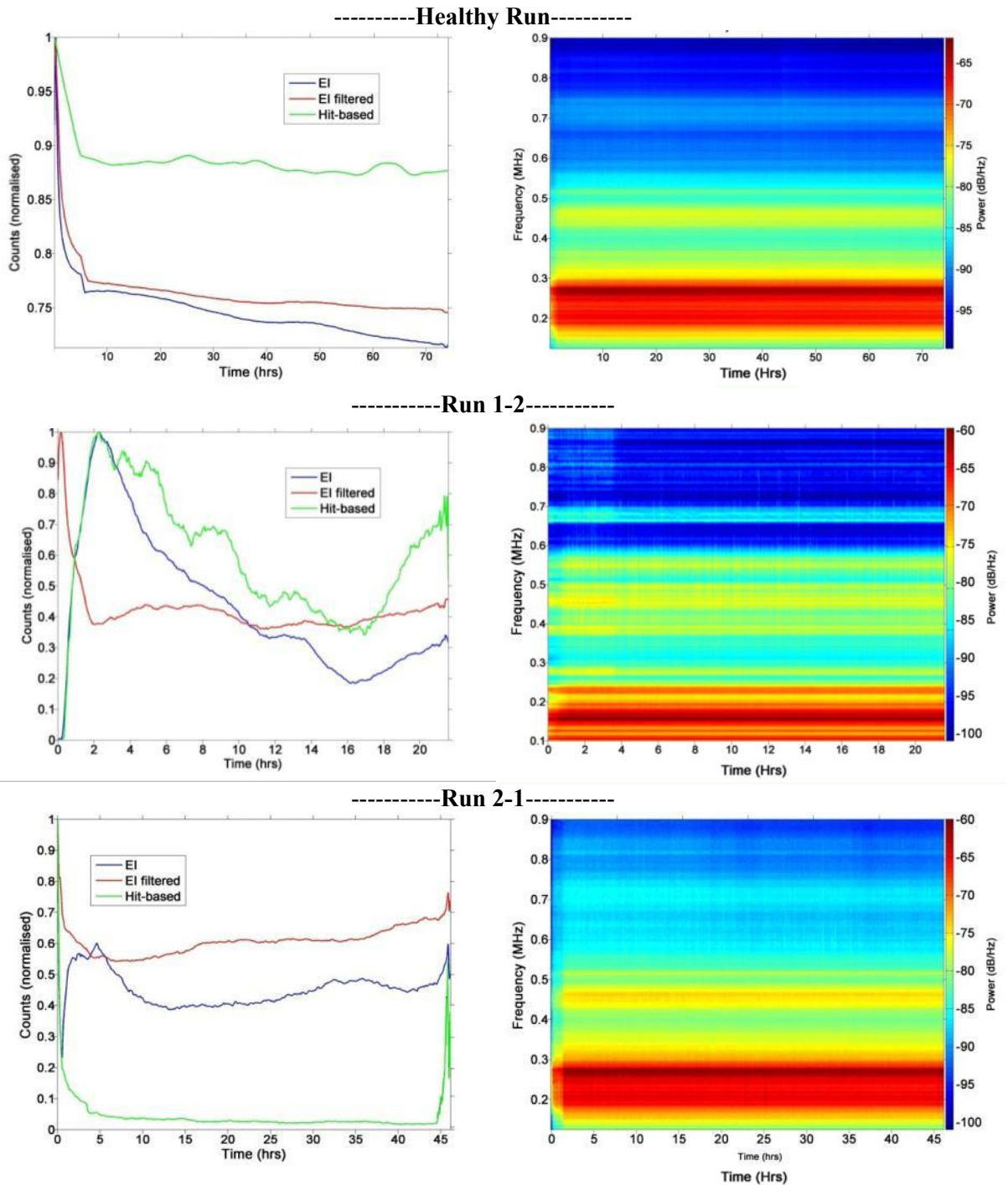
Table 8.13: Deconstruction of the fault developments in Run 2-1

Run 2-2		Time (hrs)	Start of Run	48	85	90	91.5	92.25	93.25	94
Fault 3	Event	Label		3A	3B	3C	3D	3E		
	Deconstructed	Event	Surface cracks propagate intermittently					Debris	Spall. Gouging wear in steel. Debris	
		Fault stage	Indeterminate	CRC(b)	CRC(c)	CRC(c)	CRC(d)			Large CRC(d)
			CRC(a) CRC(b)							
	Evidence	Vibration			Peak (a)		Peak (b)		Peak (c) Increased vibration in steel	
		Charge		Peak	Peaks				Strip	Strip
Fault 2	Event	Label							2A	2B
	Deconstructed	Event							Flaking	Flaking
		Fault stage	Indeterminate	F(a)					F(b)	F(c)
	Evidence	Vibration							FS- remains visible	
		Charge							FS?, Strip	Strip

Table 8.14: Deconstruction of the fault developments in Run 2-2

8.6 Acoustic Emission Sensing

Because the angular-time profiles were not available for AE, its analysis was restricted to signal-time and frequency-time profiles. The former have already been introduced in section 8.2 and are reproduced below in Figure 8.19, together with the frequency-time profiles and both types of profiles from the Healthy Run for comparison.



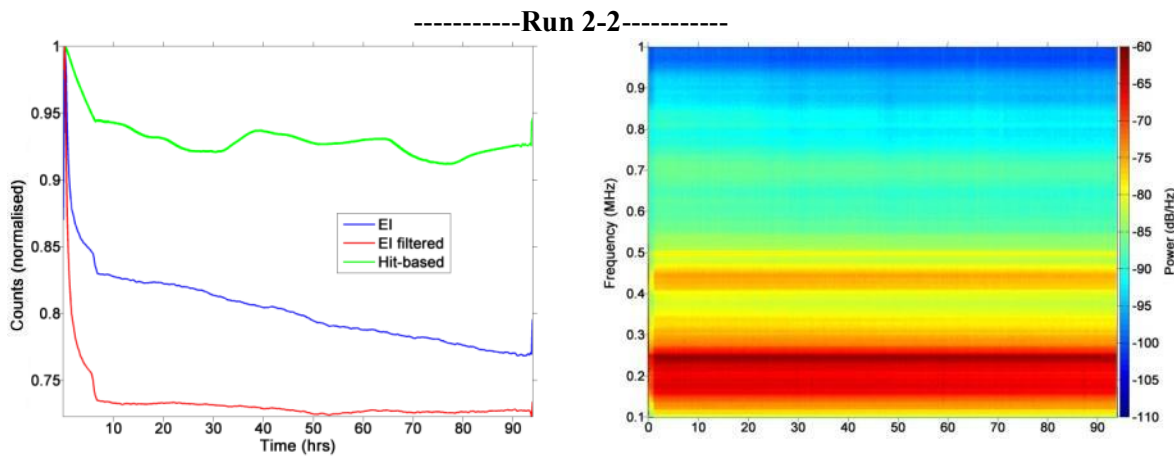


Figure 8.19: AE signal-time (left column) and frequency-time (right column) profiles.

The advanced signal processing methods developed in Chapter 5 were designed to detect fault-related events as signal values above a certain background threshold, set to a level based on the Indentation Tests and Healthy Run. Specifically, a filtering method was developed to enhance the signal features relating to crack propagation in silicon nitride and thus distinguish them from a background whose characteristics were obtained in the Healthy Run. As this method is new, profiles based on conventional processing methods are also presented above. One (Hit-based) is based on a standard threshold-based analysis removing signals below a fixed value (also determined from the results of the Healthy Run). The other (Energy Index, EI) is based on smoothing the signal over burst periods to reduce a stationary background signal and enhance transient events.

Correspondingly, the signal levels in the signal-time profiles should directly correlate with the intensity of fault-related processes in the specimens. Therefore the analysis of the AE profiles focuses on examining correlations between any prominent features of these profiles and the fault events already identified by the vibration and electrostatic analyses above.

Frequency analysis is based on the premise that AE signals generated by progressing faults produce varying frequency patterns as illustrated in section 5.2.3 on the different frequency patterns between the Indentation Tests and Healthy Runs.

By comparing the above profiles with the fault progression histories deduced from the analyses of the vibration and charge spectra, one can immediately draw a number of conclusions.

1. The AE signal-time profiles are generally smooth, much smoother than is expected from the multiple events identified in the vibration and electrostatic analyses.
2. With the exception of the very starts and ends of the runs, the general trends do not seem to follow the intensity of the fault-related activity discerned. For example the

onset of activities in Run 1-2 at 14.45 hrs is not seen; on the contrary, the EI and Hit-based profiles continue to decline after this point and the filtered-EI profile remains flat.

3. They present broad patterns that are difficult to connect with sustained periods of activities (for example the direct contact between the specimens in Run 1-2 from 17.2 hrs). Neither do the frequency-time profiles show discernible changes reflecting these activities.

These conclusions call for a re-examination of some of the premises used in setting up the signal acquisition and processing, some of which are now highlighted:

The AE signatures of crack propagation were provided by the indentation machine. Whilst there are grounds for believing that crack propagation in the course of the production runs produces similar signatures at the source (see subsection 5.2.3 above), the arrangement of the specimen in the test machine is much different from the indentation machine. Subsequently, the wave transmission path travelled by the AE signals from the source to the AE detector is significantly different. The transmission path, which can be seen for example in Figure C1 (Appendix C), encompasses the following stages: Silicon nitride specimen → steel shaft → inner race → REs → outer race → bearing housing/AE sensor mount → AE sensor. Subsequently, fault-related signals arriving at the detector are likely to be significantly attenuated. Furthermore, as different frequencies can travel through the complex media differently [154], fault signal energies will be smeared in the time domain - attenuating peak signal magnitudes increasing the difficulty in distinguishing them from the background noise profile.

An attempt to reduce these effects was made by reducing the distance from the contact to the sensor (through modification of the machine's test cell see Subsection 4.2.2.1). However, the propagation path which included rolling bearing elements was not avoidable. A bearing not only produces noise during operation (from friction effects for example), but continually alters the wave propagation path attributed to the motion of the rolling elements with respect to the sensor (as reported, for example, by Toutountzakis et al. [260]).

By measuring the effects of the transmission path on the fault signals will aid in the further development of a suitable signal processing technique, and the data already collected in the Production Runs could be used for further analysis in the future. The effects of the transmission path to the AE signal is normally achieved by a pencil lead break test but such test is not feasible on a running machine with high speed rotating shafts and a closed test cell⁴⁵. Therefore, a more sophisticated arrangement would need to be set up. Such arrangement would require the

⁴⁵ Because of the rolling element bearings, the test would have to be performed on a running machine as the transmission characteristics would be significantly different from a stationary machine.

collection of high frequency tachometer data⁴⁶ together with the AE data, necessitating a data collection system with high throughput and capacity.

Furthermore, the collection of tachometer signal data during a run (such that TSA could be performed) may further aid in the extraction of fault signal components. This has been achieved with AE monitoring, for example, by Rahman, Ohba et al. [164] for the monitoring of steel rolling contacts in a twin disc machine and also by Al-Balushi [261] on gears in a gear test rig. However, in these experimental investigations, shaft rotating speeds were restricted to 800 and 1,300 RPM respectively in order to minimise signal overlapping (from complex AE transmission paths)⁴⁷. Low speed runs would require extended operating times which were not workable in the current study (exceeding 300 hrs), and also, do not represent the typical in-service conditions experienced by hybrid bearings (for example in gas turbines rotating at high speeds).

Additional information could be gathered by installing a second sensor providing 'stereo' AE data, and a more accurate means for determining the various component sources of the signals. The two sets of data could also provide a basis for the development of a processing system based on differential patterns - enabling further distinction between signals produced at the contact (i.e. fault related) and noise (for example by bearing components).

8.7 Summary

This chapter presented the results of the analyses of the data collected in the Production Runs. Their interpretation was helped by references to the model of fault development presented in the previous chapter together with the interpretations made there of the fault snapshots taken after each run.

The main objectives were to determine not just the suitability of a medium for the detection of RCF fault in silicon nitride, but also to see to what extent can these follow the progression of the faults, and equally important, to what extent can the various signal enhancement techniques developed and described in Chapter 5 can make the various fault events more identifiable.

The analyses were based on several graphical representations of the signal data, generally plotting the development of selected variables over the course of the run. *Signal-time* profiles for the total signal intensity, *angular-time* profiles for the signal intensity at particular angular

⁴⁶ The minimum frequency threshold for the AE acquisition is 40 kHz, much higher than for the vibration and WSS ranges (0-16 kHz).

⁴⁷ Low operating speeds also minimised the noise produced from bearing operation at high speed as found by Choudhury and Tandon [116].

locations and *frequency-time* profiles for the signal intensities across various frequencies. For some detailed investigations individual *burst data* were also inspected.

The analyses were performed by considering the details of three runs, selected as representative of varieties of fault types, sizes and signal patterns. The preliminary assessment was based on inspections of the signal-time profiles, the end snapshots and the interpretations of the faults reached in the previous chapter.

The initial analysis was focused on the vibration signals as these are produced by mechanical processes only, i.e. are most closely related to the visual data analysed in the previous chapter. All vibration profiles are dominated by eccentricity components, arising from both the natural eccentricity of each specimen individually, and also from them affecting each other through contact. Both of these represent a noisy background from which the fault-related components are difficult to discern. Both TSaA and EMD were able to attenuate this background, TSaA reducing the contact eccentricities and EMD reducing the natural eccentricities.

The contributions of the individual faults were identified in the angular-time profiles. The signal characteristics expose a non-uniform fault development, showing several peaks in intensity and thus revealing sudden crack or flaking stages. As expected, Flaking type faults were more prominent in the vibration profiles than CRC.

The electrostatic profiles were found to be dominated by the eccentricity components to an even greater extent than vibration. Both eccentricity types were visible, thus demonstrating that higher traction and load levels can induce charge changes on the surface of silicon nitride. As for vibration, TSA was able to attenuate the contact eccentricity; however some natural eccentricity features remained even after filtering.

The angular-time profiles were able to show all the significant fault development events found in vibration with most of the first sights appearing earlier. CRC faults were more detectable than in the vibration profiles. Both CRC and Flaking fault types appear more prominent in the filtered profiles.

A distinct feature in the electrostatic profiles were vertical bands resulting from either a direct contact between the two specimens (thus sinking the charge by earthing through the steel specimen), or debris becoming attached to the steel specimen surface inducing charge changes on the silicon nitride in subsequent rotations. The non-conductivity of the silicon nitride specimen did, indeed, assist in the visibility of both fault- and debris- induced charge.

The ability to discern distinct developmental stages of individual faults confirmed that the angular and frequency resolutions provided by the new WSS1 sensor head were satisfactory, therefore endorsing its design and construction.

The AE profiles were found to be poor on features, both pre- and post-filtering. It was not possible to relate them to the fault-generated events identified in the visual, vibration and electrostatic analyses. This was disappointing, following the promising results from the preparatory tests. Instead the analysis turned into investigating the causes of the resulting profile patterns. This analysis was itself difficult to do with the present data as the angular-time profiles were not available in AE, due to capacity limits of the data acquisition system.

An obvious reason appears to be the actual configuration of the test machine, presenting complex propagation paths for both the fault-generated and background signals. New types of tests will have to be developed to determine the effects of the propagation paths.

Chapter 9. Conclusions and Recommendations

9.1 Project Overview

The project was set up in order to determine whether the use of sensor-based detection and processing systems can provide a reliable means of detecting wear failures in silicon nitride rolling elements of hybrid bearings.

The resolution of this problem was pursued in a series of seven test runs (Production Runs) on a twin disc machine using an industrial-grade silicon nitride specimen against a standard-steel counter specimen. In each of these tests, data from vibration, electrostatic and AE sensors were enhanced and analysed in detail.

The project was implemented in several stages, summarised below.

Preparation of Equipment

The project first involved an extensive preparatory work concerning the test machine, the specimens, the sensors, signal acquisition systems and the data processing techniques.

Following a feasibility test on a standard twin disc machine, a number of modifications were designed and implemented to accommodate a realistic silicon nitride specimen, provide a more appropriate location for the AE sensor, and to improve its configurational repeatability.

The new test cell configuration was designed to accommodate specimens made from standard industrially-produced bearing elements, notably silicon nitride balls and steel races.

A new WSS sensor head was designed and constructed to provide a sufficient sensitivity and selectivity for the detection of highly localised charge produced by RCF faults in silicon nitride.

PC-based data acquisition systems were set up to record sensor and tachometer data in bursts, providing a format suitable for multiple subsequent analyses using a variety of different processing techniques.

Calibration Tests

The resulting configuration was subjected to calibration tests designed to provide base data for the development of the signal processing methods. This included individual runs appropriate for each sensor type, designed to emulate fault generation; these were followed by a combined run to provide background, non-fault-generated, signals.

To excite the dominant resonant frequencies of the test cell for the vibration sensor, standard impact tests were performed. To emulate a fault for the electrostatic signal detection, the

modified twin disc machine was run with a purposely engineered disc specimen on which a defect was implanted using a component made from a different material (silver weld on steel). To obtain fault signals for the AE, a standard indentation machine was equipped with an AE sensor and signals produced by indentations performed on both silicon nitride and steel specimens were recorded.

Data Enhancement Systems

For the enhancement of vibration data, time synchronous averaging was applied to separate the signal patterns of the two specimens by using the timing data from their respective tachometers. In addition, it was combined with Empirical Mode Decomposition emphasising high resonant frequencies, providing a degree of adaptability to changes of dominant resonant frequencies and aimed to attenuate regular, resonant i.e. non-fault related, features.

Time synchronous averaging was also applied for the enhancement of electrostatic signals. In addition for the silicon nitride specimen, a filtering method was designed to adaptively remove a background dominated by an eccentricity profile that was gradually changing.

For the processing of AE data, the Energy Index method was adapted to attenuate the stationary background profiles and enhance transient signal features. Furthermore, an advanced filtering method was developed using a wavelet filtering technique, designed to extract crack propagation signals from a complex background profile. These methods were deployed in addition to a conventional hit-based analysis.

Production Runs Planning and Execution

To accelerate the RCF fault development processes in the Production Runs, pre-cracks were seeded on the ceramic specimens using an indentation process performed on the Vickers indentation machine. The pre-cracks, as well as the resulting faults developed from them in the course of the Production Runs, were inspected using a number of inspection tools. These were: a 3D microscope, electron microscope, profilometer, computer tomography analyser and a digital anglemeter. The measurements taken included the sizes and shapes (e.g. skewness) of the cracks. The profilometer and 3D microscope were also used to inspect the steel counterface following the runs.

The Production Runs differed in the size of the pre-seeded cracks, the machine load and the slide-roll ratios applied. Each run was terminated automatically when a dedicated machine-based sensor detected vibrations exceeding a pre-set limit.

Analysis of Results

Analysis of the runs first focussed on the results produced by the inspection tools in order to obtain a basis from which the signals could be judged. For this, a model of RCF propagation

was constructed incorporating the models of the two major ways of fault propagation reported in the literature: near-vertical and shallow-inclined cracks. The interpretation of the results was based on identifying each end-run snapshot with a stage in the model and reconstructing the mechanisms that led to its production from earlier stages, notably the initial pre-crack.

The signal analyses were based on several graphical representations of the acquired and enhanced data, generally plotting the development of selected variables over the course of the run. Signal-time profiles for the total signal intensity, angular-time profiles for the signal intensity at particular angular locations and frequency-time profiles for the signal intensities across various frequencies. For some detailed investigations individual burst data were also inspected.

The signal analyses were performed by considering the details of three runs, selected following a preliminary assessment as representative of varieties of fault types, sizes and signal patterns. The preliminary assessment was based on comparisons of the signal-time profiles, the end snapshots and the interpretations of the faults reached in the analysis of the inspection tool results.

9.2 Conclusions

The major conclusions are now presented addressing the two major aspects of this study i.e. wear characteristics and their signal production and processing.

Wear Characteristics:

1. Silicon nitride bearing elements produces two types of surface initiated RCF faults - reflecting two essentially different mechanisms in their development: cracks (developing into depth and extensively on the surface), and flaking (developing primarily on the surface by progressive material removal). The former is unique to ceramic bearing elements. As there is no recognised standard classification in the literature yet, they have been denoted in this study as C-/Ring crack Complex (CRC) and Flaking respectively.
2. These faults develop separately from an initial stage, i.e. once a fault acquires the characteristics of a CRC it will continue to develop in this form (and similarly for Flaking). Each type presents specific surface patterns and, in hybrid bearings, produces different wear characteristics on the steel counterface.
3. The recognition of the two distinct types of ceramic faults enabled the creation of a general model that unites the standard models used by various researchers to describe specific aspects of fault progressions. In addition, the combined model incorporates further advanced stages in the fault development not seen in the literature.

4. The progression of each type of fault has been found to be influenced by configurational factors which include the angle of inclination from the surface of a dominant crack, lubrication, rolling direction, load, surface traction and the reaction of the steel counterface.
5. The separate progression of each fault type indicates that the initiating stage plays a key role in determining the subsequent development. The exact parameters within this stage can at this time can only be hinted at (for example the skewness of the initiating crack) - requiring further research.
6. The work considered also the mutual interaction between the ceramic and steel counterfaces in the course of a developing ceramic fault, not reported by previous research publications. Two major mechanisms are found to be instrumental: the production of debris (including three-body abrasion) and direct contacts (without intervening lubricant). Their exact progression was found to depend on the fault type and the stage of its development.

Signal Production and Processing:

Vibration

7. EMD is able to remove background noise generated by the configurational aspects and adapt to different operational conditions (e.g. load and speed levels), therefore it is able to detect faults at an early stage – as identified by the general model. An advancement, suggested by the research, points to the use of multiple IMFs to improve the fault tracking at the advanced stages characterised by a shift in the dominant frequency values.
8. By monitoring the rotation of the bearing elements during operation using, for example, a tachometer can provide important information enabling the tracking of multiple faults and, if collected for both rotating elements, also the tracking of the interaction of the counterfaces.
9. The resulting TSaA signal profiles show that the RCF fault development is non-uniform revealing sudden crack or flaking stages and, also, the onset of direct contacts. In addition, Flaking type faults are able to produce higher vibration signals than CRC type.

Electrostatic

10. The electrostatic signals are dominated by the configurational aspects to an even greater extent than vibration. A significant portion of this noise can be reduced by a spline subtraction method using a filtering technique that adapts to the continuously reducing surface charge in silicon nitride (this is an advancement on the standard method reported in the literature of using fixed noise profiles). The post-filtered profile enables early fault detection and improved fault tracking.

11. The monitoring of the rotation of the bearing elements can provide even more important benefits in the analyses of electrostatic signals than for vibration, because of the significantly different electrostatic characteristics of the two materials (i.e. a significant charge is generated by the silicon nitride material only).
12. The resulting TSaA signal profiles not only confirm a non-uniform fault development detected in vibration (matched to the location of the fault), but also localised rapid charge changes distributed across the ceramic element. These arise in the fault development stages that include debris production and direct contacts between the counterfaces. The analysis also confirm that the electrostatic charge generation on the ceramic surface is particularly sensitive to changes in traction and load levels which can, for example, be produced by the eccentricity of the counterface.
13. In contrast to vibration monitoring, CRC faults produce more substantial charge signals than the Flaking faults and are therefore detectable earlier than for vibration monitoring. For the reasons above, electrostatic sensing is particularly well suited to monitoring the development of RCF faults in silicon nitride.

Acoustic Emission

1. The AE profiles are extremely sensitive to the configurational aspects, especially the signal transmission path to the sensor - subsequently hindering the detectability of faults and the ability to generalise findings across different configurations. Understanding of the transmission path is seen as a key issue in the development of a suitable method to extract AE fault profiles.
2. The research showed that under special and strictly controlled circumstances where the effects of the transmission path are significantly reduced, the AE fault signal profiles have high magnitude with characteristic frequency patterns and low sensitivity to load conditions.

To summarise, the findings show that both vibration and electrostatic charge signals are capable to detect RCF faults in silicon nitride. Equipment using new instrumentation and signal processing techniques can not only detect individual faults early but also track their progression.

Generally, Flaking type faults produce more prominent signatures in the vibration spectra, whilst C-/Ring Crack Complex faults produce stronger electrostatic charge profiles than in vibration.

Combining the vibration and electrostatic signal profiles provided information sufficient to reconstruct their histories and thus offers a basis for a practical and reliable fault monitoring system.

9.3 Recommendations

The following presents several recommendations based on the experiences obtained during the runs and the analysis of the results:

The indentation process employed for the runs produced a variety of pre-cracks shapes; this was an advantage because it created multiple testing conditions. However, the repeatability of this process is severely diminished through the lack of full control over the indentation results. Other researchers whose results were referred to here (for example [51]) were able to produce single types of pre-cracks through better control of crack generation processes; such ability may be required for our own future work.

The inability to assess crack features below the surface was a disadvantage in that many pre-cracks could not be reliably assessed for distinct fault types (e.g. by the relative prominence of the radial and lateral cracks). Such ability would be required e.g. to further investigate the conditions that lead to crack development from an initial indefinite fault stage to a specific CRC or Flaking type. This could be achieved by measuring the subsurface characteristics of the pre-cracks before a run using an appropriate non-destructive technique. Ultrasonic inspection methods could be adapted for this purpose, in particular surface wave acoustic microscopy has shown promising measurement resolutions to provide a suitable basis ([44] [262]).

In interpreting some of the results, notably those that involved single or double imprints, an uncertainty related to the shut-down procedure of the twin disc machine. Namely, it was possible that the (gradual) shut-down process could induce additional effects, which would therefore appear as a one-off. In the detailed analyses of the Production Run results these were dismissed on the statistical grounds (as the shut-down process was of negligible duration compared to the entire run), however future investigation could be helped with the shut-down process being instantaneous, for example involving instant disengagement of the rolling surfaces.

Further enhancement would be to develop a method for an automated shut-down of the machine on detection of a predefined, anomalous condition, for example representing a particular stage in fault development. Such an arrangement would enable inspection of the specimens at this critical point. To develop this capability would require a multi-senor analysis using several techniques such as based on Gaussian mixture models and an appropriate machine configuration.

The use of three indents per specimen was advantageous in that it was an easy way of producing multiple results from a limited number of runs and the interpretation of the steel surface effects was aided by the fact that mostly a single fault became dominant. Having established the major

conclusions from these runs, further work will however require the use of single pre-cracks per specimen to avoid cross-fault interference and to obtain statistics of fault life-times.

In order to track the development of faults over extended stages, longer runs would be required.

Further information on the development of faults can be obtained by a more detailed examination of the debris produced, for example its material, quantity and shape. This would require collection of the debris during and after a run for its subsequent inspection and analysis.

Similarly, real-time monitoring of the charge properties of the debris produced would assist in a better understanding of how the charge effects are produced in various stages of fault development. This was attempted for the Production Runs using an electrostatic oil-line sensor, however the sensors were over-whelmed by the noise arising from air bubbles in the oil. Using a test machine with an enclosed and pressurised system may remove this problem.

It was found that faults at their later, larger-size stages excite lower resonant frequencies that can potentially move beyond the sensitivity of the present filtering method. The method should therefore be enhanced to enable tracking the faults beyond this stage.

The electrostatic profiles were dominated by eccentricity profiles that, because of their non-sinusoidal shape that varied not only over the course of a run but also from run to run, were difficult to filter out. Investigating their precise pattern and progression would assist in improving the filtering method. Considering the use of techniques other than TSaA may also lead to improved ways of distinguishing between fault and non-fault related signals. In addition, investigating whether the use of a smaller sensor (with a capable amplifier) would provide better resolution of individual cracks.

The AE profiles were significantly influenced by the transmission path from the initiating fault to the AE sensor. The contribution of the signal path is normally measured by a pencil lead break test but such test was not feasible on the running machine. Therefore, a more sophisticated arrangement would need to be set up in order to determine the effects of the transmission path on the signal pattern. Such arrangement would require the collection of high frequency tachometer data together with the AE data, necessitating data collection system with high throughput and capacity. Once the effects of the transmission path are understood, the signal processing technique could be further adapted and the data already collected in the Production Runs could be used for further analysis in the future.

Further information could be gathered by installing a second sensor providing 'stereo' AE data, providing a more accurate means for determining the various component sources of the signals. The two sets of data could also provide a basis for the development of a processing system based on differential patterns.

Appendices

Appendix A: Hertzian (Elliptical) Contact Equations

For ball bearings, the ball-raceway interfaces form elliptical contact shapes. The equations describing an elliptical contact of two contacting bodies, body ‘1’ and body ‘2’ are found below⁴⁸, providing the basis for the calculations used in the Production Runs.

The eccentricity of the contact ellipse e , is calculated from the radii of the contact: a and b where a describes the major radius.

$$e^2 = 1 - \left(\frac{b}{a}\right)^2 \cong \left(\frac{R_a}{R_b}\right)^2 \quad \text{A.1}$$

R_a and R_b describe the relative radii of the contacting bodies and are found in [4]. The equivalent radius for a circular contact is constructed from the applied load P , the relative radii R_C ($R_C = \sqrt{R_a R_b}$) and the reduced elastic modulus E^* as follows:

$$c = \sqrt{ab} = \left(\frac{3PR_C}{4E^*}\right)^{\frac{1}{3}} F \quad \text{A.2}$$

with E^* calculated using the elastic moduli E and Poisson’s ratios of the two bodies.

$$\frac{1}{E^*} = \frac{1 - v_1^2}{E_1} + \frac{1 - v_2^2}{E_2} \quad \text{A.3}$$

In equation A.2, F denotes a correction factor accounting for the elliptical contact shape calculated as follows:

$$F = \left[\left(\frac{R_a}{R_b}\right)^{0.0602} - 1 \right]^{1.456} \quad \text{A.4}$$

Subsequently, the maximum and mean contact pressures p_{max} & p_0 respectively can be calculated:

$$p_{max} = \frac{3P}{2\pi ab} = \frac{3P}{2\pi c^2} \quad \text{A.5}$$

$$p_0 = \frac{2p_{max}}{3} \quad \text{A.6}$$

The tensile stresses at the contact radii a and b are calculated as follows:

⁴⁸ Equations provided by Dwyer-Joyce [3] and Hale [4].

$$\sigma_a = p_0 (1 - 2\nu) \left(\frac{b}{ae^2} \right) \left(\frac{1}{e} \tanh^{-1}(e) - 1 \right) \quad \text{A.7}$$

$$\sigma_b = p_0 (1 - 2\nu) \left(\frac{b}{ae^2} \right) \left(1 - \frac{b}{ae^2} \tanh^{-1}\left(\frac{ea}{b}\right) \right) \quad \text{A.8}$$

And subsequently $\sigma_a > \sigma_b$ and therefore maximal.

The maximum shear stress is calculated as follows:

$$\tau_{12} = p_0 \left(0.303 + 0.0855 \left(\frac{b}{a} \right) \right) \quad \text{A.9}$$

Occurring at a depth z equal to:

$$z = b \left(0.7929 + 0.3207 \left(\frac{b}{a} \right) \right) \quad \text{A.10}$$

Appendix B: Feasibility Test

A feasibility test was performed using a RBSSN disc specimen contacting an EN31 through-hardened steel disc using the instrumented test cell in the original disc-on-disc configuration.

Instrumentation

The configuration of the instrumented test cell is illustrated in Figure B.1 which is repeated here for convenience.

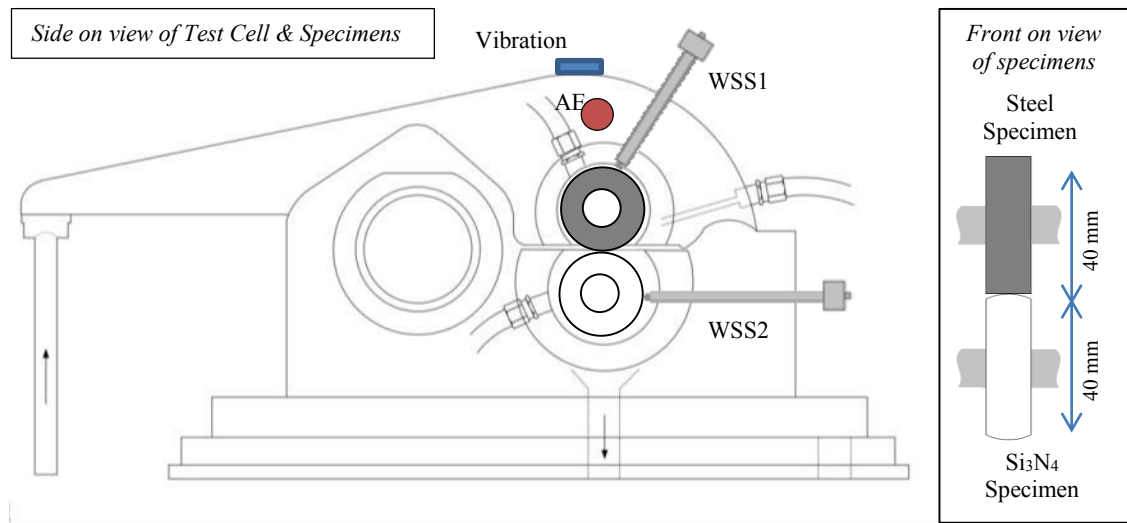


Figure B.1: Schematic of the instrumented test cell in the original disc-on-disc configuration.

▪ Mechanical Set-up

The four sensors (AE, WSS1, WSS2 and vibration) were placed in the positions illustrated in the figure above. The vibration and AE sensors were glued to the surfaces of the Load Arm with epoxy-based glue. The WSSs (with their original 4.5 mm diameter head) were incorporated into the Test Cell by holes drilled through its casing to allow them to be positioned as closely as possible to the respective specimens. The distance between the sensor face and its respective specimen was set to 0.3 mm using a sized spacer before tightening the grub screw anchoring the sensor to the casing.

▪ Data Acquisition

AE data was acquired using the PAC AE system described in section 5.2.3.

Vibration and WSSs were acquired using an Agilent *VEE* system, routinely used in nCATS.

This is composed of a Smith's Aerospace USB DTI data acquisition (DAQ) card and software installed on a PC. The card provides 8 input ports and a maximum sample rate of 100 kHz shared between them, with a resolution of 16 bits.

The sensors were connected as follows: WSS1 to Port 1, WSS2 to Port 2, Vibration to Port 3. Additional two ports were connected to tachometer monitoring the rotation of the two specimen shafts.

In order to achieve the maximum signal acquisition within the limitations of the overall sample rate capacity, data was acquired in bursts alternating between the WSS (WSS1&2 together) and the Vibration channels. Each of these bursts included also the recording of both tachometer signals.

Each sensor channel was converted to digital format with a sampling rate of 25 kHz over a burst lasting 1 second. The intervals between bursts (dead time) were set to 20 seconds. This ratio was imposed by the limitations of the card. An additional limitation was due to the throughput limitations of the PC signal processing software (VEE v6) which was overcome by buffering the incoming raw binary data and restricting the software use to a basic RMS analysis; both raw and processed data were stored on the PC as the acquired data (tachometer data were recorded in their raw form).

▪ Specimens

The steel disc specimen, measuring 40 mm in diameter was machined in-house at nCATS workshop using EN31 steel. The specimen was heat-treated in an electric furnace at 850°C with 20 mins soak before quenching in oil at ambient temperature. After this, the specimen was placed in the furnace for tempering at 150 °C over 2 hrs, achieving a measured Vickers hardness of 700 Hv30 \pm 25, comparable to bearing steels.

The RBSSN disc specimen was obtained from an external supplier, HC Starck Ltd. Its dimensions were 40 mm diameter with a crowned edge with a diameter of 40 mm, thus emulating a ball. Its Vickers hardness was 1450 Hv30 \pm 14. Its porosity was rated at 2% by the manufacturers, a level much higher than in industrial bearing elements (less than 0.01%), but was accepted as an adequate initial sample for the test. As a result any wear was produced more easily than would be in real applications.

Both specimens were sent to CMC Precision Engineering Ltd Aldershot for grinding and polishing. The ceramic disc was finished to 0.4 and 0.7 μ m in the rolling and transverse directions respectively; the steel disc was polished to 0.047 and 0.125 μ m in similar directions respectively. Both specimens are illustrated in Figure B.2.

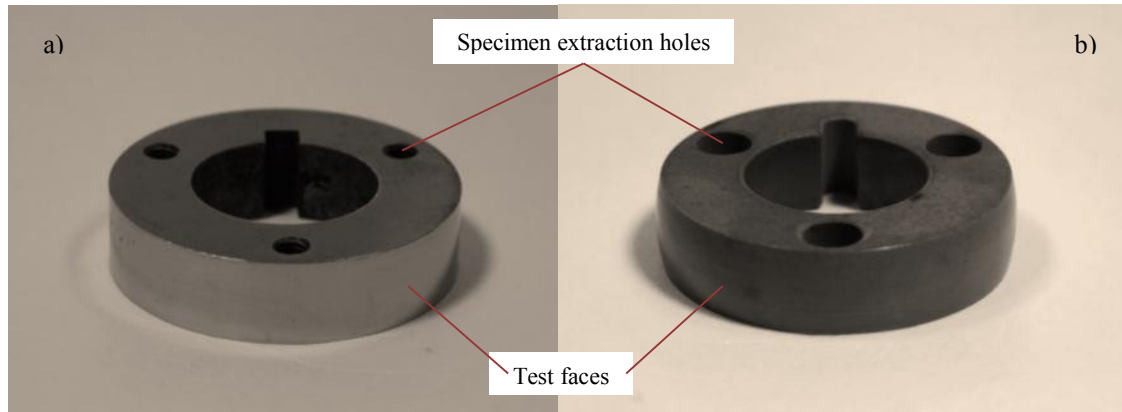


Figure B.2: Samples for the feasibility test a) steel, b) silicon nitride.

▪ Test Details

The test parameters were set up to reproduce the conditions adopted by Braza et al [41] for investigation of wear in hybrid bearings using a similar test machine. The test was started with a preload of 0.04 kN (0.95GPa), linearly ramped up to 0.15 kN (1.4GPa) at 1 hr, and held at this load for the remainder of the test. The lubricant used was the mineral oil Vitrea 32 (similar to the one used by Braza et al.) with a kinematic viscosity of 32 mm²/s at 40°C, pumped continuously at 50 ml/min at 40°C \pm 1°. Discs were driven at 1560 RPM (26 Hz) each, and due to the small difference in the specimen diameters, a low slide-roll ratio of 0.15% was produced. The test was run for 51.5 hours from start-up, corresponding to approximately 4.35×10^6 cycles. The discs were gravimetrically analysed to an accuracy of 0.001g using a Mettler Toledo XP 205 5-figure balance. Surfaces were then examined using SEM to identify any resulting wear.

▪ Results and Discussions

The objective of the test was to determine the fitness of the twin disc machine to perform wear tests on hybrid components. Therefore, to evaluate this fitness, the following criteria were used:

- Accurate and consistent set-up
- Production of wear
- Production of appropriate sensor data
- Robustness of the sensor data processing systems

These are now discussed in turn.

Accurate and Consistent Set-up

To make the machine operational for each test requires the removal of several key components from the Test Cell, fitting them with the specimens at their precise locations and reassembling them subsequently in the machine. This is hampered by a limited access due to several adjacent components.

In the course of setting up the machine for this test, some axial movement in the specimen shafts was found resulting in an inaccuracy in the set distances between both WSSs and their respective specimens. This would therefore prevent the creation of consistent conditions over multiple consecutive tests. For this one-off test this was not relevant, but it exposed an important drawback for any repeated testing.

The ceramic specimen is mounted on the shaft by pushing against a tapered part of the shaft by a locking nut. The wedge effect of the taper combined with the high brittleness of the specimen resulted in several sample breakages. This was acceptable for a one-off test, however would result in prohibitively expensive multiple tests.

To overcome the above deficiencies, several modifications were recommended to the shaft and the specimen mounting.

Production of Wear

The test was successful in generating wear in both the steel and silicon nitride specimens each of which had appearance similar to that reported by Braza et al. [41] Further analysis however indicated that the silicon nitride material used in these tests was poorer than that used in industrial bearings. It was therefore important that the subsequent tests use a bearing-relevant material.

The details of the wear generated are now described below.

Figure B.3 and Figure B.4 show the SEM images of both discs at three different magnifications. The worn surface of the *steel disc* appeared to be generally smooth but with a large number of pits (see Figure B.3a). Various cracks perpendicular to the rolling direction are also seen in some areas of the wear track, and some pits are outlined by distinguishable cracks approximately 50 μm in diameter (see Figure B.3b). Figure B.3c is a close-look of a pit on the steel disc, showing a spall outlined by multiple cracks.

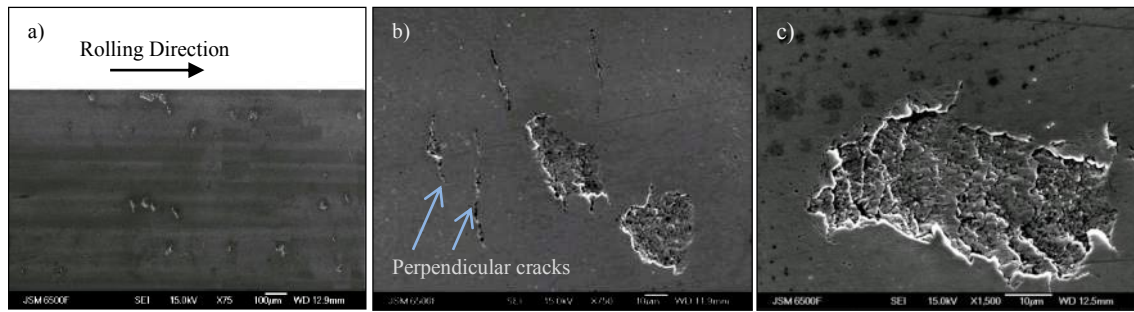


Figure B.3: SEM images of the steel disc wear track at magnifications (a) x75, (b) x750, (c) x1,500. (b): two parallel interacting cracks forming a pit. (c) shows a typical pit.

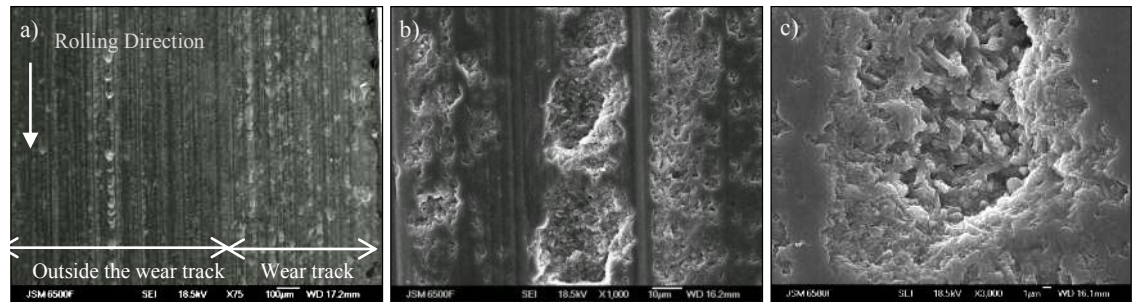


Figure B.4: SEM images of the SN disc surface at magnifications (a) x75, (b) x1,000 (c) x1,500. (a) the damage of the wear track and polishing damage, (b) untested surface- damage from polishing. (c) typical damage at the wear track

Figure B.4 shows the silicon nitride disc with several areas of apparently different wear types. Figure B.4a shows that wear was present both within and outside the roll contact path. The explanation for the latter is that it has been induced by the polishing process prior to the test and could therefore be ignored, but bearing in mind that similar wear was therefore present also in the contact path prior to the test. Fortunately, a significant amount of wear was generated on the contact path during the test, dominating the pre-test wear. This included spalls that in the virgin surface were about 30 µm in size but have grown to over 90 µm in the wear track by the end of the test.

Following this analysis, the machine was deemed suitable for producing suitable wear i.e. measurable and therefore detectable - even bearing in mind the high porosity of the sample.

Production of Appropriate Sensor Data

The AE sensor was positioned in a convenient location as close to the contact as possible, however, it was not in the optimum position as it allowed multiple AE wave propagation paths from the contact to the sensor and it became obvious that it could be positioned closer, with some modification to the Test Cell, to achieve better signal acquisition.

WSS2 monitoring the ceramic specimen, despite the high tolerance in manufacturing both the shafts and specimens (<0.02 mm), was dominated by a high amplitude, low frequency signal relating to the eccentricity in the rotation of the specimen. This produced a high noise level, potentially masking the fault signatures. It also became obvious that the relatively large sensor

head was an important contributor to the noise and that therefore substituting it with a new, smaller head should alleviate the problem.

The WSS1 and vibration sensing were judged to be fit for the use in hybrid RCF testing.

These results followed from a detailed analysis of the sensor signal data described in the following.

In order to determine whether the sensor data actually relate to the production of significant wear in the course of the test, friction coefficients, calculated throughout the test by the machine from the load and torque sensors, were acquired in parallel with the sensor data for subsequent analysis (supplementing the end-test visual inspections above).

Their values were high at the start of the test (approximately 0.14 during the first 5 hours) due to initial run-in and the gradual initial application of the load. After this, they decreased to an average of 0.09 until the 45th hour, at which point they increased to 0.1 and remained at that level to the end of the test at 51.5 hrs. This suggests that much of the wear was generated in the last period (referred to below as *run-out*).

Figure B.5 illustrates the RMS signal-time profiles from AE, WSS1, WSS2 and vibration sensors (a) in the initial 0-9 hours and (b) the final phase (42.5-51.5 hours) of the test.

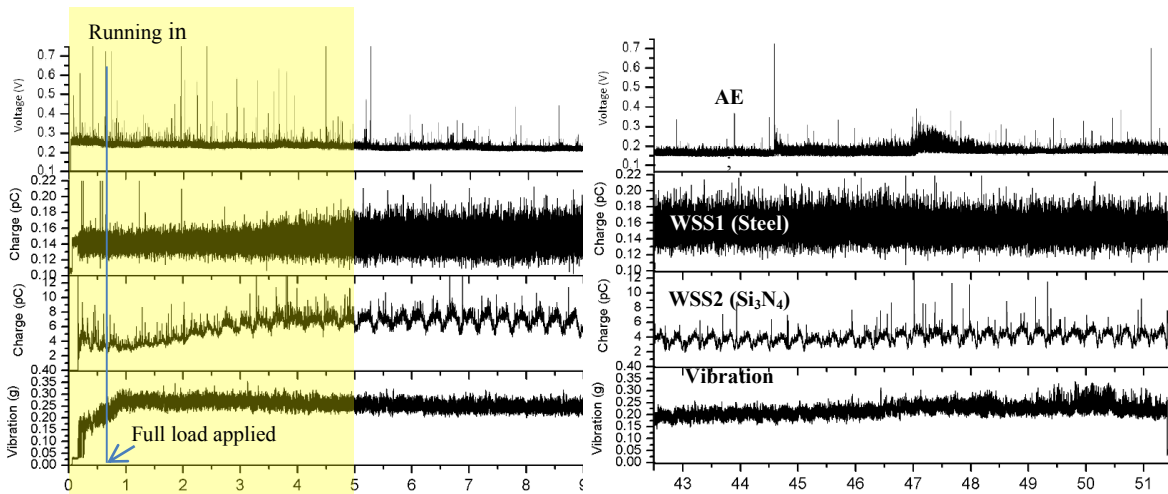


Figure B.5: RMS graphs of sensor responses split in to a): initial phase of test (0-9hrs) and b): final phase of test (42.5-51.5hrs). Both plots show the graphs of (from top to bottom) AE, WSS1, WSS2, Vibration.

All sensors responded, to a varying degree, to the running-in wear (0-5 hours defined by the friction responses) and also to the run-out.

The vibration sensor shows a gradual increase over the first hour (reflecting both run-in and increasing load), after which remains steady. An increase after 46 hours was likely the result of the large pits generated in the steel and silicon nitride discs.

WSS2 shows a more gradual running in pattern followed by oscillatory pattern with occasional peaks. The former are most likely caused by specimen rotational eccentricity (for more details see below), whereas the latter (hopefully) reflect the generation of the wear. The intensifying of the wear generation in the run-out period, identified above, is reflected by increased multiplicity of the peaks during this period.

WSS1 profile patterns are significantly smaller than those of WSS2 (note the relative scales in the figure above), reflecting the rapid dissipation of charges in steel compared to a non-conductive silicon nitride. The WSS1 profile shows in the first 2 hours multiple sharp peaks standing out from a high frequency background, which subsequently increases to a level at which it remains for much of the test. The initial sharp peaks are explained by a relatively fast rate of running-in in steel, the background is probably charge being transferred to it from the ceramic specimen. The slight drop in the background at 46 hours coincides with the vibration response mentioned above, due to the pit generation i.e. increased wear rate. Nevertheless the peaks are significantly less numerous than in WSS2, this is because no charge carrying silicon nitride debris was lodged on the steel surface, as confirmed by the visual inspections above.

The AE sensor produced multiple peaks going slightly beyond the initial run-in period, standing out from a background with occasional smaller peaks which persisted over the entire run. The run-out period is dominated by two large peaks and a significant increase in the background at 47 hours, coinciding with the generation of pits already identified in the vibration and WSS2 profiles above.

To investigate the eccentricity found in the WSS2 profile to a greater detail, TSA processing was applied to the raw WSS2 signal using the tachometer data that were acquired in parallel with the signal data. This brought out the angle dependency of the signal as shown below.

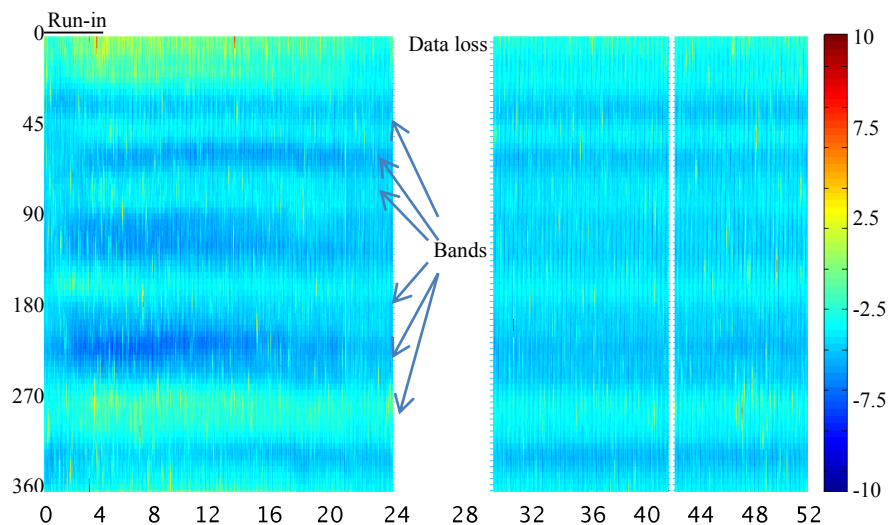


Figure B.6: TSA angular-time profile of WSS2 over the entire test. Blank vertical bands are due to raw data loss by the processing system.

The diagram shows distinct horizontal bands present throughout the test – a clear effect of eccentricity.

Robustness of the Sensor Data Processing Systems

The AE-win system installed on a dedicated PC was able to cope with the incoming AE signal data and had sufficient capacity to subsequently calculate its RMS.

The VEE system appeared to struggle with the incoming signal data, freezing at 2 points during the test with accompanying data losses. The system also has a limited functionality, allowing only basic analyses. Enquiry with the manufacturers (HP) revealed that this was an old version of an obsolete system; another system was therefore acquired for the subsequent testing.

Appendix C: Test Cell Modification Details

▪ Test Cell Modifications

The assembly consisting of the modified parts attached to the Load Arm is shown in the drawings below.

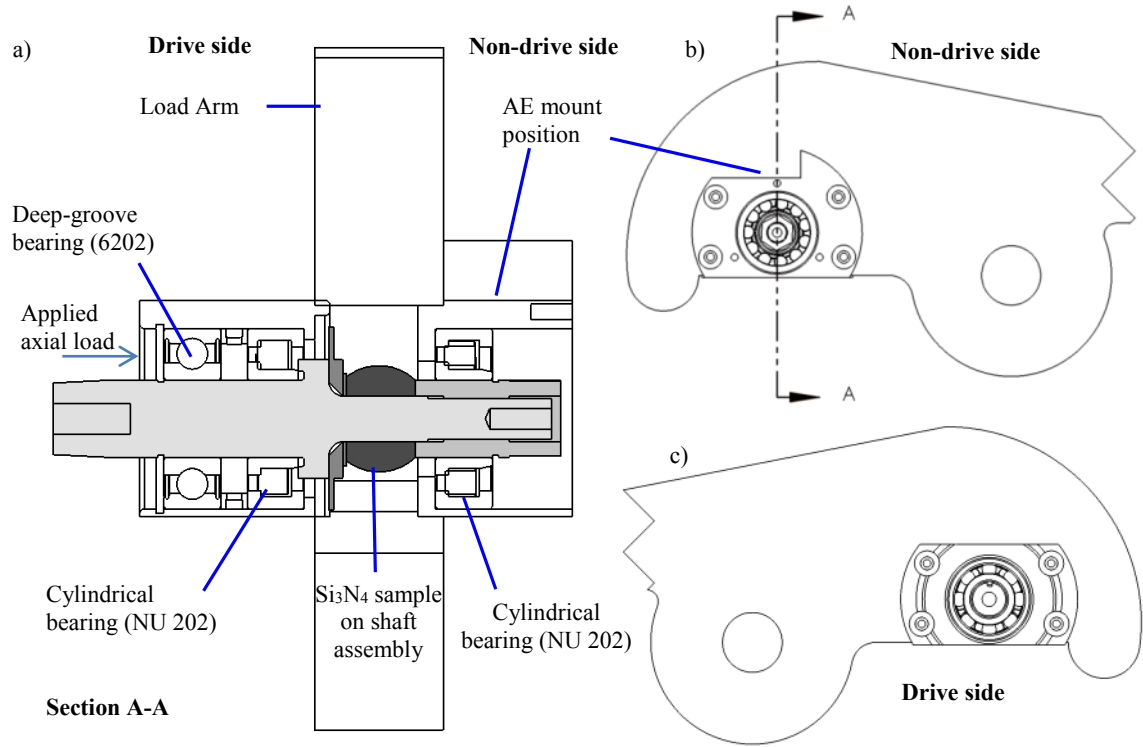


Figure C.1: Assembly of the modified upper shaft and its modified upper bearing housings attached to the Load Arm. a) front-on-view sectioned vertically at the shaft axis shown in (b). b) & c) the side-on-views.

Apart from providing the reader with a greater appreciation of the new shaft seating arrangement, the figure shows the following modification features:

- The AE mount position, located directly above the non-drive side support bearing.
- The non-drive side support bearing has been located closer to the contact to reduce the distance of the AE wave propagation to the sensor.
- The deep-groove bearing at the drive side enables an axial load to be applied to the shaft to lock it in a desired axial location with high precision, without inducing a noticeable axial movement.

▪ Silicon Nitride Specimen

An assembly of the specimen and the modified shaft are shown in the drawings below.

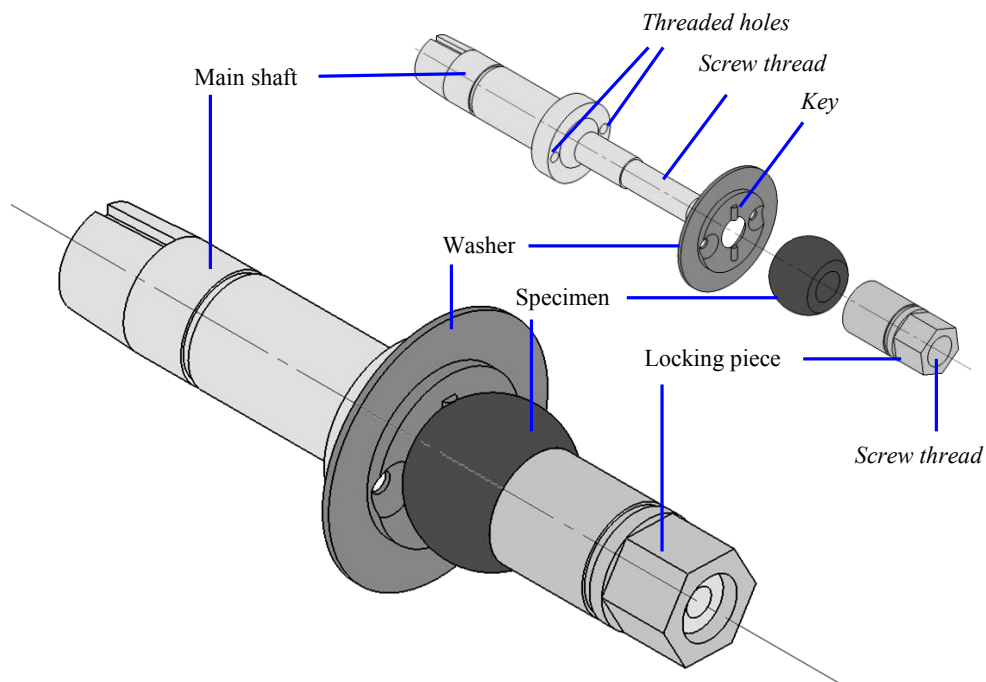


Figure C.2: Drawings of the upper specimen shaft assembly in its exploded and assembled forms.

It consists of three main parts, these are:

- Main shaft - incorporating threaded holes and a screw thread at one end (for mounting the washer and locking piece).
- Washer – (to provide a driving attachment for the specimen) incorporates semi-circular protruding keys and is fixed to the main shaft by the threaded holes using screws.
- Locking piece - fastens the specimen to the shaft assembly. It incorporates a threaded inner hole which is tightened against the screw thread of the main shaft.

▪ Steel Specimen

An assembly of the specimen is shown in the drawings below.

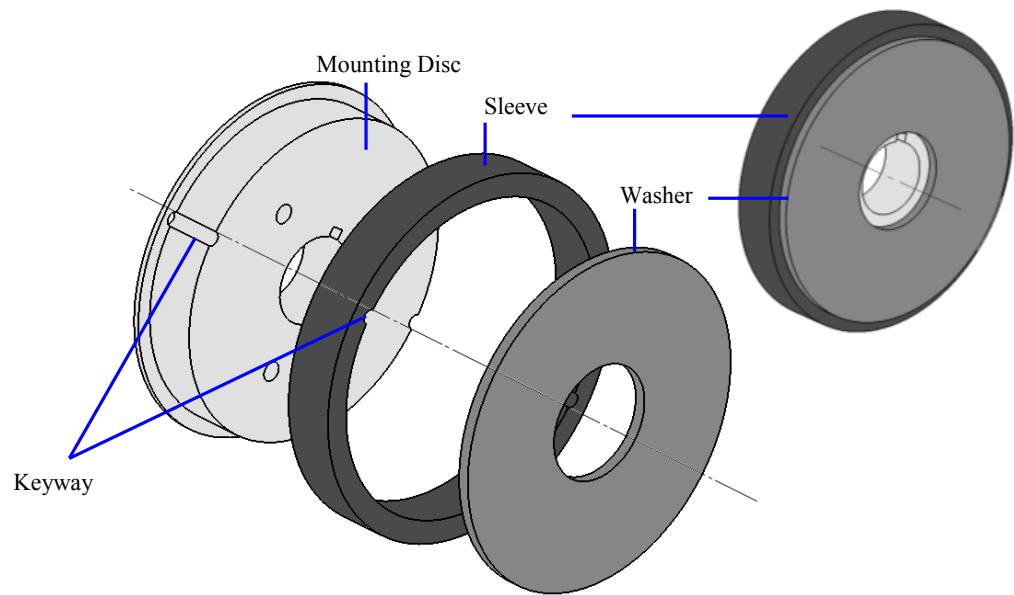


Figure C.3: Drawings of the lower specimen assembly in its exploded and assembled forms.

The specimen consists of three parts, these are:

- Mounting disc - mounts the sleeve to the original lower shaft. It incorporates two semi-circular keyways diametrically opposite one another.
- Sleeve – provides the specimen wear track. Includes keyways for locking to the mounting disc via steel pins (not illustrated in the drawing).
- Washer – fixes the sleeve to the mounting assembly as it is pushed against them with a locking nut on the lower shaft.

Appendix D: WSS Charge Modelling

WSS Electrostatic Charge Model

A charge source q_1 passing a WSS sensing face was modelled adapted from [140], and is illustrated in Figure D.1.

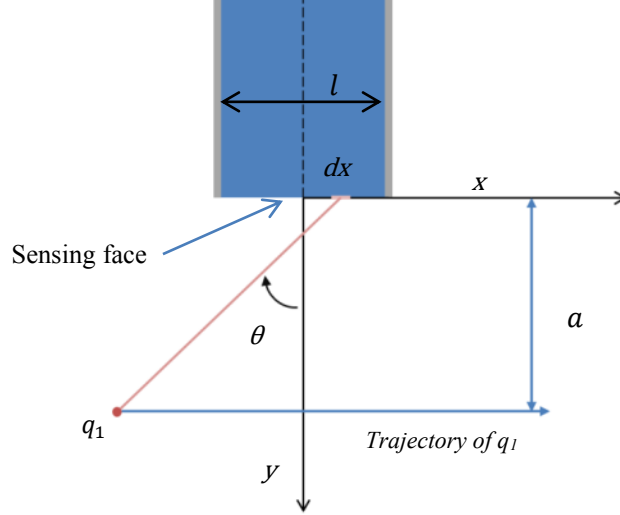


Figure D.1: Illustration of a charge source passing the WSS sensing face

Where d , the distance between the charge source q_1 and the sensor face in the y direction, dx , θ the angle between q_1 and dx with respect to x .

$$E = \frac{q_1}{4\pi\epsilon_0(y^2 + (x - a)^2)} \quad 9.11$$

For a rectangular/square sensor face with width W , the area ds at dx is calculated:

$$ds = W \cdot dx \quad 9.12$$

The induced charge on the sensor face q is therefore calculated using Gauss's theorem:

$$dq = -\epsilon_0 E \cdot ds = -\epsilon_0 E ds \cos\theta = -\frac{q_1 W y dx}{4\pi[y^2 + (x - a)^2]^{1.5}} \quad 9.13$$

Such that:

$$q = -\frac{q_1 W y}{4\pi} \int_{-l/2}^{l/2} \frac{dx}{[y^2 + (x - a)^2]^{1.5}} \quad 9.14$$

$$q = -\frac{q_1 W}{4\pi y} \left[\frac{x + a}{\sqrt{[y^2 + (x - a)^2]}} \right]_{-l/2}^{l/2} \quad 9.15$$

And therefore:

$$q = -\frac{q_1 W}{4\pi y} \left[\frac{l/2 - a}{\sqrt{[y^2 + (l/2 + a)^2]}} - \frac{l/2 + a}{\sqrt{[y^2 + (l/2 + a)^2]}} \right] \quad 9.16$$

To find the impulse response of the sensor or transfer function , the Dirac delta function or unit impulse δ of charge q_1 was modelled, by submitting a with 0 such that

$$q = -K q_1 \quad 9.17$$

where

$$K = -\frac{W}{4\pi} \frac{l}{y\sqrt{y^2 + \left(\frac{l}{2}\right)^2}} \quad 9.18$$

▪ Transfer Functions

In order to find the required modifications to the WSS1 (ceramic), a model was run that emulated four separate sensor face sizes and subsequently the signal profiles in both spatial and frequency domains were investigated. The transfer function signal profiles for each of the sensors were calculated using equation 9.16 above using various sensor lengths l and a distance of 0.3 mm between the charge source and sensing face (a). The profiles in the spatial and frequency domains (the latter computed using Fourier analysis) are illustrated in the figure below.

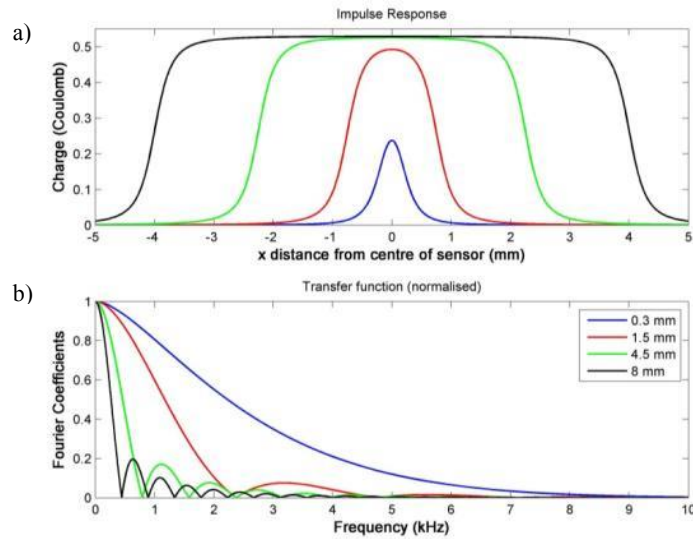


Figure D.2: Transfer profiles for WSS with various sensing face lengths, in the spatial domain (a) and in the normalised frequency domain (b).

A pay-off between the ability to distinguish precise charge sources and the sensitivity is demonstrated (compare the broadness of the profiles in Figure D.2a with their maximal values). Notably, with increased spatial resolution (i.e. smaller sensor face) a corresponding reduction in the frequency resolution is found.

The selection of the smallest sensor size (i.e. 0.3 mm) was based on the rationale that spatial resolution would assist in differentiating precise, localised charge sources such as those expected for RCF faults. Also, that the associated reduced sensitivity would be mitigated by the

high charging nature of the ceramic specimen and supplemented by increased charge amplification.

The poor frequency resolution was not considered to hinder the fault detection response at this time as such a drawback could be overcome by a suitable choice of filtering method.

To investigate the frequency profiles of the sensor to RCF faults in order that a filtering strategy could be determined, the chosen sensor was then analysed using the model above with various defect sizes. Here digital convolution between the sensor transfer function (shown in Figure D.2a above) was performed with defects of various sizes modelled as unit impulses shown in Figure D.3a (and their frequency profiles are shown in Figure D.3b). The defects modelled were comparable in length to the failures observed by Hadfield et al [85] and Kida et al [51].

The results of the computations are shown in both spatial and frequency domains in Figure D.3c&d below.

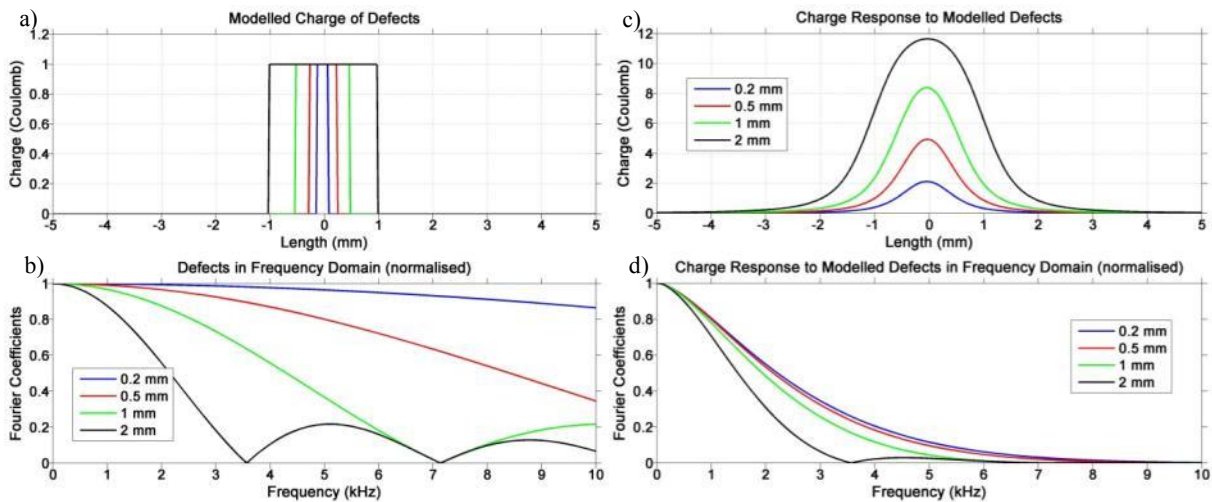


Figure D.3: Modelled defects of various sizes in the a) spatial and b) frequency domains. The signal profiles of the modelled 0.3 mm length WSS in the c) spatial and d) frequency domains

As expected, a broader and higher signal profile was observed for larger faults (refer to Figure D.3c). Also, a variation in the frequency profiles was observed with the response becoming narrower and more resolvable as the defect size increases.

Because much of the power signal is in the low frequency range (maximal at 0 Hz), and is quite broad for small faults, an overlap between the eccentricity signal waveform and the fault signal is therefore expected. Subsequently, a removal of the eccentricity waveform by an appropriate filter should avoid significant loss of the fault waveform. For this, a measurement of the eccentricity frequency should be performed before the Production Runs.

Appendix E: Indentation Tests

- **Indent Inspections**

Steel

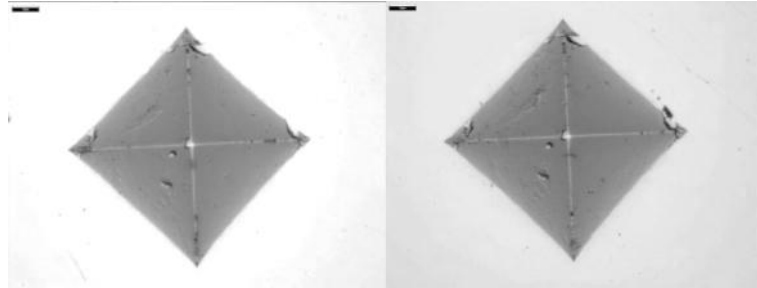


Figure E.1: Alicona optical images (20x magnification) of indentations made at 15 kgF for steel tests.

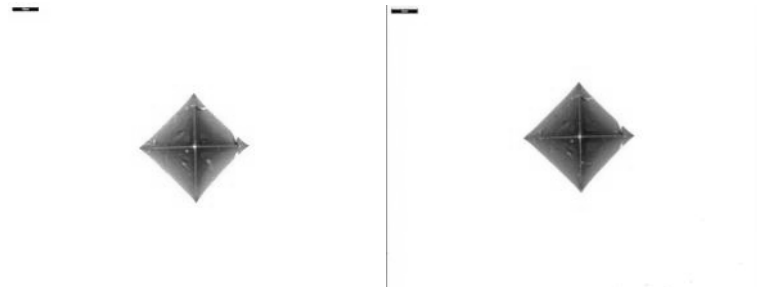


Figure E.2: Alicona optical images (10x magnification) of indentations made at 20 kgF for steel tests.

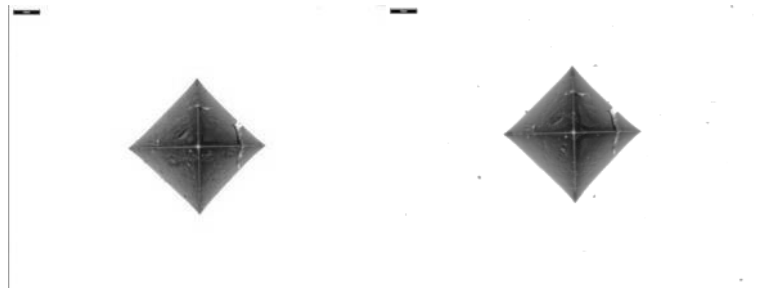


Figure E.3: Alicona optical images (10x magnification) of indentations made at 30 kgF for steel tests.

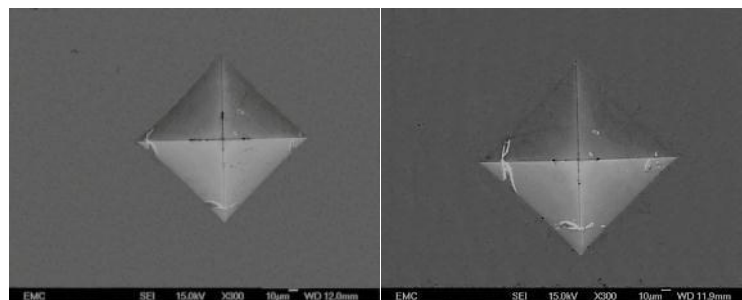


Figure E.4: SEM of indentations 15-A and 20-A.

- **Silicon Nitride**

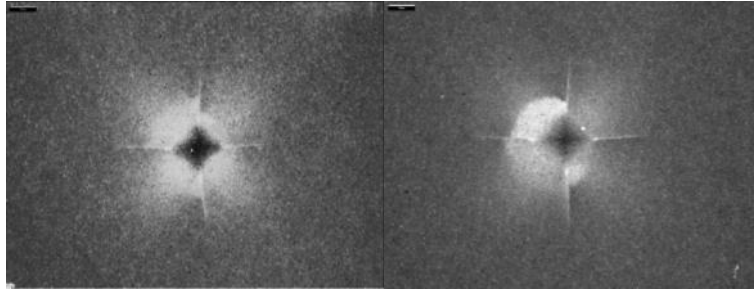


Figure E.5: Alicona optical images (20x magnification) of indentations made at 15 kgF for steel tests.

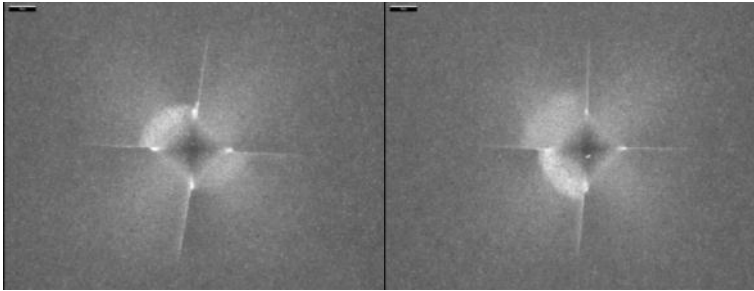


Figure E.6: Alicona optical images (20x magnification) of indentations made at 20 kgF for steel tests.

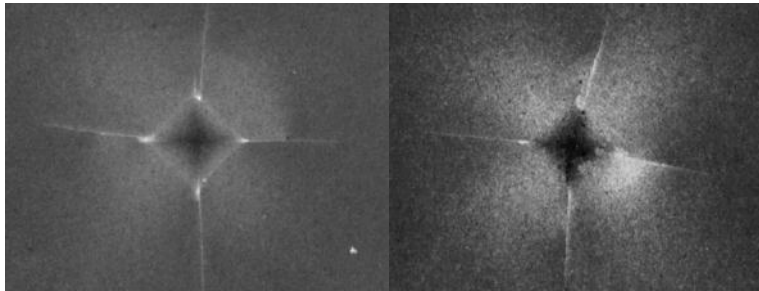


Figure E.7: Alicona optical images (20x magnification) of indentations made at 30 kgF for steel tests.

■ AE Waveforms

Background Acquisitions

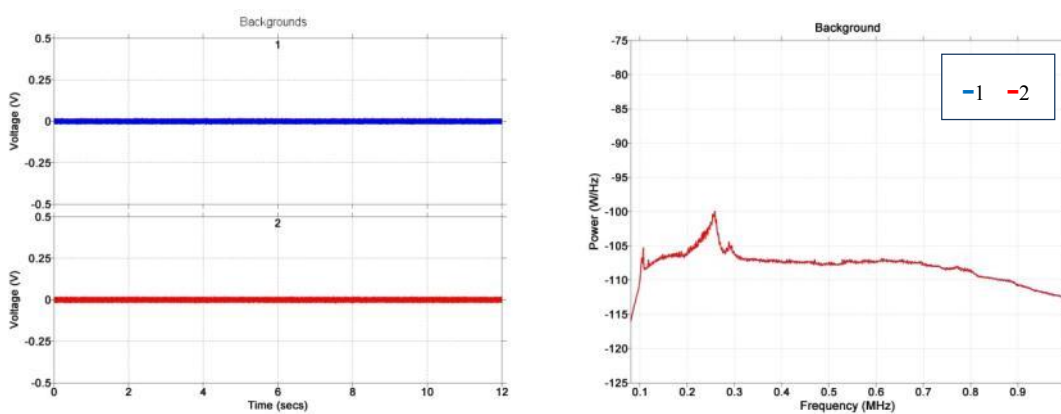


Figure E.8: a) Recorded AE waveforms for Background tests 1 (upper) and 2 (lower). b) PSD estimated of the two waveforms (indicated in the legend).

Steel Indentations

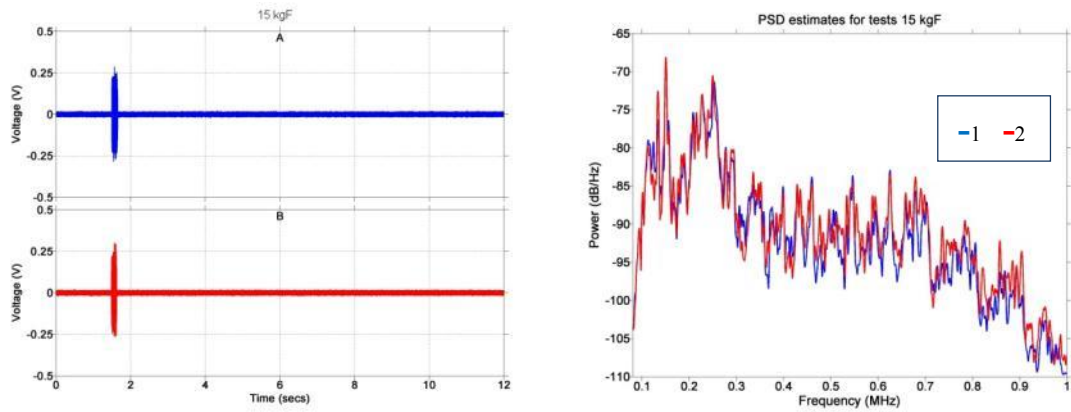


Figure E.9: a) Recorded AE waveforms for Steel indentation tests 15-A (upper) and 15-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).

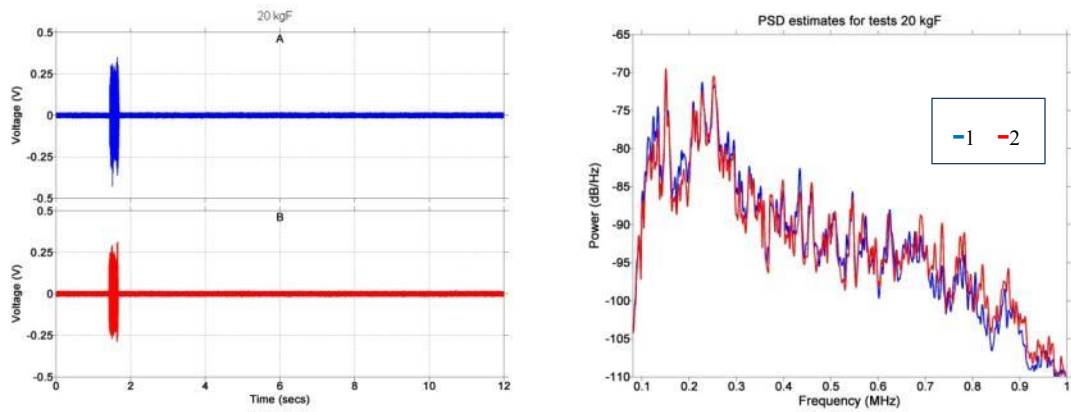


Figure E.10: a) Recorded AE waveforms for Steel indentation tests 20-A (upper) and 20-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).

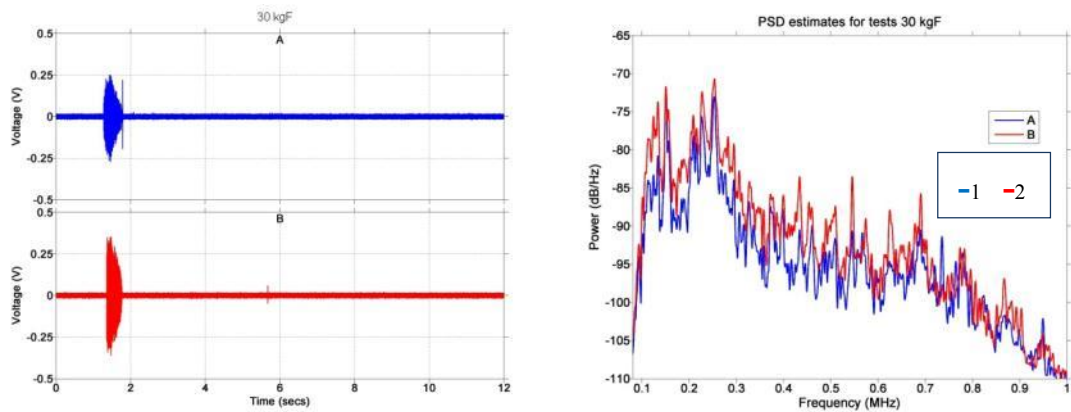


Figure E.11: a) Recorded AE waveforms for Steel indentation tests 20-A (upper) and 20-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).

Silicon Nitride Indentations

The raw AE waveforms are illustrated in the graphs below (Figure E.12 to Figure E.14). Those demonstrating significant asymmetry in lateral cracks are indicated with a ‘*’.

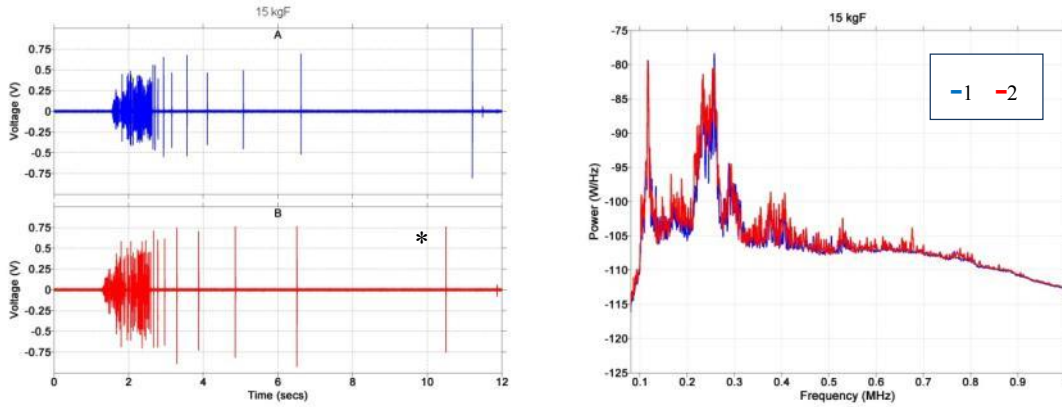


Figure E.12: a) Recorded AE waveforms for Silicon nitride indentation tests 15-A (upper) and 15-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).

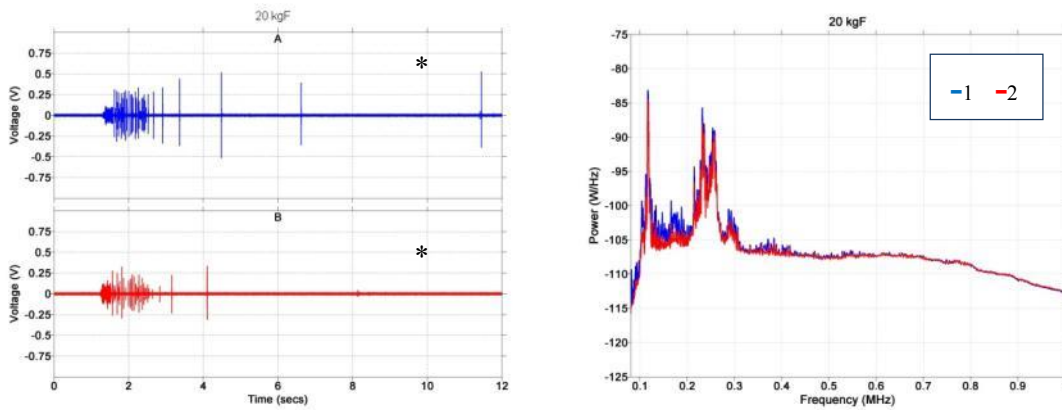


Figure E.13: a) Recorded AE waveforms for Silicon nitride indentation tests 20-A (upper) and 20-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).

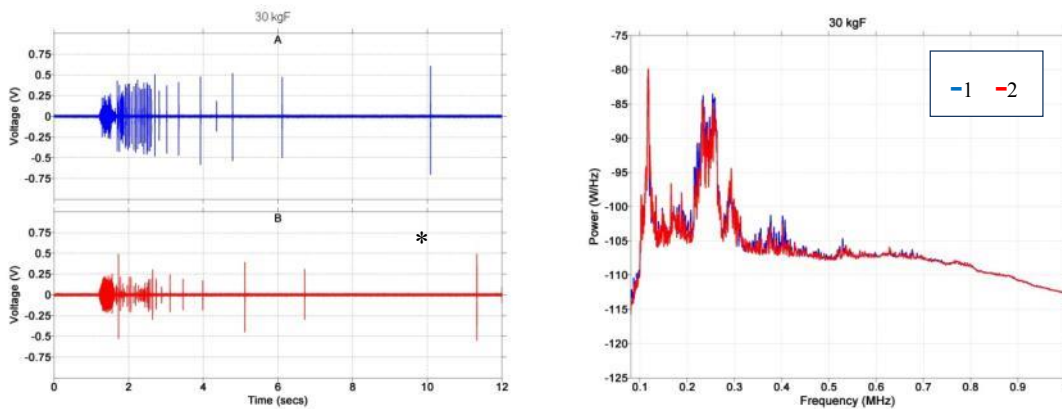


Figure E.14: a) Recorded AE waveforms for Silicon nitride indentation tests 30-A (upper) and 30-B (lower). b) Corresponding PSD estimates of the two waveforms (indicated in the legend).

■ Zoning Silicon Nitride

The zoning method for all silicon nitride indentations are illustrated from the initial detected AE activity above background to indent completion, capturing the entirety of the indentation response.

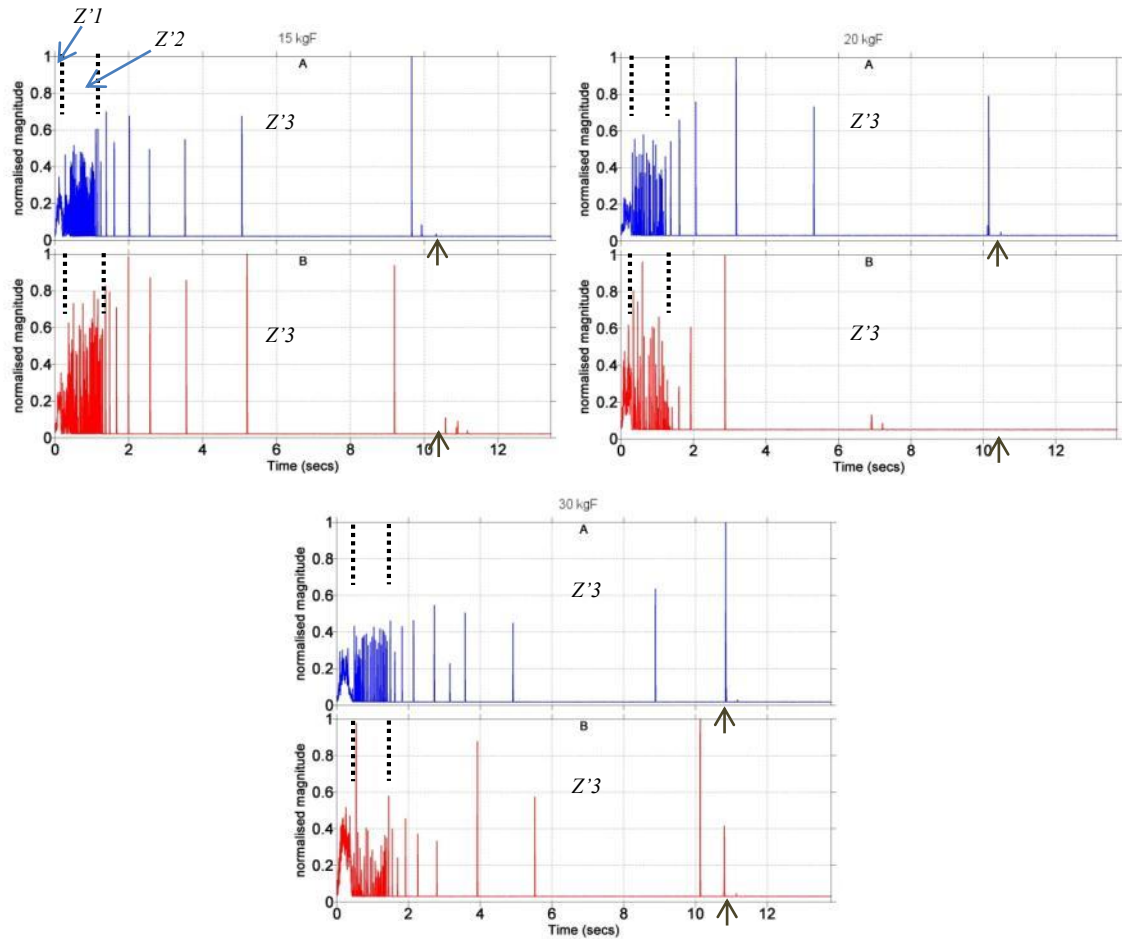


Figure E.15: The smoothed AE waveforms incorporating the total durations of Zones 1 -3 for all silicon nitride indentation tests at each of the three test loads (labelled). Up arrows indicate the full retraction of the indenter tip from the sample

Appendix F: AE Filter Development

Two stages are essential for the setup of the wavelet packet filtering method, these are discussed in the subsections following.

▪ Mother Wavelet Selection

With a variety of mother wavelet functions available, it was necessary to select one based on its ability to separate signal components with maximal amplitude and minimal overlap. To evaluate them, a CWT analysis was performed on the indentation signal (section 6.2.1) with those mother wavelets that demonstrated WPA support. Four wavelet bases, listed below in Table F.1 were selected for the CWT analysis as they were readily available in the MATLAB library of wavelets. A total of 69 mother wavelets were subsequently analysed.

Wavelet Base	Index
Meyer	1
Symlet	1-20
Coiflet	1-5
Daubechies	1-46

Table F.1: Wavelet bases and their corresponding MATLAB index range.

For each mother wavelet, the input waveform was decomposed by the CWT analysis into base wavelets representing 36 scales spanning 0.1 to 1 MHz frequency range with increments of 0.025MHz between them, and for each base wavelet/band computed a coefficient waveform. The respective frequency F_c for each of the wavelet band (scale) s was calculated using the equation:

$$F_c = \frac{F_s F_a}{s} \quad 9.19$$

where F_a is the dominant frequency component of the mother wavelet and F_s the sampling frequency (5 MHz). The suitability of each mother wavelet was evaluated based on the energy concentration and peakedness of the associated set of coefficient waveforms using the kurtosis-to-entropy ratio (KER) which is sensitive to both these features. The KER was calculated as follows:

$$KER = \frac{Kurtosis_{WT}}{Entropy_{WT}} \quad 9.20$$

Where $Kurtosis_{WT}$ and $Entropy_{WT}$ are described by the following equations [203]:

$$Kurtosis_{WT} = \frac{(N) \sum_s \sum_i \left(C(s, i) - \overline{C(s, i)} \right)^4}{\left(\sum_s \sum_i \left(C(s, i) - \overline{C(s, i)} \right)^2 \right)^2} \quad 9.21$$

$$Entropy_{WT} = \sum_{i=1}^N \frac{|C(s,i)|^2}{\sum_{i=1}^N C(s,i)^2} \cdot \log_2 \left(\frac{|C(s,i)|^2}{\sum_{i=1}^N C(s,i)^2} \right) \quad 9.22$$

where $C(s,i)$ describes the coefficient at scale s and datapoint i and N is the number of respective data points per scale.

The results are presented in the diagram below; the mother wavelet demonstrating the highest KER value (db43) was then selected for subsequent WP selection. To illustrate its properties its time-frequency profile was computed and the results are presented below in Figure F.2.

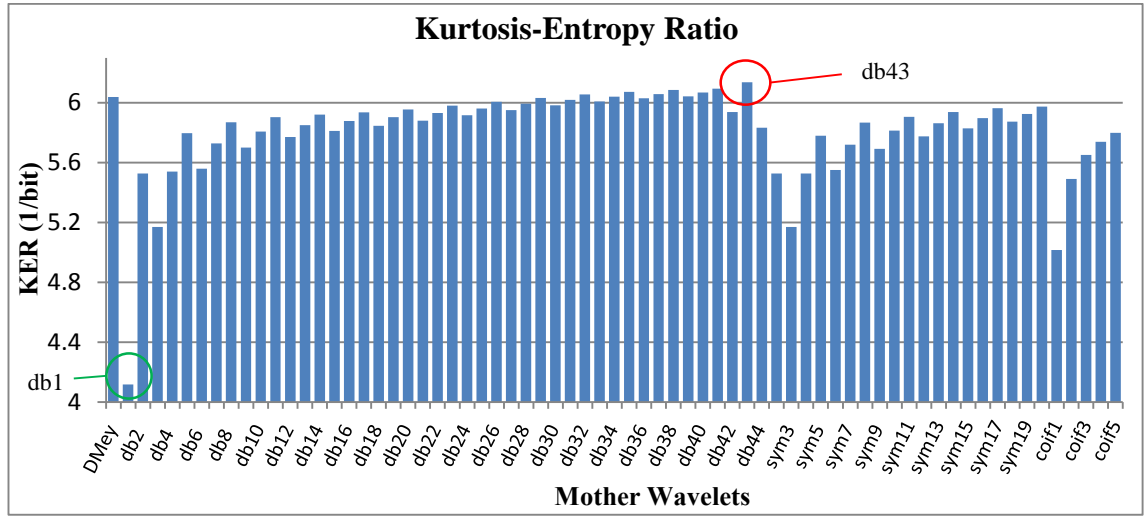


Figure F.1: KER results for various applied mother wavelets, with the wavelets corresponding to the maximum and minimum KER values indicated.

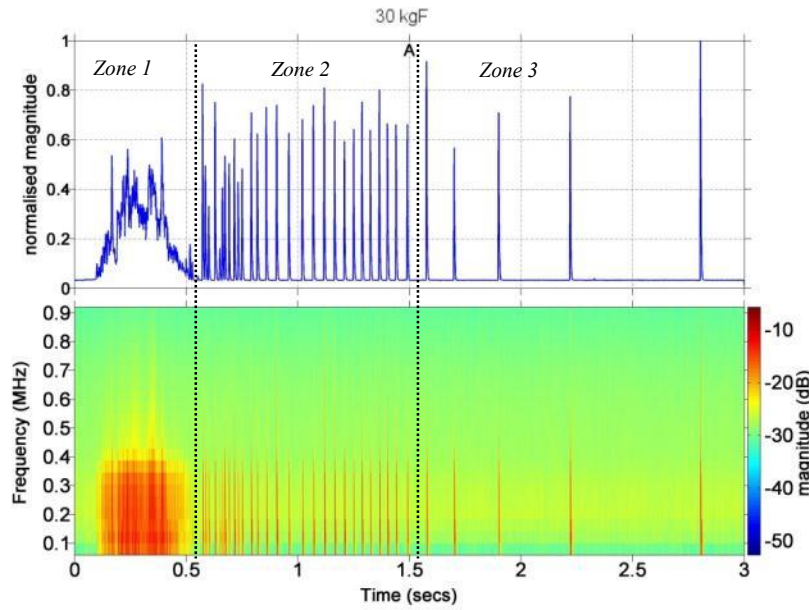


Figure F.2: The smoothed response of 30-A (upper) and the CWT time-frequency representation using the db43 mother wavelet.

High amplitude peaks were observed corresponding well to the crack propagation signal features, and peaks at the 0.1 to 0.4 MHz bandwidth can be distinguished, encompassing the major characteristic peaks highlighted in the PSD profile in Figure 5.13d. Notably no significant variation in the frequency characteristics between zones was observed in agreement with the PSD analyses.

▪ Wavelet Packet Selection

With the mother wavelet function selected, it was necessary to find a packet node representing the optimal filtering function that could separate fault signatures from noise. Gaussian white noise was added to the fault waveform to provide a mixed profile for the selection of the optimal packet node according to its ability to separate the original signal from noise.

WP decomposition was performed using the selected mother wavelet and, for comparative purposes, an additional decomposition was run of the same signal profile using another, reference mother wavelet. The depth of the decomposition (representing the bandwidth) and the relevant packet node (representing the filter centre-frequency) was evaluated based on the energy concentration and peakedness of the corresponding coefficient waveform. The KER method described in equation 3.16 was adopted to evaluate packet coefficients, and a decision tree made at each decomposition node was applied, developing the methods created by Brechet et al [220] and Hemmati [203] and summarised in Figure F.3.

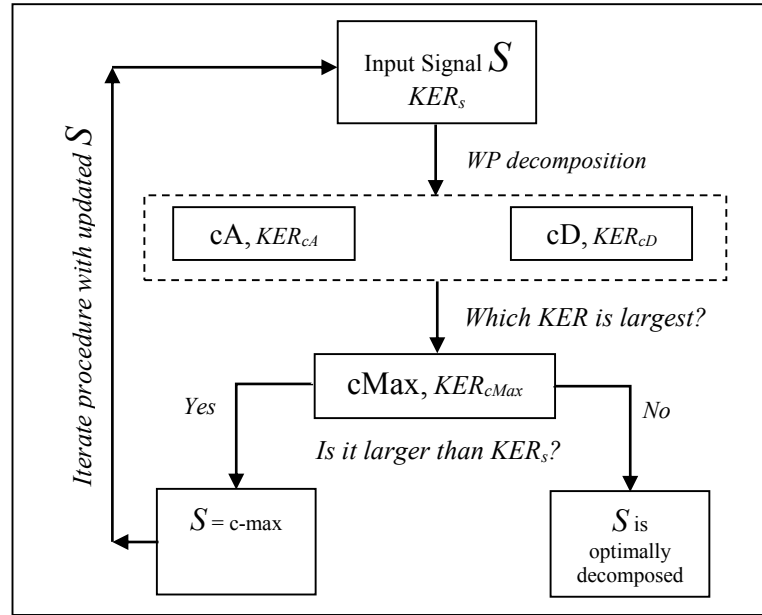


Figure F.3: Illustration of the decision tree applied to wavelet decomposition for recorded AE signals

The original waveform S , whose KER was calculated as KER_s , was decomposed into the approximation (cA) and the detail (cD) coefficient waveforms, and their corresponding KER values calculated (KER_{cA} and KER_{cD} respectively). Subsequently, the larger of the two values (KER_{max}) was selected and compared with KER_s . If $KER_{cMax} < KER_s$, then an optimal decomposition had been reached, identifying the waveform $cMax$ corresponding to KER_{max} as the resulting optimal packet node. If $KER_{cMax} > KER_s$ (meaning that an optimal packet node has not been reached), the above procedure was repeated with $cMax$ replacing the input S waveform until a resulting $cMax$ produced an optimal packet node.

The above process was performed with both selected and reference mother wavelets and their corresponding filtered waveforms, producing the selected and reference filtered waveforms; these are presented in the figure below together with the original and mixed (containing noise) waveforms.

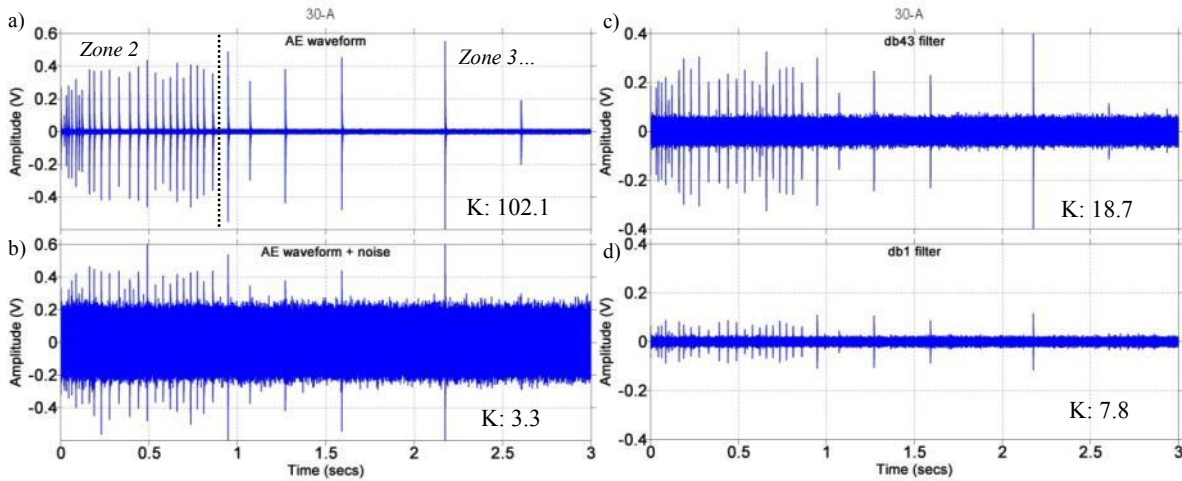


Figure F.4: Illustrations of the WP selection analyses of the raw AE waveform (zone 2 and partial zone 3) a), the waveform with added noise, b) filtered with wavelet db43 at packet (4,6), filtered with wavelet db1 at packet (6,25). Kurtosis (K) values for each of the waveforms are indicated in the bottom right hand corner of their respective graph

The diagrams demonstrate the effectiveness of the filtering process in attenuating the noise (b → c, d). Of the two filtered waveforms the selected one possesses higher kurtosis thus confirming its superiority over the reference waveform.

The spectral profiles of the original (fault) and the selected filtered waveforms are shown in the figure below.

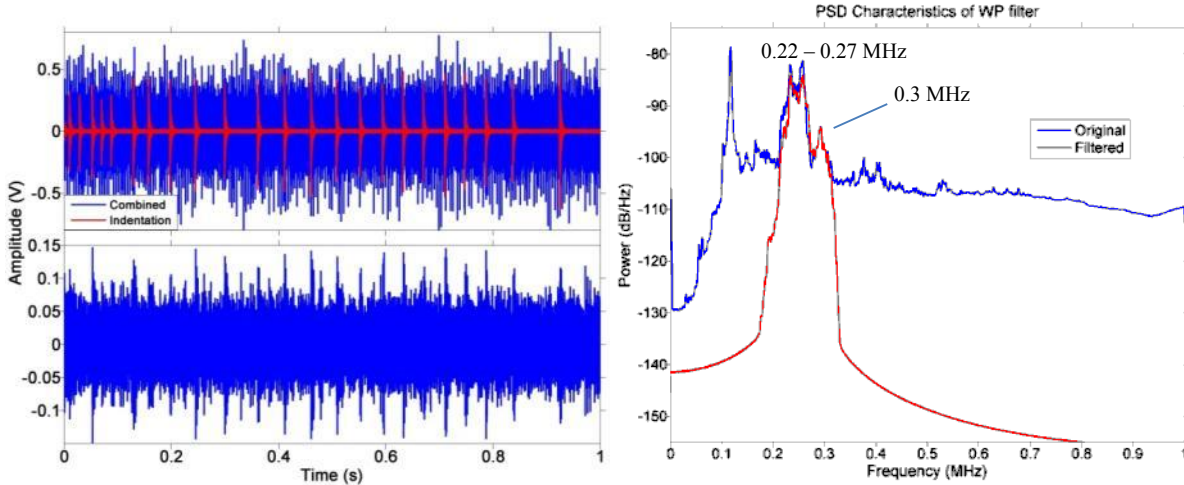


Figure F.5: The PSD characteristic of the optimal WP filter demonstrated on the original signal waveform (see Figure F.4a)

The filter provides a moderate bandwidth capturing accurately the fault peaks at 0.22 – 0.27 and 0.3 MHz.

References

1. Kadin, Y. and A.V. Rychahivskyy, *Modeling of surface cracks in rolling contact*. Materials Science and Engineering: A, 2012. **541**: p. 143-151.
2. Olver, A., *The mechanism of rolling contact fatigue: an update*. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2005. **219**(5): p. 313-330.
3. Dwyer-Joyce, R.S., *Tribological Design Data, Part 3: Contact Mechanics*. . 1997: p. 1-17.
4. Hale, L.C., *Principles and Techniques for Designing Precision Machines. Appendix C: Contact Mechanics*. MIT PhD Thesis, 1999: p. 417-426.
5. Lawn, B.R., *Indentation of Ceramics with Spheres: A Century after Hertz*. Journal of the American Ceramic Society, 1998. **81**(8): p. 1977-1994.
6. Bhushan, B., *Introduction to tribology*. 2013: Wiley.
7. England, G., <http://www.gordonengland.co.uk/hardness/hvconv.htm>. Internet Website, 2012. Independent Metallurgist and Thermal Spray Coating Consultant.
8. Ichikawa, M., et al., *Ring Crack Initiation Load of Silicon Nitride Bearing Balls*. JSME international journal of mechanics and material engineering, 1995. **38**(2): p. 226-230.
9. Levesque, G. and N.K. Arakere, *An investigation of partial cone cracks in silicon nitride balls*. International Journal of Solids and Structures, 2008. **45**(25-26): p. 6301-6315.
10. H. Cai, M., A. Stevens Kalceff, B. R. Lawn., *Deformation and Fracture of Mica-Containing Glass Ceramics in Hertzian Contacts*. Journal of Materials Science, 1994. **9**: p. 762-770.
11. Restall, P. and L. Wang, *Crack Detection for Silicon Nitride Ball Bearings using Acoustic Emission Techniques*. Univ. of Southampton Bachelor Dissertation, 2009.
12. N.K. Myshkin, C.K.K., M.I. Petrokovets, *Introduction to Tribology*. Moon Gak, 1997.
13. Ludema, K.C., *Friction, Wear, Lubrication: A Textbook in Tribology*. 2011.
14. Craig, M., *Advanced Condition Monitoring to Predict Rolling Element Bearing Wear Using Multiple In-Line and Off-Line Sensing*. Univ. of Southampton PhD Thesis, 2010.
15. Dowson, D. and G.R. Higginson, *A numerical solution to the elasto-hydrodynamic problem* Journal of Mechanical Engineering Science, 1959. **1**: p. 6-15.
16. Ahmed, R., *Rolling contact fatigue*. ASM International, 2002. **253**(11): p. 941-956.
17. G. Hamburg, C.C., *Operation of an all-ceramic mainshaft roller bearing in a J-402 Gas-turbine engine*. Lubrication Engineering, 1981. **37**(7): p. 407-415.
18. Shoda, Y., et al., *The Performance of a Hybrid Ceramic Ball Bearing Under High Speed Conditions with the Under-Race Lubrication Method*. Tribology Transactions, 1997. **40**(4): p. 676 - 684.
19. Chudecki, J.F., *Ceramic Bearings- applications and performance advantages in industrial application*. SAE Technical Paper Series SAE-TP-89104., 1989.
20. Nosaka, M., et al., *Tribo-Characteristics of Cryogenic Hybrid Ceramic Ball Bearings for Rocket Turbopumps: Bearing Wear and Transfer Film*. Tribology transactions, 1999. **42**(1): p. 106-115.
21. Aramaki, H., et al., *The Performance of Ball Bearings With Silicon Nitride Ceramic Balls in High Speed Spindles for Machine Tools*. Journal of Tribology, 1988. **110**(4): p. 693-698.
22. Schrader, S.M., *Performance of a hybrid cylindrical roller bearing*. STLE and ASME Tribology Conference, 1992. **48**(8): p. 665-672.
23. Nishimura, M. and M. Suzuki, *Solid-lubricated ball bearings for use in a vacuum -- state-of-the-art*. Tribology International, 1999. **32**(11): p. 637-647.
24. Shingo, O. and S. Mineo, *Long-Term Operation of Si₃N₄ Ball Bearings at Temperatures up to 650°C in Ultra-High Vacuum*. Tribology Transactions, 1997. **40**(1): p. 31 - 40.
25. Wang, L., R.W. Snidle, and L. Gu, *Rolling contact silicon nitride bearing technology: a review of recent research*. Wear, 2000. **246**(1-2): p. 159-173.

References

26. Ziegler, G., J. Heinrich, and G. Wötting, *Relationships between processing, microstructure and properties of dense and reaction-bonded silicon nitride*. Journal of Materials Science, 1987. **22**(9): p. 3041-3086.
27. Cláudio, C.V. and C.A. Costa, *Fracture toughness of Si₃N₄ processed by gas pressure sintering and hot pressing*. Materials Research, 2006. **9**(2): p. 143-146.
28. Zhou, L., T. Tachibana, and K. Syoji, *Form mirror grinding of inner race of ceramic bearing*. Trans. Jpn. Soc. Mech. Eng., 1997. **612**(3): p. 2905-2910.
29. Hadfield, M. and S. Tobe, *Residual stress measurements of hot isostatically pressed silicon nitride rolling elements*. Ceramics International, 1998. **24**(5): p. 387-392.
30. Jisheng, E., T.A. Stolarski, and D.T. Gawne, *Tribochemically assisted wear of silicon nitride ball*. Journal of the European Ceramic Society, 1996. **16**(1): p. 25-34.
31. Yuan, J., et al., *Research on abrasives in the chemical-mechanical polishing process for silicon nitride balls*. Journal of materials processing technology, 2002. **129**(1): p. 171-175.
32. J. R. Miner, et al., *F117-PW-100 Hybrid Ball Bearing Ceramic Technology Insertion*. Jour. of Eng. for Gas Turbines and Power, 1996. **118**: p. 434-442.
33. Lee, S.K., S. Wuttiaphan, and B.R. Lawn, *Role of Microstructure in Hertzian Contact Damage in Silicon Nitride: I, Mechanical Characterization*. Journal of the American Ceramic Society, 1997. **80**(9): p. 2367-2381.
34. Fischer-Cripps, A.C. and B.R. Lawn, *Stress Analysis of Contact Deformation in Quasi-Plastic Ceramics*. Journal of the American Ceramic Society, 1996. **79**(10): p. 2609-2618.
35. Burrier, H.I., *Optimizing the Structure and Properties of Silicon Nitride for Rolling Contact Bearing Performance*. Tribology Transactions, 1996. **39**(2): p. 276 - 285.
36. Allen, D.L., *Effect of Composition and Physical Properties of Silicon Nitride on Rolling Wear and Fatigue Performance*. Tribology Transactions, 1994. **37**(2): p. 410-414.
37. Li, M. and M. Sakai, *Mixed-Mode Fracture of Ceramics in Asymmetric Four-Point Bending: Effect of Crack-Face Grain Interlocking/Bridging*. Journal of the American Ceramic Society, 1996. **79**(10): p. 2718-2726.
38. Iwasa, M. and Y. Taibana, *Friction and wear of ceramics measured by a pin-on-disk tester*, . International Journal of High Technology Ceramics, 1986. **2**(4): p. 319-320.
39. Skopp, A., M. Woydt, and K.H. Habig, *Tribological behavior of silicon nitride materials under unlubricated sliding between 22°C and 1000°C*. Wear, 1995. **181-183**(Part 2): p. 571-580.
40. Chen, Z., et al., *Damage processes in Si₃N₄ bearing material under contact loading*. Wear, 1996. **198**(1-2): p. 197-207.
41. Braza, J.F., H.S. Cheng, and M.E. Fine, *Silicon Nitride Wear Mechanisms: Rolling and Sliding Contact*. Tribology Transactions, 1989. **32**(4): p. 439 - 446.
42. Zhou, J.L., et al., *Tensile stress fatigue life model of silicon nitride ceramic balls*. Tribology International, 2009. **42**(11-12): p. 1838-1845.
43. Hadfield, M., et al., *Failure modes of pre-cracked ceramic elements under rolling contact*. Wear, 1993. **169**(1): p. 69-75.
44. Deneuville, F., et al., *High frequency ultrasonic detection of C-crack defects in silicon nitride bearing balls*. Ultrasonics, 2009. **49**(1): p. 89-93.
45. Bashkansky, M., et al., *Subsurface detection and characterization of Hertzian cracks in Si₃N₄ balls using optical coherence tomography*. NDT & E International, 2001. **34**(8): p. 547-555.
46. Kida, K., et al., *Surface crack growth of silicon nitride bearings under rolling contact fatigue*. Fatigue & Fracture of Engineering Materials & Structures, 2004. **27**(8): p. 657-668.
47. Niihara, K., *A fracture mechanics analysis of indentation-induced Palmqvist crack in ceramics*. Journal of materials science letters, 1983. **2**(5): p. 221-223.
48. Lawn, B.R., A.G. Evans, and D.B. Marshall, *Elastic/Plastic Indentation Damage in Ceramics: The Median/Radial Crack System*. Journal of the American Ceramic Society, 1980. **63**(9-10): p. 574-581.

References

49. Hadfield, M. and T.A. Stolarski, *Observations of delamination fatigue on pre-cracked ceramic elements in rolling contact*. Ceramics International, 1995. **21**(2): p. 125-130.
50. Kida, K., *Flaking Failure in Silicon Nitride under Reciprocating Rolling Contact Fatigue*. Advanced Materials Research, 2011. **217**: p. 866-873.
51. Kida, K., M. Saito, and K. Kitamura, *Flaking failure originating from a single surface crack in silicon nitride under rolling contact fatigue*. Fatigue & Fracture of Engineering Materials & Structures, 2005. **28**(12): p. 1087-1097.
52. Wang, Y. and M. Hadfield, *The influence of ring crack location on the rolling contact fatigue failure of lubricated silicon nitride: experimental studies*. Wear, 2000. **243**(1-2): p. 157-166.
53. Zhao, P., et al., *Subsurface propagation of partial ring cracks under rolling contact: Part I. Experimental studies*. Wear, 2006. **261**(3-4): p. 382-389.
54. Levesque, G.A. and N.K. Arakere, *Critical Flaw Size in Silicon Nitride Ball Bearings*. Tribology Transactions, 2010. **53**(4): p. 511-519.
55. Hadfield, M., et al., *Failure modes of ceramic elements with ring-crack defects*. Tribology International, 1993. **26**(3): p. 157-164.
56. Cundill, R.T., *Impact resistance of silicon nitride balls*. Ceramic Materials and Components for Engines, 6th International Symposium, 1997: p. 556-561.
57. Adachi, K., K. Kato, and N. Chen, *Wear map of ceramics*. Wear, 1997. **203-204**: p. 291-301.
58. Takadom, J., H. Houmid-Bennani, and D. Mairey, *The wear characteristics of silicon nitride*. Journal of the European Ceramic Society, 1998. **18**(5): p. 553-556.
59. Cranmer, D.C., *Wear surface analysis of silicon nitride*. Lubr. Eng.:(United States), 1988. **44**(12): p. 975-980.
60. Fischer, T.E., et al., *Genesis and role of wear debris in sliding wear of ceramics*. Wear, 2000. **245**(1-2): p. 53-60.
61. Chao, L.Y., et al., *Transient wear of silicon nitride in lubricated rolling contact*. Wear, 1998. **223**(1-2): p. 58-65.
62. Akdogan, G. and T.A. Stolarski, *Wear in metal/silicon nitride sliding pairs*. Ceramics International, 2003. **29**(4): p. 435-446.
63. Wang, L., et al., *Wear performance of oil lubricated silicon nitride sliding against various bearing steels*. Wear, 2003. **255**(1-6): p. 657-668.
64. Denape, J., A. Marzinotto, and J.A. Petit, *Roughness effect of silicon nitride sliding on steel under boundary lubrication*. Wear, 1992. **159**(2): p. 173-184.
65. Akazawa, M., K. Kato, and K. Umeya, *Wear properties of silicon nitride in rolling contact*. Wear, 1986. **110**: p. 285-293.
66. Asada, S., et al., *Rolling Contact Fatigue of Engineering Ceramics*. Solid State Phenomena, 1992. **25**: p. 627-634.
67. Lakshminarayanan, R., et al., *Wear of steel in rolling contact with silicon nitride*. Wear, 1997. **210**(1-2): p. 278-286.
68. Mitchell, D.J., J.J. Mecholsky Jr, and J.H. Adair, *All-steel and Si₃N₄-steel hybrid rolling contact fatigue under contaminated conditions*. Wear, 2000. **239**(2): p. 176-188.
69. Kang, J., M. Hadfield, and R. Ahmed, *The effects of material combination and surface roughness in lubricated silicon nitride/steel rolling contact fatigue*. Materials & Design, 2003. **24**(1): p. 1-13.
70. Rosado, L., et al., *Rolling Contact Fatigue Life and Spall Propagation of AISI M50, M50NiL, and AISI 52100, Part I: Experimental Results*. Tribology Transactions, 2010. **53**(1): p. 29-41.
71. Arakere, N.K., et al., *Rolling Contact Fatigue Life and Spall Propagation of AISI M50, M50NiL, and AISI 52100, Part II: Stress Modeling*. Tribology Transactions, 2010. **53**(1): p. 42-51.
72. Burrier Jr, H.I. and C. Burk, *Ceramic Bearing Development. Vol. 1, Fatigue and Wear Behavior of NBD-200 Silicon Nitride Balls*, 1995, DTIC Document.
73. Zaretsky, E.V., B.L. Vlcek, and R.C. Hendricks, *Effect of silicon nitride balls and rollers on rolling bearing life*. Tribology transactions, 2005(48): p. 425-435.

References

74. Forster, N.H., et al., *Rolling Contact Fatigue Life and Spall Propagation Characteristics of AISI M50, M50 NiL, and AISI 52100, Part III: Metallurgical Examination* Tribology Transactions, 2010. **53**(1): p. 52-59.
75. Mosleh, M., et al., *Fatigue failure of all-steel and steel-silicon nitride rolling ball combinations*. Wear, 2011. **271**(9-10): p. 2471-2476.
76. Parker, R.J. and E.V. Zaretsky, *Fatigue life of high-speed ball bearings with silicon nitride balls*. Journal of Lubrication Technology-Transactions of the Asme, 1975. **97**(3): p. 350-357.
77. O'Brien, M.J., N. Presser, and E.Y. Robinson, *Failure analysis of three Si₃N₄ balls used in hybrid bearings*. Engineering Failure Analysis, 2003. **10**(4): p. 453-473.
78. Chao, L.-Y., et al., *Development of silicon nitride for rolling-contact bearing applications: a review*. Journal of Materials Education, 1995. **17**: p. 245-304.
79. Wang, Y. and M. Hadfield, *The influence of ring crack location on the rolling contact fatigue failure of lubricated silicon nitride: fracture mechanics analysis*. Wear, 2000. **243**(1-2): p. 167-174.
80. Zhao, P., et al., *The influence of test lubricants on the rolling contact fatigue failure mechanisms of silicon nitride ceramic*. Wear, 2004. **257**(9-10): p. 1047-1057.
81. Wang, Y. and M. Hadfield, *Rolling contact fatigue failure modes of lubricated silicon nitride in relation to ring crack defects*. Wear, 1999. **225-229**(Part 2): p. 1284-1292.
82. Wang, Y. and M. Hadfield, *A study of line defect fatigue failure of ceramic rolling elements in rolling contact*. Wear, 2002. **253**(9-10): p. 975-985.
83. Hadfield, M., T.A. Stolarski, and R.T. Cundill, *Delamination of ceramic balls in rolling contact*. Ceramics International, 1993. **19**(3): p. 151-158.
84. Hadfield, M., T.A. Stolarski, and R.T. Cundill, *Failure Modes of Ceramics in Rolling Contact*. Proceedings: Mathematical and Physical Sciences, 1993. **443**(1919): p. 607-621.
85. Hadfield, M., *Failure of silicon nitride rolling elements with ring crack defects*. Ceramics International, 1998. **24**(5): p. 379-386.
86. Pattabhiraman, S., et al., *Uncertainty analysis for rolling contact fatigue failure probability of silicon nitride ball bearings*. International Journal of Solids and Structures, 2010. **47**(18-19): p. 2543-2553.
87. Hamilton, A. and F. Quail, *Detailed State of the Art Review for the Different Online/Inline Oil Analysis Techniques in Context of Wind Turbine Gearboxes*. Journal of Tribology, 2011. **133**(4).
88. Whittington, H.W., B.W. Flynn, and G.H. Mills, *An online wear debris monitor*. Measurement Science and Technology, 1992(7): p. 656-661.
89. Edmonds, J., M.S. Resner, and K. Shkarlet, *Detection of precursor wear debris in lubrication systems*. Aerospace Conference Proceedings, 2000 IEEE, 2000. **6**: p. 73-77.
90. Miller, J.L. and D. Kitaljevich, *In-line oil debris monitor for aircraft engine condition assessment*. Aerospace Conference Proceedings, 2000 IEEE, 2000. **6**: p. 49-56.
91. Dempsey, P.J., J.M. Certo, and W. Morales, *Current status of hybrid bearing damage detection*, in *60th Annual Meeting of the Society-of-Tribologists-and-Lubrication-Engineers2005*, Taylor & Francis Inc: Las Vegas, NV. p. 370-376.
92. Dempsey, P.J., et al. *Hybrid Bearing Prognostic Test Rig*. in *Society of Tribologists and Lubrication Engineers Annual Meeting and Exhibition*. 2005. Cleveland, OH.
93. Sperring, T.P.N., T. J., *SYCLOPS-a qualitative debris classification system developed for RAF early failure detection centres*. Tribology International, 2005. **38**(10): p. 898-903.
94. Williamson, M., *The Low-Down on Particle Counters*. Practicing Oil Analysis Magazine, 2002.
95. Chendong, D., H. Zhengjia, and J. Hongkai, *A sliding window feature extraction method for rotating machinery based on the lifting scheme*. Journal of Sound and Vibration, 2007. **299**(4-5): p. 774-785.
96. Choudhury, A. and N. Tandon, *Vibration Response of Rolling Element Bearings in a Rotor Bearing System to a Local Defect Under Radial Load*. Journal of Tribology, 2006. **128**(2): p. 252-261.

References

97. Hoffman, A.J. and N.T. van der Merwe, *The application of neural networks to vibrational diagnostics for multiple fault conditions*. Computer Standards & Interfaces, 2002. **24**(2): p. 139-149.
98. J.B. Catlin, J., *The use of ultrasonic diagnostic techniques to detect rolling element bearing defects*. Proceedings of Machinery Vibration Monitoring and Analysis Meeting , Vibration Institute, Houston, 1983: p. 123-130.
99. Jack, L. and A. Nandi, *Support vector machines for detection and characterization of rolling element bearing faults*. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2001. **215**(9): p. 1065-1074.
100. Lin, J. and L. Qu, *Feature Extraction Based on Morlet Wavelet and its Application for Mechanical Fault Diagnosis*. Journal of Sound and Vibration, 2000. **234**(1): p. 135-148.
101. McFadden, P.D. and J.D. Smith, *Vibration monitoring of rolling element bearings by the high-frequency resonance technique — a review*. Tribology International, 1984. **17**(1): p. 3-10.
102. Morhain, A. and D. Mba, *Bearing defect diagnosis and acoustic emission*. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2003. **217**(4): p. 257-272.
103. Nikolaou, N.G. and I.A. Antoniadis, *Demodulation of Vibration Signals Generated by Defects in Rolling Element Bearings using Complex shifted Morlet Wavelets*. Mechanical Systems and Signal Processing, 2002. **16**(4): p. 677-694.
104. Nikolaou, N.G. and I.A. Antoniadis, *Rolling element bearing fault diagnosis using wavelet packets*. NDT&E International, 2002. **35**(3): p. 197-205.
105. Peng, Z., F. Chu, and Y. He, *Vibration Signal Analysis and Feature Extraction Based on Reassigned Wavelet Scalogram*. Journal of Sound and Vibration, 2002. **253**(5): p. 1087-1100.
106. Sawalhi, N. and R.B. Randall, *Vibration response of spalled rolling element bearings: Observations, simulations and signal processing techniques to track the spall size*. Mechanical Systems and Signal Processing, 2011. **25**(3): p. 846–870.
107. Shiroishi, J., Li, S., Kurfees, T., Danyulk, S., *Bearing condition diagnostics via vibration and acoustic emission measurements*. Mechanical Systems and Signal Processing, 1997. **11**(5): p. 693-705.
108. Smith, C., et al., *An approach to vibration analysis using wavelets in an application of aircraft health monitoring*. Mechanical Systems and Signal Processing, 2007. **21**(3): p. 1255-1272.
109. Tandon, N. and A. Choudhury, *A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings*. Tribology International, 1999. **32**(8): p. 469-480.
110. Tse, P.W., Y.H. Peng, and R. Yam, *Wavelet Analysis and Envelope Detection For Rolling Element Bearing Fault Diagnosis---Their Effectiveness and Flexibilities*. Journal of Vibration and Acoustics, 2001. **123**(3): p. 303-310.
111. Wang, W.J. and P.D. McFadden, *Application of Wavelets to Gearbox Vibration Signals for Fault Detection*. Journal of Sound and Vibration, 1996. **192**(5): p. 927-939.
112. Williams, T., et al., *Rolling element bearing diagnostics in run-to-failure lifetime testing*. Mechanical Systems and Signal Processing, 2001. **15**(5): p. 979-993.
113. Yan, R. and R. Gao, *Base Wavelet Selection For Bearing Vibration Signal Analysis*. International Journal of Wavelets, Multiresolution and Information Processing, 2009. **07**(04): p. 411-426.
114. Bin, Z., et al. *Rolling element bearing feature extraction and anomaly detection based on vibration monitoring*. in *Control and Automation, 2008 16th Mediterranean Conference on*. 2008.
115. Mohammadi, A. and M. Safizadeh, *Bearing Multiple Defects Detection Based on Envelope Detector Time Constant*. Journal of Tribology, 2013. **135**: p. 011102.
116. Choudhury, A. and N. Tandon, *Application of acoustic emission technique for the detection of defects in rolling element bearings*. Tribology International, 2000. **33**(1): p. 39-45.

References

117. Donnellan, P., *Condition Monitoring - Engineering the practice*. Industrial Lubrication and Tribology, 2002. **54**(6): p. 55-69.
118. Wowk, V., *Machinery Vibration: Measurement and analysis*. Mc Graw - Hill Professional, 1991: p. 65.
119. Dyer, D. and R.M. Stewart, *Detection of Rolling Element Bearing Damage by Statistical Analysis*. ASME J. Mech. Des., 1978. **100**: p. 229-235.
120. Broderick, J., R. Burchill, and H. Clark, *Design and fabrication of prototype system for early warning of impending bearing failure*. 1972.
121. Darlow, M.S., R.H. Badgley, and G. Hogg, *Application of High-Frequency Resonance Techniques for Bearing Diagnostics in Helicopter Gearboxes*, 1974, DTIC Document.
122. J. Nurse, C. Petch, and E. Fisher, *Engine gas path integrity monitoring*. Aerotech, 1994. **52**(6): p. C470.
123. H. E. G. Powrie and K. McNicholas, *Gas path condition monitoring during accelerated mission testing of a demonstrator engine*. 1997.
124. Cartwright, R.A., Fisher, C., *Marine gas turbine condition monitoring by gas path electrostatic techniques*. International Gas Turbine and Aero Engine Congress and Exposition, 1991.
125. Booth, J.E., et al., *Scuffing detection of TU3 cam-follower contacts by electrostatic charge condition monitoring*. Tribology International, 2008. **43**(1-2): p. 113-128.
126. Harvey, T.J., et al., *Real-time monitoring of wear debris using electrostatic sensing techniques*. Journal of Engineering Tribology, 2007. **221**(1): p. 27-40.
127. Powrie, H., et al. *Electrostatic charge generation associated with machinery component deterioration*. in *Aerospace Conference Proceedings*, 2002. IEEE. 2002.
128. Tasbaz, O.D., et al., *Electrostatic monitoring of oil lubricated sliding point contacts for early detection of scuffing*. Wear, 1999. **230**(1): p. 86-97.
129. Morris, S., Wood, R.J.K., Harvey, T.J., Powrie, H.E.G., *Use of electrostatic charge monitoring for early detection of adhesive wear in oil lubricated contacts*. Tribology: Transactions of the ASME, 2002. **124**(2): p. 288-296.
130. Harvey, T.J., et al., *Investigation of electrostatic charging mechanisms in oil lubricated tribo-contacts*. Tribology International, 2002. **35**(9): p. 605-614.
131. James, A.M. and M.P. Lord, *VNR index of chemical and physical data*. 1992: Van Nostrand Reinhold.
132. Morris, S., *Real-time electrostatic charge monitoring of the wear surfaces and debris generated by sliding bearing steel contacts*. Univ. of Southampton PhD Thesis, 2003.
133. Coehn, A., *Annals of Physics*, 1898. **64**(217).
134. Nakayama, K. and H. Hashimoto, *Triboemission of charged particles and photons from wearing ceramic surfaces in various gases*. Tribology transactions, 1992. **35**(4): p. 643-650.
135. Nakayama K, H.H. and Y. Fukuda, *Triboemission of charged particles and photons from solid surfaces during frictional damage*. 1990: p. 1141.
136. K, N., H. H, and F. Y, *Triboemission of charged particles and photons from solid surfaces during frictional damage*. J Phy D Applied Physics, 1992. **25**(2): p. 303-308.
137. Nakayama, K. and H. Hashimoto, *Triboemission of charged particles and photons from ceramic surfaces in various gases*. Tribology transactions, 1992. **35**(4): p. 643-650.
138. Enomoto, Y. and M.M. Chaudhri, *Fracto-emission during Fracture of Engineering Ceramics*. Journal of the American Ceramic Society, 1993. **76**(10): p. 2583-2587.
139. Hammer, E.A. and R.G. Green, *The spatial filtering effect of capacitance transducer electrodes (flow measurement)*. Journal of Physics E: Scientific Instruments, 1983. **16**(5): p. 438-444.
140. Zhe, K., W. Xiao-lei, and Z. Shu-jiang, *Study on the spatial filtering and sensitivity characteristic of inserted electrostatic sensors for the measurement of gas-solid two-phase flow parameters*. Flow Measurement and Instrumentation, 2013. **30**(0): p. 26-33.
141. H. E. G. Powrie, O. D. Tasbaz, and R. J. K. Wood, *Performance of an electrostatic oil monitoring system during an FZG gear scuffing test*, in *Proceedings of the International Conference on Condition Monitoring* 1999: University of Wales, Swansea.

References

142. Wang, L., et al., *Feasibility of using electrostatic monitoring for oil lubricated ceramic to steel sliding contacts*, in *Tribology and Interface Engineering Series*. 2004, Elsevier. p. 625-635.
143. Booth, J.E., et al., *The feasibility of using electrostatic monitoring to identify diesel lubricant additives and soot contamination interactions by factorial analysis*. Tribology International, 2006. **39**(12): p. 1564-1575.
144. Jun, S., et al., *Wear monitoring of bearing steel using electrostatic and acoustic emission techniques*. Wear, 2005. **259**(7-12): p. 1482-1489.
145. Wang, L., et al., *Electrostatic wear sensing of ceramic-steel lubricated contacts*, in *Tribology and Interface Engineering Series*. 2003, Elsevier. p. 711-720.
146. H.E.G. Powrie, R.J.K.W., T.J. Harvey, S. Morris, , *Re-analysis of electrostatic wear-site sensor data from FZG gear scuffing tests*. Condition Monitor, 2001. **177** p. 6-12
147. Harvey, T.J., R.J.K. Wood, and H.E.G. Powrie, *Electrostatic wear monitoring of rolling element bearings*. Wear, 2007. **263**(7-12): p. 1492-1501.
148. Craig, M., et al., *Advanced condition monitoring of tapered roller bearings, Part 1*. Tribology International, 2009. **42**(11-12): p. 1846-1856.
149. Watson, M., et al. *A comprehensive high frequency vibration monitoring system for incipient fault detection and isolation of gears, bearings and shafts/couplings in turbine engines and accessories*. in *ASME Turbo Expo*. 2007.
150. Bashir I., B.R., Probert D., *Release of Acoustic Energy during the Fatiguing of a Rolling-element Bearing*. Appl. Energ., 1999. **62**: p. 97-111.
151. Raja Hamzah, R.I. and D. Mba, *The influence of operating condition on acoustic emission (AE) generation during meshing of helical and spur gear*. Tribology International, 2009. **42**(1): p. 3-14.
152. Wevers, M. and K. Lambrighs, *Applications of Acoustic Emission for SHM: A Review*. Encyclopedia of Structural Health Monitoring. 2009: John Wiley & Sons, Ltd.
153. Meyers, M., *Dynamic behaviour of Materials*. 1994.
154. Physical Acoustics Corporation, *AEWin Software Users Manual*. Manual, 2007(Rev 1.90).
155. Faisal, N., R. Reuben, and R. Ahmed, *An improved measurement of Vickers indentation behaviour through enhanced instrumentation*. Measurement Science and Technology, 2011. **22**(1): p. 015703.
156. Kim, K.Y. and W. Sachse, *Study of brittle fracture by acoustic emission from indentation cracks*. Journal of applied physics, 1989. **65**(11): p. 4234-4244.
157. Sreekala, S. and G. Ananthakrishna, *Acoustic emission and shape memory effect in the martensitic transformation*. Physical review letters, 2003. **90**(13): p. 135501.
158. Boness, R.J., S.L. McBride, and M. Sobczyk, *Wear studies using acoustic emission techniques*. Tribology International, 1990. **23**(5): p. 291-295.
159. Al-Ghamd, A.M. and D. Mba, *A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size*. Mechanical Systems and Signal Processing, 2006. **20**(7): p. 1537-1571.
160. Wang, L. and R.J.K. Wood, *Acoustic emissions from lubricated hybrid contacts*. Tribology International, 2009. **42**(11-12): p. 1629-1637.
161. Boness, R.J. and S.L. McBride, *Adhesive and abrasive wear studies using acoustic emission techniques*. Wear, 1991. **149**(1-2): p. 41-53.
162. Yongyong He, X.Z., Michael I. Friswell, *Defect Diagnosis for Rolling Element Bearings Using Acoustic Emission*. Journal of Vibration and Acoustics, 2009. **131**.
163. Ichikawa, M., et al., *Intra-Ball and Inter-Ball Variability of Ring Crack Initiation Load of Silicon Nitride Bearing Balls*. JSME international journal. Ser. A, Mechanics and material engineering, 1995. **38**(2): p. 231-235.
164. Rahman, Z., et al., *Incipient damage detection and its propagation monitoring of rolling contact fatigue by acoustic emission*. Tribology International, 2009. **42**(6): p. 807-815.
165. Mechefske, C.K., G. Sun, and J. Sheasby, *Using acoustic emission to monitor sliding wear*. Insight, 2002. **44**.

References

166. Faisal, N.H., et al., *The Use of Acoustic Emission to Characterize Fracture Behavior During Vickers Indentation of HVOF Thermally Sprayed WC-Co Coatings*. Journal of Thermal Spray Technology, 2009. **18**(4): p. 525-535.
167. Al-Balushi, K.R., *The use of high stress waves for monitoring failing gears*. PhD dissertation Cranfield University, 1995.
168. Al-Balushi, et al., *Energy Index technique for detection of Acoustic Emissions associated with incipient bearing failures*. Applied Acoustics, 2010. **71**(9): p. 812-821.
169. Price, E.D., A.W. Lees, and M.I. Friswell, *Detection of severe sliding and pitting fatigue wear regimes through the use of broadband acoustic emission*. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2005. **219**(2): p. 85-98.
170. Navarro, E., et al., *Response of plasma-sprayed alumina-titania composites to static indentation process*. Materials Letters, 1999. **41**(5): p. 234-240.
171. Faisal, N. and R. Ahmed, *Acoustic emission analysis of Vickers indentation fracture of cermet and ceramic coatings*. Measurement Science and Technology, 2011. **22**(12): p. 125704.
172. J. W. Cooley, J.W.T., *An Algorithm for the Machine Computation of the Complex Fourier Series*. Mathematics of Computation, 1965. **19**: p. 297-301.
173. Welch, P., *The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms*. Audio and Electroacoustics, IEEE Transactions on, 1967. **15**(2): p. 70-73.
174. MathWorks Documentation Centre, *Welch Method*. Internet Website, 2012.
175. Dowling, M.J. *Application of non-stationary analysis to machinery monitoring*. in *IEEE International Conference on Acoustics, Speech, and Signal Processing*. 1993.
176. Chi, P.J. and C.T. Russell, *Use of the Wigner-Ville distribution in interpreting and identifying ULF waves in triaxial magnetic records*. J. Geophys. Res., 2008. **113**(1): p. 121-128.
177. Liu, W. and B. Tang, *A hybrid time-frequency method based on improved Morlet wavelet and auto terms window*. Expert Systems with Applications, 2011. **38**(6): p. 7575-7581.
178. Zhang, Y., Y.-J. Kwon, and H.J. Lee, *A systematic generation of initial register-reuse chains for dependence minimization*. SIGPLAN, 2001. **36**(2): p. 47-54.
179. Hang, H. *Time-frequency DOA estimate algorithm based on SPWVD*. in *Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications, 2005. MAPE 2005. IEEE International Symposium on*. 2005.
180. Hammond, J.K. and P.R. White, *The Analysis of Non-Stationary Signals Using Time-Frequency Methods*. Journal of Sound and Vibration, 1996. **190**(3): p. 419-447.
181. Khandan, F. and A. Ayatollahi. *Performance region of center affine filter for eliminating of interference terms of discrete Wigner distribution*. in *Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the 3rd International Symposium on*. 2003.
182. Sattar, F. and G. Salomonsson, *The use of a filter bank and the Wigner-Ville distribution for time-frequency representation*. Signal Processing, IEEE Transactions on, 1999. **47**(6): p. 1776-1783.
183. Tang, B., W. Liu, and T. Song, *Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner-Ville distribution*. Renewable Energy, 2010. **35**(12): p. 2862-2866.
184. Cohen, I., S. Raz, and D. Malah. *Adaptive time-frequency distributions via the shift-invariant wavelet packet decomposition*. in *Time-Frequency and Time-Scale Analysis, 1998. Proceedings of the IEEE-SP International Symposium on*. 1998.
185. Li, H., H. Zheng, and L. Tang, *Wigner-Ville Distribution Based on EMD for Faults Diagnosis of Bearings*, L. Wang, et al., Editors. 2006, Springer Berlin / Heidelberg: Fuzzy Systems and Knowledge Discovery. p. 803-812.
186. Hui, L. and Z. Yuping. *Bearing Faults Diagnosis Based on EMD and Wigner-Ville Distribution*. in *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*. 2006.

References

187. Peng, Z.K. and F.L. Chu, *Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography*. Mechanical Systems and Signal Processing, 2004. **18**(2): p. 199-221.
188. Kim, B.S., et al., *A comparative study on damage detection in speed-up and coast-down process of grinding spindle-typed rotor-bearing system*. Journal of Materials Processing Technology, 2007. **187–188**(0): p. 30-36.
189. Grossmann, A. and J. Morlet, *Decomposition of Hardy functions into square integrable wavelets of constant shape*. SIAM journal on mathematical analysis, 1984. **15**(4): p. 723-736.
190. Junsheng, C., Y. Dejie, and Y. Yu, *Application of an impulse response wavelet to fault diagnosis of rolling bearings*. Mechanical Systems and Signal Processing, 2007. **21**(2): p. 920-929.
191. Bhunia, S., K. Roy, and J. Segura, *A novel wavelet transform based transient current analysis for fault detection and localization.*, in *Proceedings of the 39th annual Design Automation Conference2002*, ACM: New Orleans, Louisiana, USA. p. 361-366.
192. Zhang, D. and W. Sui, *The optimal morlet wavelet and its application on mechanical fault detection*, in *Proceedings of the 5th International Conference on Wireless communications, networking and mobile computing2009*, IEEE Press: Beijing, China. p. 1911-1914.
193. Abhinav, S., W. Biqing, and G. Vachtsevanos. *A methodology for analyzing vibration data from planetary gear systems using complex Morlet wavelets*. in *American Control Conference, 2005. Proceedings of the 2005*. 2005.
194. M Loksha, M.M., KP Ramachandran, KFA Raheem, *Fault diagnosis in gear using wavelet envelope power spectrum*. International Journal of Engineering, Science and Technology 2011. **3**(8): p. 156-167.
195. Belsak, A. and J. Flasker, *Adaptive wavelet transform method to identify cracks in gears*. EURASIP J. Adv. Signal Process, 2010. **2010**: p. 1-8.
196. Hu, Q., et al., *Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble*. Mechanical Systems and Signal Processing, 2007. **21**(2): p. 688-705.
197. Freudinger, L.C., R. Ling, and M.J. Brenner, *Correlation filtering of modal dynamics using the Laplace wavelet*. Proceedings of the international modal analysis conference, 1998. **2**: p. 868-877.
198. Yangyang, Z., et al., *Vibration Based Modal Parameters Identification and Wear Fault Diagnosis Using Laplace Wavelet*. Key Engineering Materials, 2005. **293-294**: p. 183-192.
199. Dong, H.B., et al., *Rotor crack detection based on high-precision modal parameter identification method and wavelet finite element model*. Mechanical Systems and Signal Processing, 2009. **23**(3): p. 869-883.
200. Al-Raheem, K.F. and W. Abdul-Karem, *Rolling bearing fault diagnostics using artificial neural networks based on Laplace wavelet analysis*. International Journal of Engineering, Science and Technology, 2010. **2**(6): p. 278-290.
201. Lin, J. and M.J. Zuo, *Gearbox Fault Diagnosis Using Adaptive Wavelet Filter*. Mechanical Systems and Signal Processing, 2003. **17**(6): p. 1259-1269.
202. Wang, W.J., *Wavelets for Detecting Mechanical Faults with High Sensitivity*. Mechanical Systems and Signal Processing, 2001. **15**(4): p. 685-696.
203. Hemmati, F., *Rolling element bearing condition monitoring using acoustic emission technique and advanced signal processing*. University British Columbia Masters Dissertation 2012.
204. Dan, Z. and S. Wentao. *Bearings Fault Diagnosis Based on the Optimal Impulse Response Wavelet*. in *Information Assurance and Security, 2009. IAS '09. Fifth International Conference on*. 2009.
205. J. Lin, M.Z., *Extraction of periodic components for gearbox diagnosis combining wavelet filtering and cyclostationary analysis*. Journal of vibration and acoustics, 2004. **126**(3): p. 449-451.

References

206. Qiu, H., et al., *Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics*. Journal of Sound and Vibration, 2006. **289**(4–5): p. 1066-1090.
207. Rosso, O.A., et al., *Wavelet entropy: a new tool for analysis of short duration brain electrical signals*. Journal of Neuroscience Methods, 2001. **105**(1): p. 65-75.
208. Kankar, P.K., S.C. Sharma, and S.P. Harsha, *Fault diagnosis of ball bearings using continuous wavelet transform*. Applied Soft Computing, 2011. **11**(2): p. 2300-2312.
209. Rafiee, J., M.A. Rafiee, and P.W. Tse, *Application of mother wavelet functions for automatic gear and bearing fault diagnosis*. Expert Systems with Applications, 2010. **37**(6): p. 4568-4579.
210. Staszewski, W.J. and G.R. Tomlinson, *Application of the wavelet transform to fault detection in a spur gear*. Mechanical Systems and Signal Processing, 1994. **8**(3): p. 289-307.
211. Boulahbal, D., M.F. Golnaraghi, and F. Ismail, *Amplitude and Phase Wavelet Maps for the Detection of Cracks in Geared Systems*. Mechanical Systems and Signal Processing, 1999. **13**(3): p. 423-436.
212. Rubini, R. and U. Meneghetti, *Application of the Envelope and Wavelet Transform Analyses For the Diagnosis of Incipient Faults in Ball Bearings*. Mechanical Systems and Signal Processing, 2001. **15**(2): p. 287-302.
213. Tse, P.W., W.-x. Yang, and H.Y. Tam, *Machine fault diagnosis through an effective exact wavelet analysis*. Journal of Sound and Vibration, 2004. **277**(4–5): p. 1005-1024.
214. Daubechies, I., S.G. Mallat, and A.S. Willsky, *Special issue on wavelet transforms and multiresolution signal analysis*. 1992: IEEE.
215. Warren Liao, T., et al., *A wavelet-based methodology for grinding wheel condition monitoring*. International Journal of Machine Tools and Manufacture, 2007. **47**(3–4): p. 580-592.
216. Purushotham, V., S. Narayanan, and S.A.N. Prasad, *Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition*. NDT & E International, 2005. **38**(8): p. 654-664.
217. Chen, C., et al. *Fault diagnosis for large-scale wind turbine rolling bearing using stress wave and wavelet analysis*. in *Proceedings of the Eighth International Conference on Electrical Machines and Systems*. 2005.
218. Lei, Y., et al., *Application of an improved kurtogram method for fault diagnosis of rolling element bearings*. Mechanical Systems and Signal Processing, 2011. **25**(5): p. 1738-1749.
219. Coifman, R.R. and M.V. Wickerhauser, *Entropy-based algorithms for best basis selection*. Information Theory, IEEE Transactions on, 1992. **38**(2): p. 713-718.
220. Brechet, L., et al., *Compression of Biomedical Signals With Mother Wavelet Optimization and Best-Basis Wavelet Packet Selection*. Biomedical Engineering, IEEE Transactions on, 2007. **54**(12): p. 2186-2192.
221. Huang, N.E., et al., *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 1998. **454**(1971): p. 903-995.
222. Meeson, R.N., *HHT Sifting and Adaptive Filtering*. Institute for Defence Analyses (Alexandria, VA), DTIC Document, 2003.
223. Flandrin, P., G. Rilling, and P. Goncalves, *Empirical mode decomposition as a filter bank*. Signal Processing Letters, IEEE, 2004. **11**(2): p. 112-114.
224. Yang, Y., D. Yu, and J. Cheng, *Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings*. Mechanical systems and signal processing, 2005. **19**(2): p. 259-270.
225. Liu, B., S. Riemenschneider, and Y. Xu, *Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum*. Mechanical Systems and Signal Processing, 2006. **20**(3): p. 718-734.

References

226. Ruqiang, Y. and R.X. Gao, *Hilbert-Huang Transform-Based Vibration Signal Analysis for Machine Health Monitoring*. Instrumentation and Measurement, IEEE Transactions on, 2006. **55**(6): p. 2320-2329.
227. Li, L., L. Wenxiu, and C. Fulei. *Application of AE techniques for the detection of wind turbine using Hilbert-Huang transform*. in *Prognostics and Health Management Conference, 2010. PHM '10*. 2010.
228. He, D., et al. *Development and evaluation of AE based condition indicators for full ceramic bearing fault diagnosis*. in *Prognostics and Health Management (PHM), 2011 IEEE Conference on*. 2011.
229. Bonnardot, F., et al., *Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation)*. Mechanical Systems and Signal Processing, 2005. **19**(4): p. 766-785.
230. Combet, F. and L. Gelman, *An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor*. Mechanical Systems and Signal Processing, 2007. **21**(6): p. 2590-2606.
231. McFadden, P.D., *Interpolation techniques for time domain averaging of gear vibration*. Mechanical Systems and Signal Processing, 1989. **3**(1): p. 87-97.
232. Decker, H.J. and J.J. Zakrasjek, *Comparison of Interpolation Methods as Applied to Time Synchronous Averaging*, in No. ARL-TR-19601999: U.S Army Research Laboratory, Cleveland, Ohio.
233. E. Bechhoefer, M.K., *A review of time synchronous average algorithms*. Annual conference of the prognostics and Health management Society, San Diego, CA., 2009.
234. M. Lebold, et al., *Review of Vibration Analysis Methods for Gearbox Diagnostics and Prognostics*. Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, 2000: p. 623-634.
235. Stewart, R.M., R. M. Stewart. *Some useful data analysis techniques for gearbox diagnostics*. Technical Report MHM/R/10/77, 1977.
236. J. J. Zakrajsek, D.P. Townsend, and H.D. Decker, *An Analysis of Gear Fault Detection Methods as Applied to Pitting Fatigue Failure Data*. Proceedings of the 47th Meeting of the Mechanical Fail-ures Prevention Group, 1993: p. 199-208.
237. K. McClintic, et al., *Residual and Difference Feature Analysis with Transitional Gearbox Data*. Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, 2000: p. 635-645.
238. P. Vecer, M. Kreidl, and R. Smid, *Condition Indicators for Gearbox Condition Monitoring Systems*. Acata Polytechnica, 2005. **45**(6).
239. McFadden, P.D., *Examination of a technique for the early detection of failure in gears by signal processing of the time domain average of the meshing vibration*. Mechanical Systems and Signal Processing, 1987. **1**(2): p. 173-183.
240. Jing, Y., L. Yip, and V. Makis. *Wavelet analysis with time-synchronous averaging of planetary gearbox vibration data for fault detection, diagnostics, and condition based maintenance*. in *Mechanical and Electronics Engineering (ICMEE), 2010 2nd International Conference on*. 2010.
241. H. Li, Y. Zhang, and H. Zheng, *Angle Domain Average and CWT for Fault Detection of Gear Crack*. Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2009.
242. Baydar, N. and A. Ball, *Detection of Gear Failures Via Vibration and Acoustic Signals using Wavelet Transform*. Mechanical Systems and Signal Processing, 2003. **17**(4): p. 787-804.
243. McFadden, P.D. and M.M. Toozhy, *Application of Synchronous Averaging to Vibration Monitoring of Rolling Element Bearings*. Mechanical Systems and Signal Processing, 2000. **14**(6): p. 891-906.
244. C. N. Komgom, et al., *On The Use of Time Synchronous Averaging, Independent Component Analysis and Support Vector Machines for Bearing Fault Diagnosis*. First International Conference on Industrial Risk Engineering, 2007.
245. Galbato, A.T., R.T. Cundill, and T.A. Harris, *Fatigue life of silicon nitride balls*. Lubrication Engineering, 1992. **48**(11): p. 886-894.

References

246. Howard, I.M. and N.S. Swansson, *Demodulating High Frequency Resonance Signals for Bearing Fault Detection*. in *Australian Vibration and Noise Conference 1990: Vibration and Noise-measurement Prediction and Control; Preprints of Papers*. 1990. Institution of Engineers, Australia.
247. Eftekharnajad, B., et al., *The application of spectral kurtosis on Acoustic Emission and vibrations from a defective bearing*. *Mechanical Systems and Signal Processing*, 2011. **25** (1): p. 266-284.
248. Akbari, J., et al., *Acoustic emission and deformation mode in ceramics during indentation*. *JSME international journal. Ser. A, Mechanics and material engineering*, 1994. **37**(4): p. 488-494.
249. Faisal, N., R. Ahmed, and R. Reuben, *Indentation testing and its acoustic emission response: applications and emerging trends*. *International Materials Reviews*, 2011. **56**(2): p. 98-142.
250. Bouras, S., et al., *Study of the resistance to crack propagation in alumina by acoustic emission*. *Ceramics International*, 2008. **34**(8): p. 1857-1865.
251. Zhe, X., C. Wang, and H. Zhang, *Fracture toughness and acoustic emission in silicon nitride*. *Journal of materials science letters*, 1987. **6**(12): p. 1459-1462.
252. Okada, A., N. Hirotsaki, and M. Yoshimura, *Subcritical Crack Growth in Sintered Silicon Nitride Exhibiting a Rising R-Curve*. *Journal of the American Ceramic Society*, 1990. **73**(7): p. 2095-2096.
253. Carden, E.P. and P. Fanning, *Vibration based condition monitoring: a review*. *Structural Health Monitoring*, 2004. **3**(4): p. 355-377.
254. Ortigueira, M.D., R.T. Rato, and A.G. Batista, *On the HHT, its problems, and some solutions*. *Mechanical Systems and Signal Processing*, 2008. **22**(6): p. 1374-1394.
255. MathWorks Documentation Centre, *Border Effects*. Internet Website, 2013.
256. Wang, Y. and M. Hadfield, *Failure modes of ceramic rolling elements with surface crack defects*. *Wear*, 2004. **256**(1-2): p. 208-219.
257. Zhao, P., et al., *Subsurface propagation of partial ring cracks under rolling contact: Part II. Fracture mechanics analysis*. *Wear*, 2006. **261**(3-4): p. 390-397.
258. Ueda, T. and N. Mitamura, *Mechanism of dent initiated flaking and bearing life enhancement technology under contaminated lubrication condition: Part I: Effect of tangential force on dent initiated flaking*. *Tribology International*, 2008. **41**(11): p. 965-974.
259. Wan, G.T.Y., A. Gabelli, and E. Ioannides, *Increased Performance of Hybrid Bearings with Silicon Nitride Balls*. *Tribology Transactions*, 1997. **40**(4): p. 701 - 707.
260. Toutountzakis, T., C.K. Tan, and D. Mba, *Application of acoustic emission to seeded gear fault detection*. *NDT & E International*, 2005. **38**(1): p. 27-36.
261. Al-Balushi, K.R. and B. Samanta, *Gear Fault Diagnosis Using Energy-Based Features of Acoustic Emission Signals*. *Journal of Systems and Control Engineering*, 2002. **216**: p. 249-263.
262. Xiang, D., Y. Qin, and F. Li, *Surface Wave Acoustic Microscopy for Rapid Non-destructive Evaluation of Silicon Nitride Balls*. *Journal of Nondestructive Evaluation*, 2011. **30**(4): p. 273-281.