TECHNOLOGY SHOWCASE

INTEGRATED MONITORING,

DIAGNOSTICS

AND

FAILURE PREVENTION

Proceedings of a Joint Conference

19970529 042

Mobile, Alabama
April 22-26, 1996

Compiled by
Henry C. Pusey
and
Sallie C. Pusey

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

DTIC QUALITY INSPECTED 4

A Publication of the

Joint Oil Analysis Program Technical Support Center
University of Wales, Swansea
Machinery Failure Prevention Technology

Copyright © 1996 by
Society for Machinery Failure Prevention Technology (MFPT)
(A Division of the Vibration Institute)
4193 Sudley Road
Haymarket, Virginia 22069-2420
All Rights Reserved

Special Notice

The U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce, or allow others to publish or reproduce, the published forms of any papers in these proceedings authored by a government agency or a contractor to a government agency whenever such publication or reproduction is for U.S. government purposes.

Technology Showcase, Integrated Monitoring, Diagnostics and Failure Prevention

PLEASE CHECK THE APPROPRIATE BLOCK BELOW:

-AQ # _____

☐ _____ copies are being forwarded. Indicate whether Statement A, B, C, D, E, F, or X applies.

☒ DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

☐ DISTRIBUTION STATEMENT B:
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES ONLY; (Indicate Reason and Date). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT C:
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND THEIR CONTRACTORS; (Indicate Reason and Date). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT D:
DISTRIBUTION AUTHORIZED TO DoD AND U.S. DoD CONTRACTORS ONLY; (Indicate Reason and Date). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT E:
DISTRIBUTION AUTHORIZED TO DoD COMPONENTS ONLY; (Indicate Reason and Date). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT F:
FURTHER DISSEMINATION ONLY AS DIRECTED BY (Indicate Controlling DoD Office and Date) or HIGHER DoD AUTHORITY.

☐ DISTRIBUTION STATEMENT X:
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND PRIVATE INDIVIDUALS OR ENTERPRISES ELIGIBLE TO OBTAIN EXPORT-CONTROLLED TECHNICAL DATA IN ACCORDANCE WITH DoD DIRECTIVE 5230.25, WITHHOLDING OF UNCLASSIFIED TECHNICAL DATA FROM PUBLIC DISCLOSURE, 6 Nov 1984 (Indicate date of determination). CONTROLLING DoD OFFICE IS (Indicate Controlling DoD Office).

☐ This document was previously forwarded to DTIC on _____ (date) and the AD number is _____.

☐ In accordance with provisions of DoD instructions, the document requested is not supplied because:

☐ It will be published at a later date. (Enter approximate date, if known)

☐ Other. (Give Reason)

DoD Directive 5230.24, "Distribution Statements on Technical D **nts, as described briefly above. Technical Documents must be assigned**

Per Phone Con with →

Authorized Signature/Date

Randy Holland
Technical Director

Joint Technical Support Center
296 Farrar Road, Bldg 780
Pensacola, FL 32508
Comm: (904) 452-3191 x 101 DSN (922)
Fax: (904) 452-4145 DSN (922)
E-Mail: nrh1@navtap-emh.navtap.navy.mil

TABLE OF CONTENTS

PREFACE ix

TECHNICAL PROGRAM COMMITTEE xii

FEATURED PAPERS

An Historical View of the MFPT Society 3
H. C Pusey

A Review of Plant Maintenance Methods and Economics 17
M. J. Neale

The History and Application of the Envelope Detector 25
J. L. Frarey

GENERAL MONITORING TECHNOLOGY

Comparison of Vibration and Oil Analysis 47
T. H. Crum

Integrated Machinery Management 55
J. W. Taylor

Expert Systems, A Decade of Use for Used-Oil Data Interpretation 65
L. A. Toms

MACHINERY DIAGNOSTICS & PROGNOSTICS #1

Model-Based Sensor Selection for Helicopter Gearbox Monitoring 75
V. B. Jammu, J. Wang, K. Danai and D. G. Lewicki

Electrical Signature Analysis Applications for Non-Intrusive Automotive Alternator Diagnostics 85
C. W. Ayers

Use of the Decision Support Problem Technique for Propulsion Engine Selection Emphasizing Reliability, Maintenance, and Repair Factors: A Limited Example 95
K. Lynaugh, E. Pogue and B. Friedman

In-Line Oil Debris Monitor (ODM) for the Advanced Tactical Fighter Engine 111
D. E. Muir, B. Howe

MACHINERY DIAGNOSTICS & PROGNOSTICS #2

Application of an Envelope Technique in the Detection of Ball Bearing Defects in a Laboratory Experiment 121
H. Toersen

Systems Engineering of an On-Line Diagnostic System 133
J. S. Lin

Turbine Engine Diagnostics Using a Parallel Signal Processot 141
T. A. Bapty, A. Ledeczi, J. R. Davis, B. Abbott, T. Hayes and T. Tibbals

LUBRICANT CONDITION MONITORING #1

Alternate Techniques for Wear Metal Analysis 151
C. S. Saba and H. A. Smith

Current Technology in Oil Analysis Spectrometers and What We May Expect in the Future 161
M. Lukas and R. J. Yurko

An Introduction to the Concept of Profiles 173
G. Fogel and D. Doll

The Mini-Lab Concept as an Alternative to Conventional Oil Analysis 183
G. Fogel

LUBRICANT CONDITION MONITORING #2

Considerations in Applying Molecular Analysis of Different Fluids in a Condition Monitoring Program 197
J. R. Powell

Use of FT-IR Spectrometry as a Replacement for Physical Property Testing of Railway Lubricants 207
D.E.G. Crutcher, R. Gervais and L. A. Toms

PARTICULATE/WEAR DEBRIS ANALYSIS #1

An Alternative Coding System Defining the Total and Severity of Wear 219
M.H. Jones and J. Alberich

Development in Wear Debris Morphological Analysis at RAF Early Failure Detection Centres 227
B. J. Roylance, L. M. Jones, A. Luxmoore, A. Killingray, S. Harris and D. Hodges

Evaluation of Various Methods to Detect Metallic Wear Particles in Lube Oil Filter Debris 237
G. C. Fisher

Condition Monitoring of Aircraft by Quantitative Filter Debris Analysis (QFDA) 247
L.G.I. Bennett, S. Swanson, F. D. Koza and J. S. Poland

ON-LINE CONDITION DIAGNOSIS

Flexible Expert System for Automated On-Line Diagnosis of Tool Condition 259
E. Jantunen, H. Jokinen and R. Milne

Advances in Optical Oil Debris Monitoring Technology 269
J. Reintjes, R. Mahon, M. D. Duncan, L. L. Tankersley, A. Schultz, V. C. Chen, P. L. Howard, S. Raghavan and N. Gupta

A Non-Redundant Sensor Validation Scheme for Transient and Steady-State Condition Monitoring 277
M. J. Roemer

MACHINERY DIAGNOSTICS & PROGNOSTICS #3

A Model-Based Failure Detection, Isolation and Recovery System 291
A. Misra and J. Sztipanovits

Solid-State Active Sensing for In-Situ Health Monitoring 301
C. A. Rogers and F. Lalande

Artificial Knee Joint Controller Characterisation and Condition Diagnosis 309
C. K. Mechefske and J. T. Davenport

Comprehensive Machinery Monitoring with FFPI Sensors 319
R. X. Perez, R. A. Atkins, C. E. Lee and H. F. Taylor

Classification of Composite Defects Using the Signature Classification Development System 329
J. S. Lin, L. M. Brown and C. A. Lebowitz

MICROELECTROMECHANICAL SYSTEMS (MEMS) OVERVIEW

Liquid Viscosity and Density Measurement with Flexural-Plate-Wave Sensors 341
S. W. Wenzel, B. J. Costello and R. M. White

MEMS TECHNOLOGIES

Plasma Micromachined MEMS for Intelligent Diagnostic Sensors 353
G. J. Galvin and T. J. Davis

MEMS Hydrodynamic Bearings: Applications and Implications to Machine-Failure Prevention 363
K. L. Wood, D. Neikirk, I. Busch-Vishniac, W. Weldon, C-S. Chu, Y. Kim, V. Gupta, W. Maddox and D. Masser

A Distributed, Wireless MEMS Technology for Condition Based Maintenance 373
K. Bult, A. Burstein, D. Chang, M. Dong, M. Fielding, J. Ho, W. J. Kaiser, E. Kruglick, F. Lin, T. H. Lin, H. Marcy, R. Mukai, P. Nelson, K. S. J. Pister, G. Pottie, H. Sanches, O. M. Stafsudd, K. B. Tan, C. M. Ward, S. Xue and J. Yao

MEMS-Based Research in Integrated Monitoring and Actuation at Case Western Reserve University 381
M. A. Huff and M. Mehregany

SIGNAL ANALYSIS

Wavelet Domain Characterization & Localization of Modal Acoustic Emissions in Aircraft Aluminum 393
G. A. Gordon and R. K. Young

Overview of Wavelet/Neural Network Fault Diagnostic Methods Applied to Rotating Machinery 405
J. E. Lopez, I. A. Farber Yeldham and K. Oliver

Helicopter Transmission Diagnostics Using Vibration Signature Analysis 419
B. E. Parker, Jr., M. P. Carley, R. J. Ryan, M. J. Szabo and H. V. Poor

Multiscale Statistical Modeling Approach to Monitoring Mechanical Systems 431
K. C. Chou

Synchronous Phase Averaging Method for Machinery Diagnostics 441
J. Jong, M. McBride, J. Jones, T. Fiorucci and T. Zoladz

RUSSIAN TECHNOLOGY INTERCHANGE #1

Principles and Experience of the On-Condition Application of Oils and Working Fluids to Aeronautical Engineering 455
B. Bedrik, M. Yampolsky, J. Szydywar and Y. Ereomine

Determination of Technical Status, Causes of Wear and Failure of Machine Parts Using the Methods of Metallophysical Analysis 465
V. Lozovsky

A System of Methods for Identifying the Causes of Failures and Accidents of Aeronautical Engineering 469
Y. Korovkin

Expert - Examination Computer System for Estimating Operational Status of Technical Equipment with Accounting for the Reliability of Its Life-Critical Parts 477
V. Lozovsky and M. Yampolsky

RUSSIAN TECHNOLOGY INTERCHANGE #2

Methods for Diagnosing the Fatigue Damage Rate of Materials of Machine Elements and Estimating Their Fatigue Life 483
A. Koltunov

Hazardous Wear Prevention System for Friction Units of Gas Turbine Engines 493
V. A. Stepanov and I. F. Touloupov

Equipment for Carrying Out Technical Monitoring of Gas Pumping Units by Lubrication Oil Analysis 503
B. R. Matveyevsky and V. A. Petrosyanz

Basic Trends of Frictional Interaction and Tribodiagnostics 509
M. A. Bronovets, S. P. Zaritski and A. S. Lopatin

Effective Magnetic Method for Detecting Longitudinal Cracks in the Pipes of Pipe-Line Transport 521
G. Shelihov, V. Usoshin, V. Lozovsky and V. Rozov

LUBRICANT CONDITION MONITORING #3

A New Orbital Ball Type of Viscometer 525
A. Abnett and R. Garvey

Estimating Water Content in Oils: Moisture in Solution, Emulsified Water, and Free Water 535
R. Garvey and G. Fogel

Rapid, Portable Voltammetric Techniques for Performing Antioxidant, Total Acid Number (TAN) and Total Base Number (TBN) Measurements 549
R. E. Kauffman

A New Electronic Viscometer Based on Rayleigh Wave Mechanics 561
J. J. Kauzlarich, R. A. Ross and D. S. Abdallah

PARTICULATE/WEAR DEBRIS ANALYSIS #2

A Phenomenological Approach to Wear Debris Analysis 573
U. Cho and J. A. Tichy

Monitoring Fluid System Debris *Via* Diagnostic Filters 579
P. Madhavan

Developing an Integrated Condition Monitoring System 587
J. Tranter

VIBRATION ANALYSIS/MONITORING

A Fuzzy-Neuro Scheme for Fault Diagnosis and Life Consumption of Rotordynamic Systems 601
M. J. Roemer

Sync-Period Frequency Analysis and Its Application to the Diagnosis of Multiple Element Defects of Rolling Bearings 611
W-Y Wang and M. J. Harrap

Coherent Phase Wide Band Demodulation Technique for Turbomachinery Cavitation Detection and Monitoring 621
J. Jong, T. Nesman, W. Bordelon, J. Jones and T. Zoladz

Self-aligning Roller Bearings Fault Detection Using Asynchronous Adaptive Noise Cancelling 633
K. Nezu and Y. Shao

NON-DESTRUCTIVE INSPECTION

Eddy-Current Examination of Large-Diameter Insulated Pipes 643
J. C. Griffith

Off-Line and On-Line Motor Electrical Monitoring and Condition Analysis: Payoffs and Problems 651
J. R. Nicholas, Jr.

The Use of the Motor as a Transducer to Monitor System Conditions 661
D. A. Casada and S. L. Bunch

Subsurface Defect Detection in Ceramic Materials Using Low Coherence Optical Scatter Reflectometer 673
M. Bashkansky, M. D. Duncan, M. Kahn, D. Lewis, III and J. F. Reintjes

Poor Surfaces and Intersections of Surfaces Still Cause Trouble Just Like They Used To Do 681
D. Krashes and J. J. Scutti

FAILURE ANALYSIS #1

Gear Crack Propagation Investigations 693
D. G. Lewicki and R. Ballarini

Detecting Gear Tooth Fatigue Cracks in Advance of Complete Fracture 703
J. J. Zakrajsek and D. G. Lewicki

Metallurgical Examination of a Failed Spiral Bevel Gear 715
M. Pepi

The Development of a Cost Benefit Analysis Method for Monitoring the Condition of Batch Process Plant Machinery 725
B. S. Rajan and B. J. Roylance

Failure Analysis of a Pitch Link Self-Locking Nut MS 17825-10 737
S. Grendahl

FAILURE ANALYSIS #2

Failure Analysis of Modular Lift Link, Lug Mounting, Self-Locking Nuts 751
S. Grendahl

Multisensor Monitoring of Gear Tooth Fatigue for Predictive Diagnostics 763
G. A. Gordon, A. K. Garga and C. Moose

Prevention of Crankcase Explosions in Reciprocating Compressors 775
V. A. Cox

Experimental and Numerical Study of Vibration Monitoring Applied on Gear Transmission Systems 785
J. Mahfoudh, C. Bard and D. Play

EQUIPMENT MONITORING AND CONDITION DIAGNOSIS

Low-Speed Mini-Motor Failure Recognition Using Fuzzy Theory 797
K. C. Siong and K. Nezu

Automatic Machinery Fault Detection and Diagnosis Using Fuzzy Logic 803
C. K. Mechefske, J. Del Mar and D. Prendergast

Technology Overview: Shock Pulse Method 811
L. E. Morando

Reliability-Based Maintenance as a Breakthrough Strategy in Maintenance Improvement 821
G. Fogel and D. Petersen

APPENDIX: MFPG /MFPT Publications 833

Preface

The Integrated Monitoring, Diagnostics and Failure Prevention Technology Showcase Conference was jointly sponsored by the DoD Joint Oil Analysis Program Technical Support Center (JOAP-TSC), the Society for Machinery Failure Prevention Technology (MFPT) and the University of Wales, Swansea. The JOAP-TSC Sponsored the meeting at the Adam's Mark Hotel in Mobile, Alabama April 22-26, 1996. Allison M. Toms, JOAP-TSC representative and conference host chaired the opening session along with G. William Nickerson, MFPT Technical Program Chairman, and Mervin H. Jones of the Mechanical Engineering Department, University of Wales, Swansea.

The theme set for this conference, Integrated Monitoring, Diagnostics and Failure Prevention Technology Showcase, highlighted the joint interests of the three sponsoring organizations and provided an excellent opportunity for interchange of ideas and technologies among the equipment maintenance, condition monitoring, diagnostics and failure prevention communities. The sponsors share the common goal of maximizing service life of machinery and structures at minimum cost without compromising availability or safety. The joint nature of this meeting promoted synergistic applications of the many technologies presented and enhanced cooperation among the various military, commercial and academic organizations.

We were fortunate to have Dr. Carl Talbott, The North American Reliability Manager of M & M Mars open the conference as the Keynote Speaker. Dr. Talbott presented his ideas on *The Business End of Condition Monitoring*, discussing how industrial manufacturers will lead and facilitate the fundamental changes in content and delivery of condition monitoring technology. Henry Pusey, Executive Director of the MFPT Society followed with a paper providing *A General Overview of the MFPT Society* and the progress in mechanical failure prevention technology during it's 29 year history. Jack Hessburg of the Boeing Commercial Airplane Group discussed the *Boeing 777 Diagnostic Program* and Ted Selby, Director of R & D at Savant, Inc. presented *Civilization and Lubrication.* The morning's presentations were rounded out by Michael J. Neale, Chairman of Neale Consulting Engineers, Ltd. who presented *A Review of Progress in Plant Maintenance with Condition Monitoring.* The afternoon of the Plenary session opened with a paper by John L. Frarey of JLF Analysis, on *The History and Application of the Envelope Detector* and Vernon Westcott, Chairman of Institute Guilfoyle closed the first day's talks with a presentation on *Bio-Ferrography.* While all were excellent presentations, only papers by Henry Pusey, Michael J. Neale and John L. Frarey were included in the proceedings.

The technical program included sessions on a wide array of condition monitoring and materials and machinery characterization technologies. Several sessions on Lubricant Monitoring contained papers discussing the current state-of-the-art of machinery condition monitoring using oil analysis techniques, while three sessions in Diagnostics and Prognostics contained papers describing sensor technology, signal processing, fault characterization and prognostic modeling. A special session on Microelectromechanical

Systems (MEMS) was organized to include overviews of the technology from a military as well as a private sector perspective and included papers describing the current research in applications of MEMS sensors for condition monitoring. The MEMS session also included a breakout session for discussion of the future of MEMS Technologies. For a summary of the findings from this breakout session, interested parties are encouraged to contact the MEMS session organizer, Ed Birdsall of SynEx, Inc. Alexandria, VA. A session in Signal Processing was organized by Dr. Kam Ng of the Office of Naval Research and included papers discussing current methods of signal analysis for condition characterization.

Two sessions organized by Mervin Jones of the University of Wales highlighted Russian Technology in the area of condition monitoring techniques, fault diagnosis and material failure characterization. A session on Non-Destructive Inspection contained papers describing on-line and off-line sensors and monitoring techniques while two sessions in Particulate/Wear Debris Analysis included papers discussing methods for monitoring the wear of components. Condition monitoring and diagnosis of machinery and structural components depends on a thorough understanding of failure mechanisms. To bring these technologies together and provide a forum for exchange of materials characterization information, two session on Failure Analysis were included in this conference containing papers on failure case histories as well as several describing failure mechanisms of materials.

The primary goal of the conference was to provide a forum for information exchange between technology developers and users as well as between the various technology groups in the hopes of furthering research, development and widespread use of condition-based maintenance technology within the military and private sectors. Organization of the technical sessions and social activities provided the prime opportunities for technical exchanges in both formal and informal forums. This goal was further enhanced by the 11 training sessions and an Exhibition of Condition Monitoring and Diagnostics Products. It is believed by the conference organizers that this goal was met but the conference attendees are the final judges of how well this goal was achieved.

The JOAP-TSC serves as the cognizant engineering authority for tribology methods and instruments and monitors data measurement reliability for over 300 laboratories. The center coordinates management and technical support to the United States Armed Forces, US Government agencies, DoD contractors and Allied Nations. JOAP members strive to protect personnel and extend the life of valuable aeronautical, ground and marine resources.

The Society for MFPT, a division of the Vibration Institute, is a forum for the interchange of information on the processes related to machinery failures. The Society is an interdisciplinary group of scientists, engineers, failure analysts and maintenance specialists and acts as a focal point for those involved in technological developments that contribute to the reduction or prevention of mechanical failures. The Society is dedicated to achieving competitive advantage by linking failure mechanisms and diagnostics.

The University of Wales, Swansea, Mechanical Engineering Department, in conjunction with the Swansea Tribology Center is actively involved in research and development of equipment and fluid monitoring technologies. Since 1984, the University of Wales has sponsored several international conferences on machinery condition monitoring in both the United Kingdom and Germany. The Tribology Center provides a focal point for those involved in tribology and equipment condition monitoring and supports commercial and military equipment monitoring programs in the United Kingdom and other countries.

On behalf of the conference organizers I would like to thank the many participants and speakers who helped to make this conference a success. Without the participation of the many commercial, military and academic professionals involved in condition monitoring technology, the goal of providing a forum for information exchange could not be met.

MarjorieAnn E. Natishan, Chairman
Board of Directors MFPT Society

A Technology Showcase:

Integrated Monitoring, Diagnostics & Failure Prevention

Technical Program Committee



Allison M. Toms
JOAP-TSC
296 Farrar Road, Suite B
Pensacola, FL 32508

MFPT 50TH MEETING The Society for Machinery Failure Prevention Technology

G. William Nickerson
Pennsylvania State University/ARL
PO Box 30
State College, PA 16804

University of Wales, Swansea
CONDITION MONITORING'96

Mervin H. Jones
Mechanical Engineering Dept.
University of Wales, Swansea
Singleton Park, W. Glamorgan
UK SA2 8PP

FEATURED PAPERS

AN HISTORICAL VIEW OF THE MFPT SOCIETY

Henry C. Pusey
Society for Machinery Failure Prevention Technology
4193 Sudley Road
Haymarket VA 22069-2420

Mechanical failures are a pervasive fact of life in our society. Ranging from the failure of small items that all of us have experienced and that many of us take for granted, to the failure of a large complex structure that often becomes front page news, they have undesirable consequences for our society. The large ones many times cause loss of life or cause serious injury to many people. The minor ones sometimes also cause loss of life or injury, and they always cause frustration and anger on the part of the one to whom they occur. Always they cause loss of valuable material, and have undesirable social and economic consequences.

... Elio Passaglia, Executive Secretary MFPG, 1976

Abstract: This paper provides a general overview of developments and progress in mechanical failure prevention technology, primarily over the last three decades. It is a condensed version of a paper presented at a luncheon of the ASME Reliability, Stress Analysis and Failure Prevention Committee at the 1995 Design Engineering Technical Conferences. The principal reference sources used are the proceedings of the 49 conferences of the Mechanical Failures Prevention Group (MFPG), now the Society for Machinery Failure Prevention Technology (MFPT). Since many technical disciplines are involved in this complex technology, an attempt is made to place mechanical failure prevention in perspective. Discussion of the evolving technology is presented under four topic areas that serve to cover most of the elements of this broad-based field. These topics are Diagnostics and Prognostics, Failure Analysis, Life Extension and Durability, and Sensors Technology.

Key Words: Condition monitoring; diagnostics; failure analysis; failure modes; maintenance; mechanical failure; oil analysis; vibration analysis

INTRODUCTION: The above quotation is taken from the Introduction to the Proceedings of the 20th Meeting of the Mechanical Failures Prevention Group (MFPG). Dr. Passaglia, who was then Chief of the Metallurgy Division of the National Bureau of Standards (NBS), assumed the responsibilities of MFPG Executive Secretary after the merger of the MFPG with the NBS Failure Avoidance Program in the early 1970s. MFPG 20 was a milestone event. Sponsored by seven government agencies and ASME in May 1974, the meeting was designed to explore the various aspects of mechanical failures with the purpose of "defining the problem". The symposium organizers chose to examine the technology from the aspect of failure modes, failure consequences and the implication of failure with respect to action by the technical community. The conference program was planned accordingly; the result was a collection of informative and thought provoking papers.

Failure of machinery, equipment or structures has even greater implication today than it did twenty years ago. As our society becomes more and more mechanized, as demands on performance become more exacting, and as our technology becomes increasingly complex, each mechanical failure reaches greater significance. In terms of the economy of the country, mechanical failures represent a cost of billions of dollars every year to industry, the government and the general public. Improved public safety, an area of paramount concern, may be achieved by a higher level of understanding of the mechanical failure process coupled with innovative techniques for failure avoidance, especially when using new or emerging materials in the design of structures and equipment.

Failure prevention technology is complex and clearly involves a wide range of technical disciplines. Any in-depth history of failure prevention would result in numerous books, each covering one or more of the technologies that relate to failure and its prevention (e.g. materials science, tribology, vibration analysis, etc.). One can only examine advances in failure prevention from a broad perspective. That is the purpose of this paper. Progress in selected areas of mechanical failure prevention is assessed, particularly over the last quarter century. The references used are primarily from the proceedings of the MFPG conferences held since its inception in 1967 through its change of name to The Society for Machinery Failure Prevention Technology (MFPT) in mid 1994 and MFPT 49 in April 1995.

FAILURE PREVENTION IN PERSPECTIVE: The need to prevent failures has been with us since man began inventing functional gadgets. There has always been an unwritten rule that each gadget should operate as long as possible without breaking or malfunctioning. Taken as a whole, failure prevention is considered by many to be a technology. The definition of technology is the application of science. Thus mechanical failure technology could be defined as the application of science to mechanical failure processes. Understanding the failure process may require a study of what is failing, the nature of a practical failure and the sequence of events which leads to failure. One must then describe the failure process in terms of design or material variables.

Early in their history, the MFPG recognized that this is a complex process and that there are many ways to prevent mechanical failures in service. These actions include but are not limited to

- developing better design techniques.
- improving reliability predictions.
- providing more complete materials information.
- better understanding of the failure process.
- improved quality control.
- effective maintenance.
- improved diagnostics.
- cleanliness.
- lubrication/wear reduction.
- improved failure analyses.
- feedback of analysis results.

By their sixth meeting the MFPG Steering Committee recognized that increasingly large numbers of interested persons were attending two-day symposia. Papers were given and discussions heard covering almost all identifiable aspects of mechanical failures. Although these meetings provided the best forum for information dissemination, they did not foster structured pursuit of specific mechanical failure prevention tasks. To meet this need Technical Committees were organized on **Detection, Diagnosis and Prognosis, Mechanisms of Failure, Design (including Testing)**, and **State-of-the-Art and Applications**. With the objectives of the MFPG remaining essentially the same, the MFPG committee structure was modified slightly over the years in response to changing emphasis. The latest committee designations are **Diagnostics and Prognostics, Failure Analysis, Life Extension and Durability** and **Sensors Technology**. For convenience, these titles are used as topic areas for the material presented in this paper.

DIAGNOSTICS AND PROGNOSTICS: *Diagnosis* is the art or act of identifying a condition from its signs or symptoms. *Prognosis* is the art or act of predicting a future condition on the basis of present signs and symptoms. Any method used for identifying incipient failures and/or predicting ultimate failure of materials, structures or systems would fall within the scope of the Diagnostics and Prognostics Committee. The reader should note that there are overlapping areas of interest among all of the technical committees. For example, predictive maintenance, a subset of condition-based maintenance, involves diagnosis and prognosis. At the same time, effective application of maintenance philosophy is a proven technique for extending the life of machinery. In this section, the most commonly used techniques for diagnostics are discussed, the issue of prognostics is addressed and an attempt is made to assess our current capabilities for failure prediction.

Diagnostic Methods: Vibration signal analysis and oil analysis are treated separately as techniques for diagnosing condition and fault mechanisms in machinery. Other selected nondestructive testing and evaluation methods that are applicable to materials, structures or machinery are then described.

Vibration Analysis: It is not known when vibration signature analysis was first used as a diagnostic tool. It is clear that by the time the MFPG was organized in 1967, machinery health monitoring techniques using vibration signatures had already been in use for a number of years. The minutes of the first MFPG meeting indicate that the technical presentations all related in some way to diagnostics involving vibration. The theme of the sixth meeting was *Detection, Diagnosis and Prognosis (DD&P)* as well as that of a number of meetings that followed. With a few exceptions, DD&P was a part of the program at all MFPG meetings and vibration was always very much in evidence as a diagnostic tool.

The tenth MFPG meeting was organized with specific emphasis on the utility of vibration analysis methods in mechanical failure prevention. A presentation on time series analysis techniques clearly showed the usefulness of analytical techniques for comparing differences in vibration waveforms. The discussion at that meeting showed lack of agreement on how these techniques could be applied to machinery condition monitoring. Five papers were devoted to the use or trial of automated vibration monitoring and diagnostic systems for aircraft gas turbine engines, ships' machinery, helicopters, commercial jet aircraft, and internal combustion engines. Advantages and limitations of the various systems were discussed. At that time there were some

limitations with respect to available instrumentation as well as on the capability to identify the faults in and condition of machinery so that effective maintenance planning could be achieved.

By the time of MFPG 44 fifteen years later it was evident that instrumentation capability had increased dramatically, but that techniques for fault diagnosis had evolved more slowly. The tools were still more advanced than the techniques. Eshleman [1] addeessed this issue in some detail and examined three technical areas that must be addressed for effective fault diagnosis using vibration; these are *condition and fault mechanisms*, *modification of signal transmission paths* and *signal analysis*.

The 1994 proceedings (MFPG 48) contains several papers that reflect significant progress in the application of vibration analysis to diagnostics. Specific advances were reported on helicopter transmission fault detection and classification, alarm threshold settings for vibration monitoring of rotating machinery and pattern classification of vibration signatures using neural networks. Some preliminary ideas on the application of smart structures in conjunction with vibration signature analysis for on-line machinery health monitoring were presented. It will be interesting to see how the use of this technology evolves.

Oil Analysis: All machinery requires lubrication to minimize wear. This includes various engines (internal combustion, diesel, turbojet, etc.) and their components (transmissions, gear boxes). Everyone knows that the oil must be changed in automobiles at regular intervals to extend the life of the engine. This routine maintenance action is necessary because the oil gets dirty. Dirty, *worn out* oil not only does not lubricate well, it actually increases engine wear. Why? What is this *dirt* in the oil? How can information about it be used to ensure that engines operate safely and to identify engine components that will fail with potentially catastrophic results unless corrective action is taken? The process is called *oil analysis*, a proven diagnostic tool for mechanical failure prevention.

At the second MFPG meeting in June 1967, Ward [2] described a Navy spectrometric oil analysis program initiated in 1955. The goal was to find out whether the concept employed by the railroads for determining the condition of diesel engines by analyzing used oil samples could be applied to aircraft engines. The Bureau of Aeronautics felt that if these techniques could be applied, inflight failures could be minimized, extension of engine operating intervals could be justified and reductions in engine overhaul costs could be achieved. This was not a new idea, but the concept had been restrained for some time by concern over the many factors which could work against development of successful techniques. Consider that the wide variety of engines in service, the many sources of wear metal contamination, the many sources of lubricating oil base stock, the necessity for development of metallic contamination threshold limits, the time and cost involved in sample analysis and the sample handling and data communication problems must all be dealt with to achieve success. They were able to handle these and other problems successfully and in 1958 some positive results accelerated the program. By 1967, the Navy was working jointly with the Army and Air Force. The results were very impressive as evidenced by the very successful DoD Joint Oil Analysis Program (JOAP) coordinated by the Technical Support Center at Pensacola, Florida.

The use of oil analysis, as reported at the MFPG conferences, has been an invaluable diagnostic tool over the years. Application of the technique to internal combustion engines was discussed at MFPG 12 and to commerical aircraft at the 14th and 16th meeting. Oil analysis was the theme of MFPG 16 and advancements in both techniques and analysis equipment have been faithfully reported in the proceedings since then. At MFPG 48 the Naval Research Laboratory [3] reported on a real time, on-line, optical oil-debris monitor that is expected to provide a cumulative record of the health of engines and gear boxes as well as advanced warning of catastrophic failure.

Nondestructive Testing and Evaluation (NDE): NDE in general is the technology of measurement, analysis and prediction of the state of material systems for safety, reliability and assurance of maximum lifetime performance. It is an old technology (more accurately a set of technologies), yet it is only in recent years that engineers and managers have awakened to the true importance and great potential of NDE. Some NDE test technologies that can be effectively applied to diagnostics include acoustics, microscopy, optics, thermography, electromagnetics and radiography. The capabilities of these NDE methods as diagnostic tools for equipment and structures have increased significantly with the rapid development of advanced hardware and software. Although a detailed discussion of progress in each NDE area is not possible, it is useful to describe how some of these methods are used for fault diagnosis.

Acoustic Methods: *Tap testing* is probably the simplest, most common and inexpensive form of acoustic inspection. The inspector taps the surface of the test structure and evaluates the sound that is generated. He either listens directly to the sound or uses a specially designed receiver to analyze the sound and compare the response with one from a non-defective part.

The *acoustic emission* (AE) technique is a method whereby sound waves emitted from a growing crack or flaw in a structure are detected. These signals are then evaluated to determine the nature of the damage. The advantages of AE are that it offers global monitoring capability and real time information as to the state of damage. The problems with the technique arise from the complexity of sound propagation in solid structures which makes interpretation of the AE in terms of structural damage difficult.

The *acoustoultrasonic* (AU) technique was devised to assess diffuse discontinuity populations of the mechanical properties of composites and composite-like materials. This NDE method, also known as the *stress wave factor* technique, has been used to evaluate fiber reinforced composites, adhesive bonds, lumber, paper and wood products, cable, rope and human bone. The AU technique has been demonstrated to be sensitive to interlaminar and adhesive bond strength variations and has been shown to be useful in assessing microporosity and microcracking produced by fatigue cycling. The AU method belongs to a class of techniques that includes tap testing, dynamic resonance and structural damping measurements. These are in addition to more directly related techniques like acoustic emission and pulse echo methods.

Acoustography is a process of forming ultrasonic images in a manner similar to X-ray fluoroscopy, using a detector screen that converts ultrasonic energy directly into a visible image. A full field imaging method, *acoustography* may be used to rapidly screen parts for discontinuities. *Acoustic holography* is yet another useful NDE method which was first

discussed by MFPG participants at their ninth meeting. This method involves only the recording of a hologram and its image reconstruction. Thus, unlike optical holography, the additional interferometry step (after stressing the material) is unnecessary to produce an image depicting internal anomalies. The reconstructed images from acoustic holograms, however, do not possess the high resolution image quality of the shorter wavelength optical holograms.

Vibration analysis is often considered to be an acoustic method. This is in part because sound produced by machinery can be used in the same way as vibration for machinery diagnostics. After all, the sound radiated from machinery is produced by the vibration of the machine. There have been technical discussions at MFPG and other conferences about whether vibration or acoustic signal analysis is most useful. Either can be used effectively. The simplest way to use sound is by an experienced machine operator who will know that something is wrong when his machine sounds different than it normally does.

Infrared Analysis: Infrared or thermographic analysis provides a high-resolution, non-contact means of monitoring the condition of electrical and electromechanical equipment, roofing and wall insulation and oven refractories. Infrared scanners, similar in appearance to video cameras, detect differences in surface temperatures and highlight those differences in black and white or color images that are displayed on a television screen. These images, called thermograms, are used to analyze patterns of heat gain or loss. Infrared analysis is an effective predictive maintenance tool because mechanical or electrical breakdowns are often preceded or accompanied by changes in operating temperatures. This information can be particularly important in electrical machinery where circuits and connections may show no visible signs of deterioration until moments before a complete failure.

Motor Current Signature Analysis (MCSA): MCSA provides a nonintrusive method for detecting mechanical and electrical problems in motor driven rotating equipment. The method is the development of Oak Ridge National Laboratory [4], as part of a study on the effects of aging and service degradation of nuclear power plant components. The basis for MCSA is the recognition that an electric motor driving a mechanical load acts as an efficient, continuously available transducer (the motor can be either AC or DC). The motor senses mechanical load variations and converts them into electric current variations that are transmitted along the motor power cables. These current variations, though very small in relation to the average current drawn by the motor, can be monitored and recorded at a convenient location away from the operating equiment. Analysis of these variations can provide an indication of machine condition, which may be trended over time to provide an early warning of machine deterioration or process alteration.

Prognostics: The words *prediction* or *prognosis* appear in a number of titles of papers in the MFPG proceedings. The *prediction* papers usually deal with mathematical models for fatigue life estimation, stochastic models for cumulative damage or trending algorithms of one sort or another. Usually these papers report useful work that is consistent with the definition of *prediction* as *a way to establish beforehand the expected value of some parameter at a definite future time*. However, these methods have not yet been effectively applied by the predictive maintenance community for on-line prognostics.

Some ideas have shown promise. Dunegan [5] suggested that the combination of acoustic emission and linear fracture mechanics can provide quantitative information regarding structural failure. Wicks [6] proposes that experimental modal analysis techniques can be used as a tool in predictive maintenance programs. Modal models of a given machine provide a baseline from which trends may be monitored and evaluated. At MFPG 47, this same idea was examined with respect to structures [7].

Salter [8] stated in 1978 that on-vehicle computing instrumentation technology offers new capabilities to improve life-cycle reliability through prognostics. Prognosis enables the selection of the best time for maintenance while reducing inspection requirements, vehicle breakdowns and secondary failures. He suggested that his program enables an improved user confidence while maximizing the productivity of scarce maintenance resources. The program requires a microdata system that can automate condition trend analysis using a technique that Salter calls *Geriometry*, a process which measures the causes of vehicle degradation, whereas all other maintenance techniques measure effects. It is not clear whether this idea was made to work, but the premise was good. Effective prognosis is important to maintenance management. Some technique is necessary to choose the time for maintenance which will minimize maintenance costs while achieving a desired reliability and availability. The technique must be quantified in terms of measurable parameters. The precision and variance with which these parameters describe the subsystem life must yield decision criteria within close confidence limits. The rationale for prognosis as an effective technique for maintenance is somewhat subjective, since data to prove the contention do not yet exist. It is anticipated by many persons in the maintenance field, however, that a number of benefits will accrue with the implementation of reliable prognosis methods. For the most part, however, the MFPG papers with *prognosis* in the title are lacking in information on effective prognostic techniques. The remaining life of a machine or component must be based upon wear, the environment, and the history of stress cycles in the machine. These factors must also be considered when establishing its current condition. Only after the current condition of a machine is known can meaningful life estimates be made. Information about current condition can also be used to evaluate the effects of changes in components and wear in order to assess whether or not machine life can be extended. The techniques of prognosis involve diagnosis, condition models, and failure models. Failure models that predict time or cycles to failure have been available for materials and simple structures for years. But no failure models have been developed for factors other than stress and strength.. The development of failure models and the data to implement them is a major challenge for the maintenance community.

FAILURE ANALYSIS: The first name for the present MFPG **Failure Analysis** Committee was **Mechanisms of Failure**. The original broad objective of the Committee was to coordinate all aspects relating to the modes and mechanisms of failure as they pertained to the overall goal of mechanical failure prevention. Among other things, the pursuit of this objective involved standardization of terminology and the establishment of interrelationships among causes, modes and results of failure. The scope of the current Committee not only covers mechanisms but also methods of failure analysis and the application of *lessons learned* from failure analysis results. Collins [9] defines *mechanical failure* as *any change in the size, shape or material properties of a structure, machine or machine part that renders it incapable of performing its intended*

function. He defines a *failure mode* as *the physical process or processes that take place or combine their effects to produce failure.* Collins lists and defines 23 failure modes, five of which have several sub-categories.

Beginning with the early MFPG conferences failure modes were discussed extenxively. The important modes of fatigue, wear and corrosion were the topics for several meetings. The eleventh meeting was devoted exclusively to an examination of *mechanical fatigue* as a critical failure mechanism. Of the MFPG papers relating to *wear*, the majority deal with minimizing wear by design of surface finish or by the use of lubricants and coatings. A number of papers in the proceedings concern wear problems and solutions for specific critical components such as bearings. This is at least in part due to the extreme complexity of the wear failure mode. At least nine subcategories of wear have been defined and more than twenty variables are involved in the wear process. Progress in wear reduction is therefore better demonstrated by reporting on specific solutions to the problem. For example, MFPG 13 was concerned with the standardization of surfaces to minimize wear (or fatigue) failure. The 16th and 30th meetings were devoted to lubricants and lubrication. All of the 23rd and part of the 37th meetings dealt with coatings to improve wear resistance.

Corrosion is another important complex failure mode. As with wear, many variables are involved in the corrosion process that relate to environmental, electrochemical and metallurgical aspects. There are eleven recognized direct subcategories for corrosion; in addition, *fretting corrosion*, *stress corrosion*, *corrosion wear* and *corrosion fatigue* may be regarded as special synergistic failure modes. The 15th and 17th meetings were concerned in great part with various aspects of corrosion. The objective of MFPG 15 was to examine the state-of-the-art in the study of various corrosion mechanisms. The papers at MFPG 17 addressed environment-sensitive failure modes such as *stress corrosion cracking* and lubrication failure which can, among other things, lead to corrosion.

Cavitation as a damage mechanism was the topic for discussion at MFPG 19, with more than a dozen excellent papers covering various aspects of the problem. Peterson [10] provides a detailed discussion on how this complex, imperfectly understood phenomenon can lead to mechanical failure. M. B. Peterson [11] pointed out that the literature has been somewhat confusing concerning the *fretting* mode and proceeded to clarify the phenomenon. The theme of MFPG 35 was "Time-Dependent Failure Mechanisms and Assessment Methodologies". The principal failure modes covered were fatigue, stress corrosion cracking and creep. The papers dealt with these mechanisms in connection with a variety of applications, including gas transportation cylinders, concrete, wood and ceramics. This is consistent with the MFPG objective to examine problems related to failure of structures and machinery and to develop realistic approaches to their solution.

The analysis of the cause of failure is important to the failure prevention effort. The most vivid examples are failures that cause catastrophic accidents resulting in loss of life, such as air or rail crashes. The National Transportation Safety Board (NTSB) investigates all such accidents as a matter of routine [12] with the goal of avoiding similar failures in the future. Such cases are extraordinary, but it is always useful to determine the root cause of failure when it is

economically feasible to do so. The general topic of the 8th MFPG meeting was an examination of "Critical Failure Problem Areas in the Aircraft Gas Turbine Engine." The papers at this conference were not formal failure analyses. Rather, they report on causes of failure as determined during required engine repair or routine overhaul and maintenance. Failures in internal combustion engines were treated in a similar way at MFPG 12. Bennett [13] presented an excellent paper at MFPG 20 on the importance of analyzing service failures, what can be learned from the analyses, and how to use the information that is obtained. The theme of the 21st MFPG meeting was "Success by Design: Progress Through Failure Analysis" could be considered a response to Bennett's paper. The 21st MFPG Proceedings clearly demonstrates the value of the feedback of information gleaned from service failures into the design process. This idea is strongly supported in papers by Rieger et al [14], Pond [15] and Natishan [16].

LIFE EXTENSION AND DURABILITY: The current **Life Extension and Durability Committee** was formed only recently. Its organizational meeting was held in April 1994 at MFPG 48, thus its scope and mission is still evolving. The terms *durability* and *life extension* need some clarification. Ideally, a design engineer establishes some achievable goal for the expected service life of the system he is designing, then proceeds to design the system to meet that goal. The durability of the system is some measure of how well the system survives without failures throughout its expected service life, while functioning effectively with little or no loss in efficiency. If as a result of a special design effort, the system is capable of functioning beyond its normal expected service life, the designer will have achieved life extension by design. Life extension can also be achieved through the use of on-line condition monitoring systems capable of detecting imminent failure in time to avoid failure through effective maintenance procedures. Properly implemented, the condition based maintenance approach can ensure that machinery, for example, will operate at peak efficiency, not only through its normal expected service life but well beyond. Based on the present state of the technology, it is reasonable to assume that designers are not able to estimate the service life of their system with any degree of accuracy. Instead, it is suggested that engineering designers will develop an environmental life cycle profile for the system they are designing and use their knowledge of materials and failure modes to develop the most durable and reliable system that they can.

At MFPG 27, durability and life extension was examined with respect to the durability of consumer products. In this case, durability involves both technical and economic considerations. Thus the major problem of design is to increase durability within the constraints of economics. Lund and Denney [17] offer keen insight on the issues involved in extending the life of consumer products. Other papers dealt with product life testing methods and reasons for being concerned about improved product performance, such as materials conservation, waste reduction and product safety. The 27th meeting also included a panel discussion on the topic *Can and Should Product Life Be Extended*? From the published introductory remarks by the panelists, it appears that the debate was lively with respect to both aspects of this question.

Active service life extension programs are ongoing in the nuclear power, aircraft and petrochemical industries, as well as for bridges and highways. Manufacturers recognize the wisdom of extending the life of their machinery. A few papers at MFPT 49 dealt with life

extension, through both design and maintenance procedures. It is expected that presentations on these topics will be much in evidence at future MFPT conferences.

SENSORS TECHNOLOGY: Sensors are key elements of data acquisition systems and, since measurement has always been an important part of mechanical failure prevention efforts, sensor technology has been of prime interest to the MFPG since its inception. Whether the problem is fault detection and diagnosis, structural inspection for defects or experimental life estimation procedures, the use of sensors is usually required. For this reason it is the role of the experts on the Sensors Technology Committee to provide technical guidance for other MFPT committees on linking failure modes to appropriate sensing technologies.

In 1968, Janowiak [18] suggested that sensors used for the diagnosis of mechanical failures must be adapted to physical effects resulting from known failure modes in mechanical systems. He pointed out that a vast sensor technology exists which is applicable to diagnostics and that a comprehensive search for sensors to bridge the interface between the mechanical and diagnostic systems was required. It is not certain whether such a search was conducted in an organized way, but a search of the proceedings reveals that at least twenty different sensor applications are reported at MFPG 6 through MFPG 36; many of these were very unusual. This is evidence that innovative engineers can develop sensors to measure almost any failure related parameter that may emerge. There are other examples in later proceedings. Floyd [19] discussed four types of sensors useful for condition monitoring. One type was an advance warning ice detector for the inlet stages of gas turbine engines, a very critical problem. Redden [20] described an intriguing smart integrated microsensor system capable of real time fatigue analysis of strain data. The use of embedded fiber optic probes to measure temperature through-the-thickness of composites was described by Whitesel and Sorathia [21]. Some interesting comparisons were made on the use of laser vibrometers, parabolic microphones and accelerometers for rolling element bearing diagnostics [22]. Several other new or unique sensor developments were reported in MFPG 48 and MFPT 49 [23-29]. It is reasonable to conclude that some of the greatest advances in the mechanical failure technology arena over the past 25 years have been in sensor development.

CLOSURE: A limited discussion of mechanical failure prevention technology as reported in the proceedings of the MFPG meetings since 1967 has been attempted. Nearly 1000 papers are published in these proceedings, in whole or in part, over the 28 year period. Some of the earlier proceedings provided only abstracts or synopses for several of the papers. These proceedings represent a significant body of work in a number of different technical areas all tied together by the common thread of failure prevention.

Diagnostics and Prognostics is prominent in the proceedings, in part because it is a "hot topic" in both the public and the private sector. Successful Condition Based Maintenance (CBM) programs using D&P technology can have a very high payoff, both technically and economically. The majority of U.S. industrial firms view effective maintenance as a fundamental prerequisite to economic success and are currently using some type of Predictive Maintenance (PdM), a major component of CBM.

The MFPG/MFPT proceedings clearly indicate significant improvements in condition monitoring and diagnostics capabilities over the past 25 years. Improved techniques developed by creative engineers, coupled with marked advancements in diagnostics hardware and software have resulted in more successful PdM efforts. In many cases, faults can be detected early, the location of the fault can be identified and appropriate maintenance can be scheduled in time to avoid catastrophic failure. More and more PdM systems are automated and many employ a combination of diagnostic techniques. In many cases, time to failure can be predicted with reasonable accuracy. However, prediction of time to failure should not be confused with prediction of remaining service life. Although some progress has been made in prognostics, much work remains to be done on the development of life estimation methods. The difficulty of predicting service life increases with the complexity of the system and the number of potential failure modes.

Procedures for conducting failure analyses are well established but are usually relatively expensive for systems with any degree of complexity. Natishan [16] points out that there are many situations in which component failure has little or no impact on safety, reliability or economics. Therefore, the cost of doing a failure analysis must be weighed against the benefits to decide whether or not an analysis is justified. As a general rule, in all cases where failure has a major impact on safety or operations it is essential that the cause of failure be determined.

This paper began with a quotation from the proceedings of MFPG 20. Six papers from that proceedings [30-35] addressed the broad implications of mechanical failures for the divergent segments of our society. These papers provide useful perspectives on mechanical failure that, taken together, define the national scope of the failure prevention problem and clearly show that the impact of such failures on our society is very broad. Almost all segments of the public and private sectors are concerned in one way or another with failure prevention. Beyond that these papers strongly suggest that solutions will be found more quickly through mutual cooperation, better communication links and effective mechanisms for interchanging technical information on mechanical failure prevention technology.

REFERENCES

1. R. L. Eshleman, "Detection, Diagnosis and Prognosis: An Evaluation of Current Technology", Proc. MFPG 44, Vibration Institute, Willowbrook, IL, 1990.
2. J. M. Ward, "Spectrometric Oil Analysis", Proc. MFPG 2, Office of Naval Research, Arlington, VA, June 1967.
3. J. Reintjes et al, "Optical Oil Debris Monitor", Proc. MFPG 48, Vibration Institute, Willowbrook, IL, 1994.
4. S. F. Smith, K. N. Castleberry and C. H. Nowlin, "Machine Monitoring via Motor-Current Demodulation Techniques", Proc. MFPG 44, Vibration Institute, Willowbrook, IL, 1990.
5. H. L. Dunegan, "Using Acoustic Emission Technology to Predict Structural Failure", Proc. MFPG 22, National Bureau of Standards, Washington, DC, 1975.
6. A. L. Wicks, "Modal Analysis as Applied to Predictive Maintenance Programs and Machine Prognostics", Proc. MFPG 36, Cambridge University Press, Cambridge, MA, 1983.

7. M. H. Richardson, "Are Modes a Useful Diagnostic in Structural Fault Detection?", Proc. MFPG 47, Vibration Institute, Willowbrook, IL, 1993.
8. R. G. Salter, "Improving Vehicle Life-Cycle Reliability by Prognostic Maintenance Management Through Geriometry", Proc. MFPG 28, National Bureau of Standards, Washington, DC, 1978.
9. J. A. Collins, "Failure of Materials in Mechanical Design: Analysis, Prediction, Prevention", Second Edition, John Wiley and Sons, New York, NY, 1993.
10. F. B. Peterson, "Physics Associated with Cavitation Induced Material Damage", NBS Special Publication 394, National Bureau of Standards, Washington, DC, 1973.
11. M. B. Peterson, "Fretting", Proc. MFPG 15, Office of Naval Research, Arlington, VA, 1971.
12. J. F. Wildey, II, "Safety Issues Arising from the Metallurgical Investigation into the DC-10 Airplane Engine Fan Disk Separation, Sioux City, Iowa", Proc. MFPG 45, Vibration Institute, Willowbrook, IL, 1991.
13. J. A. Bennett, "What We Can Learn from the Examination of Service Failures", Proc. MFPG 20, NBS Special Publication 423, National Bureau of Standards, Washington, DC, 1976.
14. N. F. Rieger, T. H. McCloskey and R. P. Dewey, "The High Cost of Failure of Rotating Equipment", Proc. MFPG 44. Vibration Institute, Willowbrook, IL, 1990.
15. R. B. Pond, Jr., "Failure Prevention Through Failure Analysis", Proc. MFPG 45, Vibration Institute, Willowbrook, IL, 1991.
16. M. E. Natishan, "Learning From Failure", Proc. MFPT 49, Vibration Institute, Willowbrook, IL, 1995.
17. R. T. Lund and W. M. Denney, "Opportunities and Implications of Extending Product Life", Proc. MFPG 27, NBS SP 514, National Bureau of Standards, Washington, DC, May 1978.
18. R. M. Janowiak, "Instrumentation for Use in the Diagnosis of Mechanical Systems", Proc. MFPG 4, Office of Naval Research, Washington, DC, February 1968.
19. M. D. Floyd, "An Overview of Some Modern Sensor Technologies and Their Impact on Implementation and Utilization Concepts", Proc. MFPG 40, Cambridge University Press, Cambridge, MA, 1987.
20. P. J. Redden, Jr. "Smart Integrated Microsensor System", Proc. MFPG 45, Vibration Institute, Willowbrook, IL, 1991.
21. H. K. Whitesel and U. A. K. Sorathia, "Fiber Optic Temperature Measurements in Composites", Proc. MFPG 45, Vibration Institute, Willowbrook, IL, 1991.
22. R. L. Smith, "Rolling Element Bearing Diagnostics with Lasers, Microphones and Accelerometers", Proc. MFPG 46, Vibration Institute, Willowbrook, IL, 1992.
23. K. Meissner and P. Sincebaugh, "Preventing Mechanical Failures in Resin Transfer Molding Using Embedded Sensors and Neural Networks", Proc. MFPG 48, Vibration Institute, Willowbrook, IL, 1994.
24. C. M. Teller and H. Kwan, "Magnetostrictive Sensors for Structural Health Monitoring Systems", Proc. MFPG 48, Vibration Institute, Willowbrook, IL, 1994.
25. H. K. Whitesel and M. J. Ransford, "Fiber Optic Sensors for Machinery Health Monitoring and Control", Proc. MFPG 48, Vibration Institute, Willowbrook, IL, 1994.
26 J. Reintjes et al, "Optical Debris Monitoring", Proc. MFPT 49, Vibration Institute, Willowbrook, IL, 1995.
27. N. Gupta et al, "High Speed Image Processing in Wear Debris Monitoring", Proc. MFPT 49, Vibration Institute, Willowbrook, IL, 1995.

28. F. E. Kennedy et al, "Temperature Sensors for Detecting Failure of Tribological Components", Proc. MFPT 49, Vibration Institute, Willowbrook, IL, 1995.
29. L. Popyack and J. Kubler, "Condition Monitoring Using the Time Stress Measurement Device", Proc. MFPT 49, Vibration Institute, Willowbrook, IL, 1995.
30. J. P. Hirth, "Mechanical Failure: Implication for Science", Proc. MFPG 20, NBS Special Publication 423, National Bureau of Standards, Washington, DC, 1976.
31. H. W. Paxton, "Implications for Action for Engineering", Proc. MFPG 20, NBS Special Publication 423, National Bureau of Standards, Washington, DC, 1976.
32. W. D. Compton, "Mechanical Reliability - Implications for Engineering, Manufacturing and Design", Proc. MFPG 20, NBS Special Publication 423, National Bureau of Standards, Washington, DC, 1976.
33. J. E. Ryan, "Implications for Action - Industry", Proc. MFPG 20, NBS Special Publication 423, National Bureau of Standards, Washington, DC, 1976.
34. L. M. Kushner, "The Implications of Mechanical Failures for Consumer Product Safety - Vice Versa", Proc. MFPG 20, NBS Special Publication 423, National Bureau of Standards, Washington, DC, 1976.
35. R. W. Roberts, "Mechanical Failure - A Material Matter", Proc. MFPG 20, NBS Special Publication 423, National Bureau of Standards, Washington, DC, 1976.

A REVIEW OF PLANT MAINTENANCE METHODS AND ECONOMICS

M J Neale
OBE BSc(Eng) DIC FCGI WhSch FEng FIMechE
Neale Consulting Engineers Ltd
43 Downing Street, Farnham, Surrey UK GU9 7PH

Abstract: This paper gives a brief review of the available techniques for plant maintenance and suggests that a mix of techniques is probably the optimum for any particular piece of plant, with different techniques being used for different components. Some approximate guidance is also give on maintenance expenditure and on industries in which improved maintenance is likely to give the greatest financial return.

Key Words: Component failure; Condition monitoring; Maintenance costs; Predictive maintenance; Preventative maintenance

The maintenance of plant and equipment is an area of activity which, in recent decades, has received a lot of attention from management consultants, with the associated generation of many buzz words. It may therefore be useful to review plant maintenance methods to see how the various "types" fit together, where they are particularly appropriate to apply, and what the relative economics of them are likely to be.

Maintenance Techniques: The original plant maintenance technique was breakdown maintenance. In this case an item of plant would be repaired each time that it broke down. The problem however is that the process of failure often creates consequential damage, which calls for more extensive repairs. Also there may be delays in the repair process because spare parts and specialised skilled labour may not be immediately available.

An improvement can be obtained by moving to regular preventive maintenance where the plant is stopped at intervals, often annually, and partly stripped and inspected for faults. The problem then is that the components of machines do not fail at regular intervals, but more with a distribution as in Figure 1. To avoid all service failures the time between overhauls has, therefore, to be very short or some ongoing in service failures have to be accepted.

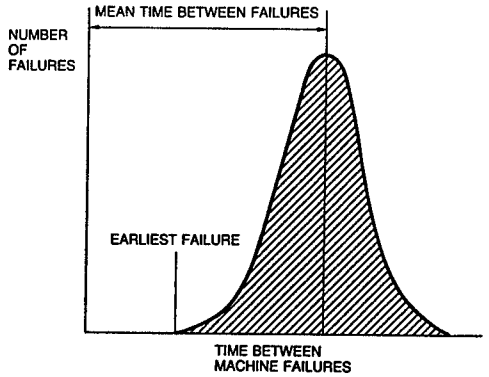


Fig. 1 The distribution of the time to failure for a typical component

The next step forward is to move to condition based maintenance where the critical components are monitored for deterioration and the maintenance is carried out just before the problem arises. The principle is to select an appropriate method of monitoring for deterioration from those listed in Table I and apply it to the machine in order to obtain a lead time of warning in advance of a failure as indicated in Figure 2.

Method	*Principle*	*Application examples*
Wear debris monitoring	The collection and analysis of wear debris derived from component surfaces, and carried away in the lubricating oil	Components such as bearings or other rubbing parts which wear, or suffer from surface pitting due to fatigue
Vibration monitoring	The detection of faults in moving components, from the change in the dynamic forces which they generate, and which affect vibration levels at externally accessible points	Rotating components such as gears and high speed rotors in turbines and pumps
Performance monitoring	Checking that the machine components and the complete machine system are performing their intended functions	The temperature of a bearing indicates whether it is operating with low friction. The pressure and flow rate of a pump indicate whether its internal components are in good condition.

Table I **Monitoring methods and the components for which they are suitable**

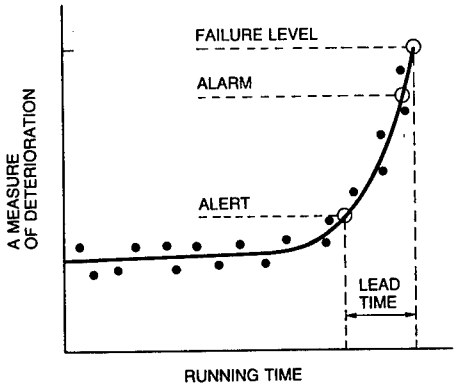
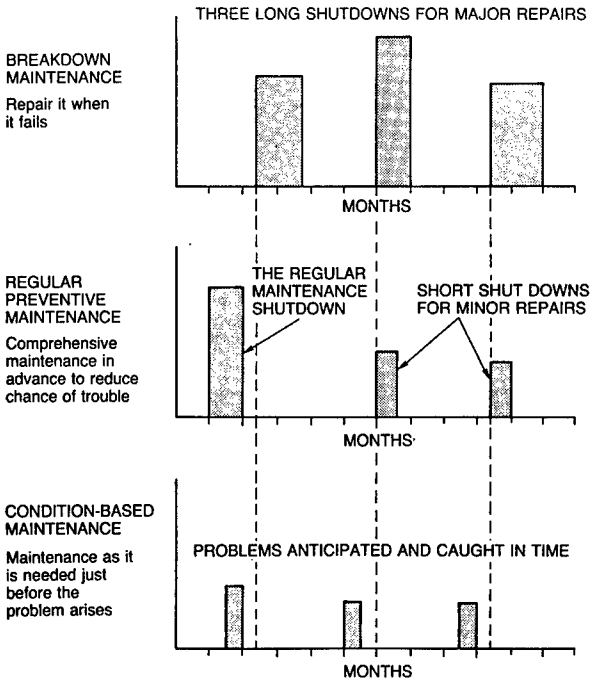


Fig. 2 **The lead time before failure provided by condition monitoring**

The expected effect of moving from breakdown maintenance through to condition based maintenance, for a plant which might have had three breakdowns in a year, is shown in Figure 3. The effect shown here is impressive and could stimulate all plant users to start monitoring all the components of their machines.

However in practice it is not as simple as this, because, not all failures can be detected by condition monitoring, and the economics of the situation limits the number of components that are worth monitoring, so a choice has to be made. Also there will be a number of components on any machine for which condition monitoring is not particularly appropriate.



The height of the bars indicates the amount of maintenance effort required.

Fig 3 The results of maintaining the same industrial plant in different ways.

The optimum mix of techniques: The various components of a machine may optimally be maintained in different ways and Figure 4 shows this in a diagrammatic form. Also the starting point of this diagram is the selection of critical machines in a plant, because not all machines will be critical. For example if machines are duplicated and can be readily repaired after failure, carefully organised monitoring and maintenance procedures may not be necessary for them.

Figure 4 could be used to provide some broad guidance on the choice of suitable maintenance methods for a critical machine. Such a machine may have some non critical components which are in this category, either because they are extremely unlikely to fail, or because, if they fail, they do not require the machine to come out of service, and can be replaced on line. The critical components are those which, if they fail, make it necessary to stop the machine, and may also cause it to suffer consequential damage.

Many of these components may be optimally maintained on a basis of their condition, as assessed by condition monitoring. However for some of these components, if their operating conditions are relative consistent and monotonous, it may be more economical to maintain them on a regular timed replacement basis.

An essential feature of components that are suitable for condition monitoring, however, is that they must have a progressive failure mode. This is because it is essentially this progression which is detected by the monitoring technique in order to give a lead time prior to the final development of the failure.

Components which, in contrast, fail with a more sudden mode, need to be replaced on the basis of life expiry. This usually involves an assessment of possible fatigue and creep, and the results of testing and experience. Also since various machine operating conditions may produce increased rates of deterioration, it will usually be necessary to record these conditions, and use them as a basis for computing the maximum allowable operating hours for the component, before it must be changed.

The majority of the critical components of most machines will, however, generally be found suitable for condition monitoring by one method or another. If all of them were to be monitored and a variety of different methods used, the cost of doing this could be unacceptable. It is therefore necessary to have some means of priority assessment so that the most critical components can be selected. This is shown in the diagram in Figure 5 which suggests that a suitable method of assessment may be to judge the required priority by looking at an index derived from the product of indicators of the likelihood of failure and the effect of any failure, together with the expected repair time.

The next decision will be to decide how far down the priority list to go, or how many components is it essential to monitor. Safety requirements are an important factor, as well as the necessary minimum size of the maintenance budget. These issues are discussed in the final section of this paper.

The monitoring techniques to be used also require consideration because there are several different techniques available, commonly with two of three that may be applicable to any one component. For efficiency, training and familiarity, at a particular plant, it is desirable to use the minimum number of different techniques. However if a plant shutdown carries very high economic penalties, it is generally also desirable to have more than one technique, dependent on different physical measurements, in order to avoid false alarms, which can arise from a single monitoring alarm. For these reasons it is suggested in Figure 5 that it is desirable to list the possible monitoring technique for each high priority component so that one or two methods that are common to these components can be selected for use on the plant.

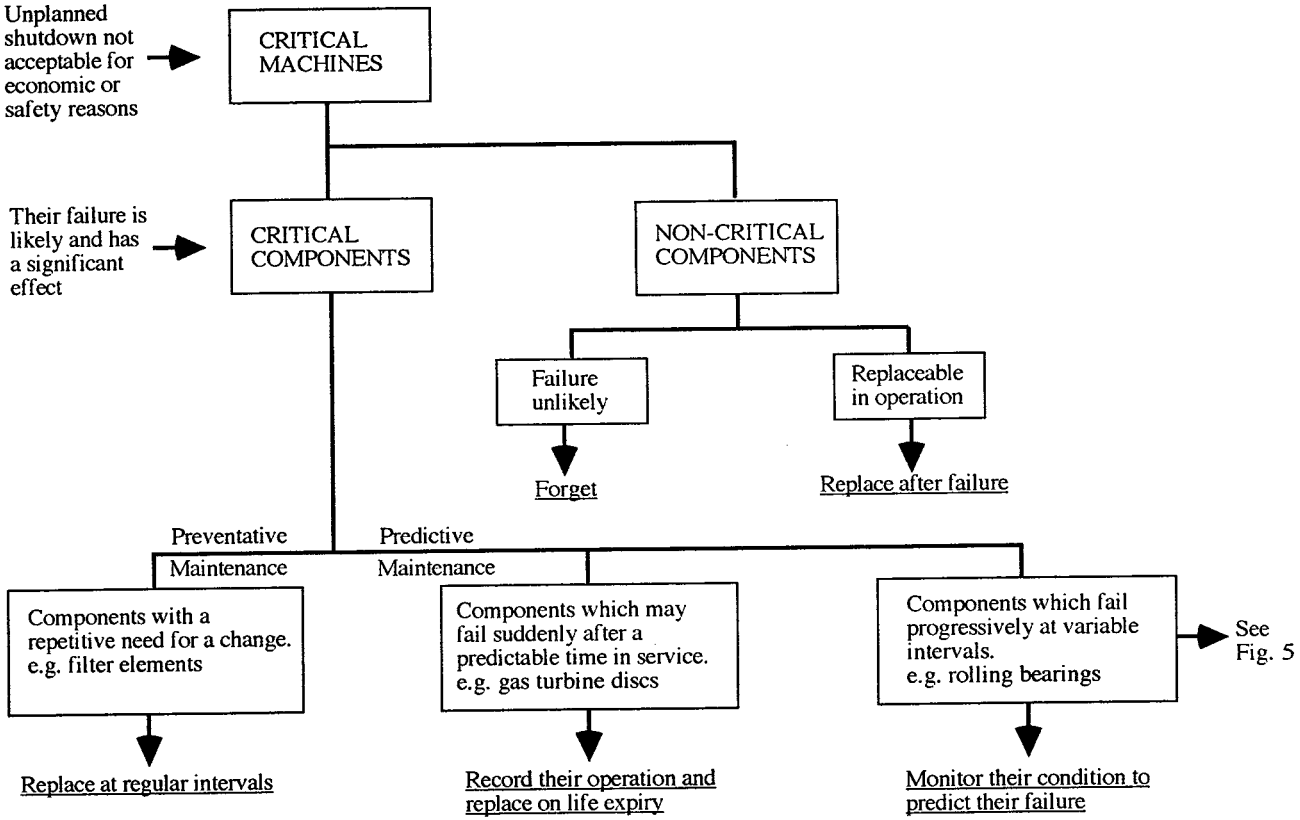


Fig. 4 Appropriate maintenance techniques for various machine components

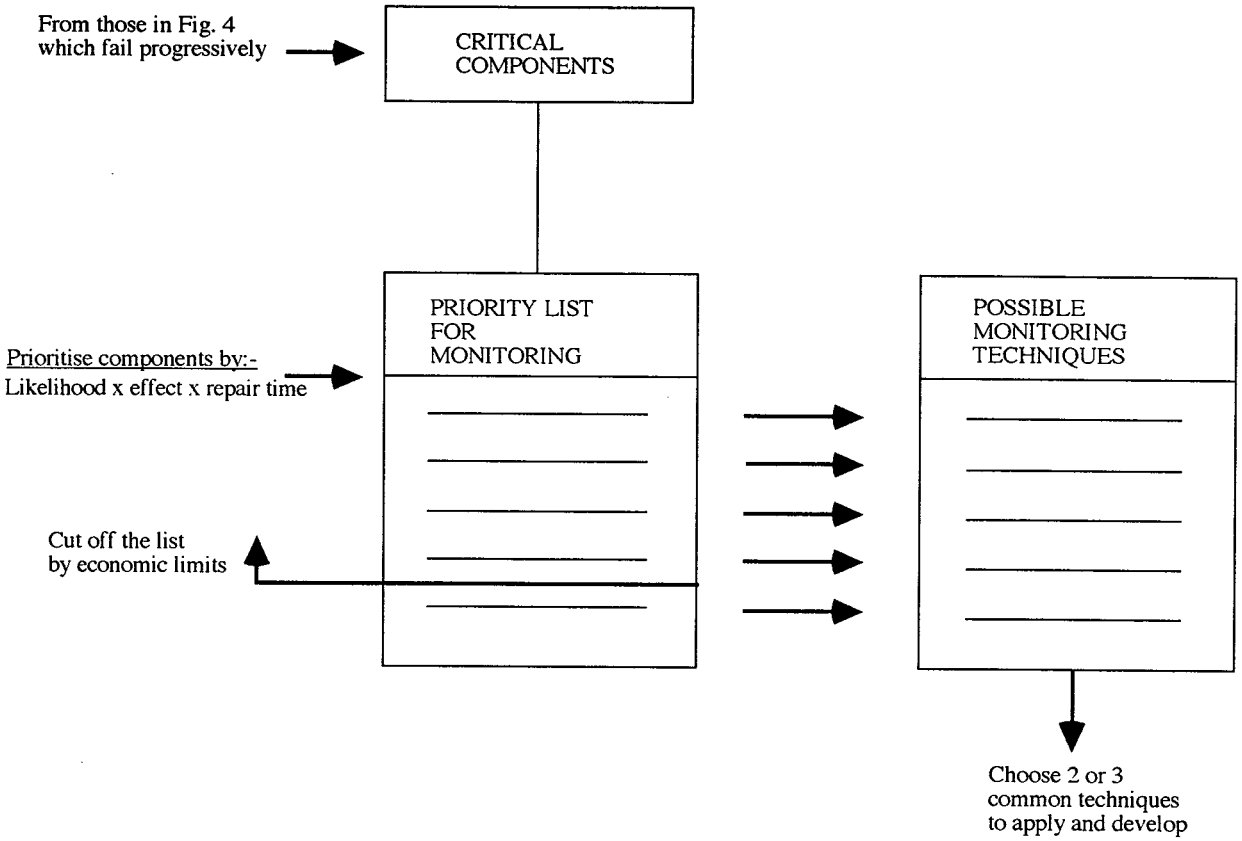


Fig. 5 The choice of components and techniques for component condition monitoring

Maintenance economics: The optimum budget is something that is very difficult to determine. One problem is that it is often difficult to persuade the financial management of a company that they should spend money in advance to avoid a problem that has not yet arisen. Expenditure on condition based maintenance may therefore be resisted, while expenditure to repair a disaster arising from a breakdown maintenance policy, will be agreed as obvious.

Another cause of problems is that very robust machinery can appear to be very tolerant of reduced and inadequate maintenance. Budgets are sometimes cut on a trial percentage basis to see if any problems arise, and robust machines often suffer as a result because it may take five years before disaster strikes. The cost of repair is then usually vastly more than the total savings achieved, from the reduced maintenance budgets. Clearly some economic guidance figures on optimum maintenance levels are required.

Some years ago a government funded study in the UK investigated the amounts that companies considered to be worth spending on setting up improved maintenance systems. This indicated that an expenditure corresponding to 1% of the capital value of the plant could achieve annual savings of the order of 1% of the added value of the company's activities, as indicated in Figure 6. Also if safety was a major consideration it was worth spending up to 5% of the capital value of the plant, which resulted in major improvements. While these figures do not give precise guidance on annual budgets, they do give some indication of the sensitivity of a system to improved standards of maintenance.

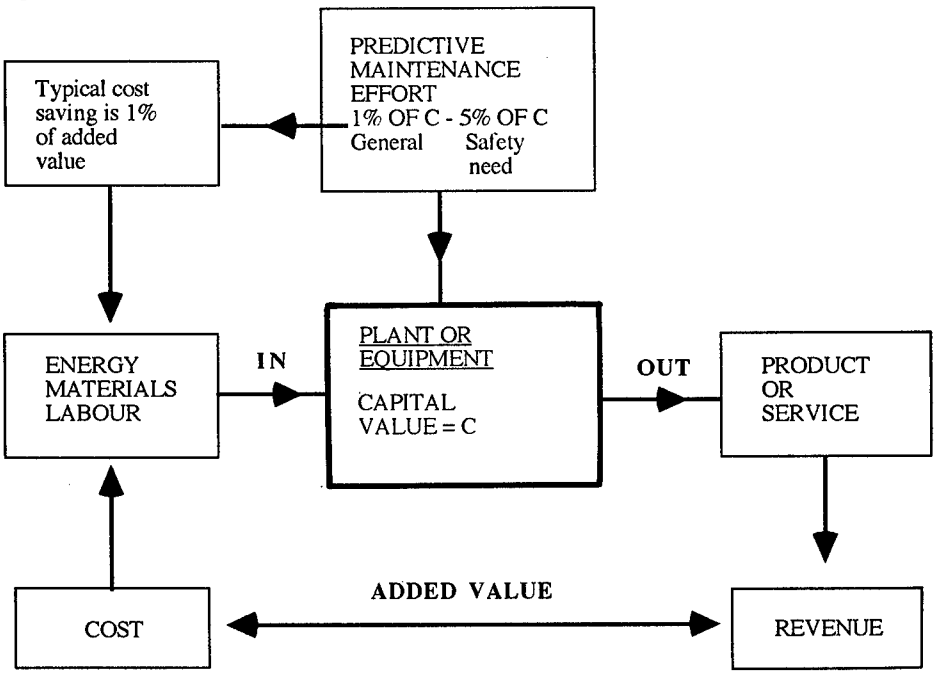


Fig. 6 An indication of the economics of predictive maintenance

In the same study an attempt was made to assess the industries that could achieve improved savings from better maintenance strategies. This was recognised as being those industries which were capital and machine intensive, and which had large enough plants for the process to be effectively organised. Figure 7 shows a range of industries plotted against these two indicators. It shows that industries towards the top right hand corner of this diagram should be able to maintain or increase their maintenance budgets with a higher level of confidence.

This paper is only an approximate overview of methods of plant maintenance. It is hoped however that it may provide some guidance and stimulation for those industrial companies who have not yet taken full advantage from some of the latest maintenance techniques.

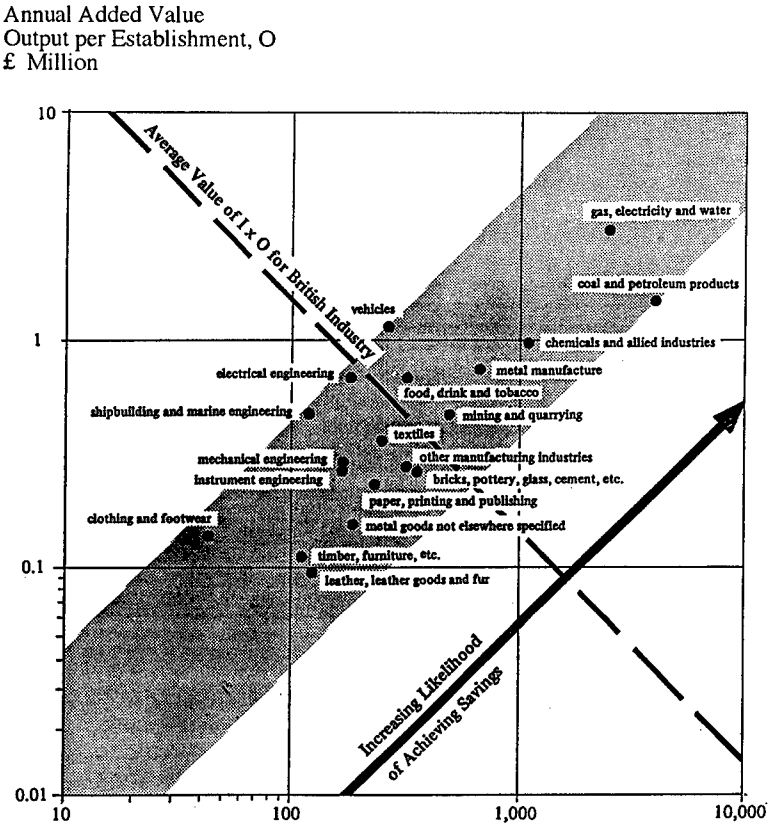


Fig. 7 The industries most likely to profit by the application of predictive maintenance

THE HISTORY AND APPLICATION OF THE ENVELOPE DETECTOR

John L. Frarey
JLF Analysis
845 Worcester Drive
Schenectady, NY 12309

Abstract: The envelope detector has become one of the standard techniques used in the detection of defective rolling element bearings. Some forms of the envelope detector are also used in gear box analysis. First uses of the envelope detector seem to date to 1971. This paper will discuss the history of the development. Data from the envelope detector will be analyzed to determine which data characteristics are due to the defective bearing and which are due to the structure and which are due to the characteristics of the envelope detector. Other forms of envelope detectors such as those employing the Hilbert transform will also be covered.

Key Words: Envelope detection; Demodulation; High frequency vibration; Hilbert transform; Machinery diagnosis; Rolling element bearings; Vibration analysis

Introduction: Almost all the early effort to diagnose defective rolling element bearings was based on detecting the classic bearing defect frequencies in low frequency (generally below 500 hz) vibration data. In 1971, with NASA - Huntsville support, a program to investigate early detection of bearing defects showed that this classical approach was not capable of detecting the onset of bearing defects. High frequency vibration and envelope detection was almost accidentally discovered in this program and was shown to be very effective in the early detection of bearing faults. This technique has enjoyed widespread use in the diagnosis of bearing defects particularly in the last five to ten years. Understanding the reason for the failure of the classical diagnostic scheme to detect early failures, which is due to the impact nature of the signal generated by a defect, is also widespread. What is not as well understood is that the amplitude of the envelope detected defect signal from the bearing is a function of the size of the defect but also, unfortunately, is a function of the characteristics of the resonance(s) excited and circuit variables of the envelope detector such as the time constant of the filter employed in the envelope detector. Understanding the role of the analysis circuitry in the defect signal generation is necessary to understand that the data produced by manufacturer A's system may be different from that generated by manufacturer B.

A software envelope detector may be developed by use of the Hilbert transform and analysis of the analytic signal produced by operating on the original and transformed signal. A more brute force approach to a software envelope detector is simply to employ a peak detector with a very fast rise time and slow decay time. A software envelope detector has some advantages, however not enough investigation has been conducted into the response of the Hilbert transform detector

when it is presented with multiple resonances excited by the impact. The use of software envelope detectors is not as mature a field as the use of hardware detectors.

Background: The classical approach to the detection of rolling element defects was to examine the low frequency vibration or sonic spectrum for the presence of defect frequencies. These frequencies were calculated by the well known defect frequency formulas that were developed to determine the repetition rate for balls striking an outer race or inner race defect. These equations are developed from the geometry of the bearing and are given below for Ball Bearings:

$$f_o = n/2 * f_r * (1 - BD/PD * \cos\alpha)$$

$$f_i = n/2 * f_r * (1 + BD/PD * \cos\alpha)$$

$$f_s = 1/2 * f_r * PD/BD * (1 - (BD/PD)^2 * \cos^2\alpha)$$

Where:

n = Number of Balls	f_o = Outer race defect frequency
BD = Ball diameter	f_i = Inner race defect frequency
PD = Pitch Diameter	f_s = Ball spin frequency
f_r = rpm/60	α = Contact angle

For roller bearings, α is zero. The equations may also be rewritten in terms of outer and inner race diameters along with the ball or roller diameter.

The diagnostic concept was simply to calculate the defect frequencies and then monitor the vibration or sonic spectrum until these frequencies appeared. While there was some success with this technique, the results were not at all consistent. In order to investigate this, a controlled test was run in the NASA R & D program mentioned above. A small defect was machined into the inner race of a test bearing and then the bearing was operated in a fully instrumented test rig. None if the instrumentation results showed the presence of this defect; these included vibration, sonic, strain torque and temperature readings. Figure 1 shows the comparison of the vibration spectrum for the undamaged and the damaged bearing. These plots are in a logarithmic scale for amplitude which is not usually used. Even in this scale it is not clear if the defect may be detected. A very weak case could be made that the bearing defect signal is just beginning to appear, however there are other spectral regions in which the random variation is as great or greater than the defect region. To this point, the result of the program was virtually nil. One day, the program team was assembled in the instrumentation room while the rig was running trying to decide what to do next. One oscilloscope happened to be set to look at the raw data and one of the team members observed the signal similar to that shown in Figure 2. The time was measured between the impact signals and when converted to frequency, matched the defect frequency for an inner race defect. When a non defective bearing was substituted, the characteristic signal of Figure 2 disappeared.

Why hadn't any of the data processing techniques used detected the presence of a defect?

The first reason could have been found if a paper by H.L. Balderston [1] given in 1968 had been found and read. He noted that defective bearings produced a large amount of energy in the high frequency region of the spectrum, above 10 to 20 kHz. The problem of locating the results of other work in an area of interest was even more difficult in 1971 than it is now. Today vibration related papers would be listed in the Shock and Vibration Digest or a literature search could be requested at a library. Spectrum analyzers used at that time usually only analyzed data to 20 KHz so if the data were at frequencies higher than 20 KHz they would not have been seen by the investigators. To investigate if the energy were above 20 KHz, the data were tape recorded and played back at half speed. When this was done, a large resonant hump was noted in the spectrum at 28 KHz. So the first answer to the question of why no difference had been noted in the data was that its frequency was above the analysis range. It also serves as a good lesson that time data, preferably from an oscilloscope should always be examined along with the spectrum.

In looking at the data shown in Figure 2 again, the information of interest is not really in the high frequency but rather in the shape of the signal. If a line could be drawn around the envelope of the signal as shown in Figure 3, then the spectrum of this envelope should show the defect frequency. In order to produce the envelope signal of Figure 3 a demodulator similar to those used in the detection of amplitude modulated radio waves could be developed for the lower frequencies. A simple demodulator is shown in Figure 4. The value of R and C establishes the cut off frequency or the 3 db down point. The cut off frequency is given by the following equation:

$$f_{co}=1/(2*\pi*R*C)$$

For example, if R is 10 Kohms and C is .08 μF, the cut off frequency is 200 hz. The cut off frequency should be about 3 to 4 times the highest defect frequency and about 1/100 of the resonant frequency. We will see later in this paper how the cut off frequency affects the analysis results.

Employing the demodulator or envelope detector on the data similar to Figure 2 and then doing a low frequency spectrum of the enveloped data, the spectrum shown in Figure 5 results. The defect signal is now clearly evident with a very good signal to noise ratio (note: the amplitude scale is linear, not logarithmic). Commercially available envelope detectors are much more sophisticated than the circuit of Figure 4; using full wave rectification and active detectors, however, satisfactory results can be obtained using simple circuits. The result of this contract work was reported in the final report [2], to MFPG (3) and in NASA Tech Briefs [4&5].

Direct low frequency defect detection insensitivity: It has been shown that the direct analysis of the low frequency spectrum for the defect signal is much less sensitive than the envelope

detected signal. In order to understand the difficulty in analyzing the low frequency spectrum for the defect signal, we should examine the energy distribution in a spectrum as a function of the shape of the signal in the time domain. Figure 6 shows this relationship for a sine wave, a square wave and two more cases where the pulse width is greatly reduced. If the nature of the defect signal were a sine wave, then all of the energy would be in the fundamental defect frequency. As the pulse changes from a sine wave to a square wave, one sees that the energy is now shared between the odd harmonics and the amplitude of the fundamental is reduced. In the final example where the defect signal is a very short impulse, one can see that the energy is almost evenly shared over a very wide range of harmonics of the defect frequency. If the manner in which a defect in the race generates a signal is reviewed , it is seen that the most likely signal appearance is the very short pulse shown in the last case of Figure 6. Much as in the case of an automobile hitting a pothole in the road, a very sharp impact results.

The signal that we are looking for in the low frequency spectrum has an energy distribution as shown in the bottom spectrum of Figure 6. In observing normal vibration measurements from machinery, there is always a noise floor due to the machine noise background. In the case of the initiation of a small defect, the harmonics of the defect frequency may easily have amplitudes below the background noise and therefore will not normally be detected. If the amplitude of the impact increases due to more and larger defects, the defect signal will eventually be seen in the low frequency spectrum. In fact the spectrum for the short pulse impact looks very much like the spectrum generated by an impact from the hammer in modal analyses tests. The similarity with modal analysis does not end here. As in modal analysis, the broad distribution of energy will excite structural resonances that may then be studied. For the case of the bearing, the wide distribution of energy will also excite resonances at much higher frequencies. These resonances may be structural or they could be the resonance of the accelerometer. The resonances excited are normally in the high frequency region since at these frequencies, the structure does not have to have much displacement to achieve a high g level. The data in Figure 2 may therefore be viewed as the excitation and decay of some higher frequency resonance due to the impact generated by the ball hitting the surface defect in the race.

The effect of envelope detecting this resonance response is to eliminate the high frequency and greatly increase the pulse width and therefore increases the energy in the defect signal fundamental frequency. In addition to the lengthening of the pulse due to the Q of the resonance, the envelope detector time constant will also play a role. Trying to find a link between the high amplitude of the demodulated spectrum defect signal and the severity of the defect, we find out that it is not only a function of the strength of the impact (or the size of the defect), but also the Q of the resonance and the time constant of the envelope detector.

Factors affecting the amplitude of the demodulated signal: Examine Figure 7 to gain an insight into the relationship between the strength of the impact, the resonance damping and the time constant of the demodulator. The top trace in Figure 7 is a simulated signal in the time domain representing the excitation of the resonance and its decay. Also shown in this time trace is the envelope of the signal produced by the envelope detector. The initial peak amplitude of the resonance is a function of the strength of the impact and the amplification factor of the

resonance. If at a later date, the amplitude has increased (say from 1 to 2), then since the characteristics of the resonance are probably the same, the impact strength has doubled (the danger here is that the system is probably not nicely linear). If one compares the bearing peak amplitude for a bearing defect in one machine with a bearing in another machine however, note that the resonance characteristics such as the amplification factor, are probably not the same. In other words, if bearing one shows a peak of 1 and bearing two shows a peak of 2, the impact in bearing two is not necessarily twice that of one since the amplification factor of the resonance may be different. If the damping of the resonance is reduced, then the ring down time will be longer and the trace will spread out. The envelope decay is set not by the ring down time of the resonance but rather by the time constant of the demodulator. In the case shown, the time constant of the decay causes the envelope to be longer than the ring down time of the resonance.

The peak amplitude of the resonance, 1 in the case of Figure 7, is often used to indicate the presence of a defect in the bearing. As stated above, one bearing cannot be compared directly to a different bearing in a different machine, but on a given machine a good bearing will have a low peak amplitude and a defective bearing will have a significantly higher value. This peak amplitude is often called HFD for High Frequency Detection. The best way to measure the HFD is either by observing the signal on an oscilloscope or on a true peak reading voltmeter. Unfortunately almost none of the walk around data collectors that offer the HFD option measure it in this way. It is much easier to make a calculation in the digital world than to make analog measurements. The middle trace in Figure 7 is the spectrum of the high frequency resonance time trace. In order to calculate the RMS amplitude of a signal from the spectrum, one can simply take the square root of the sum of the squares of all spectrum bins over some frequency. The frequency selected is some frequency that will exclude the low frequency responses due to the once per rev and other low frequency signals. As shown in the center trace, the HFD calculated in this way is 0.172 rather than one. This value is not equal to 1 first because it is an RMS value and secondly because the signal in the time trace is neither a pure tone nor steady state but rather a transient signal, so even multiplying by 1.414 would not yield the peak amplitude.

The third trace shows the low frequency spectrum of the envelope of the resonance as seen by the envelope detector with its decay time constant. P_6 is the amplitude of the bin that has the defect frequency fundamental. Note the presence and amplitude of the harmonics of this signal. Do these relative amplitudes have diagnostic significance? The answer is no because the relative harmonic amplitudes is set by the decay shape or time constant of the demodulator and not due to the defect or even the ring down of the resonance.

Figure 8 compares the amplitude of the defect frequency fundamental for three different time constants in the demodulator. The top trace simply shows the high frequency resonance and ring down along with the HFD calculated in the incorrect (or false) manner. The next three traces shows the time constant varying from very short to very long. Note the resonance is the same in all three cases but the amplitude of the fundamental varies by 2:1. If the decay is either too short or too long, the amplitude will be reduced. The optimum time constant is selected so that the envelope decays to zero just before the next impact. The envelope detector however cannot

predict the defect frequency so it simply uses a constant value. While the amplitude of the defect signal may be compared for the same bearing in the same machine from one time to another, comparisons may not be directly made for situations where the defect frequency is significantly different or where another manufacturers envelope detector (and probably different time constant) is used.

Figure 9 compares the HFD and the defect frequency amplitude Pn_6. Note that as the damping of the resonance decreases, the ring down time increases but since the envelope time constant is at the optimum setting, the defect amplitude remains essentially constant. The HFD however changes over a range of 3:1 even though the actual peak amplitude of the impact does not change.

To summarize this section, we find that if a defect signal is observed in the demodulated spectrum, then there definitely is a defect present. It is difficult to set limits for either the HFD (true or false method of calculation) or the defect signal since these amplitudes are a function of the bearing defect **AND** the resonance amplification factor **AND** the damping **AND** the envelope detector time constant.

Software envelope detectors: Software envelope detectors may be implemented by use of the Hilbert transform or by implementing a peak detector with a fast rise time and a slower decay time constant (much as in the case of the hardware envelope detector).

The Hilbert filter or transform is a function whose output is exactly the same as the input except that its phase at all frequencies has been shifted by 90 degrees. This is shown in Figure 10. The input and output signals taken as a pair may be considered to be the real and imaginary components of a complex signal. This complex signal is called the analytic signal. In the case of a spectrum, each bin has a real and imaginary component and the magnitude is computed by taking the square root of the sum of the squares of the real and imaginary components. This same process may be done for the analytic signal where each pair of signals are at one instant in time rather than for a spectrum bin. If the input signal is a constant amplitude sine wave, then the output will be a straight line. If the input signal is modulated, then the output from this process will be the envelope of the modulation. This is shown in Figure 11.

It is easy to see that this process will work so long as the input is a well defined signal such as generated by a function generator as in Figure 11 or the gear mesh signal from a gear box. (Applying the Hilbert transform to gear box signals is an entirely different topic. A good explanation of the Hilbert envelope detector and its application to gear boxes is given in reference [6]). How would it work on the resonance excited by a defective bearing. Figure 12 shows the result of this application. The envelope is clearly shown in the center trace and the spectrum of the envelope is shown in the bottom trace. In the case of the data used to generate Figure 12, the data included only one low frequency resonance. In real life, when the resonance is at a much higher frequency, the spectrum of the envelope should be zoomed to the lower frequency portion in order to accurately identify the defect frequency. I have not seen this technique explored for complex cases where several resonances are excited at once. One way of limiting the analysis to a single resonance is to first zoom on one of the resonances, then

perform the Hilbert envelope detector on the zoomed time domain signal and then display the envelope spectrum. This is shown in Figure 13. The disadvantage of this zoom process is that the time trace (actually a complex signal) does not look like the modulation. A big advantage of the Hilbert envelope detector is that there is no envelope detector time constant to modify the data. The envelope faithfully follows the shape of the time domain resonance.

A simple peak detector may be implemented in software. The original signal, the envelope of this signal and the spectrum of the envelope are shown in Figure 14. The advantage of this approach is that the envelope will be produced no matter how many resonant frequencies make up the time trace. The disadvantage is that we have reintroduced the envelope detector time constant as a variable. For high frequency resonances, the spectrum of the envelope should be zoomed to the lower frequency region.

Conclusions: By using the envelope detector technology, one can see that the sensitivity to the presence of a small initial bearing defect is many times what it is by observing only the low frequency vibration or sonic data. This added sensitivity allows the detection of the onset of surface defects in rolling element bearings. This advantage is partially offset by the fact that the HFD and defect signal amplitudes are functions of structural and envelope detector characteristics in addition to the size of the defect. One possible approach would be to monitor the envelope spectrum for the presence of defect frequencies and once detected, expand the monitoring to looking for these defect signals in the low frequency spectrum. In any case, determining exactly when to remove a bearing just prior to catastrophic failure remains a crap shoot.

More work should be done investigating the Hilbert envelope detector to very complicated signals made up of several resonances, including the accelerometer resonance. The data reported here for Hilbert envelope detectors is just an initial look at how the Hilbert envelope detector responds to typical bearing defect signals.

References:

1. Balderston, H.L., "The Detection of Incipient Failures in Bearings", Presented at the 28th National Fall Conference of the American Society for Nondestructive Testing, Detroit, Mich. Oct 14-17, 1968

2. Broderick, J.J, Burchill, R.F. and Clark, H.L., "Design and Fabrication of Prototype System for Early Warning of Impending Bearing Failure", Design Report MTI-71TR to NASA under contract NAS8-25706, January 1972

3. Burchill, R.F. "Resonant Structure Technique for Bearing Fault Analysis", Presented at the 18th Meeting of the MFPG, Gathersburg, MD 1972

4. Broderick, J.J., Burchill, R.F. and Clark, H.L. "A system for Early Warning of Bearing Failures", NASA Tech Brief B72-10494, August, 1972

5. Burchill, R.F. and Frarey, J.L. "New Detection Method for Rolling Element and Bearing Defects", NASA Tech Brief B72-10689, October, 1972

6. McFadden, P.D., "Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Frequency", Transactions of the ASME Journal of Vibration, Acoustics, Stress and Reliability in Design, April 1986, Vol. 108

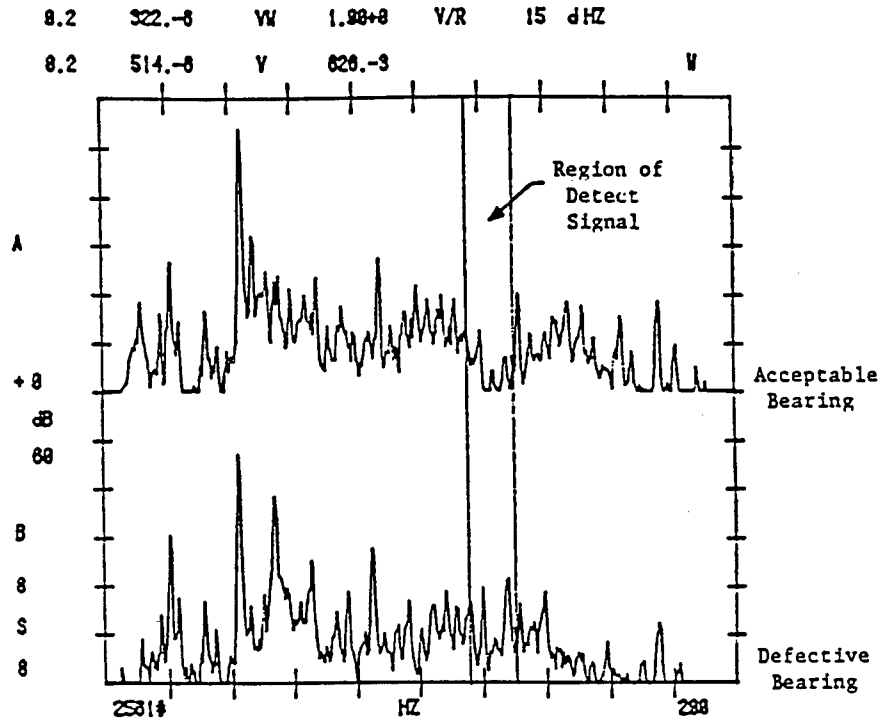


Figure 1 Comparison of the vibration spectrum for a damaged and undamaged bearing.

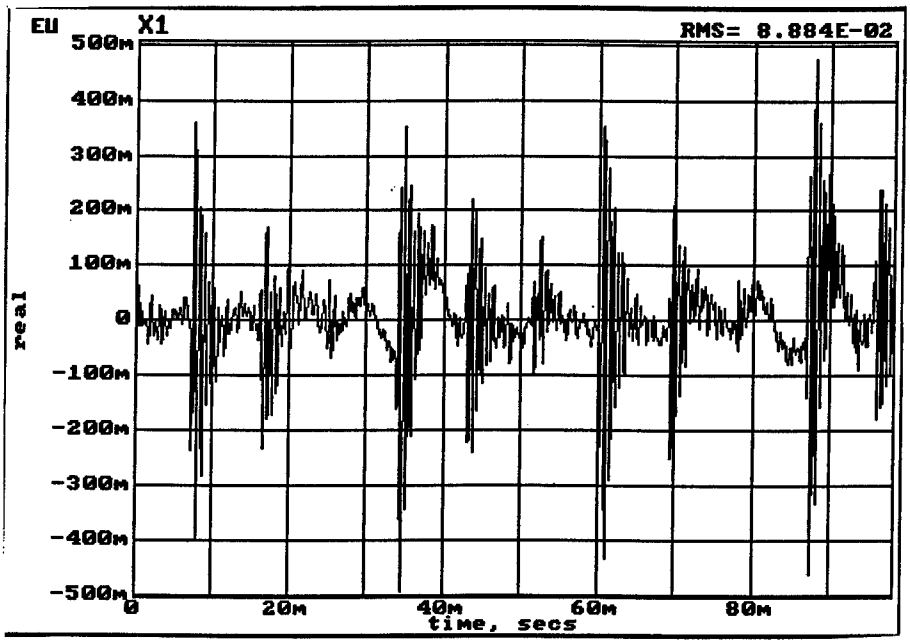


Figure 2 Time domain signal from a damaged bearing.

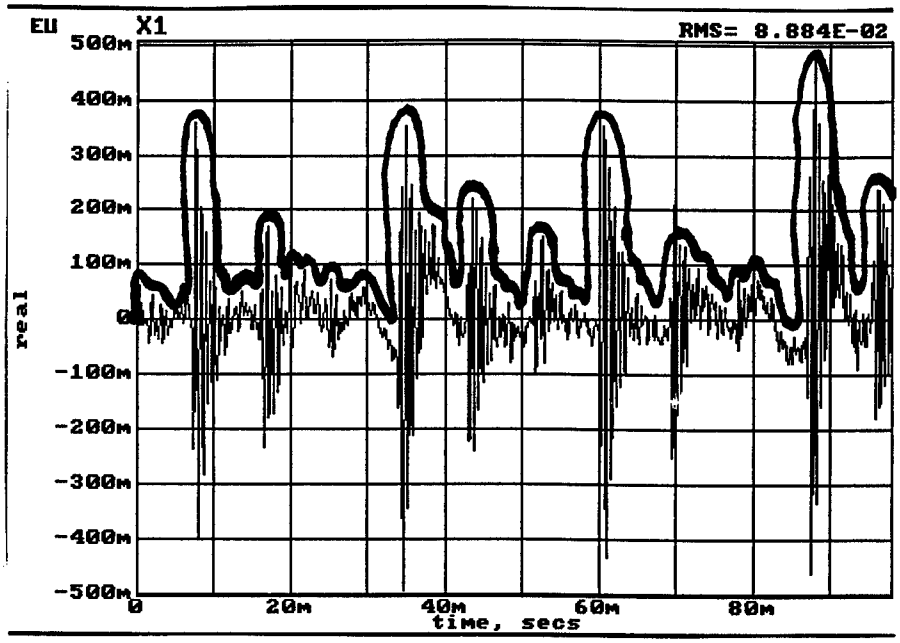


Figure 3 Time trace of damaged bearing with envelope drawn.

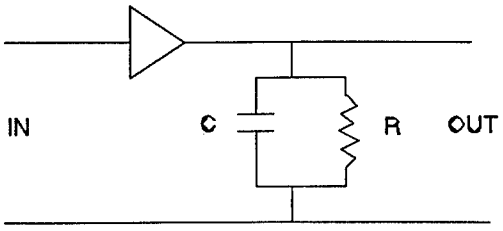


Figure 4 Simple Demodulator

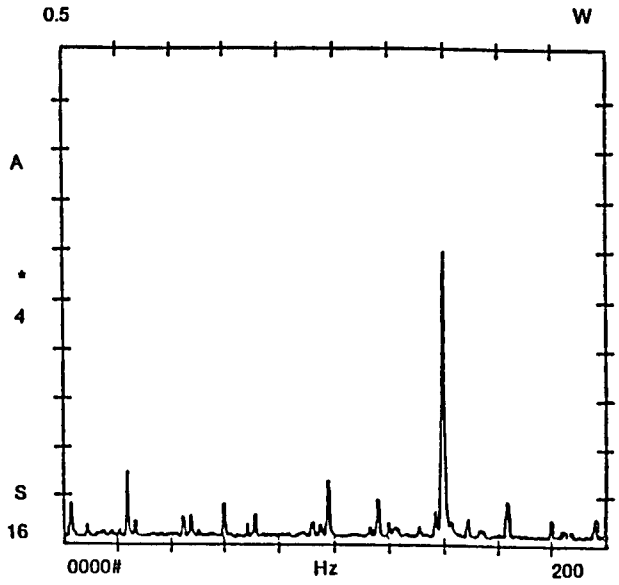


Figure 5 Spectrum of the enveloped data from the same defective bearing as shown in Figure 1.

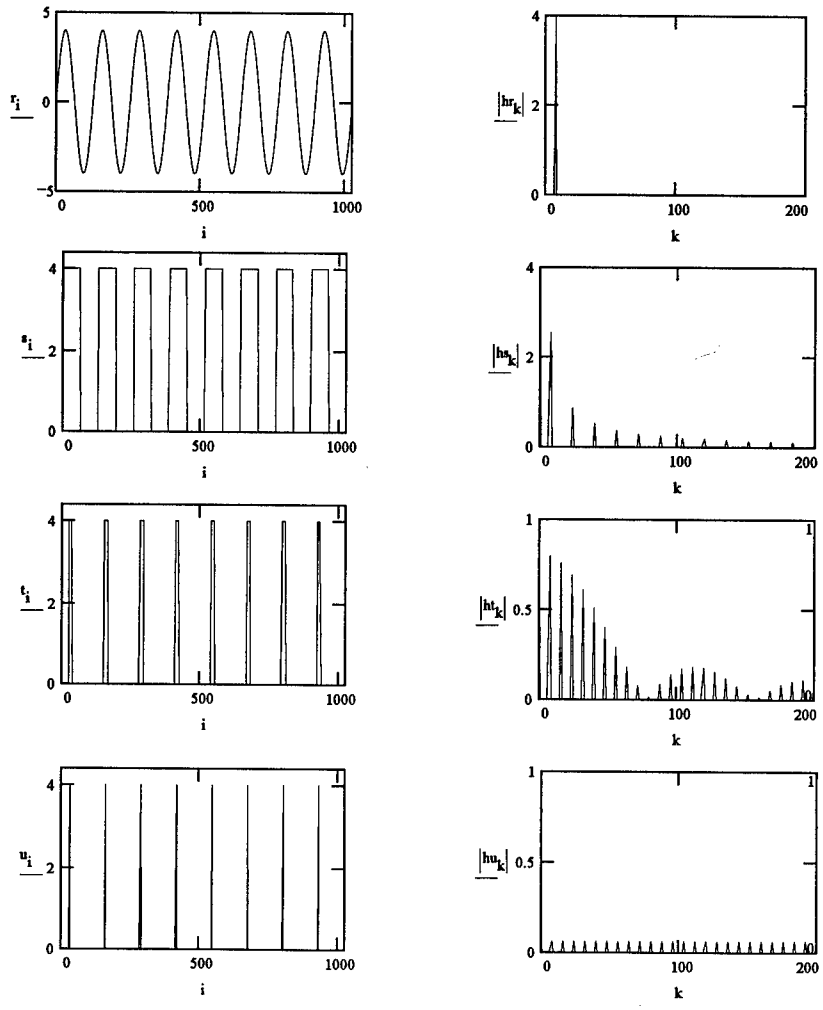
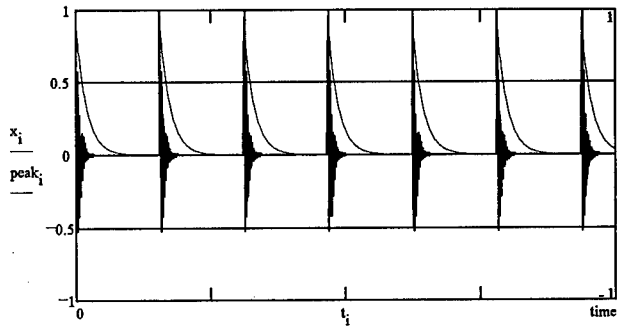
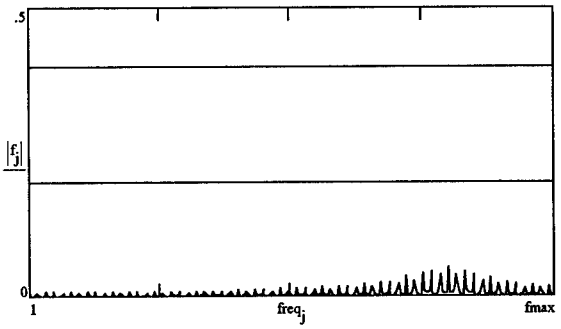


Figure 6 Spectrum energy distribution as a function of pulse shape



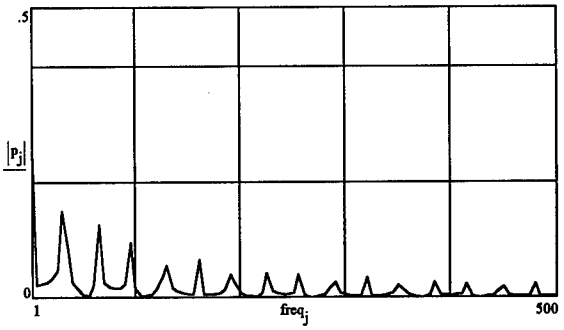
time = 0.2

hfd = 0.172



fmax = 2000

$|p_6|$ = 0.148



decay = 0.95 m = 4

Figure 7. Relationship between HFD (true & false), resonance damping and demodulator time constant

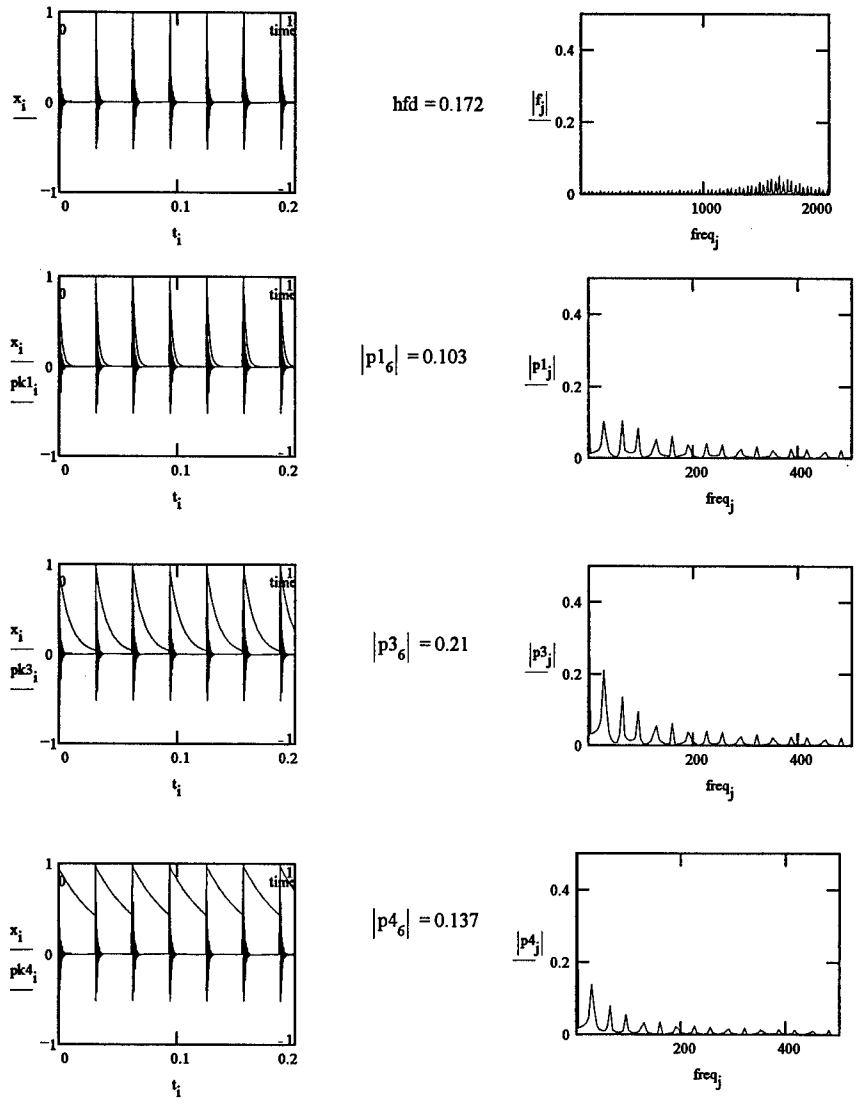


Figure 8. Relationship between the amplitude of the defect spectrum peak and the demodulator time constant

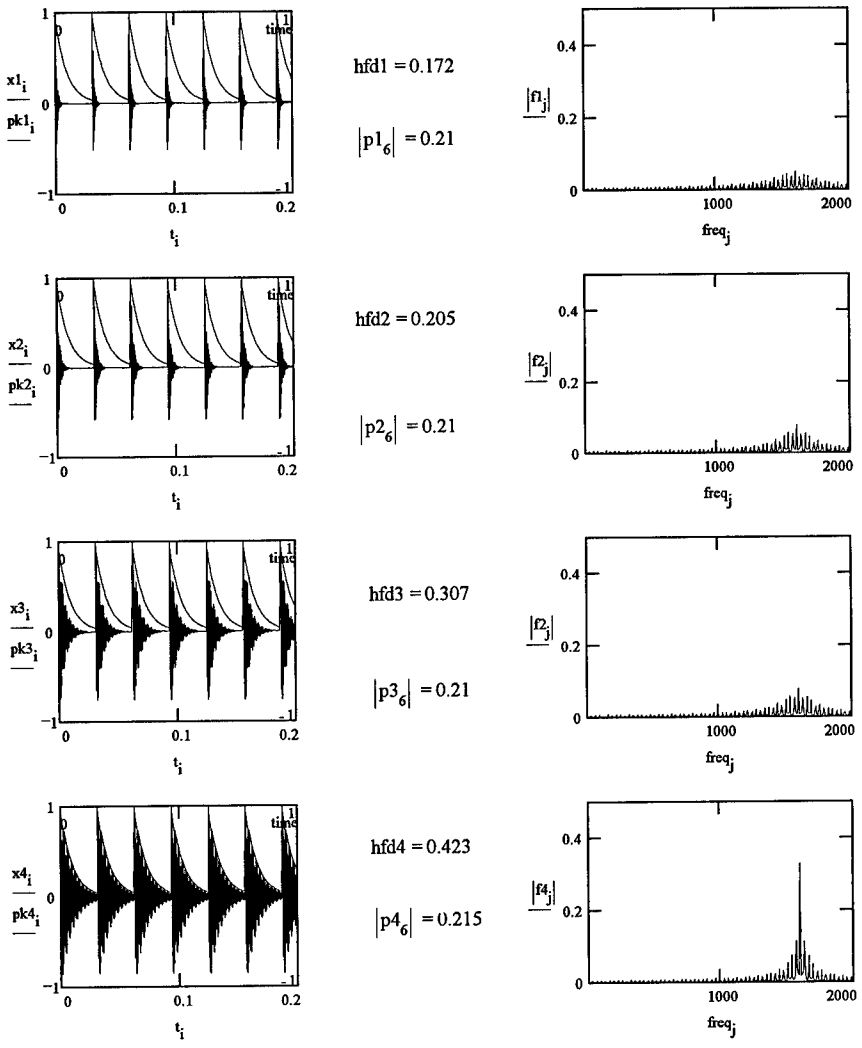


Figure 9. Relationship between the hfd(spec) and the resonance damping

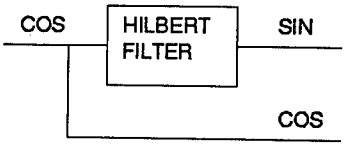


Figure 10 Hilbert filter and analytic signal.

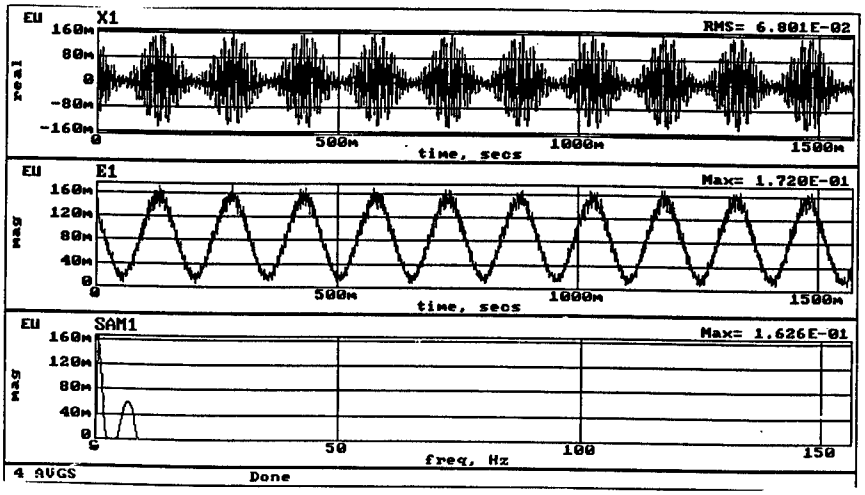


Figure 11 Demodulation by Hilbert transform of a modulated sine wave.

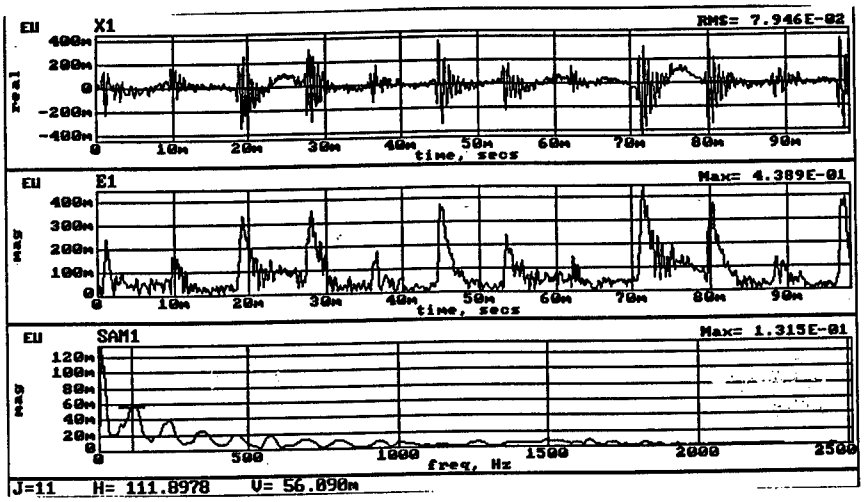


Figure 12 Envelope and spectrum of the envelope for a defective bearing using the Hilbert envelope detector.

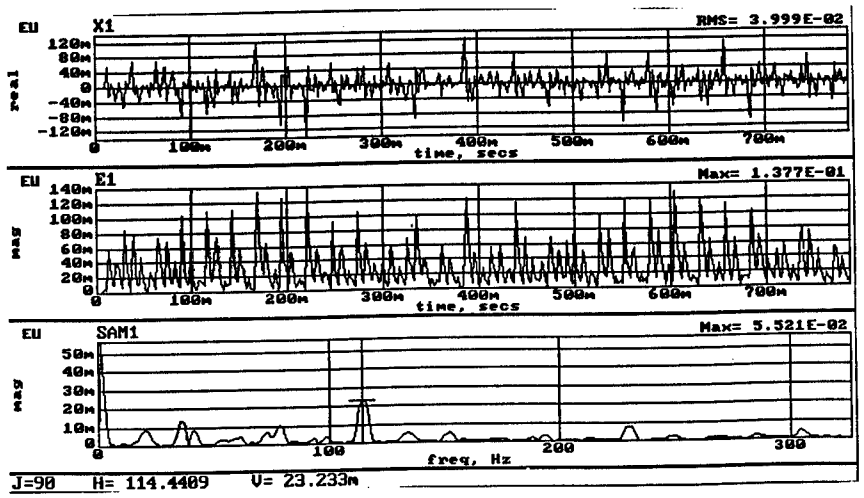


Figure 13 Spectrum formed by zooming on the resonance, then producing the spectrum using the Hilbert envelope detector.

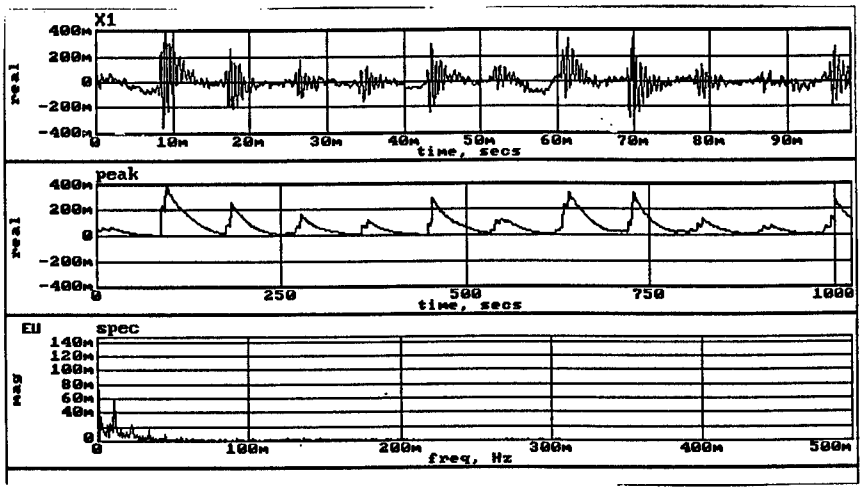


Figure 14 Peak detector used as an envelope detector.

GENERAL MONITORING TECHNOLOGY

COMPARISON OF VIBRATION AND OIL ANALYSIS

Thomas H. Crum, P.E.

C J Analytical Engineering, Inc.
R. R. #1, Box 353
Francisco, Indiana 47649

Abstract: This paper demonstrates that the use of just one Predictive Maintenance technology is handicapping the analyst. The merging of technologies can produce earlier and better fault detection. This paper covers Hard Particle Oil Analysis and Vibration Analysis on a system that uses piston type hydraulic pumps.

Key Words: Hard particle oil analysis; vibration analysis

For many years, Vibration Analysis has been the cornerstone, if not the sole component, of the majority of successful Predictive Monitoring Programs. Vibration Analysis is still the foundation, but other technologies have evolved to augment vibration data in equipment condition assessment.

The merging of Predictive Maintenance technologies allows for earlier and better machinery condition determination than the use of a single method. There are many technologies; for example : Vibration Analysis, Hard Particle Oil Analysis, Infrared Spectrography Oil Analysis, Ferrography, Thermography, Ultrasonic Analysis and Passive Electric Current Analysis that could be used together. Each technology has its nitch in detecting certain machinery faults, but when used together as a team, few problems can go undetected.

Several hydraulic systems are used on a metal rolling mill. One system maintains the proper gap of the rolls and is referred to as the SGC system on this particular machine. The SGC hydraulic system has six (6) servo valves, six (6) actuators and operates at 3650 PSI. There are three (3) piston type hydraulic pumps, of which any two (2) are required to run for proper system operation.

Table I

Number of particles per ml		R
More than	Up to & including	Range Numbers
80,000	160,000	24
40,000	80,000	23
20,000	40,000	22
10,000	20,000	21
5,000	10,000	20
2,500	5,000	19
1,300	2,500	18
640	1,300	17
320	640	16
160	320	15
80	160	14
40	80	13
20	40	12
10	20	11
5	10	10
2.5	5	9
1.3	2.5	8
0.64	1.3	7
0.32	0.64	6
0.16	0.32	5
0.08	0.16	4
0.04	0.08	3
0.02	0.04	2
0.01	0.02	1

This rolling mill performs the first four passes of the metal and then the metal goes to trimming for shipment or on to a pair of final mills for even further reduction. If the proper gap is not maintained, quality suffers on further reduction and can cause entire coils to be scraped. The good operation of the SGC system is important to this plant. In the past, the SGC system had experienced failures of the servo valves and actuators due to contamination related problems. The servo valves would be erratic or stick open. The actuators would be slow to react to changes or fail to maintain appropriate pressure.

The plant was using vibration monitoring to assess the condition of the hydraulic pumps. This technology was not suited to offer any insight into the contamination problems with the hydraulic system. The plant decided to employ the Predictive Maintenance technique of hard particle counts of the hydraulic oil on a monthly basis. This technique would tell the plant personnel that there was contamination present and quantify the level. Alarm levels (Target Cleanliness) were set to initiate action to address the contamination. Hard particle counts would identify if there was a problem and by dividing the system into subsets; such as, pumps, tank, actuators, could identify the source of the contamination.

Target Cleanliness is the measure of maximum number of contaminates allowed in the oil system being measured. This target should allow for continued good operation of the oil system. The particles are counted as greater than 5 micron and as greater than 15 micron. A micron is 0.000001 meter. A human hair is approximately 80 microns thick. The unaided eye can not see below 50 microns. The particle counts, number of contaminates present, are divided into numeric levels, Range Numbers, called ISO (International Standards Organization) codes.

Each increase in the ISO Range Numbers doubles the upper limit of particles. The ISO codes shown in Table 1 are numeric ranges. For example, an ISO code of 21 would mean the number of particles present in a milliliter (ml) of liquid would be greater than 10,000 but not more than 20,000. The Target Cleanliness is expressed by two ranges, R 5 micron / R 15 micron. The lower the R number means the cleaner the oil system is. R 5 micron is the number of particles greater than 5 micron. R 15 micron is the number of particles greater than 15 micron.

To establish a Target Cleanliness for the SGC system, several factors were taken into account. Some of the factors were; the mill is a high initial expenditure, downtime is expensive, quality is very important, safety is an issue due to fire hazard, high pressure oil, multiple servo valves with tight clearances, air borne particles minimal and possible water contamination due to coolers. An ISO code of 14/11 was chosen for the SGC system Target Cleanliness. This means that the maximum number of particles of greater than 5 micron would be 160 and the maximum number of particles of greater than 5 micron would be 20. The SGC system has an auxiliary makeup tank that is equipped with a 10 micron filter. The oil is circulated through the filter to clean the hydraulic oil before it is added to the SGC system.

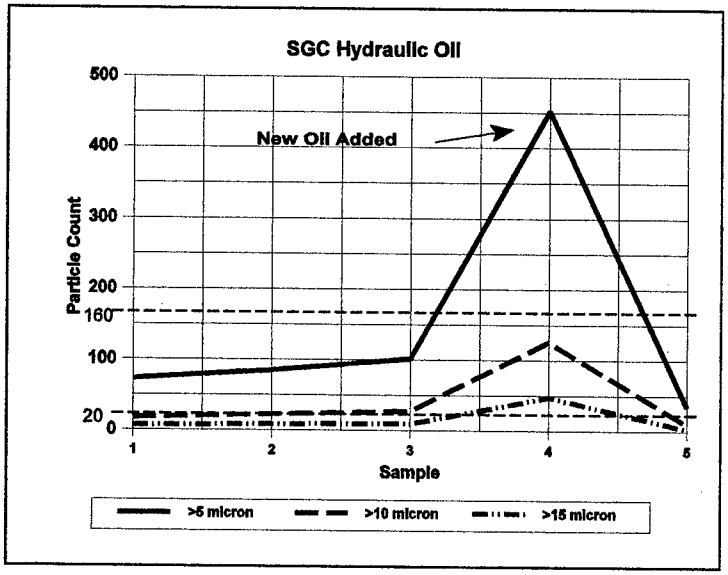


Figure 1

The first samples were within Target Cleanliness of 14/11. The fourth sample jumped into alarm with a 16/13. Figure 1 shows the trend of the hard particle counts. Samples were taken from the tank, from after the pumps, from after the filter and from the return line to try to determine where the contamination was occurring. Also, operating and maintenance records were checked for any activity involving the SGC system. The tank contents were the source of the contamination.

Before the fourth sample was taken, the system was losing hydraulic oil. On the night shift, a mechanic from another portion of the plant was assigned to top off the SGC tank. The mechanic added hydraulic oil from a new barrel to the system instead of using the oil that had been cleaned in the auxiliary tank. The new oil had an ISO code of 21/18 which is typical for hydraulic oil fresh from the barrel. Adding the new oil to the SGC system contaminated it. Administrative procedures were put in place to eliminate this source of contamination in the future.

The fifth sample showed the SGC system's 6 micron filter on the tank removed the contamination and the system returned to below a 14/11.

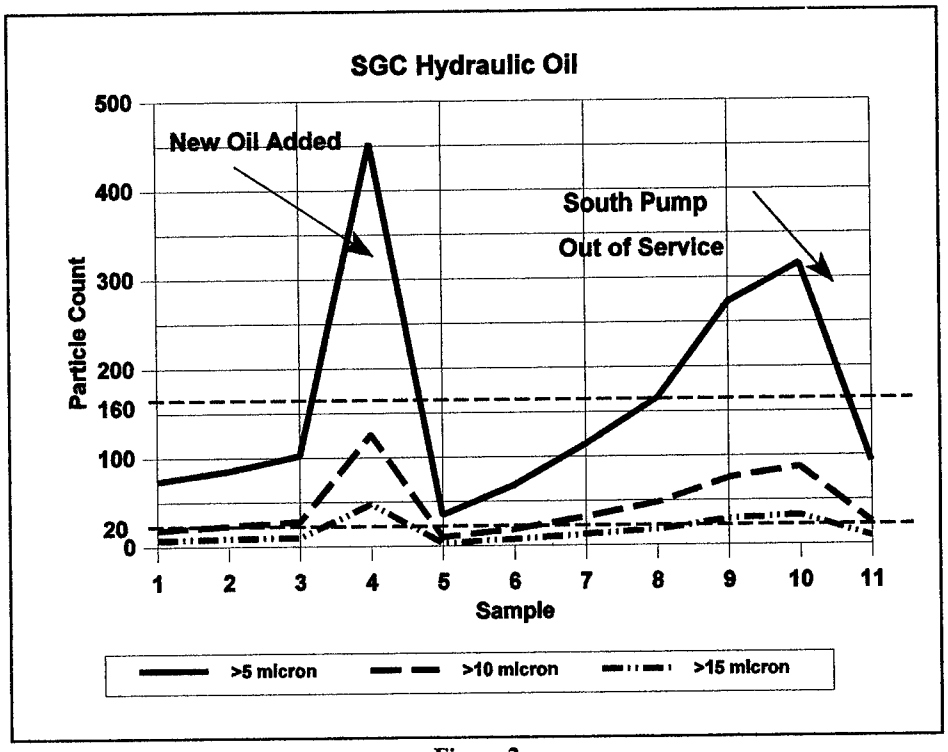


Figure 2

The particle counts increased over the next several samples and went into alarm (figure 2)

on sample 8. The particle counts continued to rise to an ISO code of 15/12. Again, samples were taken from the tank, from after the pumps, from after the filter and from the return line to try to determine where the contamination was occurring. The samples indicated the South Pump was producing particles. The vibration data was checked (figure 3) and revealed no apparent problems. The overall and bandwidth amplitudes had not increased significantly over the past few months.

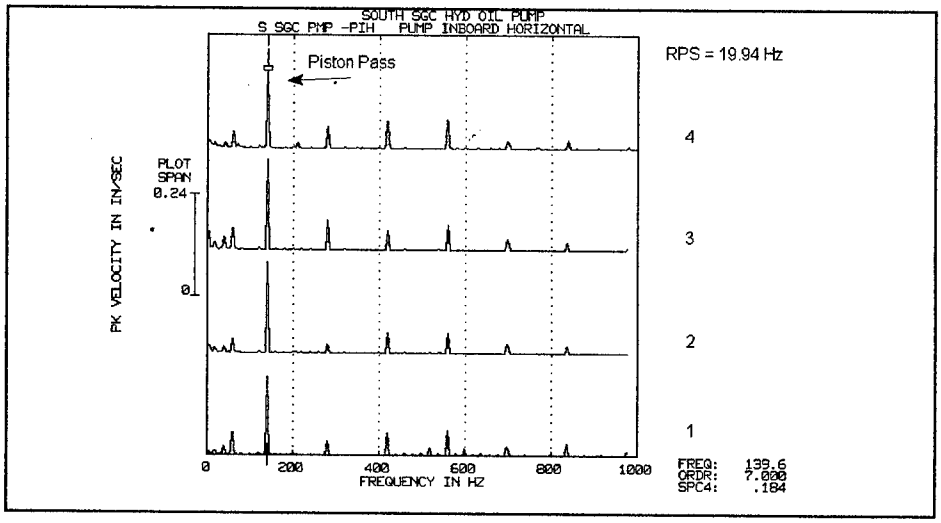


Figure 3

The Center Pump was started and the South Pump was removed from service. The particle counts at sample 11 dropped back to below an ISO code of 14/11 (figure 2). The South Pump was disassembled and scoring of the cylinder walls was found. Parts were ordered to rebuild the hydraulic pump.

The SGC system now had the Center and North Pumps in service. As stated above for reliable and safe operation two pumps were required to be in service. The particle counts started to climb again. By sample 12, the particle counts were in alarm at an ISO code of 15/12. Samples were taken from the tank, from after the pumps, from after the filter and from the return line to try to determine where the contamination was occurring. The samples indicated the North Pump was producing particles. The rolling mill was out of service for maintenance, so the SGC system was run with just the Center Pump in service. After one day the particle counts, sample 15, dropped down, though were still just in alarm (figure 4).

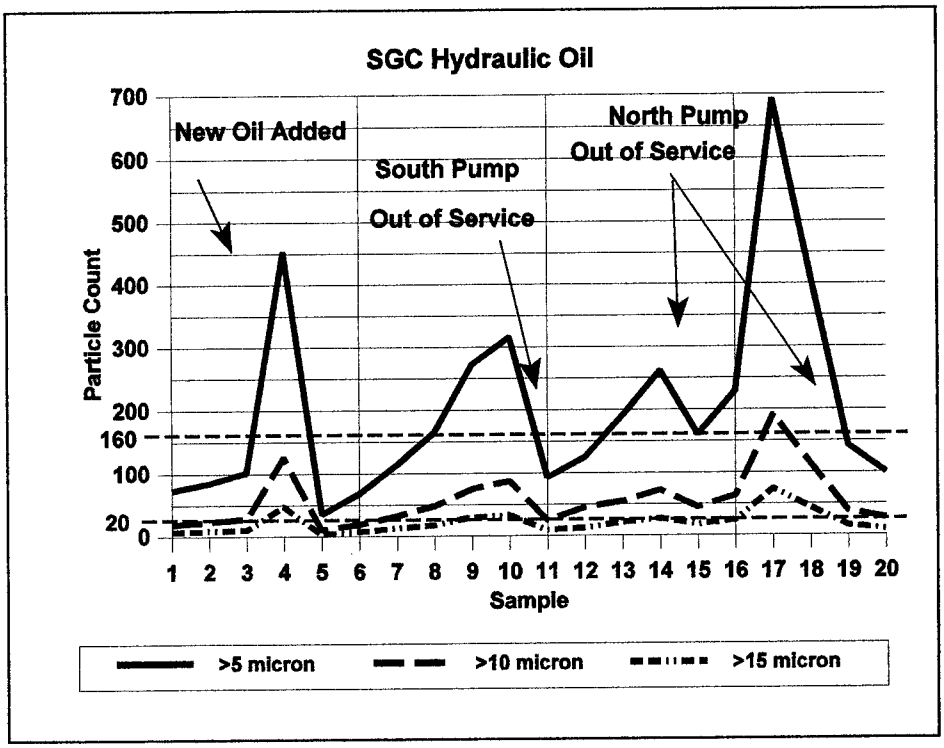


Figure 4

Since the parts had not arrived to repair the South Pump, the North Pump was returned to service. The particle counts continued to rise while the South Pump was being reassembled. After six weeks, the particle counts peaked at an ISO code of 17/14 (figure 4). Two servo valves had to be changed during this time due to contamination binding the valve stem and rendering the valve inoperable.

The vibration data began to show indications of a problem (figure 5). The vibration at piston pass frequency, 7X, increased significantly. The peak vibration increased from 0.11 in/sec to 0.31 in/sec. Also, 1X sidebands of the pump speed appeared around the piston pass center frequency.

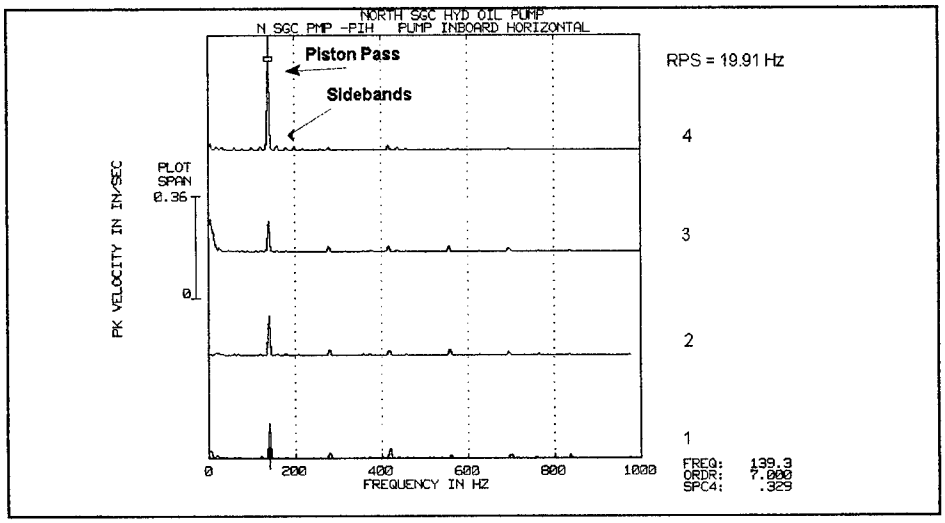


Figure 5

Vibration monitoring and analysis did show that the SGC system hydraulic pumps had a problem developing. The spectra and timewave data began to indicate there was a piston related event. The Hard Particle Counts of the hydraulic oil gave warning of the piston problem approximately six (6) months before the vibration data. The North Pump had scoring of the walls, too.

As this shows, having more than one analysis tool for the analyst to use will help identify problems sooner. Multiple Predictive Maintenance techniques work well together in an organized, well thought out program. One technique is not the answer for every problem.

Integrated Machinery Management

James W. Taylor
Machinery Management Services
1018 Cherry Lane
W. Lafayette, In 47906

Abstract: This paper proposes a holistic, integrated approach to machinery management. Such a program is comprised of four major elements. There must be a systematic planned maintenance system for routine maintenance and condition assessment. A system must be in place to insure needed spares are available without carrying excessive inventory. Most importantly, a measurement system must collect necessary data and present it to management in such a way that they understand their problems and what action to take. Once good measurements are available, there must be a continuous improvement system to make the maintenance function more effective, the equipment more reliable and efficient, the work performed more professional and less expensive, while minimizing spares inventory carrying costs.

Key Words: Continuous improvement; integrated maintenance; machinery management; predictive maintenance; preventive maintenance; reliability engineering; spares management

INTRODUCTION: Manufacturing today is becoming increasingly competitive. Production processes have been squeezed for the last pound per hour, for one more component per shift. The production equipment itself is one of the few remaining areas in plant operation where significant gains can be made. To maintain your manufacturing leadership, you must move from a defensive strategy of maintaining the status quo to a proactive, aggressive plan to improve plant reliability and capacity. You must manage the production assets in a systematic and professional way.

Much attention has been given lately to the use of modern condition monitoring techniques such as vibration analysis and oil analysis. Many claims, most of them true to some extent, have been made for their ability to spot failures before they happen. Though these methods are powerful, it's important to remember that they must be used as part of an overall system of maintenance.

Without a complete program, only minimal, short term benefit will be gained from condition monitoring.

The most progressive, competitive manufacturing organizations see management of production assets as a potential strategic advantage. They are developing integrated machinery management programs that pull all the separate technologies and management techniques together in a way that allows them to understand where the problems are and how important they are. Knowing that, they can make plans to solve those problems permanently. Until you know what the problems are, any action taken will at best result in short term solutions. In the worst case it may make the problem worse.

I want to remind you of some of the other types of information that we should consider. Each piece is important on its own. Each area has its own techniques and tools, both hardware and software. The real advantages come from the synergy that results from combining this information.

Machinery information comes from many sources. A machinery information management system (MIMS) often consists of several parts. For purposes of this discussion, the major ones are:

- Planned maintenance system (PMS)
- Spares management system (SMS)
- Production system (PS)
- History collection system (HCS)
- Continuous improvement system (CIS)

There must be a systematic Planned Maintenance System for routine maintenance and condition assessment. A Spares Management System insures needed spares are available without carrying excessive inventory. Accurate history on maintenance expenditures and experience is essential to identify the most troublesome problems. But most importantly, to achieve the potential production improvements and reductions in cost per unit, all the information available to plant management, including history, must be *used*. This is the Measurement System. Once good measurements are available, the Continuous Improvement System guides you in making the planned maintenance function more effective, the equipment more reliable and efficient, the work performed more professional and less expensive, and in minimizing spares inventory carrying costs.

The difficult part of maintenance management is to balance the maintenance expenses and needs, while minimizing overall costs. An objective of self-assessment of your maintenance processes may be to identify opportunities to improve equipment performance or reduce maintenance costs. However, these improvements cannot be accomplished without some investment in manpower or improved technology.

Done correctly, costs may go up but sales revenue and gross profit will go up faster. The payoff from putting in place an integrated machinery management program is longer Mean Time Between Failure (MTBF), shorter Mean Time To Repair (MTTR) and shorter Mean Logistics Delay Time (MLDT). Those factors mean higher reliability and availability. This pays off in higher product throughput and lower operating expenses.

The MIMS is a system in the true sense. As can be seen in Fig. 1., it has inputs and outputs. And it has feedback. They are:

Inputs:

- Maintenance actions on the machine
- Spare parts actions
- Engineering / design change actions
- Machine operation

Outputs:

- Routine and adhoc reports
- Program improvement objectives
- Machine improvement objectives
- Personnel improvement objectives

Feedback:

Continuous improvement system

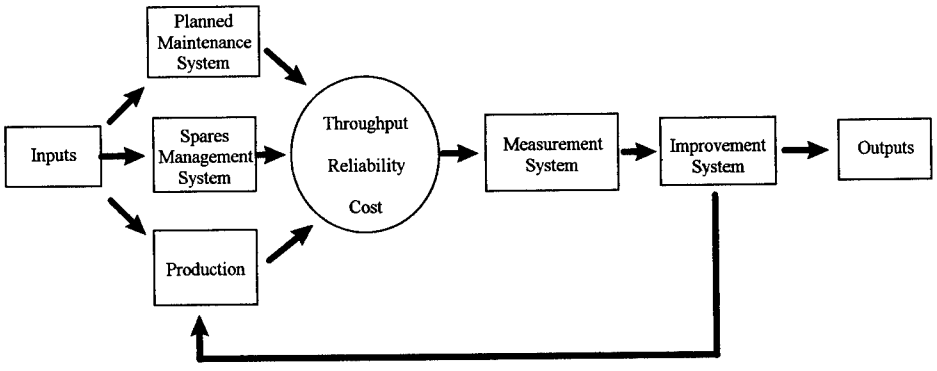


Fig. 1. Machinery Information Management System

PLANNED MAINTENANCE SYSTEM: All work is planned! If it is not pre-planned, then it is planned during execution. Pre-planning insures needed parts, materials and skills are available. Multiple trips to the tool room or store room are eliminated. Crafts are coordinated, avoiding wasted manpower caused by people standing around waiting. Work planned during execution suffers from false starts, missing parts or information and wasted manpower. Work that is not pre-planned can cost you as much as 25 times more to accomplish.

A good planned maintenance system is designed with an optimal mix of preventive (PM) and predictive (PdM) maintenance tasks. As much corrective maintenance as possible is planned to make best use of manpower and spares. The number of unplanned repairs is minimized. A good planned maintenance system will reduce the number of emergency repairs to a minimum because many of the failures that would be emergencies are found early while doing PM or PdM tasks. Because they are found before the failure occurs, they can be repaired with the least impact on production.

Preventive (Time Directed) Maintenance: FACT: Equipment that is operated can fail, and not all failures can be prevented. Preventive maintenance forms the backbone of a planned maintenance system. PM tasks are either fault finding (inspection) or preventive replacement. The routine of time directed tasks fosters discipline and focus in the maintenance organization. A well designed program has a standard daily and weekly routine that encourages systematic planning and performance of maintenance tasks.

Predictive (Condition Directed) Maintenance: A good planned maintenance system has a heavy emphasis on condition directed tasks, often called predictive maintenance. Periodic monitoring of equipment condition will optimize use of manpower, consumables and spares because work on the machine is only done when it's needed. Both installed production instrumentation and special technologies are used to monitor and trend equipment condition. Some of the most common technologies are vibration analysis, oil analysis, motor analysis and infrared scanning.

Planned Corrective maintenance: Corrective maintenance must be accomplished in the most cost effective manner possible. Because equipment problems are detected sooner by PM and PdM tasks, corrective action can be planned. Planned repairs result in much lower maintenance costs. Labor costs are lower because the crafts people start the job with all the parts and tools they need. Overtime can be reduced or eliminated. Spares costs will be lower because they don't have to be ordered on a crises basis, with the resulting high freight fees and expediting costs. And lost production cost will be lower because the repair is done during a scheduled shutdown. Even if a special shutdown has to be scheduled for the repair, it will be shorter because all parts, tools and people can be in place before production is stopped. The work will proceed more efficiently because the crafts know what they will be doing. Lost product or off quality product is minimized because production is shut down in an orderly manner rather than by a casualty. And because the line is shut down in an orderly manner, secondary, cascading failures are avoided.

Even emergency work can be planned. Some emergencies can be anticipated. For these, a preplanned work order can be on file ready to go. For repairs that are not anticipated, the savings resulting from a conscious, formal planning step can still be significant. Planning will minimize trips to the tool and store room. It can ensure all needed parts are ordered the first time, and that the correct crafts and tools are on hand.

Most CMMS systems collect general work and performance data. The various condition monitoring technologies have their own, often proprietary, supporting software. Each technology has pieces of information that are unique to it. Each of these pieces of information are important and must be captured.

SPARES MANAGEMENT (SM) SYSTEM: Spare parts are kept on hand so repairs can be made in the quickest time possible. But do you know what spare parts you have on hand? Do you know what parts you should have on hand? A good SM program has a spares allowance list (SAL) for each piece of critical equipment. The SAL differs from a Bill of Materials because it lists all the parts that should be kept in the store room to repair that machine. The lists for the individual machines are combined into a consolidated list for the whole facility. Consolidation accounts for duplicate usage of some of the parts, resulting in lower total parts inventory.

SM also provides for systematic storage, inventory control and issue of spares. As a part is issued, it is charged to the proper job order and machine. This assigns the cost of the part to the cost of maintaining the machine and is information needed to assess the usage of the part. Knowledge of usage lets you update the allowance lists as the equipment ages and conditions change. A periodic audit of spares will identify parts no longer needed or additions to the allowance because of increased usage. You will be able to adjust order points and minimum levels to insure parts are available when needed.

The Spares Management System generates several kinds of useful information. Parts should be designated as to usage and inventory management method. There can be non-stock, usage stock, insurance stock and project stock to keep track of, each with its own management method. Storeroom management needs information on location, layout, stock catalogs, and issue/receipt. There are also other spares management actions that generate information. Spares are ordered, unused parts are turned in, kits are built for specific jobs, bills of material and allowance lists for each machine must be maintained. And the whole system must be updated as a result of engineering change orders that affect the machinery.

PRODUCTION: Production generates a number of pieces of information that can be useful to maintenance and engineering. In addition to operating temperatures, pressures, speeds and other data, the planned and actual production times and run times are important. Records of change-over, bottle-necks, product produced and operating problems provide needed information to assess program effectiveness and machine improvement.

MEASUREMENT SYSTEM: One of the difficulties in the maintenance field is objectively judging the actions of those involved. The large machines in major industrial installations are inherently reliable and their natural rate of deterioration is very low. A manager can spend very little money over a period of several years without experiencing a marked decrease in availability. He often appears to do a better job than a manager who applies a more comprehensive approach. In the latter case, corresponding expenses are visible, and have an immediate negative impact on the budget. Because machinery information is so hard to gather and present consistently, senior management does not recognize the resulting aging and deterioration until too late.

A measurement system is like the control panel of an aircraft. It tells you where you are and what the conditions you are operating in are like. Your measurements should accurately represent the current conditions. Comparing measurements from period to period, looking for trends, will tell you where you are headed. By collecting the right data and displaying it properly, the beneficial impact of maintenance can be quantified in a way that is understood and believed by upper management.

The basic requirement for good measurements is good records. A common complaint of crafts people is that they are paid to work on the machinery, not to do paperwork. At first glance this seems to be true. But the paper work involved in history collection is a vital part of understanding the equipment. The cost of the time spent filling out good history reports will be more than offset by the long term savings in improved maintenance and reliability. Good records let you know what the most expensive machines are, what the most common failure modes are, what parts to keep on hand, what machines need more or less PM or PdM. Accurate detailed history gives you the information you need to do root cause analyses of failures and fix the real problem so it doesn't reoccur. Another advantage of detailed descriptions is that they can act as a planning tool if the casualty occurs again. Good records also provide you the data needed to develop and justify your budgets. It gives you a management tool to target budget cuts to the areas where they will be most effective.

The history collection system must gather at least three types of information. Statistical information must be captured for overall analysis of the maintenance function. This includes such data as man-hours expended, parts used, consumables used. It must also capture production downtime and maintenance downtime. The information contained in detailed narrative descriptions of problem and actions taken is needed for root cause analysis. It should include how the problem was discovered, what the symptoms were, trouble shooting steps taken, repair procedures used and any special problems. Accounting information - costs, badge numbers, wage rates, special shipping costs, etc. - is needed by plant management to develop and administer budgets. It also is captured as part of the life cycle cost of the machine.

Once the history is captured, it must be used. Serious analysis of maintenance should:

- take into consideration indirect maintenance costs (costs of reduced or lost production).
- Judge the evolution of economic indicators over a prolonged period.
- Evaluate the evolution of the equipment condition over the same period.

The raw data must be manipulated to find the information that's in it. MTBF, MTTR and MLDT are calculated. Costs are summarized. Effects of varying production utilization are accounted for. The resulting measurements should be a true picture of the current situation.

CONTINUOUS IMPROVEMENT PLAN: The Continuous Improvement System is where it all comes together. In this system, the history collected is used to identify areas where maintenance can be made easier and cheaper. Unneeded maintenance or additional needed maintenance is identified. Additional training requirements, changes in logistics requirements and potential equipment redesign or modification are other results of analysis of the history. The review should not be limited to large, costly or politically visible failures. Clusters of failures should be sought out. Often several minor failures can cost as much or more than one large failure. Looking for clusters of failures and eliminating their cause will pay for itself quickly.

A Maintenance Reduction System consists of several separate but related parts. They are:

- Planned Maintenance Optimization
- Precision Maintenance
- Maintainability Improvement
- Equipment Improvement
- Logistics Improvement
- New Equipment Selection

Planned Maintenance Optimization: Plant design characteristics can create a performance limit for the plant, no matter how effective the maintenance is. However, it is also true that improper, inadequate, or too much maintenance can adversely affect performance. Over-maintenance increases direct and indirect maintenance costs by increasing voluntary production losses, speeding aging due to excess dismantling and re-assembly, and increasing the risks of damage through human error. Given these limitations, how do you establish the optimal maintenance program.?

Planned Maintenance Optimization is a formal program to review the effectiveness of the Planned Maintenance System. Using data from the measurement system, reliability information can be derived. Review of this information will identify those machines with too much or too little PM or PdM. It will assess the effectiveness of the planned tasks and recommend added tasks if needed.

Finally, this review will tell management where they should apply more expensive techniques such as Reliability Centered Maintenance, condition monitoring, machine modification and others. It can identify needed training and tools.

Precision Maintenance: Precision maintenance denotes a high level of skill and training of the crafts. Such things as use of standard procedures for bearing replacement, correct preparation of foundations before setting equipment to prevent soft foot, and using vibration to verify correct installation are examples. It also encompasses using tighter tolerances than is normal. When balancing, take the little extra time to make an extra run - it will pay off in increased bearing life. Precision alignment to increase seal life, strict cleanliness to avoid later problems because of dirt, and extra care in seating surface preparation are other examples. Analysis of the history will show where the training dollars should be spent.

Maintainability Improvement: Maintenance tasks should be reviewed to determine those that take a long time to do or that require large manpower expenditure. These tasks should be investigated to see if there are changes that can be made to make the task more economical. These might be procedures changes, special tools or jigs, or modifications to the machine. All of these improve the maintainability of the machine. Better maintainability means the machine is cheaper and safer to maintain and less problems are introduced while doing maintenance.

An example might be a task requiring periodic checking of gear tooth wear on a gear set. Because of equipment design, the job takes three men (two mechanics and an electrician) 4 hours, a total of 12 man-hours. They have to remove an interfering motor and rig a cover plate off the casing. By installing an access plate in the gear casing, the job can be done by one man in 30 minutes.

Equipment Improvement: A history of frequent failures of a machine might lead to a design change to eliminate the problem. For example, review of the history reveals that maintenance is called several times a week to unjam product on the conveyer feeding the wrapper. The product is getting cocked, and following product shoves it into the rollers. Engineering redesigns the guides to prevent the cocking, eliminating the problem. The increased production and reduction in off quality product more than pay for the modification.

Logistics Improvement: The spares usage should frequently be reviewed to assess the adequacy of the SAL's. Parts may no longer be needed because of changes to the equipment, stocking levels may need to be adjusted because of usage, additional parts may need to be added to the SAL. History can help identify parts that should be vendor stocked, vendors who are not responsive either in timeliness or quality of parts, and parts that have a high failure rate and should be substituted. Incidences of not having the proper tools, drawing or manuals will also be identified by history review. Correction of these problems will reduce the delay before the repair can start (MLDT), which means less lost production and more sales dollars.

New Equipment Selection: The best plants will monitor the life cycle costs of their equipment. As your equipment ages, there may come a time when it is more effective to replace it with new than it is to continue to repair it. A recent study by Eastman Chemical Company showed that maintenance costs can be as high as 50% to 75% of total life cycle costs. A good machinery history will give you the information needed to make an overhaul / buy decision.

SUMMARY: As we work towards an open system for machinery information management, we need to think of all the aspects of machinery information. Because a majority of our members are from the condition monitoring field or from large user organizations, there is a danger that we may become too narrowly focused.

While big companies have enough people and resources to put together large projects to integrate the various sources of machinery information, most companies don't. There are 350,000 manufacturing plants in the USA, but only 35,000 have more than 100 employees. It is those 315,000 who most need our help.

One screen access to all required information will be a major step towards giving management the tools needed to intelligently set policy and improve their operation. By identifying problem areas and suggesting solutions the MIMS can start the process of continuous improvement. By collecting all pertinent information consistently, it provides a measurement system that tells management if they are on course or off course.

EXPERT SYSTEMS, A DECADE OF USE FOR USED-OIL DATA INTERPRETATION

Larry A. Toms

Larry A Toms Technical Services
5018 High Pointe Drive
Pensacola, FL 32505-1825,
TEL: (904) 494-1310

ABSTRACT

Pressed by the need to improve oil analysis performance, some equipment operators have increased sampling frequency (shortened intervals) in order to increase the probability of early fault detection. As a consequence, laboratory labor costs increased considerably--quadrupled in some cases. Over the past 10 years, expert systems have been increasingly used to compensate for the increased processing time by automatically interpreting sample data in near real time, improving evaluation reliability and minimizing the associated labor costs.

A properly designed expert system can quickly review all recorded equipment and sample data, while keeping the analysis time and costs within acceptable levels. These systems greatly increase data interpretation consistency, and can generate significant returns-on-investment. This paper presents an overview of several of these systems and the general principles utilized in their development.

KEY WORDS: Expert system; fault detection; fault library; knowledge based systems; oil condition monitoring; railway systems; statistical process control; used-oil analysis

INTRODUCTION

Used-oil analysis, pioneered by American railways and the Department of Defense, has led to significant reductions in unexpected equipment failure and has increased equipment reliability and safety of operation[1]. The traditional used-oil analysis process is based on frequent periodic samples, simple-to-use test methods and trained oil analysts to evaluate findings and advise maintenance personnel of a required action. The high sampling frequency required for reliable monitoring of diesel powered (< 250 hours) and gas turbine powered (< 25 hours) equipment requires substantial labor (and costs) for testing and data interpretation. The greater the sample frequency or the number of tests performed on each sample, the greater the costs involved. In addition, reliable interpretation of machinery and fluid condition requires assessment of many

other types of data including, oil related failure mechanisms, their symptoms, effects and costs; machine configuration changes and utilization; scheduled and completed maintenance; etc. Requiring a review of all possible data for every sample is very labor intensive and costly. As a consequence, the scope of an oil analysis program is often limited to fit the available resources. The usual result is longer sample intervals, less analysis reliability, fewer quantifiable benefits and less confidence in the program.

Pressed by the need to improve locomotive reliability after the recession of 1981/82, the Canadian Pacific Railway (CP) conducted an evaluation of locomotive engine failure modes, their effects and costs[2]. One of the outcomes of this evaluation was the conclusion that the used-oil analysis program did not always indicate the occurrence of an oil related engine failure mode prior to failure. In fact, agreement between laboratory recommendations and subsequent inspection or failure reports indicated the laboratory reliability to be about 65%. At the time, CP, as with most railways, obtained engine oil samples during Federal Government mandated inspections at either 46 or 92 day intervals. These intervals were very convenient but failed to respect failure mode duration or timing. Oil analysis at such long sample intervals identified some faults before failure, although sometimes, engine failure was the first sign of trouble. Engine reliability was not considered a serious problem as sophisticated preventive maintenance (PM) procedures kept failure rate down. However, with large fleets of 1000 plus locomotives, even a "low number" can be significant. The 40 or so engine failures per year experienced by CP amounted to several million dollars in loss and it was thought that something could be done to recover these expenditures. It was also thought that more effective monitoring would lead to longer PM intervals with a considerable reduction in maintenance costs.

The failure modes analysis conducted at CP during the mid 1980's indicated that oil contamination and metallic wear faults were the most prevalent problems and would provide the highest return on investment for an oil analysis program. The symptoms of these faults include: coolant contamination, fuel dilution, metallic wear, incorrect oil addition and bad sample recognition. While other oil related faults are possible, they were statistically non-prevalent in the CP locomotive fleet. Analysis of failure mode progression intervals suggested that the sample interval required a significant reduction if the important failure modes were to be monitored reliably. To maximize the probability of early fault detection, CP shortened the sample interval to about 200 to 250 hours of engine operation or about every 7 to 10 running days. This however, resulted in a three-to-four fold increase in the sample collection rate--and a commensurate increase in laboratory labor and consumables' costs. Offsetting this increase was very desirable and expert systems were investigated as a possible solution.

DEVELOPMENT CONSIDERATIONS

The real world of used-oil analysis presents many obstacles to the expert system developer. Consider the following problems which had to be overcome during the development of the CP expert system:

1. When the CP system development started in 1984, very little documentation or published work on "used oil" analysis was available. Consequently, the company had to document and validate all

of the factors and relationships involved in the used-oil analysis process. Since there were multiple "expert" opinions and very little science, this was a very frustrating process. Taking samples, applying some ASTM tests and reporting the data does not make a condition monitoring program. One must have a plan, tailored to some objective such as failure prevention, or maintenance costs reduction..

2. As with most traditional practitioners of used-oil analysis, CP compared individual oil measurements to a set of empirical limits and recommended an inspection in the event of a limit exceedence. Little was known about the statistical behavior of sample data, the relationships among data parameters, the relationships between data and fault mechanisms, or the impact of operational policies on data variability. Limits were arbitrarily determined by an "expert" by somewhat mysterious procedures. In summary, the data evaluation procedure depended on the intuitive response of a highly trained person. The procedure was not completely clear and could not provide a reliable expert system. The entire data interpretation process had to be re-thought and developed into a general purpose paradigm which could be encoded into the available expert system software.

3. Used-oil sample data is subject to frequent change from both reported and unreported events. When a technician encounters a fluctuating data pattern, a variety of intellectual processes can be utilized for interpretation and problem resolution. This is time consuming and very costly, but generally works. However, an expert system does not readily adapt to fluctuating data and should only be used where *data variability can be controlled, or the interpretative knowledge can be provided.* Fortunately, analysis of several years of test and maintenance history indicated that oil data variability can be brought under control by strict adherence to corporate standard operating, maintenance, sampling and testing procedures. Any remaining sample data variability is usually compensated by an adaptive trending algorithm--part of the expert system implementation.

4. In addition to evaluating sample data, reliable used-oil analysis recommendations require access to, and the evaluation of, a myriad of operational factors including equipment configuration change, maintenance and usage. However, a knowledge base that includes all of these factors, for *each different equipment* model, function and operational circumstance, will be enormous, difficult to develop and even more difficult to validate and maintain.

A PRACTICAL EXPERT SYSTEM

Consequently, CP decided on a general purpose, statistically based data analysis paradigm which would be compact, efficient and easy to maintain. The development and validation of this procedure required many months of effort, and the analysis of many thousands of used-oil samples. The new software integrated spectrographic, water contamination and oil viscosity analyses with an expert system for data interpretation. The expert system evaluated all relevant test and maintenance data and significantly improved the performance and consistency of interpretation while eliminating the labor involved. In early 1987, a statistical study of 20,000 samples taken from 1200 locomotives over a 3 month period verified the premise that simple analytical tests performed at a high frequency provided reliable indicators of engine and lubricant condition. The study placed the effectiveness of the CP expert system at 98.6%, with no engines

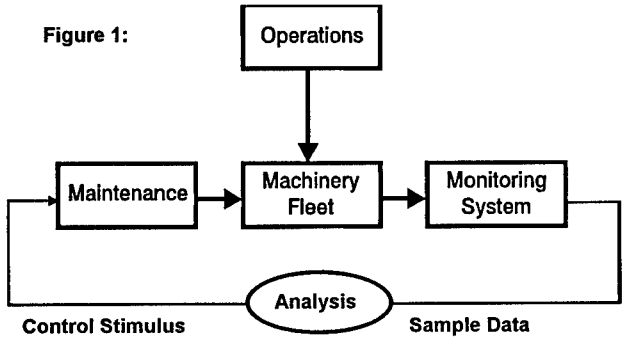
missed[2]. By comparison, the oil analysis data interpretation effectiveness before 1986 averaged less than 65%. By the beginning of 1988 oil-related engine failure occurrences at CP were nearly eliminated. In addition, the company moved to a reliability centered maintenance (RCM) program which significantly extended component life utilization. The move would not have been possible without the availability of up-to-date accurate equipment condition data. Today the shortened interval, consistent analysis and quick turnaround time permits problem engines to be inspected and repaired early, generating substantial savings in materials and labor.

The fact that the original CP expert system is still in operation, and without modification since 1989, attests to it's success. In fact, many other equipment operators, including Canadian National Railways, Chinese National Railway, CSX Transportation, Royal Canadian Navy[3] and Royal Navy have implemented oil analysis expert systems based on the general principles first used at CP Rail. While the performance statistics of the military applications are unknown, it should be noted that Canadian National and CSX Transportation also recorded marked improvements in locomotive failure rates and oil utilization after their oil analysis expert systems were commissioned. In each case, a reduction in engine failures of 40% or greater was reported at the end of the first year of operation. These benefits were a direct result of consistent, early problem indication, and improved maintenance scheduling.

GENERAL DATA INTERPRETATION PRINCIPLES

Before an equipment monitoring process can be automated by an expert system, it is necessary to completely understand the equipment system in terms of how it behaves; its inputs and outputs; it's performance, reliability and cost factors, etc. It is also necessary to reduce this information into a set of general principles which can be easily encoded into an expert system knowledge base[4,5]. If this is not done, the knowledge base grows at an exponential rate as individual machine related rules and relationships are added. Such a knowledge base is very complex and difficult to validate or maintain. Fortunately, statistical process control (SPC) procedures provide a convenient, compact and general purpose paradigm for an expert oil analysis system. In a SPC based paradigm, the machinery train or equipment fleet is viewed as a "closed loop" system as shown in Figure 1 below.

Figure 1:



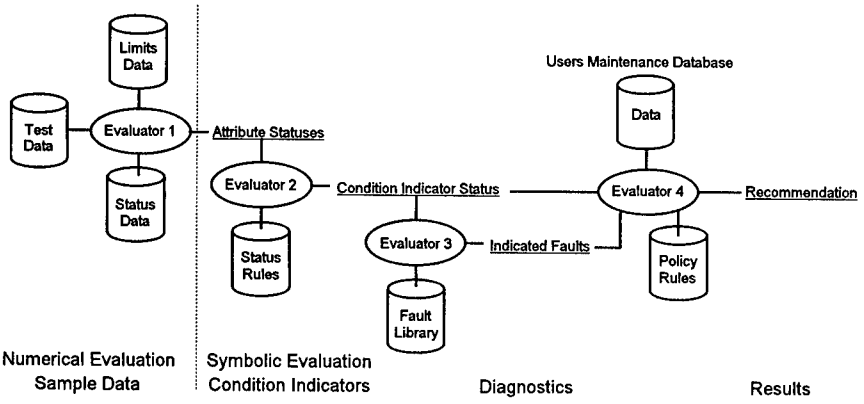
In a closed loop system, frequent oil samples provide condition indicating data (feedback); sample analysis determines the individual component condition and standard responses to direct machine rehabilitation (control). Abnormal sample data identifies failing components; once identified, they are removed, repaired or replaced and cease to become contributors of abnormal data; oil consumption, make-up addition and oil change function to restore sample data to normal levels. Thus, the system attains a *dynamic equilibrium* and under normal circumstances, a relatively *normal frequency distribution.*

When the machinery system is maintained by consistent procedures, sample data also tends to be consistent, changing only as a function of equipment usage, a fault occurrence or a maintenance action. Since condition-data variability due to usage and consistent maintenance is predicable, and usage and maintenance data is recorded, condition-data variability due to a developing fault is easily identified. In fact, data deviations caused by some maintenance procedures can be monitored by the occurrence of a particular pattern or trend and provide an additional source of useful information to the maintenance manager. In addition, historical condition-data from a normally operating equipment system can be utilized to calculate statistical alarm limits for each fault signature monitored. Thus, an SPC based condition monitoring system is simple, general purpose, machine independent, easy to automate with an expert system, and easy to maintain once deployed. However, condition-data will only exhibit a "normal" distribution if proper, consistent operating practices are implemented. Any improper practice such as incorrect sampling, improper maintenance, excessive oil change, permitting deep sumps to run low, etc. will profoundly affect system equilibrium and data distribution, with a commensurate effect on the reliability of sample data interpretation and alarm limit calculations.

STATE OF THE ART

Oil data interpretation can be performed by a simple set of data driven procedures using a divide and conquer procedure. The evaluation problem is divided into four main procedures, each with it's associated databases and rules as shown in Figure 2:

Figure 2



Evaluation procedure 1 prepares raw data and converts it into symbolic text or status values that carry the meaning imparted by the data. This process uses current sample, historical and statistical data to completely define a parameter's meaning in simple language terms. In this process, each input parameter is considered an object defined by a series of attributes. The symbolic value of each attribute is indicated by plain language, domain related terms. For example:

Input Parameter: *Percent Water*
Descriptive Attributes: Level (*Normal, Marginal, Reportable*)
Trend: (*Decreasing, Stable, Increasing, Etc.*)
Rank: (*Hi-Reader, Nominal, Low-Reader*)

For simplicity, all input data is defined by the same attributes and each attribute's value is established by statistically based limits. Attribute text phrases are chosen to impart real meaning in human terms.

Evaluation Procedure 2 combines parameter attributes to generate an overall *Condition Indicator* Status. The status of the condition indicator is related to the maintenance response required by corporate policy should the indicated abnormal status of the parameter be true. For Example:

Condition Indicator: *Percent Water*
Descriptive Attributes: *Normal, Alert, Urgent, Hazard, Danger*

Normal	No Action Required, Continue Routine Sampling
Alert	Shorten Sample Interval
Urgent	Maintenance Recommended, Deferral Permitted
Hazard	Maintenance Required, No Deferral Permitted
Danger	Shut Machine Down, Immediate Maintenance Required

Since a condition indicator can sometimes indicate an abnormal state for other reasons such as false positives, bad samples, multiple occurring fault signatures, improper maintenance, unreported maintenance or the lack of maintenance, other evaluation steps are necessary to ensure an accurate response to any indicated abnormal status, and eliminate any potential false positive.

Evaluation Procedure 3 compares all abnormal condition indicator statuses to a library of fault signatures to arrive at a diagnosis. A positive diagnosis increases the certainty that an abnormal data indication is justifiably abnormal and rates a maintenance response. The fault signature library contains signatures for all known faults, bad sample indications, false positive indications, inappropriate trend indications or any known symptom which could impact the accuracy or reliability of an intended maintenance response.

Evaluation Procedure 4 combines diagnosis and condition indicator status levels and generates an overall risk of failure indication. This module searches the users maintenance database, scheduled maintenance to-do lists for factors which would alter a maintenance recommendation based solely on condition data. For example, it would be more desirable to inspect or repair a fault at a PM

interval and knowledge of the next PM may prevent a request for shutdown and repair of a machine. Similarly, knowledge of recent component and lubricant maintenance would refine the response generated by a purely condition based recommendation. Once the maintenance data is known, the a set of corporate business rules determines the appropriate maintenance response which is printed on the output report.

An expert system based on an SPC based paradigm, such as the one shown in Figure 2 above, can evaluate sample data, rapidly and consistently, and return reliable recommendations for all samples where there is a high certainty of outcome. Favorably, this is over 98% of all samples. In the few cases where there is incomplete data or unusual external factors limiting the certainty of interpretation, the expert system can request re-tests, additional tests, additional samples, or assistance.

The expert systems used by Canadian Pacific, Canadian National and CSX Transportation are excellent examples of the level of reliability that can be achieved using a statistically based data interpretation paradigm. For example. the CN and CSX systems operate completely unattended, transmitting recommendations directly to the railways' maintenance shops. The CSX system processes over 700 samples per day and supports the integrity of a fleet of 3000 locomotives providing maximum reliability at a substantial saving in laboratory labor. This level of success can be easily achieved if the following development considerations are followed:

1. Perform a failure modes, effects and criticality analysis for each machine type to determine the failure modes which are most damaging to reliable equipment operations, which are economical to monitor, their respective fault mode symptoms (fault signatures) and the appropriate sampling interval.

2 . Select the required analytical tests to be performed on each sample from an evaluation of failure mode symptoms. Only tests which economically provide failure symptom data need to be considered. Use the failure modes analysis as the primary guidance in the selection of tests. Be wary of traditional testing conventions, many traditional tests were developed for new-oil performance testing, not used-oil condition monitoring. Data parameters that do not relate to fault indicators or bad sample/test indicators only consume resources with little probability of generating a return.

3. Develop a structured set of alarm responses in accordance with maintenance and operational policy. These responses dictate the maintenance measures to be taken when a particular alarm level is encountered.

4. Utilize SPC procedures to calculate an appropriate alarm limit corresponding to each alarm status response. Magnitude, trend and other statistical evaluations may be combined to achieve the desired evaluation matrix.

5. Develop and validate a structured set of condition indicator/failure mode relationships. These diagnostic indicators are encoded into the expert system knowledge base (fault library) to provide the basis for fault diagnosis or verification.

Lastly, integrate the evaluation steps into a logical paradigm such as indicated in Figure 2 above. Divide the data evaluation problem into simple discrete steps, in which simple plain language analysis rules can be used to solve each step. This design provides high run-time performance and is easy to modify and maintain over the long term.

Note: system maintenance and validation procedures are often performed by field grade personnel. It will be very helpful to have the expert system knowledge base and any mathematical formulae encoded in plain domain related language for easy reading and understanding.

CONCLUSION

The development and operation of the oil analysis expert systems over the past 10 years demonstrates that expert system technology is mature and can be used effectively for used-oil analysis automation. The systems implemented at major North American railways also indicate the level of sophistication that can be achieved with low cost expert system technology[4,5]. The major lesson learned from the development of these systems is that the key to the successful development of a used-oil analysis expert system requires the establishment of:

1. simple and reliable 'condition indicators' of equipment failure mechanisms based on oil wear metal and contamination data,

2. statistically based limits for the magnitude and trend of each condition indicator,

3. validated fault signatures indicating the relationships between condition indicator variations (or combinations) and specific component failure mechanisms, and;

4. application of consistent operational and maintenance practices to ensure high data integrity.

REFERENCES

[1] Toms, L., Machinery Oil Analysis - Methods, Automation and Benefits, Chapter 1, Sections 1.4 and 1.5, pages 6 to 8, Pensacola, Fl, 1995.

[2] Toms, L., "Lube Oil Analysis, A Knowledge Engineering Perspective" presented at 'Update 88' Conference of the National Association of Corrosion Engineers (NACE), Toronto, Ontario, 1988.

[3] Sams, D., " The Oil and Coolant Condition Analysis Program (OCCAP)", proceedings of the 1992 Joint Oil Analysis Program International Condition Monitoring Conference, Pensacola, Fl.

[4] Toms, L. and D.E. Muir, "Expert Systems for Engine Condition Data Reduction and Interpretation", proceedings of the Third Symposium on Expert Systems at Royal Military Collage, Kingston, Ontario, 1991.

[5] Toms L., "Laboratory Automation and Expert Systems", proceedings of the 1992 Joint Oil Analysis Program International Condition Monitoring Conference, Pensacola, Fl.

MACHINERY DIAGNOSTICS & PROGNOSTICS #1

MODEL-BASED SENSOR SELECTION FOR HELICOPTER GEARBOX MONITORING

Vinay B. Jammu, Graduate Research Assistant
Keming Wang, Visiting Scholar
Kourosh Danai, Associate Professor[1]
Department of Mechanical Engineering
University of Massachusetts
Amherst, Massachusetts
and
David G. Lewicki
Vehicle Propulsion Directorate
U.S. Army Research Laboratory
NASA Lewis Research Center
Cleveland, Ohio

ABSTRACT

A new methodology is introduced to quantify the significance of accelerometers for fault diagnosis of helicopter gearboxes. The basis for this methodology is an *influence model* which represents the effect of various component faults on accelerometer readings. Based on this model, a set of *selection indices* are defined to characterize the diagnosability of each component, the coverage of each accelerometer, and the relative redundancy between the accelerometers. The effectiveness of these indices is evaluated experimentally by measurement-fault data obtained from an OH-58A main rotor gearbox. These data are used to obtain a ranking of individual accelerometers according to their significance in diagnosis. Comparison between the experimentally obtained rankings and those obtained from the *selection indices* indicates that the proposed methodology offers a systematic means for accelerometer selection.

1 INTRODUCTION

Present helicopter gearboxes are significant contributors to both flight safety incidents and maintenance costs. For example, for large/medium civil transport helicopters in the period 1956-86, gearboxes were the principal cause of 22% of the accidents which often resulted in loss of life and the aircraft [1]. Rapid and reliable fault diagnosis of helicopter gearboxes is therefore necessary to prevent major breakdowns due to progression of undetected faults, and to enhance personnel safety by preventing catastrophic failures.

[1]To whom all correspondence should be addressed

Fault diagnosis of helicopter gearboxes, like most rotating machinery, is based upon vibration monitoring. Therefore, an important issue in helicopter gearbox diagnostics is determination of the number of accelerometers to be used for monitoring and their location on the gearbox housing. Accelerometers are generally located by experts based on their proximity to gearbox components, orientation, and ease of mounting on the housing. However, the fundamental problem with this approach is that experts often demand too many sensors to be monitored on-line by the on-board computer. Another motivation for reduction of the number of accelerometers is the cost associated with the extra mountings, cabling, signal conditioning equipment, etc. As such, experts are often faced with the dilemma as which sensor(s) to eliminate without seriously undermining diagnosability of the gearbox. This calls for a methodology for quantifying the significance of various accelerometers in diagnosis.

Ideally, the significance of each accelerometer could be determined empirically by comparing diagnostic results with and without the accelerometer. This, however, would require a comprehensive set of measurement data associated with all component faults in the gearbox, which is generally not available. As a compromise to empirical assessment of accelerometer values, a model-based methodology for accelerometer selection is proposed in this paper that is independent of measurement-fault data. The basis for this methodology is an *influence model* between component faults and accelerometer readings obtained from a lumped mass model of the gearbox. The proposed sensor selection methodology uses the following criteria to evaluate the significance of individual sensors: (1) the diagnosability of the system; i.e., the ability to diagnose faults in each component, (2) accelerometer coverage; i.e., the number of components an accelerometer can monitor, and (3) the level of redundancy between accelerometers in their coverage of various components. The above three criteria are quantified by three indices: *diagnosability*, *coverage*, and *redundancy*, which are computed using the *influence model* of the gearbox. Individual or groups of accelerometers are then ranked according to their impact on the value of these indices when accelerometers are discarded. The accelerometers whose exclusion provides the greatest loss in the value of indices are assigned the highest rank to indicate their significance in diagnosis.

The validity of the proposed sensor selection indices is evaluated experimentally. For this, measurement-fault data obtained from an OH-58A main rotor transmission are used to rank the accelerometers for their effect on diagnosability of components. A comparison between these experimentally obtained rankings and those from the selection indices indicate that the indices offer a viable means of evaluating the signification of accelerometers.

2 INFLUENCE MODEL

The proposed methodology requires an *influence model* to represent the effect of various component faults on accelerometer readings. Ideally, these *influences* should be defined so as to represent the strength of vibration caused by a component fault monitored by each accelerometer. However, the strength of vibration depends on the attenuation property of the *travel path* between the component and the accelerometer, which, in turn, is a function of parameters such as the moment of inertia, stiffness, and damping of the components in the path [2,3]. As such, computation of the vibration transfer would require consideration of all the vibration travel paths associated with each component-accelerometer pair and knowledge of the parameters associated with each path [4]. Such a detailed knowledge of the gearbox is practically infeasible [5]. For example, the gear mesh stiffness is obtained by considering the gear tooth as a non-uniform can-

tilever beam [6,7], as a function of the cross-section of the tooth at the point of loading as well as load variation due to changes in direction of load application [6,4,8], friction between the meshing teeth [9], contact ratio [10], the type of gears (spur, helical, etc.) [6,7,8], and gear errors such as profile, transmission and manufacturing errors [3,8]. Similarly, the stiffness of bearings is a time-varying, non-linear function of bearing displacement and the number of rolling elements in the load zone, as well as the bearing type (roller, ball, etc.), axial preload, clearance, and race waviness [11,13].

In view of the difficulties associated with computing the strength of vibration, as an alternative, an approximate influence model is obtained to represent the average strength of vibration across all frequencies [5]. To compute this average vibration, several simplifications are considered: (1) a lumped mass model of the gearbox is considered; (2) in the absence of accurate values for stiffness coefficients, only the average static values for the stiffness coefficients are used [6,11,14]; (3) damping ratios of bearings and shafts are neglected [6]; (4) the damping ratio of gears, estimated between 0.03 and 0.17 [15], is set at 0.1 for all gears; (5) the cross-coupling terms in the stiffness matrix are neglected; (6) only the shortest vibration travel path between component-accelerometer pairs is considered; and (7) vibration transfer through the housing is neglected. Using the above simplifications, the vibration caused by a faulty component can be simulated by considering an excitation source at the component, which consists of unit amplitude sine waves of all frequencies within the bandwidth. In this research, the average vibration transferred from each component to an accelerometer is characterized by the root mean square (RMS) value of vibration across all frequencies [5].

The simplifications used to facilitate estimation of the RMS values result in only approximate RMS values. Such approximate RMS values would, in turn, result in approximate *influence* values which may seriously affect the accuracy of diagnosis. In order to cope with the approximate nature of influences, fuzzy variables [16] are used where the influences are not treated as values but as variables with a range. The influence values are, therefore, transformed into the fuzzy variables: *nil*: (0, 0.1), *low*: (0.1, 0.4), *medium*: (0.4, 0.6), *high*: (0.6, 0.9) and *definite*: (0.9, 1), so as to define the fuzzy influences in the influence model [5].

3 SELECTION INDICES

In the proposed sensor selection methodology, three indices are considered to represent the effectiveness of accelerometers in monitoring various gearbox components. These indices, which are computed based on the *influence model*, are the *diagnosability index*, *coverage index*, and *redundancy index*. As discussed earlier, influences represent the average vibration registered by accelerometers due to component faults. Therefore, they can be used to characterize the effectiveness of each accelerometer by measuring the number of components it covers, the amount of energy it receives from these components, and its relative redundancy with respect to other accelerometers.

The *diagnosability index* represents the amount of influence each component has on all the accelerometers. As such, the *diagnosability index* is defined as the sum of the influences for that component, as

$$D_i = \sum_j \frac{u_{ij} + l_{ij}}{2} \tag{1}$$

where u_{ij} and l_{ij} represent the upper and lower limits of the fuzzy influence between component i and accelerometer j, and the summation is carried out over all accelerometers j. One possible utility of the *diagnosability index* is that it ensures a minimum level of diagnosability for each component given the combination of accelerometers selected.

The *coverage index* is the measure of reach of each accelerometer. It represents the total influence between an accelerometer and all the components within the system. The *coverage index* for each accelerometer is defined as:

$$C_j = \sum_i \frac{u_{ij} + l_{ij}}{2} \tag{2}$$

where u_{ij} and l_{ij} represent the upper and lower limits of the fuzzy influences between the accelerometer j and gearbox component i, and the summation is carried over all components i. One possible use of the *coverage index* is to ensure that each accelerometer has a minimum level of coverage.

The *redundancy index* measures the effectiveness of accelerometers by measuring their overlap with other accelerometers. In terms of fuzzy influences, an accelerometer is redundant if the components it covers are already covered with higher influence values by other accelerometers. The *redundancy index* is measured for each accelerometer as:

$$R_j = \sum_i [I_{ij} - MAX(I_{il}, \text{ for every } l \neq j)] \tag{3}$$

where $I_{ij} = (u_{ij} + l_{ij})/2$ represents the average value of each influence, and the summation is carried over all components i. As defined, a larger value of R_j indicates a smaller overlap between accelerometer j and others.

4 EXPERIMENTAL

The effectiveness of the sensor selection indices was evaluated using experimental measurement-fault data from an OH-58A main rotor gearbox. The configuration of this gearbox is shown in Fig. 1, and the location and orientation of the eight accelerometers used for vibration measurement are shown in Fig. 2. Vibration data were collected at the NASA Lewis Research Center as part of a joint NASA/Navy/Army Advanced Lubricants Program [17]. The vibration signals were recorded from eight piezoelectric accelerometers (frequency range of up to 10 KHz) using an FM tape recorder. Two magnetic chip detectors were also used to detect the debris caused by component failures. Accelerated fatigue tests were performed where the gearbox was run under a constant load and was disassembled and inspected periodically, or when one of the chip detectors indicated a failure. A total of five tests were performed, where each test was run between nine and fifteen days for approximately four to eight hours a day. A total of eleven failures occurred during these tests. They consisted of three cases of planet bearing pitting fatigue, three cases of sun gear pitting fatigue, two cases of top housing cover cracking, and one case each of spiral bevel pinion pitting fatigue, mast bearing micropitting, and planet gear pitting fatigue. In order to identify the effect of faults on the vibration data, the vibration signals obtained from the five tests were digitized and processed by a commercially available diagnostic analyzer [18]. For analysis purposes, only one data record per day was used to represent gearbox vibration for each test [18].

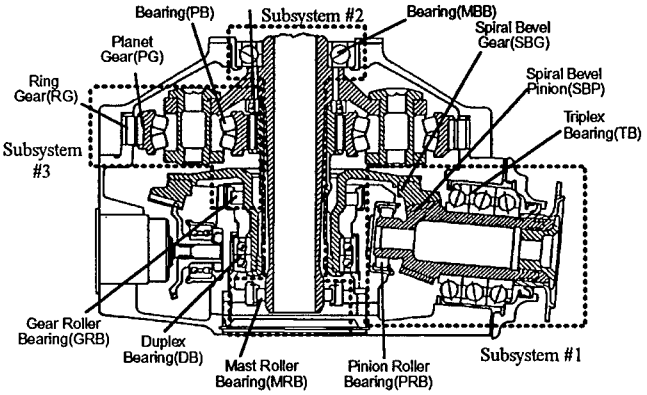


Figure 1: Configuration of the OH-58A main rotor gearbox, with the components classified into subsystems for diagnosis.

5 EVALUATION

The first step in the evaluation process was determination of the influence model. For the OH-58A gearbox five primary vibration travel paths were identified: (1) Duplex Bearing-Spiral Bevel mesh-Triplex Bearing, (2) Duplex Bearing-Sun-Planet mesh-Ring Gear (3) Mast Roller Bearing-Main Shaft-Mast Ball Bearing (4) Ring Gear-Planet Bearing-Mast Ball Bearing, and (5) Duplex Bearing-Sun Planet mesh-Mast Ball Bearing. The first travel path was in connection to accelerometers #4, #5, and #6, where as all the other paths were connected to #1, #2, #3, #6, #7, and #8. The RMS values of vibration were then computed using the lumped mass model of these paths, and used as the basis for defining the fuzzy influence model. It should be noted that due to fuzzification of influences, components which are adjacent to each other are likely to have the same fuzzy influences and hence, become indistinguishable for ranking purposes. To cope with this problem, the OH-58A gearbox was divided into three subsystems (see Fig. 1): Subsystem #1 (the input subsystem), Subsystem #2 (the output subsystem), and Subsystem #3 (the planetary subsystem), and the influences between the subsystems and accelerometers were obtained by averaging the influences of components within each subsystem (Table 1).

Using the influence model in Table 1, the three sensor selection indices were computed for the complete suite of 8 accelerometers. Next, in order to evaluate the effect of individual accelerometers on the value of these indices, the row associated with each accelerometer in Table 1 was removed and the three indices were recomputed for the resulting suites of 7 accelerometers. The estimated loss in the quality of diagnosis with various suites of 7 accelerometers was defined as:

$$\hat{L}_s = \sum_{i=1}^{3}(D_{i8} - D_{is}) + \sum_{j=1}^{8}(C_{j8} - C_{js}) + \sum_{j=1}^{8}(R_{j8} - R_{js}) \quad (4)$$

where D_{is} denotes the *diagnosability index* of Subsystem i with suite s, C_{js} represents the *coverage index* of accelerometer j with suite s, and R_{js} denotes the *redundancy index* of accelerometer j

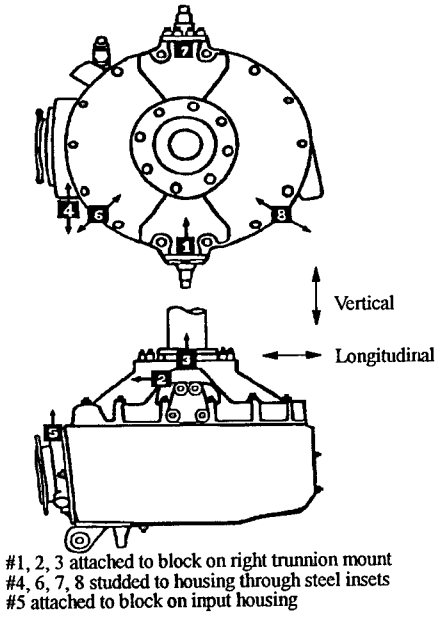


Figure 2: Location of the accelerometers on the test stand.

with suite s. This loss value was then used to classify accelerometers into one of four categories, to signify their value in diagnosis. The accelerometer whose exclusion resulted in the highest loss was classified into the first category, while the accelerometer with the least loss was placed in category 4. The above strategy was also used to classify groups of accelerometers, by computing the loss values associated with suites of 6, 5, 4, 3, 2, and 1 accelerometer(s).

Experimental evaluation of the rankings from the three indices was obtained via a model-based diagnostic system [5]. The overview of this model-based diagnostic system is presented in Fig. 3. The inputs to this system are the vibration features obtained by processing the raw vibration signal from the OH-58A accelerometers. These vibration features are first input into an unsupervised fault detection network to identify the presence of faults in the gearbox. Once the presence of a fault is prompted by the fault detection network, fault diagnosis is performed by the *Structure-Based Connectionist Network* (SBCN) whose weights comprise the fuzzy influences. The inputs to SBCN are abnormal features, which are separately identified by an unsupervised pattern classifier, referred to as the *Single Category-Based Classifier* (SCBC) [19]. The SCBC is designed to identify the degree of abnormality in individual features by comparing them against their normal-mode values. To perform diagnosis, these abnormality-scaled features are propagated through the weights of SBCN to yield as outputs fault possibility values p_i between 0 and 1 for individual gearbox subsystems.

The SBCN was used to quantify the significance of individual accelerometers in diagnosis of

Accelerometer #	Subsystems		
	Input	Output	Transmission
1	-	M	H
2	-	M	H
3	-	M	H
4	H	-	L
5	H	-	M
6	M	M	H
7	-	M	H
8	-	M	H

Table 1: Influences of the three subsystem on the eight accelerometers. The influences shown are: '-' Nil, M Medium, H High.

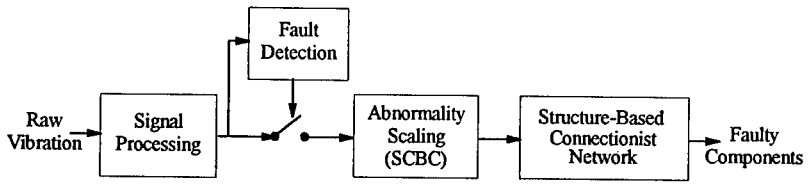


Figure 3: Overview of fault detection and diagnosis in the proposed model-based diagnostic system for helicopter gearboxes.

the OH-58A gearbox. For quantification purposes, a *performance index* P was defined to represent the accuracy of the diagnostic results obtained from SBCN. This *performance index* had the form

$$P_i = \sum_{faults} p_i \tag{5}$$

where p_i represents the fault possibility value of the ith subsystem and the summation was carried out over all the available faults that had occurred during the OH-58A experiments. In SBCN, to study the significance of each accelerometer on the diagnostic results, P was recomputed for various accelerometer suites. The *loss function* L_s was then defined to represent the differential in the P value for each of the suites relative to its value for the complete set of 8 accelerometers as:

$$L_s = \sum_{i=1}^{3} \left[\frac{P_{i8} - P_{is}}{P_{i8}} \right]^2 \tag{6}$$

where P_{is} represents the *performance index* of the ith subsystem obtained for suite s, and P_{i8} denotes the *performance index* of this Subsystem with the complete suite of 8 accelerometers. The summation is carried over the 3 subsystems to obtain the total loss. The L_s obtained for suites of the same size were then normalized to have values between 0 and 1 and used for ranking purposes. As with the selection indices, the excluded accelerometer that caused the greatest loss

was assigned to the first category. The normalized loss values and the associated ranks for suites of 7 accelerometers are included in Table 2 along with the estimated loss and ranks obtained from the selection indices. The rankings indicate that accelerometer #5, the elimination of which resulted in the highest loss, is the most important for diagnosis. The results for these suites indicate that for accelerometers #1, #2, #4, #5, and #8 the estimated and diagnostic ranks match, and that there is a mismatch by 1 rank for accelerometer #3, #6, and #7. A summary of matches and mismatches for suites consisting of a smaller number of accelerometers is given in Table 3. The results indicate that out of 254 possible suites, 150 suites have an exact match in estimated and diagnostic ranks, 103 mismatch by 1 rank and 1 suite mismatches by 2 ranks.

Accelerometer # Excluded	Indices		Diagnostic	
	Loss	Rank	Loss	Rank
1	0.425	3	0.347	3
2	0.353	3	0.464	3
3	0.353	3*	0.100	4
4	0.100	4	0.149	4
5	1.000	1	1.000	1
6	0.518	3*	0.126	4
7	0.425	3*	0.223	4
8	0.425	3	0.371	3

Table 2: Ranks obtained from the selection indices and diagnostic ranks for suites of 7 and 6 accelerometers. A '*' indicates a mismatch between indice-ranks and diagnostic ranks.

Suites of	Match Exactly	Mismatch by		
		1	2	3
7	5	3	0	0
6	16	12	0	0
5	35	21	0	0
4	40	29	1	0
3	32	24	0	0
2	17	11	0	0
1	5	3	0	0
Total	150	103	1	0

Table 3: Summary of comparison of ranks obtained from the three sensor selection indices and diagnostic ranks for suites of 7, 6, 5, 4, 3, 2, and 1 accelerometers.

The results summarized in Table 3 indicate that the selection indices are effective in assessing the value of individual (suites of) accelerometers. Using the influence model as the common point between the selection indices and SBCN de-emphasizes the effect of modeling errors. The influence

model alone, however, does not ensure a perfect match between the two rankings because the diagnostic rankings from SBCN also depend on the experimental data as well as the performance of SBCN's other components (*detection network* and *SCBC*). The experimental data, although one of the most complete sets available in the industry, are still not as comprehensive as required for a complete evaluation of the method. The main limitation is the absence of faults that could signify the value of some accelerometers which otherwise may have been assigned a lower rank. For example, there is only a single fault in Subsystem #2 (mast bearing micropitting), therefore, accelerometers that are important in isolating other faults within this Subsystem may appear as unnecessary. Although the rankings obtained from the selection indices agree well with those obtained from the diagnostic results, they can be improved further to enhance their effectiveness. One possibility is the redundancy index which may be refined further so as to include spatial redundancy of accelerometers as well.

6 CONCLUSION

A methodology is introduced for evaluating the significance of accelerometers in diagnosis of faulty gearbox components. This methodology, which is based on the influences between components and accelerometers, uses three indices for evaluation of accelerometers. The *diagnosability index* represents the coverage each gearbox component receives from all the accelerometers, the *coverage index* denotes the reach of individual accelerometers, and the *redundancy index* characterizes the overlap between the accelerometers. These indices were evaluated together in estimating the diagnostic quality of various accelerometer suites for an OH-58A gearbox. The results indicate that the rankings provided by the indices agree well with the actual rankings obtained from a diagnostic system.

ACKNOWLEDGEMENTS

This work is supported in part by the Vehicle Propulsion Directorate, U.S. Army Research Laboratory, and Lewis Research Center of the National Aeronautics and Space Administration under Grant No. NAG3-1458, and Sikorsky Aircraft.

REFERENCES

[1] D. G. Astridge, "Helicopter transmission - Design for safety and reliability," *Proc. of Inst. of Mech. Engrs.*, pp. 123–138, vol. 203, 1989.

[2] R. H. Lyon, "Structural diagnostics using vibration transfer functions," *Sound and Vibrations*, pp. 28–31, Jan. 1995.

[3] J. D. Smith, *Gears and their Vibrations.* Newyork, NY: Marcel Dekker, The Macmillan Press Ltd., 1983.

[4] F. K. Choy and W. Qian, "Global dynamic modeling of a transmission system," NASA CR-191117, Army Research Laboratory ARL-CR-11, Lewis Research Center, Cleveland, Ohio, April 1993.

[5] V. B. Jammu, K. Danai, and D. G. Lewicki, "Fuzzy connectionist network for fault diagnosis of helicopter gearboxes," in *Proc. of International Mechanical Engineering Congress and Exposition*, San Francisco, CA, Nov. 1995.

[6] H-H. Lin, R. L. Huston, and J. J. Coy, "On dynamic loads in parallel shaft transmission: Part I - Modelling and Analysis," *Journal of Mechanisms, Transmission, and Automation in Design*, pp. 221–229, vol. 110, June 1988.

[7] L. S. Boyd and J. Pike, "Epicyclic gear dynamics," *AIAA Journal*, pp. 603–609, vol. 27, no. 5 1989.

[8] W. D. Mark, "Use of the generalized transmission error in equations of motion of gear systems," *J. of Mechanisms, Transmission, and Automation in Design*, pp. 283–291, vol. 109, June 1987.

[9] B. Rebbechi, F. B. Oswald, and D. P. Townsend, "Dynamic measurements of gear tooth friction and load," NASA, TM-103281, AVSCOM TR-90-C-0023, Lewis Research Center, Cleveland, Ohio, Oct. 1991.

[10] R. W. Cornell and W. W. Westervelt, "Dynamic tooth loads and stressing for high contact ratio spur gears," *J. of Mechanical Design*, pp. 69–76, vol. 100 1978.

[11] M. F. While, "Rolling element bearing vibration transfer characteristics: effect of stiffness," *J. of Applied Mechanics*, pp. 677–684, vol. 46, 1979.

[12] T. A. Harris, *Rolling Bearing Analysis.* Newyork, NY: John Wiley and Sons, Inc., 1966.

[13] T. L. H. Walford and B. J. Stone, "The sources of damping in rolling element bearings under oscillating conditions," *Proc. of Inst. of Mech. Engrs.*, pp. 225–232, vol. 197c, Dec. 1983.

[14] W. A. Tuplin, "Dynamic loads on gear teeth," *Machine Design*, pp. 203–211, vol. 25, Oct. 1953.

[15] R. Kasuba and J. W. Evans, "An extended model for determining dynamic loads in spur gearing," *Journal of Mechanical Design*, pp. 398–409, vol. 103, April 1981.

[16] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-1," *Information Sciences*, pp. 199–249, vol. 8 1975.

[17] D. G. Lewicki, H. J. Decker, and J. T. Shimski, "Full-scale transmission testing to evaluate advanced lubricants," NASA TM-105668, AVSCOM TR-91-C-035, 1992.

[18] H. Chin, K. Danai and D. G. Lewicki, "Pattern classifier for fault diagnosis of helicopter gearboxes," *J. of IFAC Control Eng. Practice*, vol. 1, no. 5, pp. 771–778, 1993.

[19] V. B. Jammu and K. Danai, "Unsupervised pattern classifier for fault detection of helicopter power train," in *Proc. of Noise and Vibration*, (Venice, Italy), pp. 797–810, April 1995.

ELECTRICAL SIGNATURE ANALYSIS APPLICATIONS FOR NON-INTRUSIVE AUTOMOTIVE ALTERNATOR DIAGNOSTICS

Curtis W. Ayers

Oak Ridge National Laboratory
P.O. Box 2009
Oak Ridge, TN 37831-8038

Abstract: Automotive alternators are designed to supply power for automobile engine ignition systems as well as charge the storage battery. This product is used in a large market where consumers are concerned with acoustic noise and vibration that comes from the unit, as well as overall quality and dependability. Alternators and generators in general are used in industries other than automotive, such as transportation and airline industries and in military applications. Their manufacturers are interested in pursuing state-of-the-art methods to achieve higher quality and reduced costs.

Preliminary investigations of non-intrusive diagnostic techniques utilizing the inherent voltage signals of alternators have been performed with promising results. These techniques are based on time and frequency domain analyses of specially conditioned signals taken from several alternators under various test conditions. This paper discusses investigations that show correlations of the alternator output voltage to airborne noise production. In addition, these signals provide insight into internal magnetic characteristics that relate to design and/or assembly problems.

Key Words: Alternator; analysis; diagnostics; electrical; generator; magnetic; signature

Summary: Electrical Signature Analysis (ESA) was developed originally at Oak Ridge National Laboratory to allow diagnostic related information to be gathered from electric motor driven devices via a remote non-intrusive current measurement on a motor power cable. The resulting signal with appropriate processing is sensitive to mechanical and electrical loads or anomalies occurring in the motor and driven machine or system. This technique has been developed for application on many different types of motors and driven machines. Recently the ESA technique has been extrapolated to evaluate alternators and generators and the perturbations seen in their voltage output signals.

The primary goal of this investigation is to determine if alternator electrical signals such as the phase

output voltage or rotor excitation voltage and current can be indicators of alternator acoustics problems believed to be related to the magnetic performance of the alternators. In addition we have used ESA to study the internal magnetic performance of the alternator in order to evaluate the manufacturing quality and operating condition of the unit.

Sound chamber testing is used for analysis of alternators to identify noise sources. These tests are complicated and time consuming to perform, and do not always produce consistent results due to the highly variable mounting configurations from vehicle to vehicle. Sound performance for a given alternator type has been known to vary considerably from vehicle to vehicle. Information that would quantify the root sources of alternator sound from an electrical or magnetic viewpoint would be very helpful and would remove the mounting configuration variables from the analysis process. This would enable design changes or field fixes to reduce sound.

Analyses in this study were conducted in several steps. The first and most obvious step was to perform a quick study of alternator acoustic behavior using the recorded sound data from two alternators. Based on these results, a comparison of various alternators' phase output voltage signals was initiated, using signal conditioning techniques on the phase output voltage. Other parameters such as rotor excitation current and rotor excitation voltage were also reviewed. Throughout the investigation, time domain and spectral analysis techniques were used, with the most interesting results arising in the time domain.

Test Setup Description: Data was collected on automotive alternators at a manufacturer's facility in the controlled environment of an alternator sound testing chamber. This chamber was designed to allow full control of alternator speed and load, with instrumentation for vibration and sound measurement. To this was added instrumentation required to collect electrical signatures from the various parts of the alternator. Vibration, sound, and electrical data were recorded for subsequent review and signal conditioning.

Data was collected using a test stand for mounting and running the alternators. The test stand supplied variable-speed belt-driven power to the alternator, a load for the alternator output, and computer control of speed, output current, and output voltage. Data was acquired from two microphones, an optical tachometer for alternator speed, an accelerometer, and electrical taps to read current and voltage from the test units. Test alternators were taken from a pool of units exhibiting sound and bearing problems. An alternator of "good quality", as designated by the manufacturer, was used for comparison purposes.

The following signals were collected using an 8 channel Digital Audio Tape (DAT) recorder:

Channel 1	optical tachometer pulse waveform
Channel 2	accelerometer mounted on the top of the alternator case (on the stator iron)
Channel 3	vertically mounted microphone
Channel 4	45 deg mounted microphone
Channel 5	output voltage, 3 phase, rectified, from alternator output terminals, divided by 12

Channel 6	output current using a current transformer (CT) clamped on power output wire *
Channel 7	regulator input current using CT probe **
Channel 8	Unrectified phase voltage (1 of 3 phase circuits, before rectifiers), divided by 100
*	for some tests, this channel recorded rotor excitation voltage
**	for some tests, this channel recorded rotor excitation current

The alternators used in these tests are designated as A17 (good unit), A16 (noisy unit), A6 (quietest style unit), A17R (A17 with induced rub), and A17S (A17 with slip ring damage). Data was collected for steady speed (2000 rpm & full load), speed ramp-up (1000-6000 rpm), and 20 amp and 40 amp controlled current at 2000 rpm.

Sound Analysis: Test results from A6 and A16 were compared for acoustic noise. A6 is a version of the A16 style, empirically found to be quieter by the manufacturer. PC based virtual instrument software programs were used extensively in this research along with some specially designed electronic signal conditioning circuits. A computer program was written to cascade (or waterfall) 30 acoustic spectra during ramp-up of the alternators. A cascade plot stacks sequential spectra of a given signal from bottom to top to provide a visual glimpse of changing frequency content during a period of time. Ramp-ups for these tests were started at 16 Hz rotational speed (RS) and monitored as the speed increased to ~100 Hz. The ramp data was used to determine which running speeds were the noisiest. Sound cascade plots of A6 and A16 are shown in Figure 1. A16 showed high sound production at 52 Hz RS. A6 was comparatively quiet with only a small peak at 43 Hz RS. Figure 1 demonstrates that 36xRS is the predominant component of sound for both units, but the sound cascades clearly show A16 to be noisier.

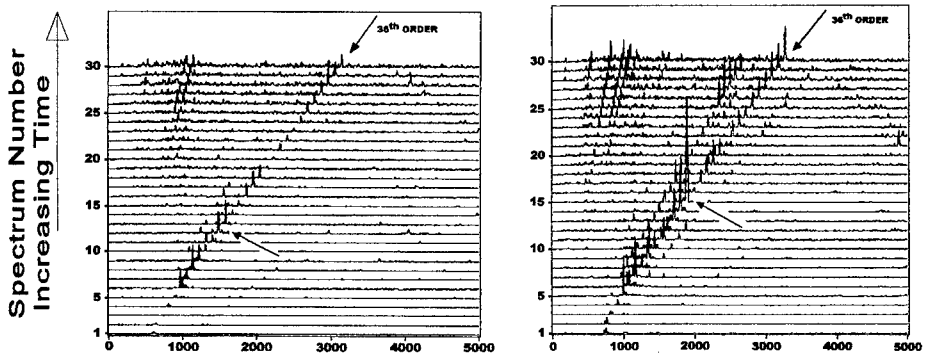


Figure 1 - Sound cascade plots during ramp tests for alternators A6 (left) and A16 (right)

An order tracking program was written to enable us to look at vibration and sound occurring at different multiples (orders) of running speed. This program tracks the speed of the alternator during ramp-up or coast-down using the optical tachometer output. A spectrum of data such as noise or vibration is acquired and sampled for a pre-determined order of rotational frequency. This

component of the frequency spectrum is plotted versus the RS for that sample. For the two alternators, 36xRS was examined in the vibration and the microphone signals to look for correlations when the noise was high, and to correlate sound behavior between the alternators. The readings taken using the order-tracking program showed the highest *vibration* for RS<40 Hz, while the alternator *sound* on all units peaked at 45-50 Hz (data not shown). This implies that the component creating the sound is not the component on which the vibration readings were taken (the external surface of the stator iron laminations). Figure 2 shows the magnitude of sound at the 36th order plotted against RS of A6 and A16. These order-tracked graphs of sound (Figure 2) correlate well with the cascade graphs (Figure 1) of sound, with order track frequency peaks matching the peaks seen in the cascade plots. The data above suggest that sound production is not necessarily related to speed, but possibly to mechanical resonances in the alternator case and supports. This is consistent with the observation that alternators sound different on different vehicles. All data analyzed showed

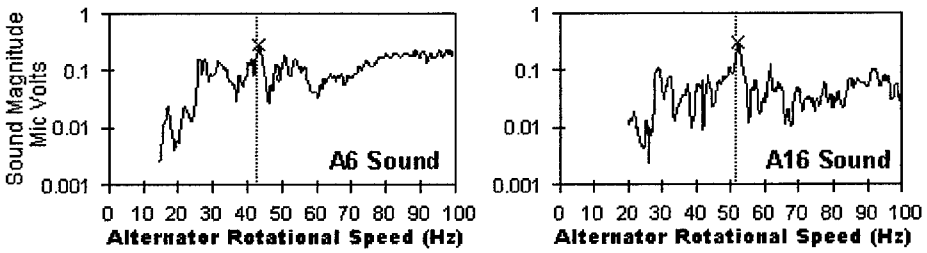


Figure 2 - Sound Magnitude tracked at the 36th order of turning speed for A6 and A16

that 36xRS was the sound driver, where 36xRS corresponds to the number of slots in the alternator stator (each rotor rotation "sees" 36 stator slots pass by).

Rotor and stator magnetic saturation conditions are believed to affect alternator transducing capability. To eliminate this possible interference, it was decided to look at lower alternator speeds

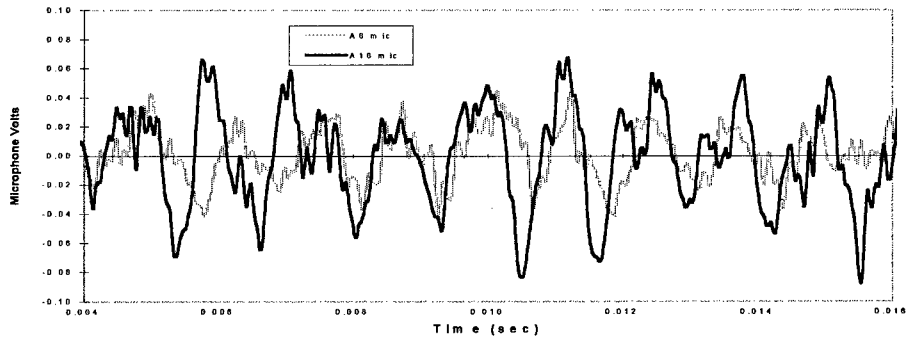


Figure 3 - Sound time waveforms from alternator A6 (gray) and A16 (bold) at 22 Hz

using the low end of the ramp data. Figure 3 shows raw time waveforms of the microphone signals taken from A6 and A16 with data acquired at 22 Hz RS. This speed was chosen because it was the lowest stable speed, and is well below magnetic saturation conditions in the iron. A6 (gray) shows lower sound amplitude than A16 (bold). The easily discernible main component in the wave is calculated to be at 36 x 22 Hz (786 Hz, 36th order). The 36th order is seen to be the sound driver here as well. This data demonstrates again that A16 is louder than A6, and is supported by data at other speeds.

Figure 4 shows voltage readings taken from the alternator's rectified outputs. This plot shows the first ESA-based evidence for differences between A16 and A6. It can be seen that the A6 voltage

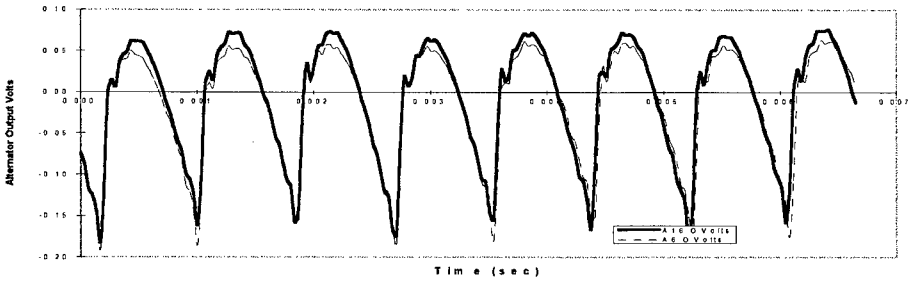


Figure 4 - Output Voltage Readings taken from A6 and A16

peaks are consistently lower than A16 voltage peaks. A similar analysis on the output current yields the same results (data not shown). The lower voltage and current may be related to lower internal magnetic force dynamics on the stator teeth and rotor poles.

Sound production is very likely related to magnetic "plucking" of these mechanical components inside the alternator, and some evidence of this is discussed later (see Pulse Ringdown). A6 has higher radial clearances between the rotor and stator teeth, and a large chamfer on the trailing edge of each rotor pole. This geometry affects the voltage and current output, suggesting the A6 unit is less efficient than units with tight geometry. The closer the radial clearance and the smaller the chamfer, the higher will be the rise rate in magnetic flux in a given stator tooth. This higher flux rise rate will result in a greater force in the stator tooth and greater mechanical deflection. A16 should experience higher dynamic forces due to it's tighter physical geometry. These forces will cause greater mechanical deflection in components such as stator teeth and rotor pole pieces, and thus more sound pressure generation.

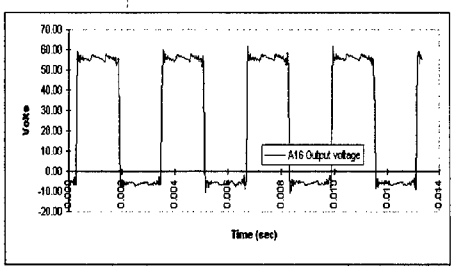


Figure 5 - Alternator Output Volts

Phase Voltage Analysis: The electrical signal

predominantly used for this study was the unrectified phase voltage (channel 8). This signal was acquired for all alternators and contains information related to stator and rotor magnetic behavior. In its unconditioned state the Phase Voltage (PV) consists of a square wave (Figure 5) having a peak-to-peak amplitude of about 60 volts. The frequency of this square wave is 6xRS generated from the 6-pole rotor passing the stator windings. Some broadband frequency content is carried on this wave, and is seen as ripples on the top and bottom of the PV wave. A circuit was developed to remove the large amplitude square wave component and amplify the other information in the signal, increasing the dynamic range of the "ripple" on the PV. This new signal is called the conditioned phase voltage (CPV, Figure 6), and was used to compare the alternators electrically. Two cycles of the CPV waveform are derived from one square wave of output voltage, the first cycle extracted from the top of the square wave and the second from the bottom.

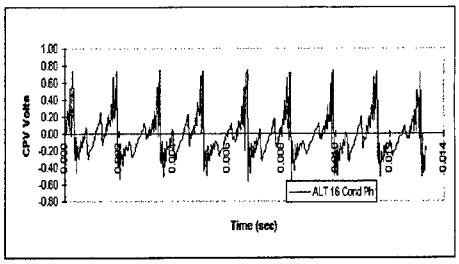


Figure 6 - Conditioned phase voltage

Figure 7 compares the CPV for four alternators examined in this study. The CPV waveforms of the alternators are very similar, to the extent that it was initially questioned whether the signal would contain any valuable alternator diagnostics. It is seen with the implanted rub and slip ring defects that the CPV clearly demonstrates diagnostic capability. All these units are running at a 2000 rpm steady state full load condition. It can be seen that nearly every deviation in the wave is matched by all four alternators. The running conditions for these tests are such that the iron in the rotor and stator are probably magnetically saturated creating a stabilizing but filtered effect.

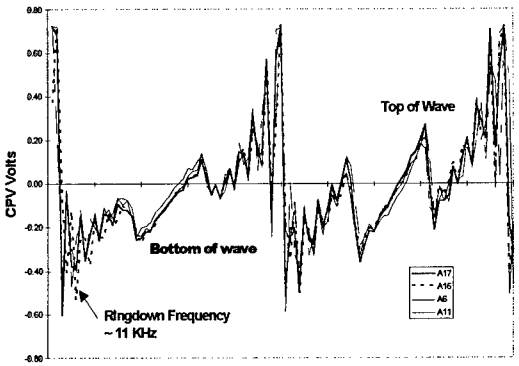


Figure 7 - Comparison of Alternator CPV

Rub and Slip Ring Damage Evaluation: Defect conditions were evaluated by using specially prepared hybrid parts along with A17. By distorting the stator a "rub" defect was implanted and is designated A17R. A "dent" placed on the slip ring produced the defect designated as A17S. These data analyses proved to be educational and indicate valuable diagnostic potential in these signals.

Figure 8 shows two time waveforms of sound (top) and CPV (bottom) for A17R. The sound wave shows pulses at running speed (52 Hz), which are the result of the rotor/stator rub. In the CPV wave

the arrow tails indicate disturbances in the amplitude that closely correlate to sound wave pulses (at arrow head). The CPV amplitude drops and there is a subsequent sound pulse at an estimated time lag of 10-13 milliseconds. Through several similar tests, the time lag was determined to be the same.

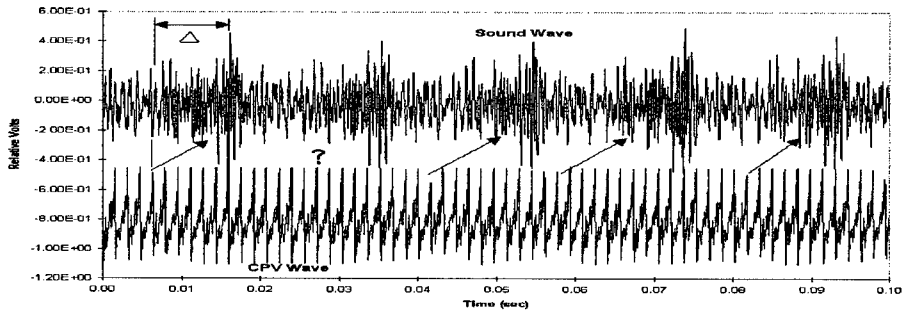


Figure 8 - CPV anomalies corresponding to sound pulses

The dented slip ring data shows interesting results and produced many areas of potential study on the nature of the alternators and the signals they produce. The pulse caused by the slip ring dent served as a "signal generator" that helped evaluate several aspects of the alternator behavior. Figure 9 shows a plot of sound and high pass filtered rotor excitation current (REC) for A17S. There is an obvious correlation between the sound wave and the REC signal showing brush "clicks" as the brush passes over the dent in the slip ring. The REC is supplied through the slip rings and should be fairly sensitive to slip ring related defects. This particular test was performed at 52 Hz RS. The sound pulses correlate to a once-per-rotation frequency, as do the REC pulses. The 3rd and 5th REC pulses shown in Figure 9 are undersized and are believed to be due to brush float after striking the dent in the slip ring. Rotor excitation voltage data (not shown) also supports the brush floating theory. A transient is also visible in the CPV data (not shown) as well when the REC spikes are large, while no CPV transient is observed when the REC pulses are small (as for the 3rd and 5th pulses).

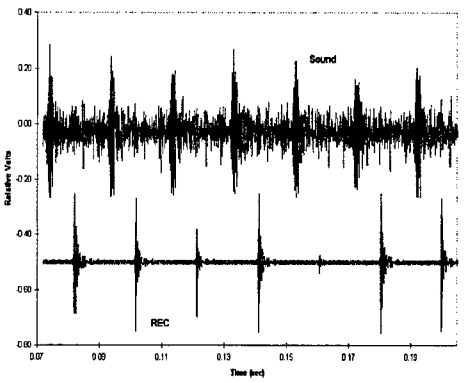


Figure 9 also shows a time lag from the electrical REC pulse to the sound pulse received at the microphone. This time is estimated at 12 milliseconds, approximately equal to the time lag mentioned previously in

Figure 9 - Correlation of REC to Sound Pulses

the rub discussion. There is acoustic time lag due to sound travel from the alternator to the microphone (2 feet in air), but this equals ~1.5 milliseconds, leaving 10.5 milliseconds of time unaccounted for. The plot in Figure 10 compares sound, *vibration* and CPV. The calculated acoustic time lag (1.5 ms) correlates well with this plot; the vibration pulse of the stator outer surface (accelerometer location) occurs 1.5 ms before the microphone signal shows it. Looking back in time in the CPV we find the characteristic pulse caused by the brush click at 10 - 11 ms before the vibration pulse. This implies that there is a 10.5 ms lag *inside the alternator*, possibly related to mechanical impedance or inductance effects. At this particular speed the lag is equal to about 50% of one rotor rotation. This significant amount of time lag is indicative of sound propagation behavior inside the alternator. In the case of the 36xRS driven sound, the time lag may be due to vibration traveling down into the rotor from a vibrating rotor pole (magnetically excited at 36xRS), through the shaft and bearings, and *then* into the alternator case (and accelerometer) where it acoustically couples to the air.

It has been observed that, given two similar alternators, one may be noisy and the other relatively quiet. Our data suggests that it is possible that *good manufacturing tolerances* in stator slot or rotor pole spacing and magnetic symmetry could cause increased sound production. Excellent symmetry (i.e. good manufacturing tolerances) could cause several poles to be "plucked" *in-phase* creating higher sound pressure.

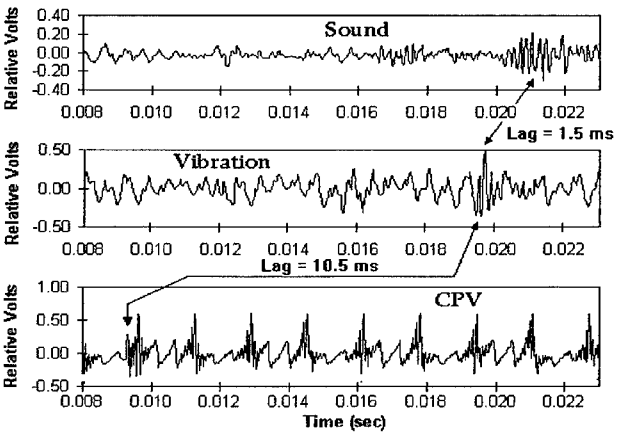


Figure 10 - Sound and vibration pulse comparison

Pulse Ringdown in the CPV: Each cycle of the CPV shows three periodic events (see Fig 5 & 7) having the appearance of ringdowns with an event period of 36xRS. The subsequent ringing could be related to inductive resonance effects, or to mechanical effects of components moving inside the alternator. The ringdown frequency seen in the time domain data (Fig 7) is estimated at 11 kHz.

In keeping with the theory mentioned in the rub and slip ring section about rotor pole vibration, a cursory impact test on one of the rotor poles on the sample alternator brought back from the manufacturer was performed. The impact was performed at the tip of the pole with the accelerometer located near the base to reduce the accelerometer mass loading effect on the pole's natural frequency. The test configuration is shown in Figure 11A and the spectral results are shown in Figure 11B. The primary peak seen in this impact test (10.6 kHz) is very close to the 11 kHz ringdown estimate (above) and suggests a relationship to the waveform ripple seen in Figure 7.

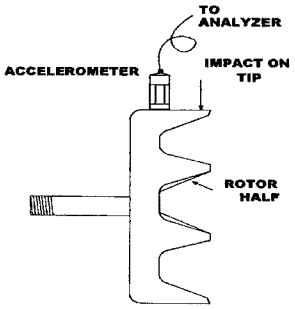


Figure 11A - Rotor Impact Sketch

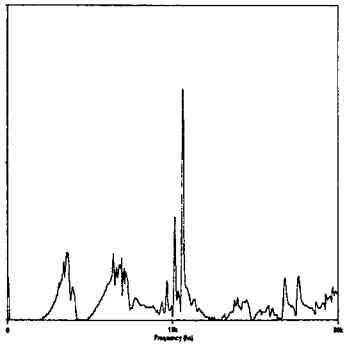


Figure 11B - Impact response Spectrum

Output Volts Analysis: Figure 12 shows two waves which are taken from the same continuous output voltage waveform. The solid wave has 36 stator tooth passing pulses representing one rotation of the alternator. The dashed data is the next consecutive rotation of the alternator, showing the similarity of the magnetic performance of the rotor/stator from one rotation to the next (the dashed data is a sequential group of 36 pulses superimposed on the previous 36 pulses and shifted ½ cycle to the right for clarity). The shape and magnitude of each peak is directly comparable, and shows stator magnetic "shape" or magnetic "trueness" within one rotation of the rotor. This analysis allows one to evaluate winding insulation degradation on a slot-by-slot basis, a weak stator section, machining quality, rotor eccentricity, winding inductive balance, and other degradation phenomena. As-assembled "magnetic" condition can also be evaluated with this method.

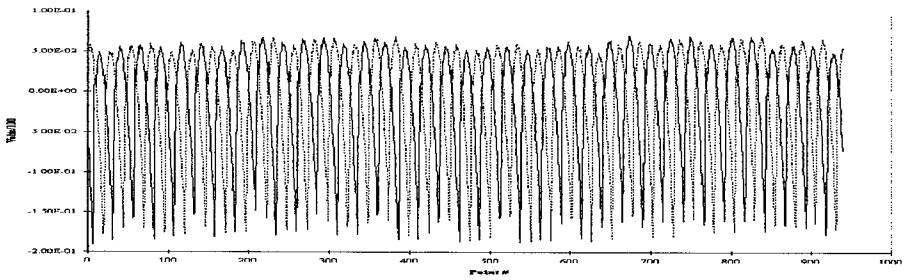


Figure 12 - A6 Output volts, Overlay of Two Full Rotations

Since alternators are mainly electrical in nature, ESA techniques provide good insight into the internal performance. Figure 13 demonstrates the ability of ESA to "look" inside the alternator, showing a plot of output voltage, with the data taken at 22 Hz RS. A6 shows a "softer" more rounded off peak than the other alternators. Particularly A16 shows a very sharp, square wave shape; the other alternator shown exhibits a similar sharp shape, especially in the leading edge. This is a good indicator of the magnetic flux "wave shape" and is related to sound production in the alternator.

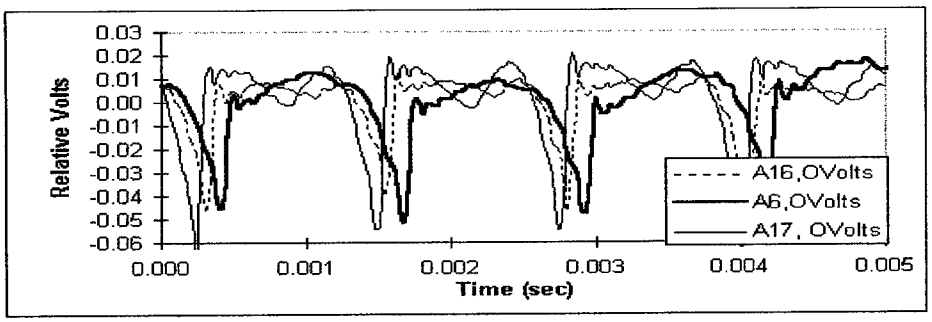


Figure 13 - Alternator wave shape comparison

Conclusions: Microphone sound data was reviewed and indicated A16 to be the noisier unit. It was found that sound correlates well with the output volts wave shape, and this could lead to design aids that would improve alternator noise levels.

Variability between units of similar design has been a source of problems to the testing efforts of manufacturers. Data seen here agrees that indeed this is a problem, and this study proposes ways to avoid the difficulties encountered with conventional methods. Electrical signature analysis is inherently less sensitive to mechanical resonances and external vibrations and more sensitive to the electrical drivers inside the alternator.

ESA was shown to be highly sensitive to rotor-to-stator rub and the slip ring defect, and reveals much about the magnetic processes inside the alternator. Future work will lead to a better understanding of the magnetic and mechanical interaction of stator slots and rotor poles.

A fundamental understanding of the sound source will lead to improved design and manufacturing efforts and ESA provides a non-intrusive view into the alternator's behavior. The output voltage wave analysis shows the magnetic shape or trueness stemming from the manufacturing process.

The impact and time lag analyses suggest that the predominant sound at 36xRS is generated by vibration of the rotor poles. Further study is indicated to determine the precise sound source, with the ultimate goal being to improve the design or manufacturing process to reduce noise at the source.

This study opens the door to analysis of many aspects of alternators, and would be easily extrapolated to generators and similar devices such as electric clutches and brakes. Significant impact can be realized in the area of manufacturing quality, as well diagnostics and prognostics of operating units and units in repair shops.

USE OF THE DECISION SUPPORT PROBLEM TECHNIQUE FOR PROPULSION ENGINE SELECTION EMPHASIZING RELIABILITY, MAINTENANCE, AND REPAIR FACTORS: A LIMITED EXAMPLE.

Kevin Lynaugh
Naval Surface Warfare Center
Carderock Division, Bethesda, MD

Eric Pogue and Bruce Friedman
ERP Enterprises
506 Sylview Drive
Pasadena, MD 21122

Abstract: A Decision Support Problem Technique (DSPT) is used for the selection of a slow speed diesel engine for the propulsion of a proposed commercial cargo vessel satisfying strategic sealift requirements, and emphasizing reliability, maintenance, and repair factors. A two step procedure is utilized. In the first step, the initial set of five engines and engine variants is reduced to three candidate engines on the basis of generalized criteria. In the second step, engine attribute weights are found through the application of quality function deployment (QFD). Combining the attribute weights with engine attribute ratings generates merit function values for each engine. The analysis is performed with three of the seven attributes related to a reliability, maintenance, and repair criterion: component consumption rate, replacement part cost, maintenance cost(overhaul). The engine chosen is the MAN B&W 5 cylinder K90MC MK VI diesel.

Key Words: customer criteria; decision support problem technique; designer attributes; slow speed diesel engines.

Introduction: The design of sealift ships has been an on-going preoccupation of the U.S. Navy. In support of these efforts, the Maritime Administration (MARAD) of the United States Department of Transportation accomplished the preliminary design of a ship referred to as PD-337. The objective of this study was to develop a commercially viable ship design which would also satisfy the mid-term strategic sealift requirements of the U.S. Navy, via incorporation of "National Defense Features" paid for by the U.S. Government.(1)

The Navy is currently pursuing a broad-based Sealift Ship Technology Development Program and the Naval Surface Warfare Center, Carderock Division (NSWCCD) has a major role in this effort. One component of this program is the Engine Room Arrangement Modeling (ERAM) Project. The purpose of ERAM is to determine a suitable propulsion engine and associated machinery and to integrate these components into an effective and efficient engine room. In support of the MARAD effort a Sealift Research and Development Program was established at the Naval Surface Warfare Center Carderock Division (NSWCCD). One component of this program is engine room arrangement modeling (ERAM). The purpose of ERAM is to determine a suitable

propulsion engine and associated machinery and to integrate these components into an effective and efficient engine room.

Objective: The objective of this study is to examine a limited special case of ERAM, namely, that of selecting for a PD-337 sealift ship a single propulsion engine from an initial set of slow speed two-stroke diesel engines. Emphasis will be placed upon ascertaining the impact of the variation of the reliability, maintenance, and repair factors upon the selection process and its results. (The continuous service rating (CSR) of the engine for the PD-337 concept 24,948 BHP (brake horsepower, metric) with the propeller rating at 89 rpm, with the maximum continuous rating (MCR) being 28,700BHP.)(1-3)

The MAN B&W engines comprising the initial set are the 5K90MC MK VI, with high efficiency turbochargers, the 7L80MC with conventional turbochargers, the 7L80MC with high efficiency turbochargers and turbo compound system (TCS), the 8S70MC with conventional turbochargers, and the 8S70MC with high efficiency turbochargers and TCS.(4-8) The first number in the designation for each engine describes the number of cylinders in that engine and the first letter represents the stroke/bore ratio. (k, short; L, long; S, super long) (4-7)

Approach: The approach to be used to make the preliminary selection is based upon the work of Mistree and others. This work has generated the Decision Support Problem Technique (DSPT). The selection process is derived from the DSPT as a special case. The problem of selection is divided in two phases, the Preliminary Selection Decision Support Problem (PSDSP) and the Final Selection Decision Support Problem (FSDSP).(2,3,8,9)

Discussion and Results: The solution of the PSDSP commences with a description of the concepts. This is the first step. In this case the concepts are the MAN B&W diesel engines. All of the engines are described in literature available from this manufacturer.(4-7)

The second step for solving the PSDSP involves the determination and description of generalized criteria and their respective sets of specific criteria. Four generalized criteria (categories) have been ascertained for this problem:(1) engine operating characteristics, (2) installation, (3) reliability, maintenance, and repair, (4) impact of operations. The specific criteria associated with category (1) are specific fuel oil consumption (SFOC), power match, engine/propeller speed match, and lubrication oil consumption. For category (2) the specific criteria are engine and installation price and ease of installation (size and dimensions). The category (3) specific criteria are the component wear rate, part replacement logistics, time between component failures, spare part cost, and overhaul cost. The category (4) specific criteria are exhaust gas amount, NO_x concentration, vibrations, and noise.(2,3,8)

The SFOC is measured in g/BHP (grams/brake horsepower). By power match is meant the difference between the power generated by the engine at the MCR power level and at its L_1 point. The L_1 point is a point on the engine layout diagram. This diagram shows the layout area within which a combination of engine power and speed can be selected that is optimum for the ship and the expected operating profile. The L_1 point denotes the nominal maximum continuous rating (NMCR) of the engine. Engine/propeller speed match refers to the difference between the rate at

which the propeller rotates at the CSR level (89 rpm) and the rotational rate produced by the engine at NMCR. Lubrication oil consumption refers to the sum of the system oil and cylinder oil consumptions. System oil is measured as kg/cylinder-24 h and cylinder oil is measured as g/BHPh. Engine and installation price will vary with the engine. The ease of installation is a function of the engine weight, volume, and dimensions.(4)

The manufacturer tabulates the average expected consumption of wearing parts over a ten year service period at one service year intervals for a given new engine with a service year assumed to be 6000 hours in duration. Part replacement logistics involves the time to acquire such parts if they are not included in the inventory of the ship operator. Literature from the engine manufacturer gives some information allowing the time factors to be deduced on a relative basis. A similar deduction can be made for time between component failures. Spare part and overhaul cost information are given by the manufacturer in absolute terms.(4-7)

The amount of exhaust gas generated by the various engines and engine variants is presented by the manufacturer for L_1 point operation. Estimates for exhaust gas amount are then calculated for operation at the CSR power level.(4-7) It is assumed that the NO_x concentration would be proportional to the amount of exhaust gas.

The major vibrations of concern are those arising from first and second order moments and the torsional vibrations. It is assumed that the exhaust noise is proportional to the exhaust flow rate and that the structural noise is proportional to the engine vibrational energy. (4-7)

At the outset, it is assumed that all of the generalized criteria are equally important with each other and all of the specific criteria are equally important with each other.

In the third step of the solution of the PSDSP, a datum is chosen with which all of the concepts are to be compared. The concepts comprise the set of MAN B&W engines and engine variations under consideration. It does not matter which engine is selected first, since, in turn, the PSDSP solution will be attained using each engine as the datum.(2,3,8)

For the fourth step, the different engines are compared. The mix of qualitative and quantitative information required is illustrated by the comparisons undertaken utilizing the specific criteria which are found in the domain of the generalized criterion of reliability, maintenance, and repair.(2,3,4,8)

The component wear rate criterion among the five candidate engines has the lowest wear rate assigned to the 5K90MC MK VI. The 7L80MC engines with and without the TCS are taken to be about equal in wear rate and assigned to be between the 5K90MC MK VI and the two 8S70MC engines. The 8S70MC engines are assigned the highest wear rate. The wear rate is taken to be proportional to the number of cylinders within a given engine. The absence or presence of a TCS is assumed to make a negligible difference.(4-7)

The part replacement logistics criterion has equal values assigned to the three engines not having the TCS. The two engines with the TCS also have equal values but these values are lower than

the value assigned to the non-TCS engines. The TCS involves more exhaust piping and a fifteen ton gas turbine and shroud at the base of the engine, thereby covering the side and end of the engine, making work on the scavenger unit and high efficiency turbochargers and related components more difficult.

The time between component failures criterion has equal values assigned to all five engines. On a relative basis the equal value assignment was made because all of the engines have similar MAN B&W parts neglecting the absence or presence of the TCS.(4-7)

The spare part cost criterion rating is assigned to each engine on the basis of quantitative information from the MAN B&W MC program engine selection guidebook. The 5K90MC MK VI gets the highest rating while the two 8S70MC variations get the same lowest rating with the two 7L80MC variations having the same middle rating. The situation with respect to the overhaul cost criterion is similar to that of the spare part cost criterion.(4-7)

Table (1) shows the preliminary selection scores and ranks for the engine comparisons using the 5K90MC MK VI engine as the datum. In Table (1), "-1" denotes "worse than the datum", "0" denotes "same as the datum," and "1" denotes "better than the datum This table of engine comparisons is the product of the fourth step of the solution of the PSDSP. (2,3,8,9)

The normalized score in Table (1) is also referred to as the merit function value (MFV). The scores are normalized using equation (1):

$$R_{ij} = (A_{ij} - A_{ij}^{min}) / (A_{ij}^{max} - A_{ij}^{min}) \quad (1)$$

where i denotes one of the engines, j denotes a specific criterion, and A_{ij}^{min} and A_{ij}^{max} represent the lowest and highest possible scores, respectively, of the engine rating A_{ij}. In this Table, the generalized criteria are taken to be of equal weight.(2,3,8,9)

Table (1) then incorporates the fifth step for the PSDSP solution which is the obtaining of the MFVs for each generalized criterion.(2,3,8,9)

In Table (1) the overall score for each engine is the sum of its four MFVs. The ranks for the engines are then obtained by comparing the overall scores. It should be noted that the sum of the MFVs does not constitute an overall MFV (OMFV). The OMFV appears within the context of the sixth step of the PSDSP solution.(2,3,8,9) At the bottom of Table (1), the sum of the scores of the generalized criteria establishes the ranking of the engines. The three top ranked engines are the 5K90MC MK VI, the 8S70MC, and the 8S70MC with TCS.

The preliminary selection table using the 7L80MC with TCS as a datum yields the same three top ranked engine although the 5K90MC MK VI is in third place. For the 7L80MC with TCS as datum, the results from the corresponding table are similar to those for the 7L80MC datum. For the 8S70MC as datum, the results are similar to those for Table (1). For the 8S70MC with TCS datum, the preliminary selection table has the 5K90MC MK VI and 8S70MC engines tied for first place with the 8S70MC with TCS in second place.(2,3,8,9)

In the sixth step of the PSDSP solution interactions between the generalized criteria are incorporated into the analysis, the OMFVs are calculated, and the overall ranks are determined. Scenarios are defined by weight assignments. Within a given scenario, each generalized criterion is assigned a weight. The sum of the weights is equal to one. By multiplying each generalized criterion by its corresponding weight and summing over the four products for each engine an OMFV is obtained. For a particular datum and a particular scenario, the OMFVs can be compared with each other to yield a ranking of the engines. The reliability, maintenance, and repair generalized criterion has been assigned weights ranging from .1 to .5 to explore the significance of variations of importance of this criterion to the ascertaining of engine OMFVs and rankings.(2,3,8,9)

Table (2) shows the OMFVs for the 14 scenarios using the 5K90MC MK VI engine as the datum.(2,3,8,9) It is seen that the top rated engine under all of the scenarios is the 5K90MC MK VI. Except for scenarios 6,9, and 13, the second place engine is the 8S70MC with TCS. The third place engine is the 8S70MC, except for scenarios 6, 9, and 13. In scenario 6, the 8S70MC and the 8S70MC with TCS are tied for second place. In scenario 9, the 8S70MC is in second place with the 8S70MC with TCS being in third place. In scenario 13, the 8S70MC is in second place with the 7L80MC being in third place. Except for scenario 13, the top three rated engines under the 5K90MC MK VI datum are the 5K90MC MK VI, the 8S70MC with TCS, and the 8S70MC. (2,3,8,9)

Tables similar to Table (2) are constructed using the other engines and variants as datums. Combining the information generated by these tables selects three engines for the completion of the preliminary selection process. These alternative engines to which the final selection process is to be applied are 5K90MC MK VI, 8S70MC and 8S70MC with TCS.

In the first step of the solution of the FSDSP the alternatives are described. The second step in the solution of the FSDSP requires that the attributes be described and that their relative importances with respect to each other be specified Within the context of this particular problem the **criteria** are taken to be primarily defined by the customer such as the ship operator. The **attributes** are taken to be primarily defined by the designer. The designer attributes are generated by the designer in order to ascertain the significant engineering parameters and characteristics that are involved in satisfying the customer criteria.(4-8)

Table (3a) presents the customer criteria. The two basic types of criteria are the generalized criteria and the specific criteria. The four main headings in Table (3a) constitute the generalized criteria with the specific criteria being listed under these headings. Table (3b) presents the designer attributes listed under the same four headings as in Table (3a). Some of the attributes are the same as the corresponding customer specific criteria. To refine the set of customer criteria into a more manageable set of designer attributes, recourse is made of Quality Function Deployment (QFD) as a translating device. In Tables (3a) and (3b), the three and four character letter and number combinations at the end of a line denote the abbreviations for the factor on that line.(8)

Table (4) is a QFD matrix. The numbers in the first column before the column of customer criteria abbreviations are the relative importances of those criteria on a scale of 0 (no importance) to 5 (extremely important). The boxes starting after the second column contain values that indicate the strength of the relationship between a given criterion and a given attribute listed in the top row. In these boxes, no relationship is denoted by 0, a weak relationship is denoted by 1, a medium strength relationship is denoted by 3, and a strong relationship is denoted by 9. Table (4) is labeled as being scenario #1. The relative importance values in the first column define a scenario. Twenty seven other scenarios were evaluated but are not presented herein due to space limitations. The numbers in the bottom row of Table 4 are importance ratings. These ratings are the sums of products. A product is obtained by multiplying the relative importance rating of a particular criterion by its relationship strength with a given attribute. The sum is obtained by adding up all of the products generated by the same given attribute.(8)

Upon examining all of the twenty eight scenarios, it was found that the same seven attributes are the most important in going from one scenario to another. These seven attributes are :(1)SFOC, (2) oil consumption (including system oil and cylinder oil), (3) rate of component consumption, (4) replacement (spare) part cost, (5) maintenance cost (overhaul), (6) exhaust gas amounts, (7) NO_x amounts within the exhaust.

Referring to Table (3a), the four customer generalized criteria are: (1) engine operating characteristics, (2) installation, (3) reliability, maintenance, and repair, (4) impact of operation on crew, ship, and the environment. Attributes (1) and (2) are classed under criterion (1). Attributes (3), (4), and (5) are classed under criterion(3). Attributes (6) and (7) are classed under criterion (4). Of particular interest for this investigation are the three reliability, maintenance and repair attributes(3), (4), and (5). The seven attributes having the highest importance ratings are those chosen for utilization in the final engine selection process.

Table (5) presents the normalized relative importance ratings (NRIRs) for the seven most important attributes for each scenario. An NRIR is obtained by dividing the importance rating of the particular attribute by the sum of the importance ratings of all seven attributes for a given scenario. The NRIRs of Table (5) represent the completion of the second step of the solution of the FSDSP. The method by which the NRIRs are determined is a variation of the ranking method for the determination of the weights for the relative importance of the attributes. The weights are taken to be identical to the NRIRs. In the ranking method, the attribute ranks and weights are proportional to their respective importances with the weights being normalized prior to their utilization in the analysis.(8)

For the component consumption rate attribute NRIR over all of the scenarios, the values vary from 0.119 to 0.160. For the replacement part cost attribute, the NRIR range is from 0.103 to 0.132. For the maintenance cost (overhaul) attribute, the NRIR range is from 0.417 to 0.211.

For a given scenario, the smallest range in NRIRs among the three attributes associated with the reliability, maintenance, and repair criterion is found in scenario #3 (.035) while the largest range is found in scenario #27 (.078). Referring to Table (3a) for the customer criteria, it is found that the specific criteria C3a, C3b, C3c, C3d, and C3e comprise the reliability, maintenance, and repair

generalized criterion. In scenario #3, the QFD relative importance values for these specific criteria are 3,2,1,5, and 4, respectively. In scenario #27, the QFD relative importance values for these specific criteria are 5,5,5,5, and 5, respectively. These two sets of values are indicative of the possibility that increasing the value of the customer criteria is related to an increase in the NRIR value range for the relevant designer attributes.

In the third step of the solution of the FSDSP, the scales of the attributes are specified and the alternative engines are rated with respect to each attribute. The interval scale was selected for all of the seven attributes.(8)

Tables (6a)-(6g) present the criteria for interval scale creation for the seven attributes. Attributes (1), (2), (4), (5), and (6) are associated with more quantitative information while attributes (3) and (7) are associated with more qualitative information. Two of the attributes related to the reliability, maintenance, and repair generalized criterion have associated more quantitative information while one related attribute has associated more qualitative information.

The upper and lower limits for each of the attribute criterion scales is determined by the limits of the three alternative engines with respect to each of the attributes. As part of the third step, each alternative engine is rated with respect to each attribute.

In the fourth step of the solution of the FSDSP, the ratings are normalized. The normalization for any set of attribute values is performed by dividing the given attribute value by the maximum attribute value. Under the component consumption rate attribute all three engines have a normalized rating of 1. Under both the replacement part cost and maintenance cost attributes, the 5K90MC MK VI has a rating of 1 with the 8S70MC and the 8S70MC with TCS both having ratings of 0.1

In the fifth step, the merit function is calculated for each alternative engine. A linear model is employed. A term is formed for each attribute comprising the product of the NRIR of that attribute and the normalized rating of the engine for that attribute. The sum of these products over all seven attributes is calculated to get the merit function value for the engine. Table (9) presents the merit function values of the three alternative engines under the 28 scenarios considered. For all of the scenarios the merit function values in descending order are those for the 5K90MC MK VI, the 8S70MC with TCS, and the 8S70MC with conventional turbochargers.

Conclusion: It is concluded that the top ranked engine is the 5K90MC MK VI, the second ranked engine is the 8S70MC with TCS, and the third ranked engine is the 8S70MC with the conventional turbochargers. Considering the broad range of scenarios encountered, the engine to be selected for the propulsion of the sealift cargo vessels is the 5K90MC MK VI.

Acknowledgments: Software that was used in part of the preliminary selection was supplied by F. Mistree and J.K. Allen of the Systems Realization Laboratory, Georgia Institute of Technology, Atlanta, Georgia.

References

1. Charles B. Cherrix and Mark P. Lasky, "A Commercial Cargo Ship for Sealift in the Year 2000," SNAME Transactions, vol. 100, 1992, pp. 449-485.

2. F. Mistree, W.F. Smith, B.A. Bras, J.K. Allen, and D. Muster, "Decision-Based Design: A Contemporary Paradigm for Ship Design," SNAME Transactions, vol. 98,1990, pp. 565-597.

3. Warren F. Smith, Saiyid Kamal, and Farrok Mistree, "The Influence of Hierarchical Decisions on Ship Design," Marine Technology, vol. 24, no. 2, April 1987, pp. 131-142.

4. MAN B&W Diesel A/S, Copenhagen, MC Programme Engine Selection Guide: MAN B&W Diesel A/S Two Stroke Engines, 3rd edition, March 1993.

5. MAN B&W Diesel A/S, Copenhagen, K90MC MK VI Project Guide: Two Stroke Engines, 3rd edition, May 1994

6. MAN B&W Diesel A/S, Copenhagen, L80MC Project Guide: Two Stroke Engines, 2nd edition, February 1993.

7. MAN B&W Diesel A/S, Copenhagen, S70 MC Project Guide: Two Stroke Engines, 2nd edition, June 1992

8. Thomas Orsak, Bruce Keckley, Janet K. Allen, Farrokh Mistree, Zahed Siddique, David Tibbets, and Douglas Nelson, DSPT Workbook, (Systems Realization Laboratory, Georgia Institute of Technology, Atlanta, Georgia)

9. Thomas Orsak, Bruce Keckley, Janet K. Allen, Farrokh Mistree, Zahed Siddique, David Tibbets, and Douglas Nelson, Users Manual for the DSPT Workbook, (Systems Realization Laboratory, Georgia Institute of Technology, Atlanta, Georgia)

Table 1

Preliminary Selection: scores and ranks

General Criteria	Specific Criteria	Concepts				
		DATUM 5 K-90MC	7 L-80 MC	7 L-80 MC (TCS)	8 S-70 MC	8 S-70 MC (TCS)
ENGINE OPERATING CHARACTERISTICS						
	SPECIFIC FUEL OIL CONSUMPTION	0	-1	1	-1	1
	POWER MATCH	0	-1	-1	1	1
	ENGINE/PROPELLER SPEED MATCH	0	1	1	1	1
	LUBE OIL CONSUMPTION	0	-1	-1	-1	-1
	Score	**0**	**-0.5**	**0**	**0**	**0.5**
	Normal Score	**0.5**	**0**	**0.5**	**0.5**	**1**
INSTALLATION						
	ENGINE AND INSTALLATION PRICE	0	0	0	0	0
	EASE OF INSTALLATION (SIZE)	0	1	1	1	1
	Score	**0**	**0.5**	**0.5**	**0.5**	**0.5**
	Normal Score	**0**	**1**	**1**	**1**	**1**
RELIABILITY MAINTENANCE AND REPAIR						
	COMPARATIVE WEAR RATE	0	-1	-1	-1	-1
	PART REPLACEMENT LOGISTICS	0	0	-1	0	-1
	TIME BETWEEN COMPONENT FAILURES	0	0	0	0	0
	SPARE PART COST	0	-1	-1	-1	-1
	OVERHAUL COST	0	-1	-1	-1	-1
	Score	**0**	**-0.6**	**-0.8**	**-0.6**	**-0.8**
	Normal Score	**1**	**0.25**	**0**	**0.25**	**0**
IMPACT OF OPERATIONS						
	EXHAUST GAS AMOUNT	0	-1	-1	-1	-1
	NOx CONCENTRATION	0	-1	-1	-1	-1
	VIBRATIONS	0	1	1	1	1
	NOISE	0	0	0	0	0
	Score	**0**	**-0.25**	**-0.25**	**-0.25**	**-0.25**
	Normal Score	**1**	**0**	**0**	**0**	**0**
OVERALL SCORES AND RANKS						
	Sum of Scores	**2.5**	**1.25**	**1.5**	**1.75**	**2**
	Ranks	**1**	**5**	**4**	**3**	**2**

Table 2

Overall Merit Function Values

K-90 DATUM

Acronym	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7
5 K-90 MC	**0.75**	**0.7**	**0.7**	**0.75**	**0.75**	**0.8**	**0.8**
7 L-80 MC	0.2	0.175	0.15	0.15	0.175	0.2	0.175
7 L-80 MC with TCS	0.25	0.3	0.3	0.25	0.25	0.2	0.2
8 S-70 MC	**0.35**	**0.375**	**0.35**	**0.3**	**0.325**	**0.3**	**0.275**
8 S-70 MC with TCS	**0.4**	**0.5**	**0.5**	**0.4**	**0.4**	**0.3**	**0.3**

Acronym	Scenario 8	Scenario 9	Scenario 10	Scenario 11	Scenario 12	Scenario 13	Scenario 14
5 K-90 MC	**0.625**	**0.8**	**0.7**	**0.6**	**0.65**	**0.85**	**0.75**
7 L-80 MC	0.3125	0.225	0.125	0.375	0.25	**0.225**	0.225
7 L-80 MC with TCS	0.375	0.2	0.3	0.4	0.35	0.15	0.25
8 S-70 MC	**0.4375**	**0.325**	**0.325**	**0.475**	**0.4**	**0.275**	**0.375**
8 S-70 MC with TCS	**0.5**	**0.3**	**0.5**	**0.5**	**0.5**	0.2	**0.4**

Table 3a ***Customer Criteria:***

1. ENGINE OPERATING CHARACTERISTICS
   a. SFOC (CSR, L_1) **C1a**
   b. power match (CSR, L_1) **C1b**
   c. engine/propeller match (CSR, L_1) **C1c**
   d. lubricating oil consumption **C1d**
2. INSTALLATION
   a. cost of engine itself and cost of installation **C2a**
   b. ease of installation, (volume and weight) **C2b**
3. RELIABILITY, MAINTENANCE AND REPAIR
   a. component wear rate **C3a**
   b. part replacement logistics **C3b**
   c. mean time between and mean duration of overhauls **C3c**
   d. spare part cost **C3d**
   e. maintenance cost (overhaul) **C3e**
4. IMPACT OF OPERATION ON CREW, SHIP AND THE ENVIRONMENT
   a. exhaust gas amount **C4a**
   b. No_x concentration in exhaust **C4b**
   c. vibration **C4c**
   d. noise **C4d**

Table 3b ***Designer Attributes:***

1. ENGINE OPERATING CHARACTERISTICS
   a. SFOC, CSR, g/BHPh **A1a**
   b. power match (CSR/L_1) **A1b**
   c. engine propeller speed match (CSR/L_1) rpm **A1c**
   d. 1 system oil consumption Kg/day **A1d1**
      2 cylinder oil consumption CSR/BHPh **A1d2**
2. INSTALLATION
   a. 1. engine cost, $ **A2a1**
      2. installation cost, $ **A2a2**
   b. 1. engine mass in tons **A2b1**
      2. H3 (electrical jib crane)+A, mm **A2b2**
      3. L1 (minimum length), mm **A2b3**
3. RELIABILITY, MAINTENANCE AND REPAIR
   a. number of components consumed in given time period, number of parts divided by number of hours **A3a**
   b. 1. average length of time to obtain parts, hours **A3b1**
      2. longest time to obtain a part, hours **A3b2**
   c. 1. mean time between breakdowns, hours **A3c1**
      2. mean duration of breakdowns, hours **A3c2**
   d. spare part cost, $ **A3d**
   e. maintenance cost (overhaul), labor $/(nominal Bhp x 6000hrs.) **A3e**
4. IMPACT OF ENGINE OPERATION ON CREW, SHIP, AND ENVIRONMENT
   a. exhaust gas amount at CSR **A4a**
   b. No_x concentration, kg/hr **A4b**
   c. 1. external unbalanced moments, PRU,Nm/KW, (CSR) **A4c1**
      2. guide force moments, CSR, KNM **A4c2**
      3. axial vibrations, CSR, sec^{-1} **A4c3**
      4. torsional vibrations, CSR, sec **A4c4**
   d. 1. exhaust gas noise, CSR, dB **A4d1**
      2. airborne noise, CSR, dB **A4d2**
      3. structure-borne noise excitation, CSR, dB, **A4d3**

table 4

QFD Matrix

scenario #1

rel. importance	user\designer	A1a	A1b	A1c	A1d1	A1d2	A2a1	A2a2	A2b1	A2b2	A2b3	A3a	A3b1	A3b2
2	C1a	9	3	3	3	3	3	0	1	1	1	1	0	0
2	C1b	3	9	3	3	3	3	0	1	1	1	1	0	0
2	C1c	3	3	9	3	3	3	0	1	1	1	1	0	0
2	C1d	3	3	3	9	9	3	0	1	1	1	3	0	0
5	C2a	1	1	1	1	1	9	9	3	3	3	1	0	0
5	C2b	1	1	1	1	1	3	3	3	9	9	1	0	0
4	C3a	1	3	3	3	3	3	0	1	1	1	9	0	0
4	C3b	0	0	0	0	0	1	0	0	0	0	0	9	9
4	C3c	1	1	1	3	3	3	0	1	1	1	3	0	0
4	C3d	0	0	0	0	0	3	0	3	1	1	9	0	0
4	C3e	1	1	1	3	3	3	0	3	1	1	9	3	3
3	C4a	9	1	1	3	3	3	1	1	1	1	3	1	1
3	C4b	9	1	1	3	3	3	1	1	1	1	3	1	1
3	C4c	3	3	3	3	3	3	1	1	1	1	3	1	1
3	C4d	3	3	3	3	3	3	1	1	1	1	3	1	1
	imp. ratings	130	90	90	118	118	172	72	82	96	96	178	60	60

rel. importance	user\designer	A3c1	A3c2	A3d	A3e	A4a	A4b	A4c1	A4c2	A4c3	A4c4	A4d1	A4d2	A4d3
2	C1a	1	0	0	1	9	9	1	1	1	1	1	1	1
2	C1b	0	0	0	1	3	3	3	3	3	3	1	1	1
2	C1c	0	0	0	1	3	3	3	3	3	3	1	1	1
2	C1d	3	0	9	9	9	9	3	3	3	3	1	1	1
5	C2a	0	0	0	0	1	1	1	1	1	1	1	1	1
5	C2b	0	0	0	0	0	0	1	1	1	1	1	1	1
4	C3a	3	1	9	9	3	3	1	1	1	1	1	1	1
4	C3b	0	9	0	9	1	1	0	0	0	0	0	0	0
4	C3c	9	9	3	9	1	1	3	3	3	3	3	3	3
4	C3d	0	1	9	9	3	3	1	1	1	1	1	1	1
4	C3e	3	9	3	9	1	1	1	1	1	1	1	1	1
3	C4a	1	0	3	3	9	9	0	0	0	0	9	0	0
3	C4b	1	0	3	3	9	9	0	0	0	0	0	0	0
3	C4c	1	0	3	3	3	3	9	9	9	9	0	9	9
3	C4d	1	0	3	3	3	3	9	9	9	9	9	9	9
	imp. ratings	80	116	150	240	161	161	108	108	108	108	96	96	96

Table 5 **Normalized Relative Importance Ratings from the QFD scenarios**

	scenario #'s									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Specific Fuel Oil Consumption	0.114	0.145	0.13	0.125	0.122	0.142	0.119	0.124	0.131	0.128
Lube/Cylinder Oil Consumption	0.104	0.111	0.099	0.115	0.114	0.114	0.113	0.112	0.111	0.123
Component Consumption Rate	0.156	0.136	0.16	0.12	0.133	0.119	0.152	0.149	0.124	0.125
Replacement Part Cost	0.132	0.103	0.125	0.119	0.124	0.116	0.125	0.121	0.124	0.116
Maintenance Cost (Overhaul)	0.211	0.176	0.18	0.201	0.191	0.165	0.197	0.192	0.184	0.18
Exhaust Gas Amounts	0.141	0.164	0.154	0.16	0.158	0.172	0.147	0.151	0.163	0.164
NOx Amounts within Exhaust	0.141	0.164	0.154	0.16	0.158	0.172	0.147	0.151	0.163	0.164

	scenario #'s									
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Specific Fuel Oil Consumption	0.119	0.143	0.128	0.119	0.119	0.119	0.119	0.141	0.141	0.141
Lube/Cylinder Oil Consumption	0.108	0.115	0.111	0.108	0.108	0.108	0.108	0.114	0.114	0.114
Component Consumption Rate	0.137	0.12	0.136	0.138	0.138	0.138	0.138	0.119	0.119	0.119
Replacement Part Cost	0.128	0.115	0.122	0.128	0.128	0.128	0.128	0.116	0.116	0.116
Maintenance Cost (Overhaul)	0.199	0.164	0.188	0.199	0.199	0.199	0.199	0.166	0.166	0.166
Exhaust Gas Amounts	0.154	0.172	0.158	0.154	0.154	0.154	0.154	0.172	0.172	0.172
NOx Amounts within Exhaust	0.154	0.172	0.158	0.154	0.154	0.154	0.154	0.172	0.172	0.172

	scenario #'s							
	21	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Specific Fuel Oil Consumption	0.121	0.12	0.121	0.126	0.149	0.153	0.115	0.131
Lube/Cylinder Oil Consumption	0.113	0.112	0.112	0.123	0.102	0.117	0.102	0.108
Component Consumption Rate	0.132	0.132	0.132	0.115	0.13	0.113	0.149	0.151
Replacement Part Cost	0.125	0.125	0.125	0.12	0.103	0.108	0.132	0.125
Maintenance Cost (Overhaul)	0.194	0.194	0.194	0.194	0.178	0.147	0.21	0.174
Exhaust Gas Amounts	0.158	0.158	0.158	0.161	0.169	0.178	0.146	0.156
NOx Amounts within Exhaust	0.158	0.158	0.158	0.161	0.169	0.178	0.146	0.156

Table 6 Criteria for interval scale creation

6a

ATTRIBUTE #1	**Specific Fuel Oil Consumption**	
	(g/BHPh)	RATING
Lowest consumption	119	10 to 9
		8 to 7
	↓	6 to 5
		4 to 3
maximum consumption	126	2 to 1

6b

ATTRIBUTE #2	**Lube / Cylinder Oil Consumption**	
	kg of cylinder oil (lube oil consumption equal)	RATING
lowest consumption	70 kg	10 to 9
		8 to 7
	↓	6 to 5
		4 to 3
maximum consumption	72 kg	2 to 1

6c

ATTRIBUTE #3	**Rate of Component Consumption**	
	# of components consumed	RATING
minimum component loss	resultant expenditures insignificant	10 to 9
		8 to 7
	↓	6 to 5
		4 to 3
maximum component loss	significant resultant expenditures	2 to 1

6d

ATTRIBUTE #4	**Replacement Part Cost**	
	cost (per BHP x 6000hrs.)	RATING
minimum part cost	8 DKK	10 to 9
		8 to 7
	↓	6 to 5
		4 to 3
maximum part cost	11 DKK	2 to 1

criteria for interval scale creation cont'd

6e

ATTRIBUTE #5 **Maintenance Cost (Overhaul)**

	cost (per BHP x 6000hrs.)	RATING
minimum overhaul cost	7.5 DKK	10 to 9
	↓	8 to 7
		6 to 5
		4 to 3
maximum overhaul cost	12 DKK	2 to 1

6f

ATTRIBUTE #6 **Exhaust Gas Flow**

	flow (kg/hr)	RATING
minimum amount	139,000	10 to 9
		8 to 7
	↓	6 to 5
		4 to 3
maximum amount	144,000	2 to 1

6g

ATTRIBUTE #7 **NOx Concentration within exhaust**

	concentration	RATING
minimum concentration	minimal threat to environment and no money needed for fines/reduction of NOx	10 to 9
		8 to 7
		6 to 5
	↓	4 to 3
maximum concentration	concentration is threat to environment and	2 to 1

Table 7 Merit Function Values

SCENARIO #'S

ALTERNATIVES	1	2	3	4	5	6	7	8	9	10	11	12	13	14
5 Cylinder K-90 MC	0.931	0.912	0.924	0.925	0.927	0.915	0.929	0.926	0.921	0.923	0.928	0.915	0.924	0.929
8 Cylinder S-70 MC	0.343	0.342	0.357	0.325	0.338	0.336	0.344	0.344	0.33	0.332	0.335	0.334	0.338	0.337
8 Cylinder S-70 MC with TCS	0.437	0.472	0.474	0.437	0.448	0.46	0.451	0.456	0.448	0.447	0.442	0.463	0.453	0.444

SCENARIO #'S

ALTERNATIVES	15	16	17	18	19	20	21	22	23	24	25	26	27	28
5 Cylinder K-90 MC	0.929	0.93	0.928	0.915	0.916	0.914	0.927	0.927	0.927	0.924	0.911	0.902	0.931	0.922
8 Cylinder S-70 MC	0.337	0.338	0.336	0.333	0.316	0.332	0.334	0.334	0.334	0.321	0.315	0.331	0.341	0.35
8 Cylinder S-70 MC with TCS	0.444	0.445	0.443	0.46	0.449	0.459	0.443	0.442	0.443	0.435	0.449	0.469	0.444	0.468

IN-LINE OIL DEBRIS MONITOR (ODM) FOR THE ADVANCED TACTICAL FIGHTER ENGINE

Dave Muir
GasTOPS Ltd.
1011 Polytek St.
Ottawa, Canada, K1J 9J3

Brad Howe
BFGoodrich Aerospace
Panton Road
Vergennes, VT 05491

Abstract: The development of an in-line, full flow oil debris sensor for the F119-PW-100 Advanced Tactical Fighter Engine is described. The sensor continuously counts and sizes wear metal particles, both ferrous and non-ferrous, in the main oil supply line of the engine. The design requirements, principle of operation, mechanical design features and electronic design features of the sensor are discussed. The performance characteristics of the sensor as measured during development testing are also presented.

Key Words: Advanced Tactical Fighter, F119 engine, F22 aircraft, in-line, oil debris sensor, wear metal detection.

INTRODUCTION: The Pratt and Whitney F119-PW-100 engine which powers the USAF Advanced Tactical Fighter (ATF) aircraft will be the first production military engine to incorporate an in-line, full flow oil debris detection capability in the lubrication system of the main engine bearings. This capability will be provided by an advanced technology Oil Debris Monitor (ODM) sensor system currently being developed by GasTOPS Ltd. for Pratt and Whitney. The ODM will be an integral component of the ATF Engine Monitoring System (EMS), which provides remote indication of the F119 engine sensor signals, comprehensive on-engine diagnostics and ground-based fault isolation using flight recorded data.

The ODM, shown in Figure 1, has successfully completed the pre-flight Engineering Test and Evaluation phase of the F119-PW-100 Full Scale Development Program. Initial Flight Release hardware is presently being provided to Pratt and Whitney for flight qualification testing. First flight of the ODM is planned for some time in 1997.

GasTOPS has also entered into a formal partnership agreement with BFGoodrich Aerospace for ongoing development and production of the ODM for aerospace applications.

This paper discusses the general requirement for in-line, full flow oil debris detection and summarizes the specific design requirements of the F119 ODM. The principles of operation of the ODM are described and a brief description of the sensor mechanical and electronic design features are provided. The results of bearing test rig experiments conducted by GasTOPS and Pratt and Whitney are also presented, demonstrating the capability of the ODM to monitor failure progression and characterize the quantity and nature of wear debris released during failure.



Figure 1 Oil Debris Monitor

REQUIREMENT: The main bearings of an aircraft engine operate under severe conditions of load, speed and temperature. Under these conditions, component damage can progress very rapidly from the point of an initial flaw to catastrophic failure; sometimes in a matter of only a few hours of operation. As well, main bearing failure frequently results in considerable secondary damage to the engine.

Some form of continuous monitoring is therefore required to detect the onset of failure and provide for safe shutdown of the engine. Magnetic chip detectors placed in the return oil flow line(s) from the bearings traditionally have been used for this purpose. These devices, however, rely on the oil-borne wear debris particles to make contact with the sensing element and remain trapped by its magnetic field. In conventional usage, the detection efficiency of a chip detector is quite poor. Moreover, chip detectors are not capable of detecting non-ferrous wear metal particles and are prone to false alarms due to the build up of fine ferrous debris.

The F119-PW-100 engine design provides for main engine bearing protection through the use of an inductive oil debris sensor which mounts directly onto the outlet of the main oil supply pump and provides continuous, in-line, full-flow wear metal debris detection. Typical performance requirements of the sensor include:

Debris Detection Size:	125 microns minimum
Debris Detection Efficiency:	100%
Debris Type:	ferrous and non-ferrous

The sensor must also meet the rigorous environmental requirements of an on-engine sensor including oil temperature (over 350°F), oil pressure (over 300 psia), entrained air (50% worst case), vibration and EMI requirements.

The output of the F119 ODM sensor will be fully integrated into the Engine Monitoring System (EMS) of the ATF. The EMS includes the Full Authority Digital Engine Control (FADEC) unit of each engine, a Comprehensive Engine Diagnostic Unit (CEDU) and the various sensors used for engine control and diagnostics. Processing of the ODM sensor signals is performed by the CEDU using electronic circuitry provided by GasTOPS.

SENSOR OPERATION: The ODM is a through flow device which installs in the main oil supply line and allows the entire oil flow to pass without obstruction. The sensor incorporates a magnetic coil assembly which is capable of detecting and categorizing wear metal particles by size and type (ferrous and conducting non-ferrous). The minimum detectable particle sizes are determined primarily by the inner diameter of the coil assembly. For the F119 application, these minimum sizes are approximately 125 microns for ferrous particles and 250 microns for non-ferrous particles.

The magnetic coil assembly consists of three coils which surround a magnetically inert section of tubing. The two outside field coils are driven by a high frequency alternating current source such that their respective fields are nominally opposed or cancel each other at a point inside the tube and just under the center sense coil. Disturbance of the magnetic fields caused by the passage of a particle results in a characteristic sense coil voltage as shown in Figure 2. The amplitude and phase of the output signature is used to identify the size and type of particle. The amplitude of the signal is proportional to the mass of the particle for ferrous materials and to the surface area of the particle for non-ferrous materials. The phase of the signal for non-ferrous particles is opposite to that of ferrous particles, allowing a distinction to be made between the two types of wear metal materials.

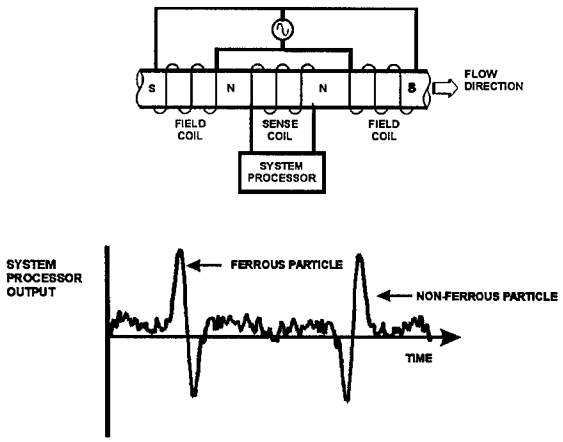


Figure 2 ODM System Operation

Signal conditioning using a threshold algorithm categorizes the particles that pass on the basis of size and type. Several size categories can be defined above the minimum particle size thresholds. Once a particle has been detected and its size determined, this information is stored in registers or bins which record the total number of particles of a given type and bin size that have been detected.

MECHANICAL DESIGN: The F119 ODM is installed on the housing of the lube oil pump. The oil enters the sensor through a 3 bolt flange connection and exits (approximately 8 inches away and rotated by 90°) through a piloted o-ring port.

To accommodate this installation the ODM was designed as a 2 piece unit including a sensor module (Figure 3) and elbow component. The sensor module contains the working elements of the sensor in a welded stainless steel housing designed to withstand the operating environment of the engine and to contain the oil under flame exposure. The unit is designed to minimize exposure of the coils to distortion due to external forces, vibration and oil pressure fluctuations. The elbow component of the sensor serves as an adapter to the oil pump. It accommodates the 90° rotated flange and allows for flexibility between the two components to accommodate mis-alignment between the sensor and pump. A spring, combined with the force of the oil pressure, keeps the sensor module held in place in the o-ring port during operation. A flexible ground strap and capture mechanism is used to provide the necessary electrical bonding to the engine structure and keep the two sensor components together during installation and removal.

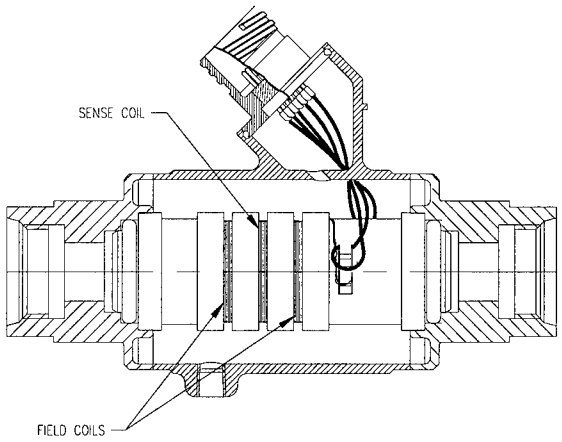


Figure 3 ODM Sensor Module

ELECTRONIC DESIGN: The ODM sensor signals are processed by electronics located in the Comprehensive Engine Diagnostic Unit (CEDU). The design that is currently undergoing Initial Flight Release testing includes two semi-custom integrated circuits which perform the analog and digital ODM processing functions. The analog functions are performed by an Application-Specific Integrated Circuit (ASIC) designed by GasTOPS and built by Harris Semiconductor. The ASIC includes electronics to amplify and filter the sensor signals, remove the residual sensor imbalance signal, and extract and detect particle signatures. The digital circuit, a Field-Programmable Gate Array (FPGA), processes the signature outputs, determines the type of particle detected (ferrous/non-ferrous), records particle counts, and interfaces with the CEDU microprocessor. The two integrated circuits also work together to provide comprehensive Built-In Test (BIT) functions which provide both fault detection and isolation.

The electronics design is being improved for the Initial Service Release version of the F119 ODM. Some of the particle detection functions are being moved into a Digital Signal Processor already present in the CEDU. This change will both simplify the hardware and provide enhanced particle sizing capability.

BEARING TEST PROGRAM: In order to establish reliable criterion for warning, alarm or shutdown conditions based on the ODM output readings, a test program has been undertaken to investigate the failure of aircraft engine bearings. The goal of this program is to quantify the debris released from bearings during failure and to evaluate the capability of the ODM to monitor failure progression. The program involves the running of bearings past the point of normally accepted failure and monitoring the debris released from the bearings as the failures progress.

Small Scale Bearing Tests: A large number of steel (i.e. ferrous) bearings (over 40 in total) have been run to failure in a test rig specially designed and instrumented for small scale (2 inch diameter) ball and roller bearings tests, as shown in Figure 4. The rig incorporates a fine mesh screen which captures the debris released during each failure for subsequent comparison to the ODM readings.

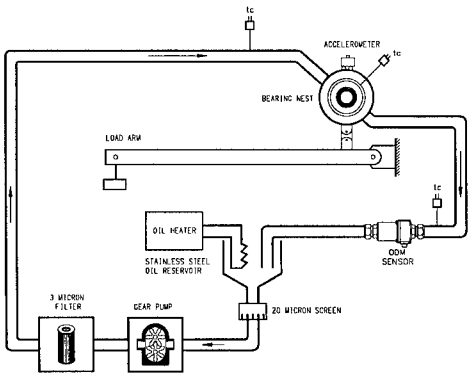


Figure 4 Bearing Test Rig

Detailed results of these tests have been presented in an earlier publication [1]. As illstrated in Figures 5 and 6, both the sensor measurements and the findings of the filter debris analysis verify that large numbers of wear metal particles within the detectable size range of the sensor are released; starting with the first spall, continuing as the bearing is kept in operation and increasing in rate as damage reaches an advanced state.

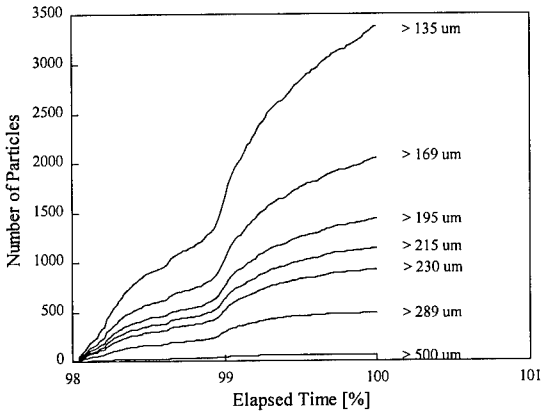


Figure 5 Sensor Output During Bearing Test

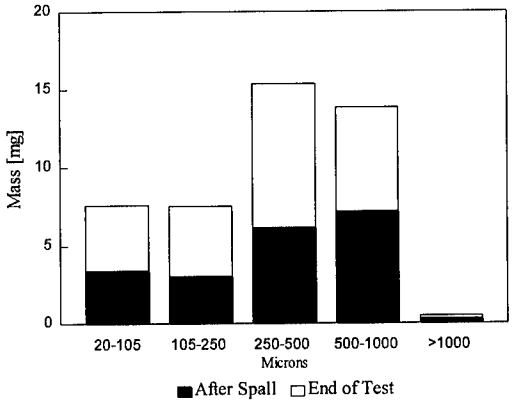


Figure 6 Filter Backwash Results

It has been established from these tests that the severity of the damage can be correlated to the mass of the wear metal material removed from the bearing contact surfaces, as shown in Figure 7. This suggests that reliable caution and alarm thresholds can be established based on the cumulative mass of debris measured by the ODM.

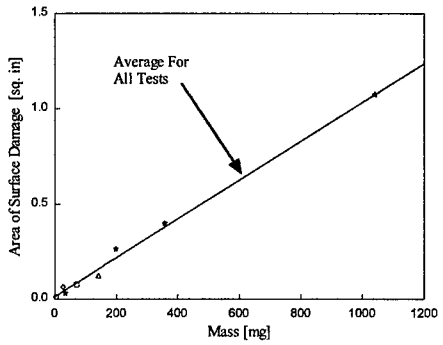


Figure 7 Correlation Between Sensor Output (Mass) and Observed Surface Damage

A single test has been conducted by Pratt and Whitney on a large diameter thrust bearing from an aircraft engine application. The bearing material was steel and failure was initiated by intentionally damaging each of the functional surfaces of the bearing (inner race, outer race and balls) using 0.025 inch diameter indents. The rig was operated at maximum bearing load conditions until vibration levels reached approximately 50 g's, at which time the rig was shut down and the bearing inspected. Inspection of the bearing following the test revealed extensive inner race spalling, outer race distress, ball spalling and flaking and cage fracture.

Figure 8 shows the number of particles generated above each of the ferrous particle thresholds used for the test. As indicated, the sensor detected the initial spall and counted significant numbers of wear metal particles greater than 200 microns throughout the failure. In total, over 150,000 ferrous particles greater than 200 microns and more than 25,000 particles larger than 700 microns were counted. Also of note was the rapid increase in the rate of particles counted during the latter portion of the test.

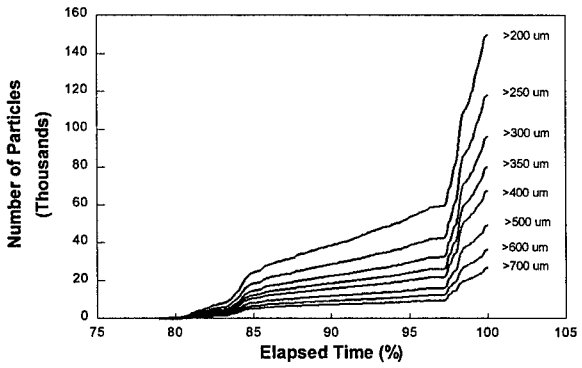


Figure 8 Large Scale Bearing Test Results

Hybrid Bearing Test: A single test has also been conducted on an advanced technology hybrid bearing for aircraft engine application. Failure of this bearing, which incorporated ceramic balls and steel races, was initiated by a series of 0.02 inch diameter stress concentration pits created on the inner race of the bearing. The bearing was run progressively from the point of initial spall to a point at which severe spalling damage was evident over roughly 10% of the race circumference.

Figure 9 shows the number of ferrous particles generated above each of the size thresholds used for the test. Once again, over the course of the test a loop number of wear metal particles were counted by the sensor (over 7,000 particles greater than 200 microns and over 100 particles greater than 700 microns). As in the case of the large scale bearing test, a significant increase in the rate of particles counted during the latter stages of the failure was evident.

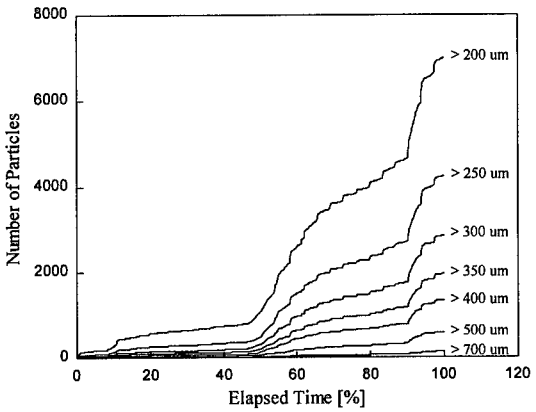


Figure 9 Hybrid Bearing Test Results

SUMMARY: An in-line full flow Oil Debris Monitor has been developed for the F119-PW-100 Advanced Tactical Fighter Engine and is presently undergoing flight qualification testing. The ODM is based on an advanced inductive sensing technology which is capable of detecting both ferrous and non-ferrous wear metal particles under full flow conditions. The ODM is fully integrated into the design of the F119 engine. The mechanical elements of the system are incorporated into the design of the engine's oil supply pump and the signal processing circuitry is fully integrated into the engine's Comprehensive Engine Diagnostic Unit. Bearing failure progression testing conducted by GasTOPS and Pratt and Whitney has verified that the ODM is capable of detecting the initiation of bearing damage and monitoring the progression of this damage towards failure.

REFERENCE:

[1] I. Hughes and D. Muir,
"On-Line Oil Debris Monitor For Aircraft Engines",
JOAP Conference, November, 1994.

MACHINERY DIAGNOSTICS & PROGNOSTICS #2

APPLICATION OF AN ENVELOPE TECHNIQUE IN THE DETECTION OF BALL BEARING DEFECTS IN A LABORATORY EXPERIMENT

Henk Toersen

Eindhoven University of Technology
Faculty of Mechanical Engineering
Post Office Box 513
5600 MB Eindhoven, The Netherlands.

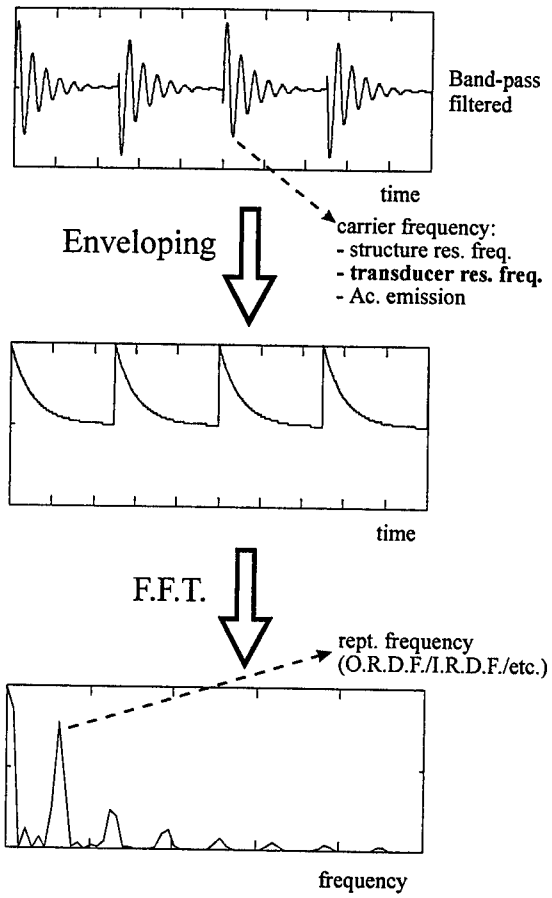
Abstract: To judge the diagnostic capabilities of a envelope technique (Bearcon Signature™; Carl Schenck AG, Germany) artificially damaged ball-bearings (SKF 6207) were run under oil-lubrication at different radial loads and speeds in a laboratory experiment. Respectively outer race, inner race and ball damages of different sizes were introduced by spark erosion and the response of the measuring system was analysed. The results showed, that vibration spectrum components of outer race defects showed up in the envelope spectra of undamaged bearings and that detectability of defects was primarily limited by the speed of the bearing, sometimes the automatic scaling facility of the analyser was a limiting factor. Inner race and ball defects showed the well known effects of load modulation in the defect-contact zône, which can become dominant in the envelope spectra at high loads and speeds. This phenomenon could be confirmed by computersimulation. The detection of ball-damage was hampered by the fact, that overrolling of the defect in purely radially loaded bearings takes place incidentally. This draw back could be mitigated by applying a trigger-technique to the envelope signal.

Key Words: Ball bearings; bearing damage; detection limits; diagnostics; envelope technique; triggering; vibration analysis.

INTRODUCTION: The widespread application of R.E. bearings and the fact that the failure modes of these components often show considerable lead times have in the past intiated many successful efforts to develop diagnostic equipment for R.E. bearing damage. Many surveys of the literature on this subject are given of which [1] is an example.
In recent years much emphasis is given to the application of envelope techniques to the analysis of R.E. bearing vibration signals.[2-6]. It offers the opportunity to reduce interference of other vibration sources in the machinery under surveillance with the initially often relatively weak vibration signals, originating from bearing defects. Moreover it enables us to diagnose in more detail the location of the defect in terms of outer ring, inner ring or rolling element by means of the defect overrolling frequencies.

For the maintenance engineer the latter is not all that important as long as he knows which bearing in a train is suffering. However defect overrolling frequencies can be very specific for a bearing and hence enable us localise the bearing that shows an impending failure.
In routine monitoring practice it is important to understand the detection limits of the monitoring equipment in order to be in a better position to estimate the rest-lifetime of a damaged bearing.
For this reason the experiments described in this paper have been set up, obviously for a very limited set of operating conditions and with the application of Carl Schenck's Vibroport 41 vibration analyser with an enveloping facilty (Bearcon-Signature™). As the experiments have been run in a laboratory environment, it can be doubted whether the results can be improved in industrial practice.

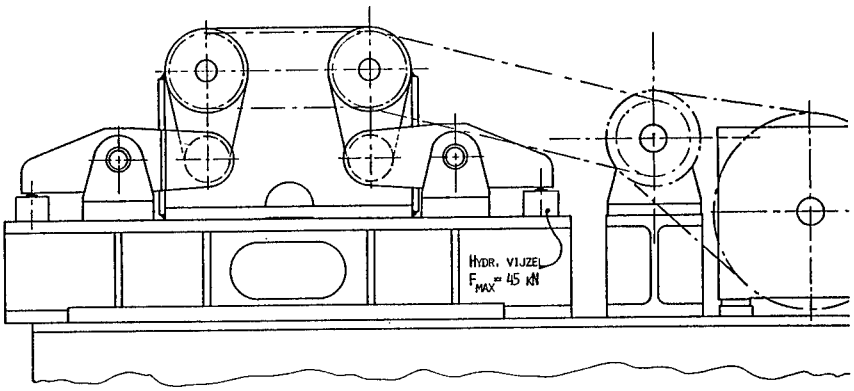
THE "ENVELOPE TECHNIQUE": The overrolling of a defect in a race of a ball bearing will give rise to impulsive action on the bearing itself, on the surrounding mechanical structure and on a vibration pick-up in case this pick-up is properly connected to the structure. The impulsive action may result in resonancies in these three elements, or may become the origin of elastic stress-waves that are emitted into the structure as acoustic emission. Due to damping the resonances and acoustic emission will die out, until a next impuls takes place.
By making use of the resonance frequencies or the frequency of the acoustic emission as a carrier frequency, the repetion rate of the impulses can be shown. To do so, it is important to eliminate other, very often stronger vibration components in the signal caused by e.g. unbalance, gear mesh, mis-alignment etc., which normally will be found in the lower frequency regions. This can be done by proper band-pass filtering. The result is shown in the upper part of fig.1.
To determine the repetition frequency of the impulses the band-pass filtered signal is rectified and subsequently the envelope of the rectified signal in the time domain is produced. (see fig.1 - middle).
The next step is a Fast Fourier Transform of the envelope, which will result in a number of spectrum components in the frequency domain. The frequency components will represent the repetition frequency of a bearing defect overrolling and harmonics of this frequency. (see fig.1 - under)
Sometimes, due to modulation of the maximum amplitudes of the band-pass filtered signals, the modulation frequency itself and side-bands of the repetition frequency will appear at distances equal to the modulation frequency.
When geometry and speed of the bearing are known, defect overrolling frequencies can be determined for outer ring defects (ORDF), inner ring defects (IRDF) and rolling element defects (REDF) by means of bearing supplier resources, e.g. the SKF-Atlas computer program.
Thus comparing the envelope spectra with the calculated frequency values, enables us to diagnose the bearing defect.



As is already mentioned, different carrier frequencies are being used. The Bearcon Signature™ measurement utilises the resonance frequency of the vibration pick-up and the band pass filter is set from 13 kHz to 65 kHz.

Fig.1 Three important steps of signal processing in the envelope analysis.

THE EXPERIMENTAL SET-UP: The experiments are carried out on a test-rig, which is schematically shown in fig.2. It is extensively described in [5] and [7].
The bearing under test is a deep-groove ball bearing (SKF 6207), which is radially loaded in a static way bij means of a hydraulic actuator. Originally the set-up provides for 4 test-bearing positions of which only one is used for reasons of reproducability.
The bearings are lubricated with circulating oil (Shell Tellus 32) and kept at a constant temperature of approx. 40 °C.
Bearing defects were artificially introduced by spark-eroding a cylindrical hole through the inner and outer ring in the



Sideview

Fig.2 Test rig; schematically.

center of the races. A blind hole was eroded into a ball straight through the side of the metal cage.
The diameter of the holes varied from 0.2 mm to 2.0 mm for the outer rings and from 0.4 to 2.0 mm for the inner rings and the balls. The size of the defect needs to be judged in relation with the ball diameter, which is 11 mm, and with the load dependent contact area.
Two loads were chosen, viz. 0.5 and 5.0 kN. The dynamic load ratio of the bearing was 25.5 kN and the static load ratio 15.3 kN, hence we used a relatively low and high load on the bearings.The characteristic defect overrollingfrequencies for the bearings, related to one damage in a bearing, are given in table I for the three speeds, used in the experiments.

Table I. Characteristic defect-frequencies of the bearings at speeds used in the experiments.

Inner ring speed (r.p.m.)	500	1500	4500
Inner ring speed (Hz)	8.33	25.00	75.00
Outer ring defect freq. (Hz)	29.71	89.14	267.41
Inner ring defect freq. (Hz)	45.29	135.86	407.59
Rolling element defect freq. (Hz)	38.40	115.20	345.59
Rolling element rot. speed (Hz)	19.20	57.60	172.79
Cage rot. speed (Hz)	3.30	9.90	29.71

Bearing SKF 6207; Contact angle 0.00°; Source SKF-Atlas-1989

Before the damages were introduced zero-measurements were taken on the same, undamaged bearings under identical conditions as compared with the damaged bearings. All measurements were taken in duplo, which meant that after the first measurement the bearing was dismounted and re-mounted for the duplo-measurement, taking care that the outer ring was mounted and fixed in the same angular position as before. The outer ring damage was always located in the center of the loaded zône.
Before measurements were taken the test-bearing was run in after mounting and re-mounting during 2 hours at a speed of 1500 r.p.m and a load of 5.0 kN. Measurements were taken with the Vibroport 41 of Carl Schenck AG and with an acceleration pick-up type AS-20, which was screwed in a position directly above the loaded zône of the bearing.

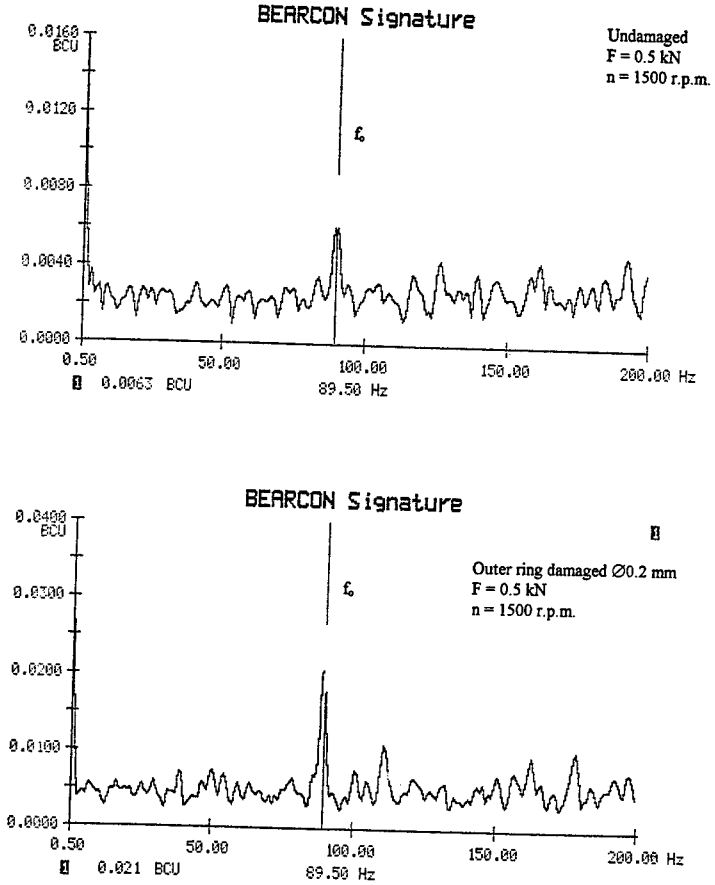


Fig.3 Envelope spectra of an undamaged and a slightly damaged (outer ring damage) bearing at similar operational conditions

RESULTS:

The Zero-Measurements: Obviously these measurements were taken as reference for the measurements on the damaged bearings. Sometimes we observed, particularly with the higher speeds, weak spectrum-components in the envelope spectrum (Bearcon Signature™), which normally are indicative of outer ring damage. Fig.3 shows two envelope spectra of an undamaged bearing (upper) and of a slightly damaged bearing (lower) for the same operational conditions.
Although the signal level of the outer ring defect frequency f_o of the slightly damaged bearing is somewhat larger than the f_o of the undamaged bearing, it is quite clear, that it is difficult to distinguish between the two and the indicated damage can hardly be diagnosed.
It appeared that this phenomenon was more dominant when geometrical discontinuities were present in the loaded zône of the bearing housing.

Bearings With Outer Ring Defects: Independent of the applied load on the bearing and of the size of the damage (up to 2 mm diam.) it was hardly possible to obtain clear damage indications at a speed of 500 r.p.m. At higher speeds the influence of the bearing load was small; in some cases an increase in damage indication was observed at the lower loads (0.5 kN). With higher speeds the influence of increasing damage diameter shows up clearly, which can be illustrated by comparing fig.3 (under, damage diam. 0.2 mm) with fig.4 (damage diam. 2.0 mm). The number of higher order frequency components (harmonics) in the latter case is striking.

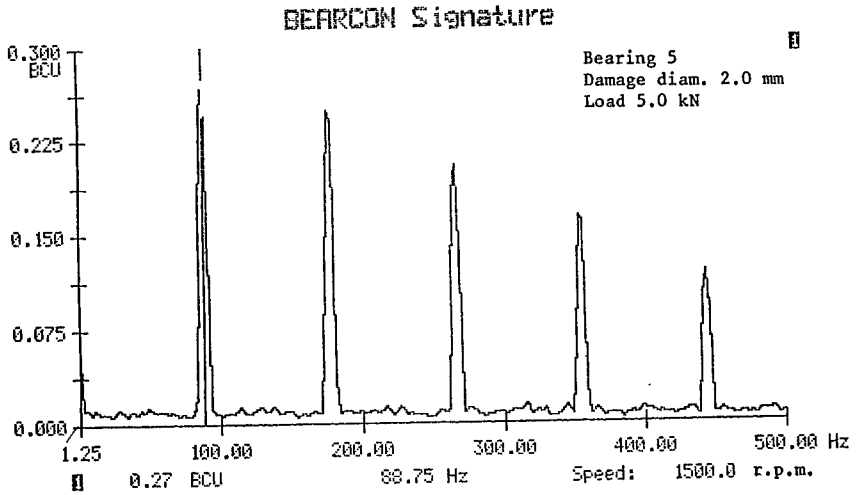


Fig.4 Envelope spectrum for outer ring defect diam. 2.0 mm

Plotting the levels of the outer ring defect frequency component f_o against damage diameter (fig.5) the influence of the increasing damage diameter is shown again. In fig.5 the results of the duplo-measurements are also shown as well as the results of the zero-measurements.

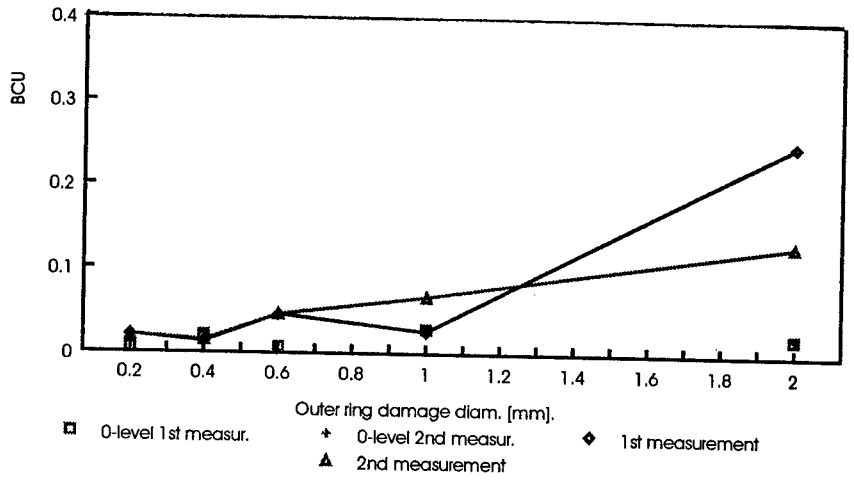


Fig.5 Vibration level of the f_o-component versus damage diameter at a load of 0.5 kN and a speed of 1500 r.p.m.

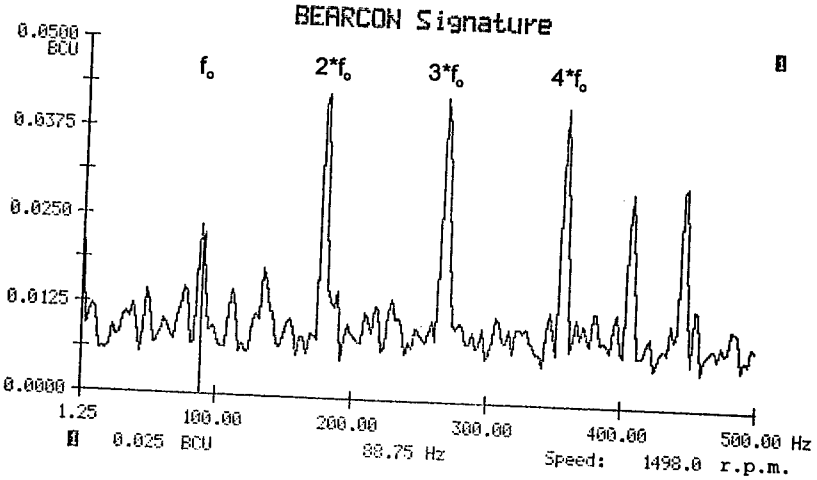


Fig.6 Envelope spectrum of an outer ring damage of 0.4 mm diam. at a load of 0.5 kN and a speed of 1500 r.p.m.

One would be inclined to conclude from fig.5 that it is not possible to detect outer ring damages of 0.4 mm diam. and smaller at a speed of 1500 r.p.m. However it appeared that sometimes harmonics of f_o had a much higher signal level, as is illustrated in fig.6, enabling clear detection of the damage. Summarising it can be concluded that the detection limit with regard to the outer ring damage lies somewhere about 0.3 mm damage diam. under the given conditions, provided the speed of the bearing is not to low (over approx. 300 r.p.m.). (See also part on ball damage)

Bearings With Inner Ring Defects: The results of the experiments with the outer ring damages prompted to choose 0.4 mm as the smallest damage diameter for the inner ring damage, as weaker signals from inner ring damages were to be expected. Here again it was difficult to determine inner ring damages at speeds lower than 500 r.p.m. irrespective of the size of the damage (up to 2.0 mm diam.).(See also part on ball damage) Apart from the spectrum components that represent the inner ring defect frequency (f_i) and its harmonics, frequency components show up that correlate with the rotating frequency (f_n) of the bearing and side-bands of the f_i appear at a distance of f_n (see fig.7)

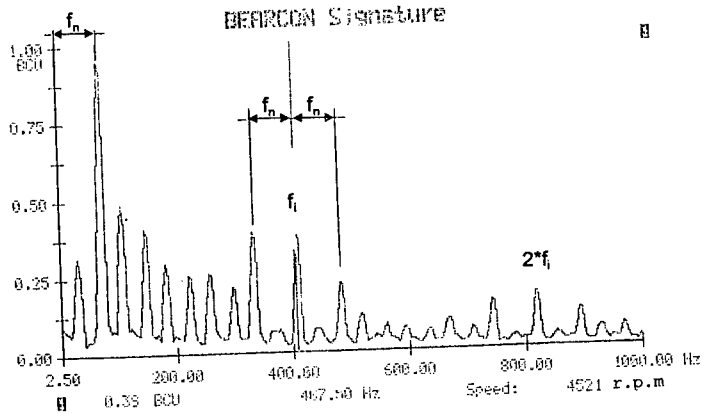


Fig.7 Envelope spectrum for an inner ring defect of 0.7 mm diam. at a load of 5 kN and a speed of 4500 r.p.m.

The f_n-component may completely dominate the envelope spectrum, particularly with the higher loads and speeds. This can be explained by the fact that the excitation level of the carrier frequency will be modulated by the rotational frequency f_n as the defect overrolling passes through the loaded zône into the unloaded zône of a statically loaded bearing, with the rotating speed of the inner ring. Hence the effect will be stronger with higher bearing clearances and higher bearing loads.
The phenomenon can be easily simulated on a PC with e.g. a

Matlab program. The result of such a simulation is given in fig. 8. It shows in the upper part the excitation modulation of the amplitude of the carrier frequency, the corresponding envelope of the signal in the middle part and the resulting envelope spectrum in the lower part.

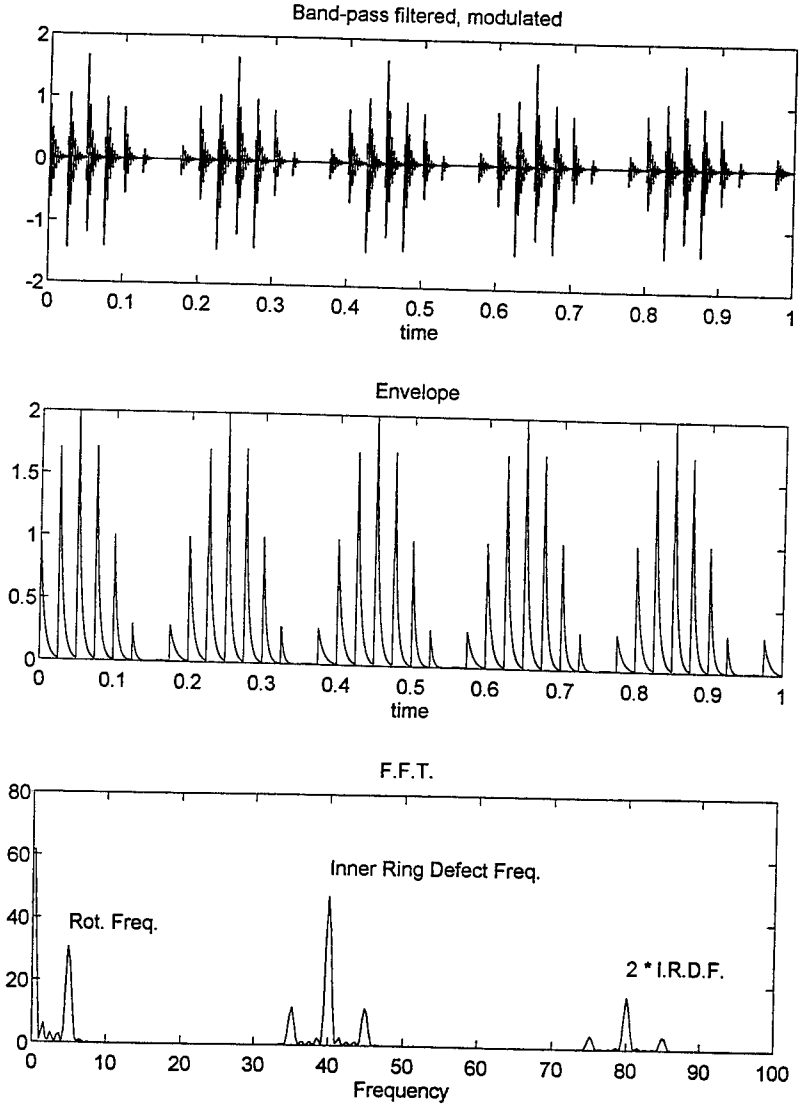


Fig.8 Simulation of an envelope spectrum with amplitude modulation of the carrier frequency with a Matlab program.

The inner ring defect frequency and its harmonic are flanked by side-bands at rotating frequency (f_n) distance and the f_n-component itself shows up clearly. The effect will be stronger with stronger modulation of the maximum amplitude of the carrier frequency, which was confirmed both in the simulation and in the real experiments. Similar effects can be expected with outer ring damage in a bearing supporting a shaft with high out of balance loading, where again modulation takes place with the frequency f_n, and with ball damage in a statically loaded bearing due to modulation with the cage rotation frequency (See part dealing with ball damage).
Returning to the detection limits it can be summarised from the experiments that:
- at the lower speed no detection of an inner ring defect was possible, irrespective of the size of the damage (up to 2.0mm).(See remarks in the part on ball damage)
- at a speed of 1500 r.p.m. inner ring damages of 0.7 mm diam. and larger could be detected and at a speed of 4500 r.p.m. damage with a diam. of 0.4 mm could easily be seen.
- sometimes the rotational frequency is a better damage indicator than the inner ring defect frequency. This holds particularly for the higher loads.

Bearings With Ball Defects: In a purely radially loaded ball-bearing with a ball defect it is a matter of coincidence when the defect will be overrolled and therefore the measuring technique had to be adapted to capture such an event. To do so the transducer signal was band-pass filtered and enveloped in a special module before feeding it to the analyser, which was set in the FFT-mode with a trigger level of 20% of f.s.d. and a short pre-trigger time.
It was observed, that the frequency of occurrance of defect-overrolling could be increased by applying light fluctuating axial loads on the bearing.
By doing so envelope spectra of bearings with a ball damage could be captured in a single shot and fig.9 gives an example of such a spectrum. The ball-defect frequency (f_b = 38 Hz) and four harmonics show up clearly and they are all flanked by two side-bands at a distance of the cage rotating frequency f_c. Moreover the f_c itself shows up very well. The whole picture shows a striking analogy with the simulated spectrum in fig.8 and with fig.7, the envelope spectrum for a bearing with inner ring damage. This obviously is not surprising as in the case af a ball defect in a statically loaded bearing the rotation of the cage, by pushing the damaged ball from the unloaded into the loaded zône, modulates the amplitude of the impacts that occur, when a defect is overrolled.
Again it could be shown experimentally that this modulating effect increases with higher loads and speeds and then very often the f_c-component in the envelope spectrum is the better indicator of a ball damage.
A striking difference with the spectra obtained with outer ring and inner ring damage is the far better sensitivity of the method used with the ball damage. Here a ball damage with

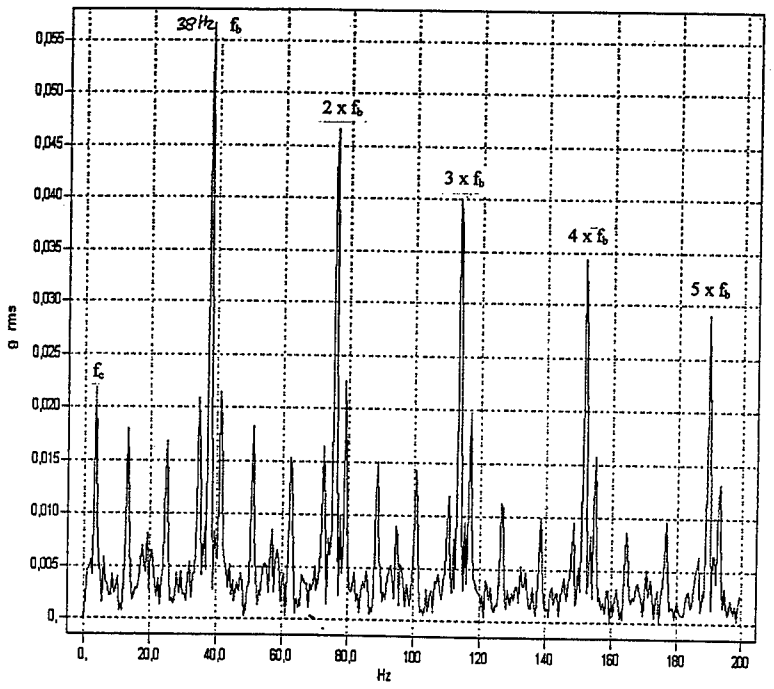


Fig.9 Triggered single shot envelope spectrum for a ball damage diam. 0.7 mm, load 0.5 kN and speed 500 r.p.m.

a diam. 0.7 mm and in a bearing running at 500 r.p.m. could be detected very clearly (fig.9), which was not possible with the outer ring and inner ring defects.
The explanation of this phenomenon could be, that the standard set-up of the Bearcon Signature™ measurement employs automatic scaling of the envelope spectra and in case of a relatively high starting peak in the spectrum (see e.g. fig.3) combined with a low f.s.d.-value, this starting peak determines the sensitivity. This was actually not the case with the single shot, triggered spectra.
It indicates, that the potential of the envelope technique with respect to detection limits, is better than reported here for the outer ring and inner ring damages.

CONCLUSIONS:

1. Signals of outer ring damage in undamaged bearings limit the detection of very small outer ring damage.

2. There is a low sensitivity of damage detection at low speeds. This is probably caused by the automatic scaling facility of the analyser.

3. Detection limits are hardly load dependent. Sometimes better results are obtained with the lower loads.

4. In a purely radially loaded ball bearing ball defects might be overrolled incidentally. This is a serious handicap for routine monitoring these defects. This handicap can be mitigated by applying a triggering technique.

5. Due to amplitude modulation the inner ring and cage rotational frequencies can be better indicators for respectively inner ring and ball defects than the defect overrolling frequencies.

REFERENCES:

[1] Khan, A.F. - Condition monitoring of rolling element bearings: a comparative study of vibration based techniques.
Ph.D.- thesis submitted to the University of Nottingham (UK), May 1991.

[2] Ratcliffe, G.A. - Condition monitoring of rolling element bearings using the envelope technique.
Proc. Inst. Mech. Eng. Seminar on condition monitoring; Jan.9, 1990.

[3] McMahon, S.W. - Condition monitoring of bearings using envelope signal processing.
Condition monitoring and diagnostic technology, Vol.2, nr 1, July 1991.

[4] Scheithe, Dr. Rer. Nat. W. - Schwingungsmessung, ein Werkzeug zur Früherkennung von Wälzlagerschäden.
Schenck Schwingungspraxis 13.

[5] Smith, H.C., Toersen, H. - Opsporing van wentellagerdefecten voor toestandsafhankelijk onderhoud.
Mechanische Technologie, Jaargang 2, 6/7 1992.

[6] Berryman, F., Michie, P., Smulders. A., Vermeiren, K. - Condition Monitoring, a new approach. A method of monitoring machines using a high frequency acoustic emission technique. S.K.F.-publication.

[7] Toersen, H. - Enkele resultaten met conditiebewaking van wentellagers.
In situ contrôle van wentellagers; studiedag Polytechnisch Tijdschrift; 29.01.92.

SYSTEMS ENGINEERING OF AN ON-LINE DIAGNOSTIC SYSTEM

Jeffrey S. Lin
The Johns Hopkins University Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20723-6099

Abstract: On-line diagnostic systems primarily use data acquisition systems and expert systems to implement the diagnostic reasoning of a human expert. But many technologies must be integrated for the development of a successful system. These technologies include:

expert systems, data acquisition, and sensor validation,
data communications,
human/computer interfaces (HCI) and uncertainty management, and
system integration testing.

The success of the system will be diminished if any of these technologies, or tests of their implementation or integration, are not planned for from the beginning of the project. This paper reviews the lessons learned during the development and testing of the Air Compressor Diagnostic System.

Key Words: expert systems; monitoring; on-line diagnostics; sensor validation; systems integration; uncertainty.

INTRODUCTION: The Johns Hopkins University Applied Physics Laboratory (JHU/APL) recently completed and delivered the Air Compressor Diagnostic System (ACDS) to the U.S. Navy for the Integrated Diagnostics Demonstration (IDD). ACDS is an on-line monitoring and diagnostic system for reciprocating high- and low-pressure air compressors (HPAC and LPAC) [1].

The primary goal of automated diagnostic systems like IDD is to reduce machinery life-cycle costs by minimizing the costs of maintenance and repair operations. The savings result primarily from reducing the manpower needed to perform required tasks. A timely and accurate diagnosis of machinery condition prevents cascade failures from causing further equipment damage, and assures that the correct problem is addressed by the repair crew. A carefully constructed human/computer interface (HCI) helps insure that the diagnosis is understood and accepted by the user.

Other benefits can be expected from the application of automated diagnostics. If failing machines are identified more quickly and repaired, the operational availability and efficiency of the equipment will increase, improving shipboard operations and reducing energy costs. In addition, logistical tasks can be automated as part of the response to the detected failure.

The HPAC and LPAC versions of ACDS are two of several subsystems designed to diagnose various machines on a DD-963 class ship as part of the IDD. The subsystems communicate their diagnostic results to the IDD Core System, which integrates the results into a common interface for the user. The subsystems must first identify failing or failed machinery components, and then provide the user with the information necessary to respond to the failure.

The successful development of ACDS demonstrates the importance of systems engineering. The ultimate goal of systems engineering should be the satisfaction of the needs of the user. With this in mind, ACDS was developed to provide the U.S. Navy with a usable and useful tool for performing air compressor diagnostics.

DIAGNOSTIC METHODOLOGY: The data-acquisition and expert-system modules of a diagnostic system provide the core of the system functionality. Data acquisition is a mature technology and many good commercial-off-the-shelf (COTS) systems can be found for a new application. An aspect of data acquisition that is frequently ignored, however, is the validation of the sensor data [1,2]. Indeed, two diagnostic subsystems are generally required for each diagnostic system: one for the target machine, and one for the diagnostic system hardware itself. These both can be implemented in the same expert-system software.

An entire spectrum of COTS expert systems is available that offers widely varying capabilities. In general, the designer of an on-line diagnostic system should select a flexible, fully-capable expert system. Such an expert system includes a knowledge-representation capability, mathematical computation capability, and programming interfaces to external programs. Neuron Data's NEXPERT OBJECT™ package, which provides these capabilities, was used to develop the ACDS knowledge base.

Many software designers new to on-line diagnostics assume that diagnostic rules are simply an if-then construct that can be implemented in C or Pascal. This may be true for small diagnostic systems. As the number of diagnostic rules and the complexity of the diagnostic system increase, however, the more the software engineer will appreciate the functionality of a well developed knowledge-base-development environment.

The object-oriented nature of NEXPERT OBJECT assists greatly in the encapsulation of the representation of the world and the diagnostic rules. For instance, the ACDS knowledge base has a class called *component*. The class *component* has several attributes: faulted, severity, and confidence. Every air-compressor component that is diagnosed by ACDS is represented by an object of this class. Each of these objects inherits the attributes of the class *component*. Therefore the *oil_pump* object has several attributes. *oil_pump.faulted* is a Boolean hypothesis that can be TRUE, FALSE, UNKNOWN or NOTKNOWN. *oil_pump.severity* indicates a

warning or alarm level of the condition of the oil pump. *oil_pump.confidence* holds the confidence factor, low, moderate, or high, that *oil_pump.faulted* is TRUE.

Not all expert systems provide an UNKNOWN or NOTKNOWN value to a hypothesis. The distinction between these values is subtle, but important. A hypothesis has a value of UNKNOWN if the rule has not yet been evaluated. A hypothesis has a value of NOTKNOWN if the rule has been evaluated, but insufficient information exists to make a judgment. Examples of their use may help show their usefulness and difference.

ACDS loads only a subset of the rules for a given diagnosis, depending on how long the machine has been on or off. The value of the hypotheses of the rules not loaded are UNKNOWN. The following scenario illustrates the importance of this value.

Suppose that ACDS finds that the oil pump is faulted because the oil temperature is too high. Then, the air compressor, and with it the oil pump, happens to cycle off. The oil temperature will drop when the machine is off for a while. If the oil-temperature rule were loaded and evaluated at this time, the hypothesis that the oil pump is faulted would evaluate to FALSE, indicating that the pump is healthy. The pump was faulted, however, when the air compressor was on, so it is most likely still faulted when the machine turns off. To improve consistency, ACDS does not load this rule when the machine is off. The hypothesis that the oil pump is faulted will be UNKNOWN. The best guess for the health of the oil pump is the last known verified hypothesis, namely that the oil pump is faulted.

The NOTKNOWN value of a hypothesis is used when ACDS detects a sensor failure. If, during the sensor validation routines, a sensor value is found to be completely unreliable, its value is set to NOTKNOWN. The hypothesis of any rule that uses this sensor value then also evaluates to NOTKNOWN. ACDS informs the user of the sensor failure, and uses the last known verified hypothesis for those rules that are affected.

The object-oriented nature of NEXPERT OBJECT, through meta-slots, also allows actions to be defined when a specific data value changes. ACDS uses this feature extensively to keep data consistent within the knowledge base. All of the air-compressor components are grouped into air-compressor subsystems in a hierarchical fashion. For instance, in the HPAC ACDS, the 3rd stage inlet valve, discharge valve, relief valve, and rings comprise the 3rd stage subsystem. The five stage subsystems form the air subsystem. The air, lube/drive motor, cooling water, condensate drain monitor, and temperature monitor subsystems are parts of the overall HPAC system. This hierarchy is duplicated graphically at the HCI. The health of each system and subsystem depends on the health of the constituent subsystems at the lower level. ACDS updates the health of each subsystem through the meta-slots provided with the knowledge base.

Diagnostic-system designers who select weak expert system tools severely limit their ability to provide a robust, flexible, and useful product. Not all required functionality or features of an expert system may be known at the beginning of the development cycle. It can be cheaper to buy a fully-featured tool, and use only 30% of the capabilities than require 130% of the capabilities

of a less expensive tool. Procedural computer code can be written to compensate for deficiencies in the expert system, but that code must be written, tested, and maintained.

DATA COMMUNICATIONS: Unless the on-line diagnostic system resides with the target machine, some form of data communication will be necessary. Machinery environments may accumulate several diagnostic systems, using a variety of hardware and software that are tied together into a central computer. ACDS, for example, communicates data and diagnostic results to the IDD Core System for display to the user. The successful communication of information requires detailed specifications and early and frequent testing.

The specifications must include more than a hardware interface and a communications protocol. JHU/APL worked closely with the IDD Core System developer to define the timing, format, and content of the communications between ACDS and the Core System. Several tests of increasing complexity assured the successful communication. The first tests verified the hardware interface and communications protocol. Later tests verified the correct timing and successful communication of data. Most communication tests had actually been performed before the final message content and format were fixed.

One important lesson learned was to test every statement in the specification document at the earliest possible time. Some tests were forgotten at early tests for ACDS that required non-trivial software changes during later tests. Fortunately, we were able to make and test the changes without any delay in the schedule.

HUMAN/COMPUTER INTERFACE: An effective HCI for a diagnostic system must perform several tasks. First, all HCIs viewed by a given operator must be consistent. Second, the HCI must filter the information displayed to the user, allowing quick, clear communication of significant data or instructions. Additionally, an effective uncertainty-management technique must be used to qualify the information presented, so that the user understands the quality and reliability of the information.

The requirement for a consistent HCI can be met by either one common interface for all diagnostic systems, or a well defined specification and strict enforcement of that specification. Otherwise, the operator must be trained to use a new interface with each addition of a diagnostic subsystem. Little tricks or features that one diagnostic subsystem supplies may appear useful at first, but will eventually confuse the operator when it is not available in other subsystems. Apple Macintosh™ users long have enjoyed and extolled a consistent user interface, whereas users of IBM-compatible personal computers have only recently enjoyed the standard interface supplied by Microsoft Windows™.

The IDD imposed one common interface, which was supplied by the Core System, for all diagnostic subsystems. The result is transparent switching between diagnostic subsystems at the HCI. The user can learn to investigate a failure in one subsystem, and immediately transfer this knowledge to another subsystem.

JHU/APL worked with the Core System developer to define the user interface for the IDD. JHU/APL used cognitive models and task analysis to develop a user-centered approach to define the user-interface requirements for diagnostic systems [3,4].

The HCI requirements for an expert diagnostic-system such as ACDS are numerous. Air compressors are mechanically complex and, as a result, the diagnostic algorithms are sophisticated. An appropriate level of abstraction and representation of the system information is required. In addition, an HCI should provide:

an explanation and justification for a diagnosis and recommended action,
guidance for conducting additional tests and querying the system,
the ability, appropriately restricted, to alter decision limits of diagnostic rules, and
methods for obtaining access to diagnostic and repair procedures and documentation.

The HCI design must support the users, which in the case of ACDS are sailors with a wide range of knowledge and experience. JHU/APL built a prototype user interface with a hierarchical structure allowing multiple levels of interaction at the IDD user interface. Surface-level interactions allow the quick communication of problems and suggested remedies. Deeper-level interactions allow detailed investigation into the data and diagnostic algorithms.

Given the complexity of the air-compressor systems and the ACDS diagnostic algorithms, JHU/APL developed a model of diagnostic uncertainty [1] to assist the user in interpreting ACDS's diagnoses. The uncertainty in a diagnosis includes uncertainty associated with: the validity of the sensor data, the decision threshold limits, the efficacy of the rules, and confirmation or lack of confirmation by additional rules. The combination of these uncertainties is transparent to the user. The HCI only displays the resultant low, moderate, or high confidence in the final diagnosis. Additional resolution or detail in the level of confidence will only serve to distract or confuse the user. The HCI provides all data and explanations at the deeper-level interactions, for the user to further investigate and form his/her own opinions.

JHU/APL recommended an HCI implementation and test plan, that included prototype walk-throughs, user feedback, and usability studies. Although funding limitations prevented the plan from being executed for IDD, the execution of such a plan will improve and validate the effectiveness of any diagnostic system HCI.

INTEGRATION TESTING: Lastly, system-integration testing is a vital part of the development of all complex systems, including on-line diagnostic systems. System integration was required at two levels for the IDD. At the subsystem level, the data acquisition, diagnostic, and communication hardware and software systems were integrated and tested. At the IDD system level, each subsystem had to be integrated with the Core System and fully tested. For ACDS, these two levels of integration testing were performed concurrently.

The first ACDS functionality tested was the communications. Simulated data files were transmitted from one computer to another using the communication protocol specified by the Core-System developer. These tests were followed by tests with a Core-System simulator

written by the Core-System developer. Several tests were performed, each exercising successively larger portions of the communications capability. All test procedures performed at one integration test with the Core System were repeated during later tests, to assure that no functionality was lost.

Concurrently, the HPAC diagnostic knowledge-base was tested against existing data taken from a fault-insertion test program. This data does not exhaustively cover the faults that ACDS diagnoses, and had to be supplemented with artificially-generated data. The Dresser-Rand Company, which supplied the diagnostic expertise for ACDS, used simulation codes to generate the additional data. No existing fault-insertion test data is available for the LPAC, therefore Dresser-Rand Company supplied similar artificial data to test all of the rules in the LPAC ACDS. The sensor-validation rules also were exhaustively tested by simulated data.

Once the knowledge base was verified in the NEXPERT OBJECT development environment, it was integrated with the ACDS main program, which already was proven to communicate with the Core System. The main program is written in C and accesses the knowledge base through an application programmers interface with C function calls. Simulated sensor-data was stored in data files, read in by the C program and passed to the knowledge base. The results of the knowledge-base execution were then read back into the C main program, and transmitted to another computer. Once this capability was demonstrated at JHU/APL, we again tested the integration of ACDS with the Core System.

The next step was to perform the data acquisition to obtain real-time sensor data. The data is acquired through C function calls to a library of routines supplied by the vendor of the data-acquisition hardware. First, the data acquired was verified independent of the rest of ACDS. Next, ACDS sent the data to the Core System, where it was verified as accurate. Due to cost constraints, realistic sensor-values could not be simultaneously simulated across all data acquisition channels inputs. We could not, therefore, control the diagnostic output of ACDS. Since the arbitrary data of open data-acquisition channels violate many sensor validation checks, only sensor failures are detected. Therefore, for subsequent tests with the Core System, all data that was acquired was overwritten by data from test files before being sent to the knowledge base.

The full diagnostic functionality of ACDS was demonstrated during an integration test with the Core System. Concurrently, the full shipboard-ready hardware/software integration of ACDS was verified independent of the Core System.

The last level of testing before shipboard installation was originally planned to be a land-based test. All of the subsystems were to be installed on the targeted machinery and connected to the Core System. The land-based test would have served as an excellent validation of IDD. The cost estimates for the land-based test, however, soon became prohibitive.

To replace the land-based tests, an IDD simulation/stimulation system was planned to demonstrate sensor-to-HCI connectivity. A simulation/stimulation test would verify each subsystem functionality, and the successful integration of the subsystems and the Core System.

Unfortunately, program funding limitations prevented the completion of this final level of integration testing.

CONCLUSIONS: Developing a successful on-line diagnostic system requires early attention to the performance requirements as well as the testing requirements. Software tools should be selected to allow flexibility in the implementation of the knowledge base. Integration, verification, and validation tests are time consuming and expensive but necessary. In an operational environment where many separate diagnostic systems must coexist, these test are imperative. Without these tests, the user cannot be insured of getting a system that solves the right problem, works with installed systems, and presents the right information to the user in an understandable, consistent format.

ACKNOWLEDGMENTS: The author wishes to acknowledge Mr. Augie DiGiovanni and Mr. Mike Good of the Carderock Division of the Naval Surface Warfare Center for their support and assistance with the many technical issues in this work. The author also acknowledges Mr. Scott Delmotte of the Dresser-Rand Company for providing his expertise in compressor diagnostics.

REFERENCES:

[1] J. S. Lin, S. Delmotte, "On-Line Diagnostics of Reciprocating Multi-Stage Air Compressors," Advanced Materials and Process Technology for Mechanical Failure Prevention, 48th Meeting of the MFPG, pp. 369-378, 1994.

[2] J. S. Lin, C. L. Resch, J. V. Palmer, "Sensor Validation for On-Line Diagnostics," Proceedings of the 40th International Instrumentation Symposium, pp. 519-528, May 1994.

[3] B. G. Coury, J. S. Lin, F. B. Weiskopf, "System Representations, Cognitive Modeling and the Design of User Interfaces for an Expert Diagnostic System," Proceedings of the IEEE/SMC'93 Conference, pp. 26-31, October 1993.

[4] B. G. Coury, R. D. Semmel, F. B. Weiskopf, J. Jackson, "Attention Direction in Uncertainty Management," in M. Mouloua and R. Parasuraman (Eds.), Human Performance in Automated Systems: Current Research and Trends, (Hillsdale, NJ: Lawrence Erlbaum), pp. 270-276, 1994.

TURBINE ENGINE DIAGNOSTICS USING A PARALLEL SIGNAL PROCESSOR

Theodore A. Bapty, Akos Ledeczi, James R. Davis
Measurement and Control Systems Laboratory
Department of Electrical and Computer Engineering
Vanderbilt University
Nashville, TN 37325

Ben Abbott
Department of Electrical and Computer Engineering
Utah State University
Logan UT

Terry Hayes, Thomas Tibbals
Sverdup Technology Inc., AEDC Group
Arnold AFB, TN

Abstract: For the past several years, the Measurement and Computing Systems Laboratory has been working in close cooperation with the United States Air Force at Arnold Engineering Development Center (AEDC), Arnold AFB, to develop techniques for large-scale instrumentation systems. In-depth, on-line analysis of test data from turbine engine testing is critical to ensuring an accurate, timely evaluation and diagnosis of engine performance. Given the complexity of the analysis algorithms and the quantity of data, the computations overrun the capability of the fastest supercomputers.

This paper describes the development of Computer Assisted Dynamic Data Monitoring and Acquisition System (CADDMAS). The CADDMAS is a 48 channel, 50 KHz full-time analysis system, capable of flexible analysis of signals in the time and frequency domains. Data is presented on real-time displays, showing, for example, spectrums, Campbell Diagrams, engine-order tracking. The system (both hardware and software) is synthesized using a novel model-based technique. The approach has been used to generate several systems used for on-line military and commercial turbine engine data analysis at Arnold AFB and for analysis of the SSME for NASA. On-line analysis has had a significant impact on turbine engine testing, reducing the time necessary to meet testing objectives and improving the quality of testing results. Substantial savings have been demonstrated by allowing immediate access to reduced data.

Keywords: Instrumentation, Parallel Processing, Turbine Engine Diagnostics, Turbomachinery, Model-Based Systems, High-Performance Computing, Digital Signal Processing, Real-Time Systems.

INTRODUCTION: For the past several years, the Measurement and Computing Systems Laboratory has been working in close cooperation with the United States Air Force at Arnold Engineering Development Center (AEDC), Arnold AFB, to develop techniques for large-scale instrumentation systems. AEDC is the premier aerospace ground testing facility in the United States. Propulsion testing is of particular importance. Propulsion test facilities allow aircraft engines to be tested at a range of altitudes, mach numbers, temperatures, etc. The entire flight envelope of an engine can be simulated.

A tremendous quantity of data is generated during turbine engine testing. Stress testing exemplifies this problem, where hundreds of sensors must be analyzed over frequencies up to tens of kilohertz.

In-depth, on-line analysis of this data is critical to ensuring an accurate, timely evaluation and diagnosis of engine performance. Given the complexity of the analysis algorithms and the quantity of data, the computations overrun the capability of the fastest supercomputers. Sustained rates of several billion floating point operations per second are required. This computation is in addition to the demands of acquiring and maintaining the incoming data. The solution to this problem is to apply parallel processing. Off-the-shelf hardware is available, allowing construction of massively parallel machines. Unfortunately, the software technology lags behind.

This paper describes the development of Computer Assisted Dynamic Data Monitoring and Acquisition System (CADDMAS). The CADDMAS is a 48 channel, 50KHz full-time analysis system, capable of flexible analysis of signals in the time and frequency domains. Data is presented on real-time displays, showing, for example, spectrums, Campbell Diagrams, engine-order tracking. The flexible architecture allows arbitrary analysis to be added.

DESIGN FACTORS: Construction of large scale instrumentation systems is complex. The factors that drive this complexity are as follows:

1. Many sensors are necessary to effectively instrument a large system. High measurement bandwidths with large numbers of sensors generate a significant computational load. As a result, large-scale parallel processing systems are required to supply the necessary computational resources.
2. Instrumentation systems have real-time specifications, both hard and soft. Building real-time systems, especially using parallel hardware, is a complex task.
3. System requirements change rapidly in many applications. The system performance must scale gracefully over a wide range. Functionality must meet changing requirements. The systems must be upgraded in-place, with minimal down-time. This hardware and software scalability incurs a cost in complexity.
4. System implementation cost is also a driving factor. The increase in cost as a system is scaled-up must be modest. To meet this goal, the computational platform is typically a

low-cost parallel DSP platform. Software development and maintenance costs must be minimized as well.

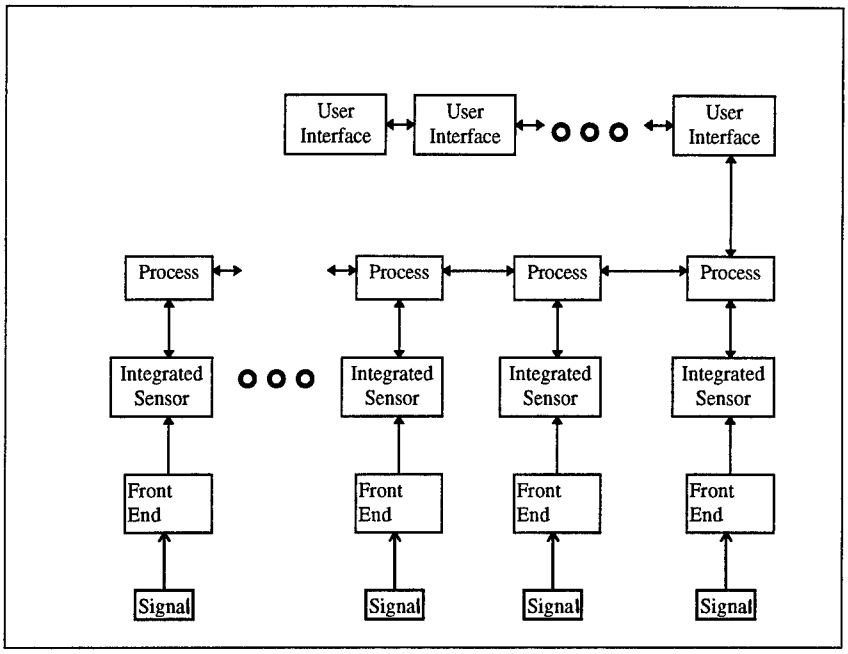
5. The users of instrumentation systems are typically not instrumentation engineers. Interaction with the system must mask the underlying complexity, since the domain users are unlikely to have experience in complex, real-time, reactive, large-scale, scalable, parallel systems. The user is concerned with specifying system requirements and with using the resultant system. Interaction should be in terms of familiar domain concepts.

These factors must be addressed if a successful system is to be constructed. The design strategy must encompass requirements specification, software design, and hardware design in a real-time environment. Cost performance requirements also specify that the target system is a parallel processor.

TARGET DIAGNOSTIC SYSTEM HARDWARE ARCHITECTURE: The systems are constructed with an application specific network of Texas Instruments TMS320C31/40 Signal Processors and/or Inmos transputers. A large number of different systems have been constructed to meet specific processing requirements. The largest system is built using 48 TMS320C31's, 12 TMS320C40's, and 48 Motorola 56002's. The TMS320C31 is a standard Digital Signal Processor, augmented with a primitive communication facility. The TMS320C40 is a newer DSP with integrated communication hardware. The TMS family processors are capable of 40 MFLOPS performance and do the bulk of the analysis, using single precision floating point arithmetic. The Motorola processors operate on integer data and serve as a data acquisition front-end to the system.

The processors are connected in a modified pipeline topology, closely matching the information flow within the application. Figure 1 shows the general architecture.

- The User Interfaces are Intel 486 or Pentium based PC's. They serve to display information, print hardcopies, and control the processing options of the system.
- The "Process" nodes accumulate information for Campbell Diagrams, perform Bicoherence computations, and route information throughout the system.
- The "Integrated Sensor" nodes perform the low-level operations: FFT, Magnitude Spectrum, Spectral Peaks, Engineering Unit Conversion, etc.



MODEL-BASED DESIGN METHODOLOGY: The system (both hardware and software) is synthesized using a novel model-based technique. Using this technique, the engineer building the system represents the goals of the system in terms of **Models** - such as signal flow graphs and processing requirements. These models are interpreted to automatically generate the wiring diagrams of the hardware and the executable software files.

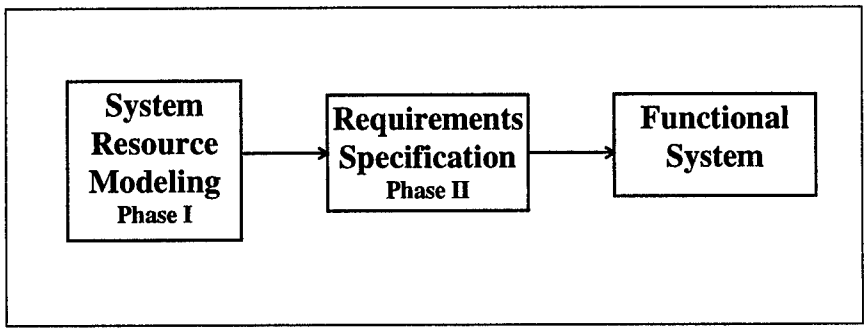
In order to manage the complexity of the effort while satisfying the flexibility and remaining within cost, a careful, structured approach must be used. We are approaching the class of parallel instrumentation systems using multiple-aspect modeling. The overall system architecture for parallel real-time instrumentation is shown in Figure 2. The basic thrust in the model-based approach is to model the system at appropriate levels, in a domain-specific manner. The domain specific nature allows us to represent the system design in concepts that are familiar to the end-user.

The information capture process is a combination of graphical editing and/or textual menu-based forms where appropriate. These models are interpreted various ways to generate the executable systems and other tools. The proper definition of the paradigms is critical to the success of the approach. In this section we describe the paradigms developed for the domain of parallel real-time instrumentation systems.

The organization of the modeling paradigm is driven by the structure of the entire testing process. This includes the process of specifying and building the systems as well as the users involved. The user interactions and processes of building instrumentation systems can be broken into three phases. In the first phase, the range of system capabilities and resources are defined by the **Instrumentation System Engineer**. The phase I paradigm is called **System Resource Modeling**, where parameterized models specify the available processing & output classes along with implementation structure and parameter options. In addition, the available hardware resources and system inputs are defined.

In the second phase of system construction, **Requirements Specification**, the **Domain Engineer** uses editors automatically generated from phase I to specify requirements for the desired instrumentation system implementation. The characteristics of the system are selected here: inputs, outputs (i.e. plots, alarms, databases), system structure, timing requirements, etc. Note that the domain engineer is not necessarily trained in signal processing, computer architecture, or parallel processing.

The third set of users of the system are the operators, who interact with the automatically generated system resulting from the interpretation of the phase II requirements specification and phase I System Resource Models. The user interface is a window into the capabilities specified and built into the system by the domain engineer, allowing interactive application of these capabilities.



System Resource Modeling Paradigm: The system resources can be broken into two groups: the algorithmic and software implementations & structure and the physical system components. The conceptual computations are defined in **the Signal Flow Aspect**, while the hardware components used in constructing the computational platform, as well as the connections to the external environment is defined in the Physical Resource Aspect.

Signal Flow Aspect: The signal processing engineer uses the Signal Flow Modeling aspect to define the mechanisms for generating the user plots. The processing is described using a signal flow graph representation, commonly used to describe many digital signal processing algorithms. The primary concepts employed in this paradigm are as follows.

- Processing Nodes, representing the elemental large grain signal processing operators, with attributes, specifying run-time behavior, resource requirements (execution time),
- input/output characteristics, etc., information necessary for process allocation and scheduling.
- Connections represent the flow of information between processing nodes.
- Timing Constraints indicate the critical time path in component computations.
- Local Parameters allow control of the behavior of the processing nodes from the high-level builder.

Hierarchy and iterators are used to manage the complexity of large processing structures and to manage highly regular computational structures.

These concepts are extended further to allow the packaging of various signal flow graphs into definitions of user plots. The actual implementation of the processing structure used to generate a specific plot may vary depending on the requirements. Consider, for example, spectral analysis. Many methods exist for computing spectra. The best algorithm depends on the goals. Multiple implementations can be specified, along with the methodology for selection of the proper option and configuration of the internal processing. We therefore add the following concepts:

- The **System Output Class Definition** consists of a grouping of signal flow graphs implementing similar functions. These represent the design alternatives available when generating system outputs. The implementations may differ in computational expense, resolution, accuracy, or time response.
- **Implementation Option Selection Criteria** represent the alternate methods for implementing a function. Criteria are specified for choosing the best implementation given a set of specifications. This information allows the builder to choose between, for instance, an FFT or a filter bank implementation of a spectrum computation. The selection criteria is specified as an algorithm, a function of the set of configuration variables. These configuration variables are specified in this model.
- **Implementation Configuration Methods** allow the builder to tune the implementation to match specifications. For example, in an FFT application, the FFT block size, window function, and overlap ratio must be specified. The implementation configuration is also specified as an algorithm, since this process is specific to the signal processing
- method.

Graphical modeling is a natural paradigm for this system aspect. Signal Flow Graphs have been used historically to design and conceptualize signal processing systems.

Physical Resource Aspect: Hardware Architecture models for defining the set of DSP's, Transputers, A/D converters, etc. used in the physical construction, allowing the user to specify where signals enter the system. **Processor Nodes** correspond to a self contained processor, memory, I/O, communications, and special resources. The node attributes to be specified include *processor type, processor speed(MIPS/MFLOPS), memory size,*

communication bandwidth, etc. **Connections** define the physical connections between node. Combined with the processor nodes, this defines the system topology. **Signals** are the raw data, the sensors digitized into the system.

Hierarchies of these nodes are used to define preconfigured nodes and connections: sub-assemblies (boards), chassis, racks, and systems. A graphical model is appropriate in this paradigm, since it is natural to draw pictures of system architectures and network topologies.

System Requirements Modeling Paradigm: At the top level, the user of the system interacts only with familiar concepts, directly reflecting the user's view of the system. These concepts include: **Inputs:** The signals available for processing. **Outputs:** The output processing requirements of the system. These include *Plots*, *Result Files*, and *Alarms*. For each of these outputs, the user specifies relevant information, such as type, accuracy's, range, update rates, maximum latencies.

From this view, the user can specify the capabilities and performance of the system that is required for a particular test. The complexities of parallel software, signal processing, and computer hardware are invisible at this level. This is required, since the end-user of the system is an aerospace engineer, with little concern or training on these issues.

MODEL INTERPRETATION: The model interpretation is a multiple phase operation. In the first phase, the signal processing expert models the available classes of system structures and outputs, and the instrumentation engineer defines the signal inputs. These models are interpreted to generate the System Requirements Definition Model Editor. This editor is used by the end-user (the domain engineer) to model/specify the system requirements. The second phase of interpretation takes these requirements and generates an architecture-independent specification for the software structure of the system. The final phase integrates the architecture-independent specification and the hardware model to generate a hardware/software design and implements an executable system. Under typical operation, the first phase will be used infrequently, whenever new features/plots are added to the system. The most common design loop will use phase II and III to generate systems of various sizes, configurations, and hardware platforms.

CONCLUSION: The approach has been used to generate several systems used for on-line military and commercial turbine engine data analysis at Arnold AFB and for analysis of the SSME for NASA. On-line analysis has had a significant impact on turbine engine testing, reducing the time necessary to meet testing objectives and improving the quality of testing results. Substantial savings have been demonstrated by allowing immediate access to reduced data.

BIBLIOGRAPHY:

Bapty, T.A., "Model-Based Synthesis of Parallel Real-Time Systems", PhD Dissertation, Vanderbilt University, 1995.

Abbott, B, Bapty T. Biegl, C. Karsai, G., Sztipanovits, J. ``Model-Based Approach for Software Synthesis", IEEE Software, pp. 42-53, May, 1993

Karsai, G. ``A Visual Programming Environment for Domain-Specific Model-based Progrmming.", IEEE Computer, May 1995.

Ledeczi, A., Abbott B.A., ``Parallel Systems with Flexible Topology", Proc. of the Scalable High Perfomace Computing Conference, pp271-276, Knoxville, TN 1994.

Sztipanovits, J., Abbott, B., Bapty, T., Misra, A., "Model-Based Synthesis of Complex Embedded Systems", Proc. of the 1994 Complex Systems Engineering Synthesis and Assessment Technology Workshop, Washington D.C. July 19-20, 1994

Bapty, T.A., Abbott, B.A., Biegl, C. Ledeczi, A. "Parallel Turbine Engine Instrumentation System", Proc. of the 9th AIAA Conference on Computing in Aerospace, San Diego, CA, October, 1993

LUBRICANT CONDITION MONITORING #1

ALTERNATE TECHNIQUES FOR WEAR METAL ANALYSIS

Costandy S. Saba and Hoover A Smith

University of Dayton Research Institute
Dayton, Ohio 45469-0166

Abstract: Wear metal measurements were performed on various type lubricant samples obtained from operating engines and laboratory prepared samples using AE, AA, ICP, graphite furnace AA, ferrography and a particle size independent method. Also, the effect of 3-micron filtration on the analytical capability of the various methods was investigated. Considering the data from all samples, all spectrometric techniques were Fe particle size sensitive. None of the spectrometers offered any significant improvement over AE with respect to analyzing large particles, monitoring capability with and without 3-micron filtration, analysis time or analysis cost or person-power. The data also indicated that 3-micron filtration could have small effect on spectrometric oil analysis results.

Key Words: Ferrography, Fine filtration, Oil analysis, Particle size, Spectrometers

Introduction: Spectrometric oil analysis programs for determining wear metals in used aircraft turbine engine lubricants have been used for the last three decades for detecting those engines experiencing abnormal wear and the removal of those engines from service prior to catastrophic failure. The program requires lubricant samples to be periodically taken from engines and analyzed in a laboratory for various wear metal concentrations. Abnormal operating engines are identified by the level and/or rate of change in specific wear metal concentrations. Many reports and papers have been published on the success (or failure) of the various monitoring techniques and programs. Many different methods and techniques have been developed and used for monitoring the wear metals in used lubricants depending on such factors as the type equipment being monitored, monitoring organization, equipment usage, etc.

For many years oil filters were used having nominal filtration capabilities of 35 to 50 microns and in some cases much greater than 50 microns. In recent years "finer" filtration has been investigated for the purpose of inhibiting secondary wear caused by the primary or initial wear particles and external contamination. Currently, finer filtration is being considered for use in aircraft turbine engine lubrication systems. These "fine" filters have the potential of greatly reducing the metal content of SOAP samples and current techniques for oil analyses may prove unsatisfactory for use in monitoring turbine engines equipped with these filters.

The objective of this investigation was to conduct an evaluation and comparative analysis of currently used AE and AA techniques with various wear metal analysis techniques such as inductively coupled plasma (ICP) spectrometry, graphite furnace atomic absorption (PWMA), ferrography and particle size

distribution using the acid dissolution method (ADM). The effect of 3-micron filtration on the analysis capability of the various methods was investigated using a test rig equipped with a 3-micron operational "in-depth" type oil filter and using parameters of pressures, temperatures and flow rates typical of operating turbine engines.

Microfiltraion Test Rig: A detailed description of the microfiltration test rig (MFR) has been previously reported [1] and only a brief description of the test rig and filter will be given in this paper. The test rig consists mainly of a 5 gallon capacity conical bottom oil reservoir, a constant speed gear pump for oil circulation and a 3/4 inch stainless steel oil circulation system incorporating a turbine flow meter, in-line thermocouples and pressure transducers and a pressure relief valve. A 7 gallon seamless stainless steel container is used for collecting the fluid after passing through the filter. A small scavenger gear pump is used for transferring the filtered fluid back into the oil reservoir for subsequent passes through the filter. The test fluid can be circulated in a by-pass mode prior to filtering for obtaining an uniform (mixed) sample. An "upstream" filter sample can be taken either from the oil reservoir or sampling port while the filtered sample is obtained immediately after filtering from the 7 gallon collection container using a precleaned vacuum flask. The filter elements (3 micron absolute, β_3 >200) were operational type "in-depth" elements capable of withstanding temperatures in the range of -65°F to 350°F. The pressure drop across the filter at a rated flow of 4 GPM is 4 psi at 100°F and has an element collapse differential pressure of 100 psi.

Wear Metal Analysis Techniques: Two different atomic emission (AE) spectrometers were used for determining the trace metal concentrations of the various samples. Both instruments employ a rotating disk lower electrode, no sample dilution, AC spark excitation, multi-element simultaneous analysis, data reporting and each using SOAP oil standards R-19 for instrument calibration. Normal SOAP procedures were used during the operation of both spectrometers.

Trace element concentrations of the samples were made using a single element mode atomic absorption (AA) spectrophotometer and only iron determinations were conducted on this instrument since iron was the most prevalent metal present in most of the samples. These analyses were made using a 1 part sample to 4 parts methlyisobutyl ketone (MIBK) dilution, nitrous oxide-acetylene flame, and 1:4 diluted SOAP oil standards for instrument calibration.

Inductively coupled plasma (ICP) analyses were made using 1 part sample and 9 parts kerosene dilution, spray nebulizer using argon gas, multi-element simultaneous analysis and diluted SOAP standards for calibration.

The portable wear metal analyzer (PWMA) is a graphite furnace atomic absorption spectrophotometer. It is a microprocessor controlled automatic sequential multielement instrument that will analyze for nine elements (Fe, Cu, Al, Cr, Ag, Mg, Ni, Si and Ti) using electrothermal element atomization. Analyses made with the PWMA required no dilution. Regular oil analysis standards in MIL-L-7808 lubricant were used for calibration.

The acid dissolution method (ADM) has been previously reported in detail [2] and only a brief summary of the method will be given in this paper. The appropriate amount of sample is combined with a HNO_3/HCL (1:3) acid mixture and hand shaken for 10 seconds. The mixture is then agitated in an ultrasonic bath for 5 minutes at 40°C (65°C if Mo analysis is required). The mixture is then diluted

with a long chain alkoxy alcohol/MIBK (or kerosene) solvent and analyzed by AA or ICP for selected wear metals.

Particle size distribution of iron wear debris was determined using a microfiltration technique. Aliquots of the sample were filtered through 12-, 8-, 5-, 3-, 2-, 1- and 0.4-micrometer membrane filters. The filtrate was then analyzed for iron by the ADM using an AA spectrophotometer.

Ferrography was used for magnetic separation and collection of wear debris (primarily iron) from lubricating fluids for the subsequent evaluation of the debris with respect to the amount and morphology (particle size, shape, source or type wear, etc.) of the debris [1]. All ferrograph analyses referenced in this report were conducted on the Analytical Ferrograph which involves depositing the debris onto a glass slide and subsequent microscopic evaluation as to particle morphology and densitometer measurements. The densitometer measurements provide a relative concentration of the various size particles deposited down the slide from which the ratio of large (L) to small (S) particles can be calculated.

Test Lubricants: The various type lubricant samples used in this investigation were obtained from operational engines, laboratory prepared samples using new lubricant blended with commercially purchased metallic powders and with new lubricant blended with wear debris generated by a pin-on-disk wear test rig. Some of the operational engine samples were obtained specifically for this program. In all cases, the samples were newly shaken and sonicated prior to analysis. Many of the samples consisted of such small volumes that only limited analyses could be conducted. In other cases, two or three samples were combined for providing sufficient sample for microfiltration rig studies.

Results and Discussion: Samples used to evaluate the analytical capabilities of the ICP, A/E JA, PWMA, AA, ADM, and ferrography and to perform particle size distribution analyses were heated and sonicated prior to analysis. A/E35U-3 emission analyses were conducted by SOAP laboratories using normal procedures. In general the data obtained by the various analysis techniques have been tabulated and arranged in order from high to low for initial iron content and percent iron loss. This approach provided for determining the relative rankings of the various analysis techniques to ADM values with respect to total iron content and iron loss due to three micron filtration. Correlation between all analysis techniques including ferrography were made only for MFR samples since no ferrography measurements were conducted on the SOAP samples.

Previous research [3] has shown that determined iron concentrations of ester base oils can be up to 2.5 times the actual iron concentration when determined using the A/E 35U-3 spectrometer calibrated with mineral oil base standards. This same research also showed decreasing emission sensitivities for particles above approximately eight microns. The data obtained during this study have shown similar matrix effects and changing of particle size sensitivity when using atomic emission analysis techniques. As shown in Table 1 samples having small iron particle sizes (less than 3 microns such as sample MFR-8) have much higher atomic emission values than corresponding ADM values. For samples having very large iron particles such as sample MFR-4 the ADM values are much higher than the corresponding emission values. Particle size distribution on MFR samples confirms the above findings. The original iron content and iron loss due to three micron test rig filtering (Table 1) show a wide variation between the various analysis techniques for both the original iron content and percent loss. As expected, the spectrometers' values diverge greatly from the ADM values at higher concentrations

where large particles exist. Samples having low initial iron concentrations can have as large of variations between percent loss analyses as high iron content samples due to the presence of large metallic particulates. Particle size can also have a greater effect on percent iron loss due to filtering than on initial iron concentration.

Table 2 gives a summary of rankings based on the highest iron content determined in the MFR samples by the analysis techniques. For example ADM analyses ranked in the highest position (6) two of twelve analyses, in fifth position three of the twelve analyses, etc. These data show that ADM, PWMA, A/E35 and A/EJA rank 3 and above 83% of the time while ICP, and AA ranks 2 and below 81% of the time. Table 3 gives similar type data for percent iron loss due to filtering. Again ADM, PWMA, A/E35 and A/EJA ranked 3 and above 73% of the time and ICP and AA ranked 2 and below 70% of the time. Table 4 gives the rankings for each analysis technique for MFR samples not filtered. In this case ADM, A/EJA, A/E35 and PWMA ranked 4 and above 72% of the time while ICP and AA ranked 3 and below 94% of the time.

Figure 1 represents all original iron test data using six different type analyses. Since the ADM analyses are particle size independent, various values obtained using the other analysis techniques are plotted against the appropriate ADM values. There is much data scattering at higher concentration but nevertheless the deviation from ADM values is quite evident. The slopes of these lines show AE instruments to be the highest followed by ICP, PWMA and AA. A slope of 1.00 would indicate a perfect correlation with ADM. It seems that A/E JA has a good correlation , slope of 1.02, but one should bear in mind that a correction factor has not been applied to the values which would significantly decrease the slope. In addition, particle size plays a significant role in the slopes. In order to minimize the particle size effect, samples having iron concentrations of < 15 ppm were plotted in Figure 2. Much smaller scatter is seen. Similar conclusions could be made from this plot since the two AE instruments gave the highest slope values which correspond to 2.0-2.5 accounting for the matrix effect. ICP slope also correlates well with ADM and AA is a distant last. PWMA shows a higher correlation due to sampling problems.

Figure 3 shows similar type curves for the 3-micron filtered samples. Again, a slope of one means perfect correlation with ADM. Slopes similar to those of Figure 2 were obtained when discounting MFR-5 high value. This indicates that all instruments did correlate well with ADM and each other for samples having small particle size.

A summary of the ferrographic analysis of the MFR filtered samples is shown in Table 5. Ferrographic data provide a comparative rating of the quantity of iron present and a comparative rating of large (entry position reading) to small (50 mm position reading). These data show several interesting points. First the level of iron content is ranked the same when using the percent area covered for the entry position, percent area covered for the entry plus 50 mm positions or the total of the percent area covered for the entry, 50, 40, 30, 20 and 10 mm positions. Secondly, these rankings are very close to the ADM iron content rankings considering the small differences in the iron content of some of the samples. Sample MFR-5-A appears to be the only sample out of order in ferrographs ranking. This could be due to this sample being the only automotive mineral oil having a very high iron content consisting of small (less than 3 micron) particles. The initial L/S rankings do not correlate to the percent area covered rankings which would be expected but do correlate to particle size [4]. It should be noted that the L/S (Initial) ranking correlate very well with the L/S loss due to filtration ranking.

Figure 4 illustrates the correlation between the ADM values (ratio of Fe concentrations before and after 3-micron MFR filtering indicated by L/S (ADM)) and decreases in L/S Ferrograph values. Slopes show good correlation between decreases in L/S Ferrograph values and percent loss due to filtration for most of the samples. Similar correlation is shown between initial ADM iron content and L/S ferrograph values for the same samples before MFR filtering.

Analysis of Wear Metals Other Than Iron: Very few MFR samples contained any significant concentrations of wear metals other than iron [4]. Silicon metal was present in several of the samples with the concentrations not being reduced by filtration. However, other studies have shown most silicon values are due to silicone contamination which is not filterable. Sample MFR-5-A (automotive oil) had very large quantities of Al, Cu, Mg and Pb with the concentrations of these metals not being affected by filtration. Samples MFR-6-A through MFR-9-A contained 3 to 30 ppm Mg and with samples MFR-6-A, MFR-8-A and MFR-9-A containing 1 to 3 ppm Cu. Again, filtration did not reduce the concentrations of these metals. This could be due not only to small particle size but part of the metals being in solution after reaction with oil breakdown products. Some of the SOAP samples contained 1 to 3 ppm Ag, 1 to 30 ppm Mg, 1 to 10 ppm Cu and 1 to 20 ppm Pb. These metals may have been dissolved since the three micron membrane filtration had only a very slight effect on reducing the concentration of the values for any samples. The above data indicate that the use of 3 micron absolute filters would primarily affect only the iron concentrations of lubricant systems.

Wear Metal Trending of SOAP Samples: Four hundred eighty four residual SOAP samples were submitted by the base level operating activities including their AE spectrometric analyses for additional studies. ICP concentrations were also determined on these residual SOAP samples to see if "trending" could be established by using either of the two analysis techniques and for comparing data obtained on lubricant systems having "fine" filtration when using both AE and ICP spectroscopy. These samples were obtained from 9 type of engines and from 2 transmission systems and two gearbox systems. Complete test data for these samples including system serial numbers, hours since overhaul and hours since oil change were reported previously [4]. Based upon information provided by the operating activities all engine lubricant systems utilized 10 micron oil filters except for the F404-GE-400 engines which utilized 5 micron filters. The SH-60B helicopter transmission lubricant system and the F404-GE-400 engine lubricant system were the only systems from which a significant total number of samples were obtained or systems from which consecutive samples were obtained. No increasing iron trends occurred for any of the consecutive sample series which should not be surprising since none of the oil systems being monitored were reported as having any problems. The average ratio of AE iron to ICP iron values of the 484 samples is very close to the average of the AE and ICP iron ratio values for the MFR samples. These AE/ICP ratios were approximately two for the MFR samples and SOAP monitoring samples with the one exception of the F404-GE-400 engines which had an AE/ICP iron ratio of 4.4. This ratio is probably high since many of the AE analyses were conducted on the high range setting of the AE spectrometer (the ICP iron range was 0.00 to 0.63 for these engines).

Concentrations of other trace elements such as Ag, Al, Cr, Cu, Mg, Ni , Si and Ti for all the 13 lubricant systems monitored are low using either the AE or ICP analysis technique with the AE analysis usually being slightly higher and much higher for Ti. Analyses for Pb and Sn were inaccurate due to a small number of intermittent unexplained very high ICP values of Pb and the Sn enhancement when using AE spectroscopy for specific formulations of ester base lubricants. Overall, monitoring of the 13 lubricant systems gave low analysis values for all metals, no iron "trending" data for either analysis

technique and with the similarity between the AE data and ICP data being about the same as that shown by the corresponding data obtained on the MFR samples

Conclusions: The comparative study of the various analysis techniques has indicated that microfiltration could have a small effect on spectrometric oil analysis results. Considering the data from all samples, all analysis techniques investigated (except ferrography and the acid dissolution method) were iron particle size sensitive with none showing significant improvement over the currently used emission spectrometric technique with respect to analyzing large particles. None of the analytical analysis techniques investigated offered any improvement over the currently used emission technique with respect to monitoring capability with or without microfiltration, analysis time or analysis cost, or man-power. Although the analytical ferrograph showed good correlation with iron particle size as well as total iron concentration it could be useful in supplementing the current SOA programs where specific lubricant related problem areas exist.

Recommendations: This study has shown that future research for improving the monitoring capability of lubricant systems would be best directed towards the following areas. Abnormal operating engines or lubricant systems which were not detected by SOAP should be drained and all the drained oil submitted to an appropriate laboratory for an in-depth evaluation including wear particle size distribution determination. Associated lubricant filters should be included for analyses. The data obtained would identify the reasons for the SOAP misses and identify specific type of measurements or data evaluation techniques which would reduce the number of SOAP misses.

Research effort should be directed towards improving the currently used atomic emission spectrometric technique. These improvements would include such factors as instrument and calibration stability, reduced instrument down time, reducing repair costs and equipment modifications such as incorporating computers for updating data acquisition and data evaluation capability.

Acknowledgment: This work was conducted for WL/POSL under USAF Contract No. F33615-C-87-2714. The authors thank Robert E. Kauffman of UDRI for determining the ICP concentrations and Griffin L. Jones of SA-AFLC/LDEN, Kelly AFB for his cooperation and support.

REFERENCES

1. Saba, C.S., Smith, H.A., Keller, M.A., Jain, V.K. and Kauffman, R.E., "Lubricant Performance and Evaluation," Interim Technical Report AFWAL-TR-87-2025, DDC No. AD A183881, June 1987.

2. Rhine, W.E., Saba, C.S. and Kauffman, R.E., "Atomic Absorption Analytical Procedures," Final Technical Report AFWAL TR-82-4021, April 1982.

3. Rhine, W.E., Saba, C.S., Kauffman, R.E., Brown, J.R. and Fair, P.S., "Evaluation of Plasma Source Spectrometers for the Air Force Oil Analysis Program," Final Technical Report AFWAL-TR-82-4013, February 1982.

4. Saba, C.S., Smith, H.A. and Kauffman, R.E., "Alternate Spectrometric Oil Analysis Techniques," Final Technical Report WL-TR-92-2017, April 1992.

TABLE 1

ORIGINAL IRON CONTENT AND IRON LOSS DUE TO THREE MICRON TEST RIG FILTERING

Sample	Orig. Fe & Loss due to Filt.	Method of analysis, ppm ICP	A/E 35	A/E JA	PWMA	AA	ADM
MFR-1-A-1	Orig. Fe	9.28	25	25	60	7.5	31.8
	Filt. Loss	3.01	16	11	51	2.3	21.0
	% Loss	32	64	44	85	31	66
MFR-2-A-1	Orig. Fe	32.7	62	93	61	15.0	60.4
	Filt. Loss	10.9	22	39	27	3.0	30.7
	% Loss	33	37	42	44	20	51
MFR-3-A-1	Orig. Fe	11.1	22	32	15	9.0	13.8
	Filt. Loss	0.7	2	5	2	0.6	2.3
	% Loss	6	9	16	13	7	17
MFR-4-A-1	Orig. Fe	6.43	2	17	40	4.0	26.0
	Filt. Loss	5.90	1	17	38	2.4	25.8
	% Loss	92	50	100	95	60	99
MFR-5-A-1	Orig. Fe	96.1	94	110	54	89	110
	Filt. Loss	14.8	5	2	5	(3)	11.3
	% Loss	15	5	2	9	(3)	10
MFR-6-A-1	Orig. Fe	6.57	11	13	10	5.1	7.41
	Filt. Loss	1.67	1	2	5	1.0	2.79
	% Loss	25	9	15	50	20	38
MFR-7-A-1	Orig. Fe	0.49	2.0	0.80	0.8	0.6	0.61
	Filt. Loss	0.02	0.5	0.50	0.1	0	0.20
	% Loss	4	25	62	12	0	33
MFR-8-A-1	Orig. Fe	12.50	27	28	11	11	10.1
	Filt. Loss	(0.01)	6	6	0	0	(1.7)
	% Loss	0	22	21	0	0	(17)
MFR-9-A-1	Orig. Fe	3.61	5.6	6.2	2.7	2.2	3.41
	Filt. Loss	0.32	(0.1)	1.4	0.3	0.3	(0.2)
	% Loss	9	(2)	23	11	14	(6)
MFR-10-A-1	Orig. Fe	2.69	4.0	6.1	2.6	1.6	6.14
	Filt. Loss	0.78	2.0	2.8	0.8	0.4	2.81
	% Loss	29	50	46	31	25	46
MFR-18-A-1	Orig. Fe	2.30	3.2	8.6	2.4	1.9	5.31
	Filt. Loss	1.50	1.4	5.2	1.4	1.2	3.68
	% Loss	65	44	60	58	63	69
MFR-22-A-1	Orig. Fe	8.56	17	23	8.6	6.2	8.57
	Filt. Loss	1.20	2	4	1.6	1.1	1.39
	% Loss	14	12	17	19	18	16

Values in ( ) show ppm and % increase in value after filtering

TABLE 2

SUMMARY OF RANKINGS FOR EACH ANALYSIS TECHNIQUE AND MFR FILTERED SAMPLES BASED ON IRON CONTENT (12 SAMPLES EXCEPT AS NOTED)

	Number of Times Ranked in Position High to Low Readings					
Measuring Technique	6	5	4	3	2	1
ADM	2	3	0	5	1	1
PWMA	2	1	4	1	2	1
A/E35	1	6	3	0	1	1
A/EJA	7	2	2	1	0	0
ICP	0	0	2	3	4	3
AA	0	0	0	0	0	12

TABLE 3

SUMMARY OF RANKINGS FOR EACH ANALYSIS TECHNIQUE BASED ON PERCENT LOSS DUE TO 3 MICRON FILTERING OF TEST RIG SAMPLES (12 SAMPLES EXCEPT AS SHOWN)

	Number of Times Ranked in Position (High to Low Readings)					
Measuring Technique	6	5	4	3	2	1
ADM	4	4	1	1	1	1
PWMA	3	1	3	2	3	0
A/E35	2	0	2	2	2	4
A/EJA	2	3	3	2	0	2
ICP	1	1	0	3	3	4
AA	0	2	1	0	2	7

TABLE 4

SUMMARY OF RANKINGS OF EACH ANALYSIS TECHNIQUE FOR MFR SAMPLES NOT FILTERED AND BASED ON IRON CONTENT

	Number of Times Ranked in Position (High to Low Readings)					
Measuring Technique	6	5	4	3	2	1
ADM	2	1	3	1	1	0
A/EJA	4	3	0	0	1	0
A/E35	2	4	1	0	1	0
PWMA	0	0	3	4	1	0
ICP	0	0	1	2	3	2
AA	0	0	0	1	1	6

TABLE 5

SUMMARY OF ANALYTICAL FERROGRAPH DATA AND EFFECTS OF MFR FILTRATION ON THE RATIO OF LARGE TO SMALL PARTICLES

		Ferrograph Data				
Sample	ADM Iron ppm	Entry Pos. % A.C.[4] Value	E[1]+50 mm Pos. % A.C. Value	Total[2] % A.C. Value	L/S3 (Init) Value	L/S (Filt. Loss) Value
MFR-5-A	110	444[5]	856[5]	1289[5]	1.08	0.05
MFR-2-A	60.4	1407	2217	3453	1.74	0.57
MFR-1-A	31.8	771	1086	1398	2.45	1.29
MFR-4-A	26.0	717	834	1323	6.13	4.79
MFR-3-A	13.8	25.9	62.7	267	0.77	(+0.11)
MFR-8-A	10.1	26.5	39.8	82.2	1.99	0.84
MFR-22-A	8.57	21.6	47.6	177	0.83	0.27
MFR-6-A	7.41	40.2	76.6	190	1.10	0.30
MFR-10-A	6.14	13.3	24.9	75	1.15	0.36
MFR-18-A	5.13	27.3	54.8	109	0.99	0.46
MFR-9-A	3.41	9.8	19.3	37.3	1.03	(+0.01)
MFR-7-A	0.61	5.4	7.3	11.5	2.84	1.84

[1] E= Ferrogram entry position; [2] Total of % area covered readings at the entry, 50, 40, 30, 20 and 10 mm ferrogram positions; [3] L/S Ratio of large (Entry) particles to small (50 mm); [4] A.C. = Area Covered;
[5] Values adjusted to normal 3 mL sample size

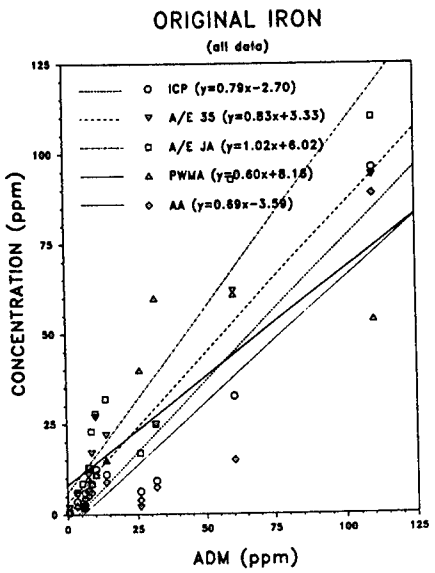


Figure 1. ADM Values of Iron Before Filtration from Five Different Type Analyses Against ADM Values

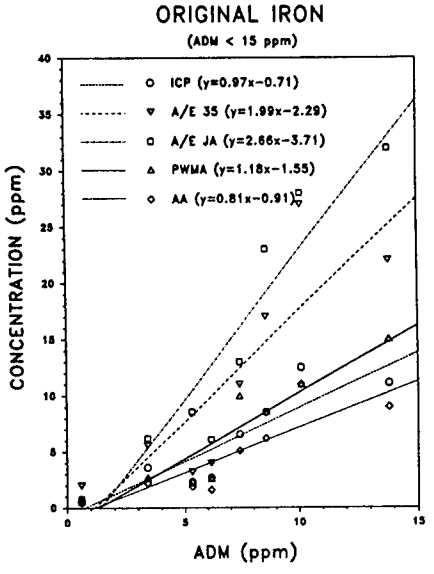


Figure 2. Data Similar to Figure 1 Except Original Iron Concentrations of < 15 PPM Were Considered

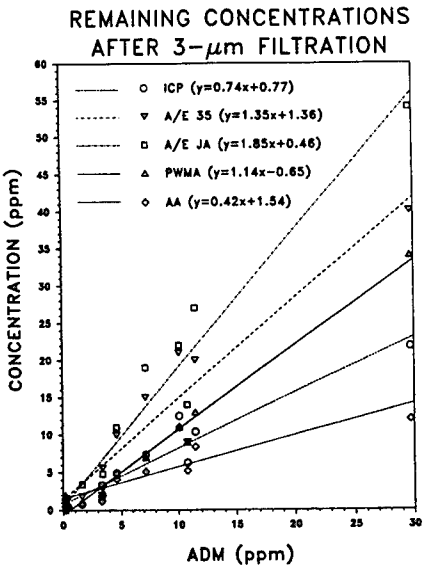


Figure 3. Iron Concentrations from Samples in Figure 1 Filtered Through Three Microns

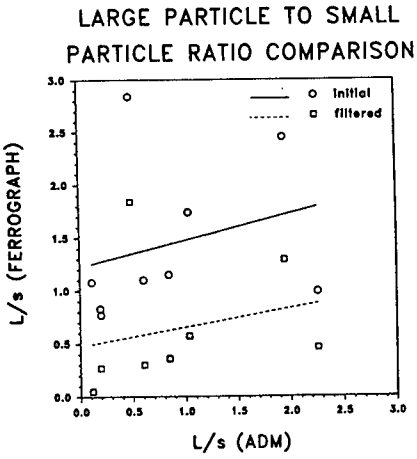


Figure 4. Large to Small Particle Size Ratios Before and After Filtration from Ferrographic Analyses Plotted Against Their Respective ADM Values

CURRENT TECHNOLOGY IN OIL ANALYSIS SPECTROMETERS AND WHAT WE MAY EXPECT IN THE FUTURE

M. Lukas and R. J. Yurko
Spectro Incorporated
160 Ayer Road
Littleton, MA 01460 U.S.A.

Abstract: Oil analysis spectrometers have been in use for the analysis of wear metals, contaminants and additives in lubricating oils for almost 50 years. They have become the mainstay and primary analytical tool of most machine condition monitoring programs based on oil analysis. Spectrometers have evolved from large instruments that take up the better part of a laboratory, to smaller table top instruments. Analysis times have decreased from hours to seconds, and the instruments no longer have to be operated by experts to obtain excellent analytical results

This paper traces the history of the atomic spectroscopy technique through current instrumentation and capabilities. It will close with recent advances in technologies that have created remarkable opportunities to develop a new generation of oil analysis spectrometers

Key Words: Atomic absorption; Atomic emission; Inductively coupled plasma; Machine condition monitoring; Oil analysis; Rotating disc electrode; Spectrometers

HISTORY OF SPECTROSCOPY: Principles of spectroscopy go back close to four hundred years. However, modern spectroscopy has its origin in the early nineteenth century.

As with many of today's technological advances, spectroscopy can also trace some of its early days to Newton. While carrying out tests in 1666 with a glass prism, Newton noticed that sunlight is divided into spectral colors (rainbow colors), Figure 1. This basic fact of science is still the operational basis of all modern spectrometers.

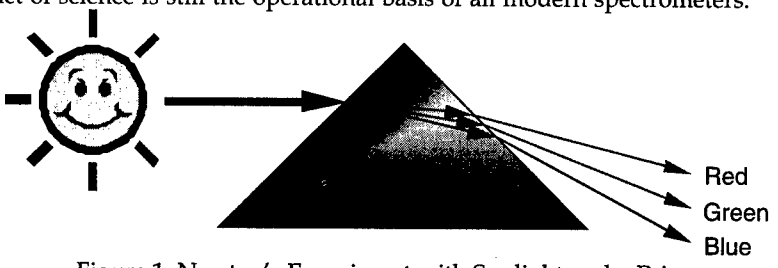


Figure 1, Newton's Experiment with Sunlight and a Prism

The nineteenth century started with further experimentation that eventually resulted in simple spectrographic instruments. In 1802 Wollaston found dark lines in the spectrum of the sun when he used an entrance slit as an aperture [1]. Fifteen years later, Fraunhofer produced the first diffraction grating and constructed a simple spectroscope to study the dark lines indicated by Wollaston. He found 700 of these lines in the spectrum of the sun and similar lines in the stars. In 1826 Talbot and Herschel introduced sodium, potassium, lithium and strontium salts into a spirit flame and observed the colors emitted by means of a spectroscope. The emission lines lay where Fraunhofer found the dark lines in the spectra of the sun and stars. Talbot recommended that this method be used to determine the presence of specific substances instead of an expensive chemical analysis.

In 1860 Bunsen and Kirchhoff extended the observations made by Talbot and Herschel. They recognized that spectral lines were connected with the atoms and not with their chemical compounds. The various atoms emit spectral lines that are unique to the elements excited in the flame. Emissions of these unique spectral lines can thus be used as evidence for the presence of these elements in the sample. This in essence, was the foundation stone for spectrochemical analysis. Kirchhoff also postulated on emission and absorption by stating that matter absorbs radiation at the same wavelength that it emits radiation. It was thus possible to explain the Fraunhofer lines in the spectrum of the sun as due to absorption by atoms in the cooler outer zones of the radiating source.

In 1869 Ångstrom produced the first reflection grating by scratching a mirror with a diamond to create parallel lines. Sample introduction was the area of study for Gouy and less than ten years later, he introduced the first pneumatic nebulizer to introduce liquids into flames.

At this time spectroscopy was a qualitative analytical technique. Spectra were initially observed visually using a spectroscope, but soon thereafter, they were photographed. Around the turn of the century, in the early 1900's, the spectroscope became a spectrograph when photographic plates were fitted to the image plane of the optics. It was also determined that few spectral lines lie in the visual part of the spectrum, and the majority are located in the ultra-violet region. Gas filled optical systems were thus developed to observe spectra below 190 nanometers which otherwise would be absorbed by air (oxygen).

In the 1930's emphasis shifted to alternative excitation sources because flame temperatures are too low to vaporize and excite most lines. Direct current (DC) arcs, alternating current arcs (AC), interrupted arcs and high voltage sparks were developed because they vaporize all elements and excite several spectral lines for each.

Modern spectroscopy was born in 1936 when Thanheiser and Heyes reported the first photoelectric detection of a spectrum using photocells. A few years later, the

development of photomultiplier tubes let to the introduction of today's modern "direct-reading spectrometers". Rotating disc electrode (RDE) systems replaced many flames as the preferred means to introduce the sample into an arc or spark for excitation. The main building blocks of spectrochemical analysis instruments became the excitation source, optical system (spectrometer), and readout system.

Since that time advances have been made in grating production, excitation sources and computerization. Gratings were expensive to produce and each was individually manufactured. When it was learned to copy gratings using a photographic replication process, they became less expensive to produce and prisms were eliminated from most spectrometers.

From 1965 to 1970 Greenfield and Fassel developed the inductively coupled plasma (ICP) source which exhibits much improved capabilities to excite spectral lines, and without most spectral interferences. This led to a renaissance for spectrochemical analysis and the ICP is still the most important excitation source for environmental analysis and research applications.

Starting with the late 1960's, computers started to be used to convert spectrometric measurements into actual concentrations. Analysis times were shortened, but most importantly, the human element was eliminated from much of the analysis process.

In 1983 fiber optics were introduced to transmit light from the excitation source to the spectrometer optics. This eliminated the need for complex optics and made it possible to manufacture stable and modular instruments. Until this time, spectrometer systems were restricted to laboratory use due to their need for environmental controls and complex operating requirements. The fiber optic which simplified construction, and computerization which simplified operation made it possible to develop mobile instruments. The entire industry changed, and it was possible for the first time to bring the instrument to the site where the analysis was required, rather than having to take the sample to a laboratory.

The Joint Oil Analysis Program is a direct beneficiary from these developments. Today oil analysis spectrometers are mobile, easier to use and more reliable.

TODAY'S SPECTROMETERS FOR OIL ANALYSIS: Spectrometers used for debris and lubricant analysis fall into two categories, atomic absorption and atomic emission. Atomic absorption spectrometers consist of flame instruments (FL/AAS) and atomic emission spectrometers can be either rotating disc electrode (RDE/AES) or inductively coupled plasma (ICP/AES). Each technique will be described briefly in this section.

Atomic Absorption Spectroscopy (AAS): In the flame atomic absorption spectroscopy (FL/AAS) technique, Figure 2, a small amount of diluted oil sample is fed into a controlled flame normally fueled by an acetylene-oxygen or nitrous oxide mixture. The flame vaporizes the sample and stimulates the atoms into an elevated energy

state so they can readily absorb light. A special hollow cathode lamp emits a high intensity light beam corresponding to the exact energy required to cause the element of interest to be raised to an excited energy state. Since the lamp emits light at the specific wavelength of the element to be analyzed, a separate lamp is required for each element. The monochromator is tuned to the desired wavelength and because of the chopper, receives a continuous reference signal from the hollow cathode lamp and an alternating absorbed signal from the flame. The absorbed light can be related to the concentration of the element present in the oil [2], and a microprocessor or computer then separates the signals and converts them into concentrations.

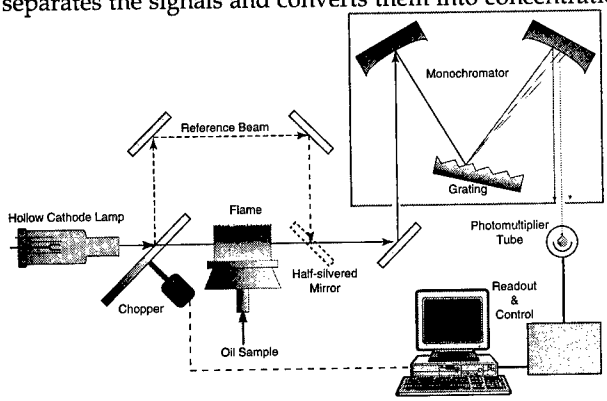


Figure 2, Schematic diagram of an atomic absorption spectrometer (AAS)

The FL/AAS technique has been in use for over 35 years in thousands of instruments for elemental analysis but is applied most often for aqueous solutions, rather than oil samples. It is a single element at a time technique that is good for quantitative, not qualitative, analysis. Improvements in the past 10 years have been in alternate means to atomize the sample [3].

The FL/AAS technique is relatively simple to operate, offers good analytical performance and is the least costly instrument to acquire. It is, however, relatively slow, involves sample preparation for most oils, and requires a different light source for each element analyzed. The sample introduction system and flame temperature limit the wear particle size detection capability to about 5 μm [4]. Table 1 summarizes advantages and disadvantages of the AAS technique.

Table 1, Advantages and Disadvantages of the AAS Technique

Advantages	Disadvantages
• simple to operate	• slow, one element at a time
• analytical performance	• sample preparation
• virtually free of spectral interferences	• inefficient for particles > 5 μm
• low purchase price	• requires gases & light sources

Analytical instruments using the AAS technique are considered ideal starter systems and are used in applications where few samples have to be analyzed at any one time.

Atomic Emission Spectroscopy (AES): Most atomic emission spectrometers employed in used oil analysis are either of the rotating disk electrode type (RDE) or of the inductively coupled plasma (ICP) type. In the RDE/AES technique, Figure 3, a rotating disc electrode brings a continuous sample into a gap between the disc and a stationary rod electrode. A high voltage arc is then struck between the disc and rod electrodes causing the individual atoms in the sample to give off light or radiant energy.

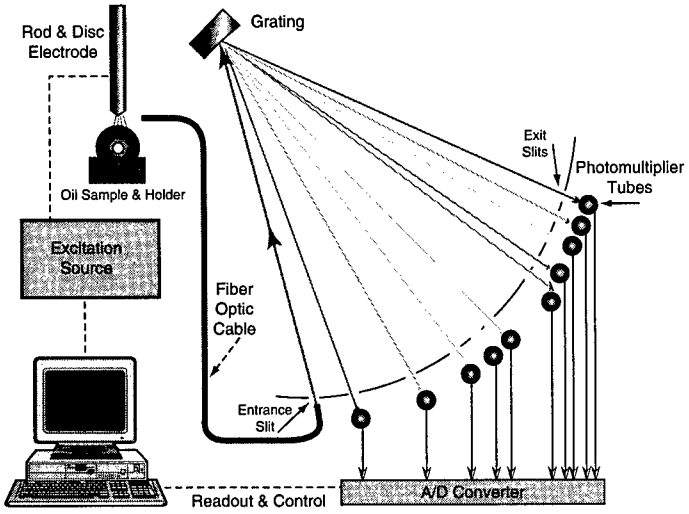


Figure 3, Schematic diagram of an RDE atomic emission spectrometer

The ICP/AES excitation technique employs an electrode-free plasma created by a flow of inert gas (argon). The gas passes continuously through the plasma torch which is located inside a two or three turn R.F. coil connected to a radio frequency alternating current generator, Figure 4. The argon gas acts as the secondary winding of a transformer, becoming highly ionized and extremely hot. The ionized argon is referred to as a plasma and reaches temperatures of 8,000 to 10,000°C. The oil sample is aspirated through the center of the torch and into the plasma where the atoms present are totally dissociated and excited sufficiently to give off radiant energy.

In RDE/AES and ICP/AES systems, a lens or fiber optic is then used to gather and focus the radiant energy from the excitation source onto a concave diffraction grating which disperses the light into the various spectral lines associated with the elements present. Photomultiplier tubes are positioned behind exit slits to detect and convert radiant energy into electrical currents which give a measure of the concentration of

each metal present in the sample. This result is displayed on a video screen, sent to a printer or stored on a hard disc. While the ICP technique was developed for water analysis, improvements have continued for various applications including oil analysis.

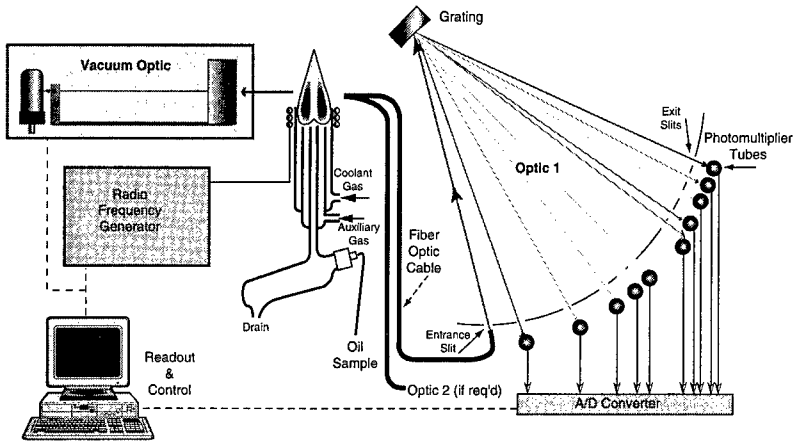


Figure 4, Schematic diagram of an ICP atomic emission spectrometer

For lubricating oil and fuel analysis, the RDE technique continues to be preferred for its simplicity of operation and reliability. It also has the ability to more efficiently analyze the larger particulates in used oil samples ranging from light distillate fuels to heavy gear oils and even greases which can be analyzed without dilution. The technique is very robust, has few moving parts and requires little maintenance. It has been the only viable choice for field deployable spectrometers for commercial and military applications. Table 2 summarizes the advantages and disadvantages of the RDE Technique.

Table 2, Advantages and Disadvantages of the RDE Technique

Advantages	Disadvantages
• simple to operate	• matrix effect
• no sample preparation	• only oil based samples
• robustness	• inefficient for particles >10μm
• does not require cooling water or gases	
• little maintenance	

The advantages and disadvantages of the ICP technique are summarized in Table 3. The main advantage of the ICP technique is its performance: accuracy, precision and detection limits are excellent. Matrix effects are minimal since samples must be diluted prior to analysis and the diluent becomes the primary base material. The ICP also lends itself to analysis of almost any material that can be dissolvable so that a fine spray is created prior to analysis. Automation of the sample introduction process is widely practiced in some cases; unattended operation is thereby possible.

Table 3, Advantages and Disadvantages of the ICP Technique

Advantages	Disadvantages
• performance	• requires sample preparation
• limited matrix effect	• requires laboratory environment
• flexibility (other materials)	• inefficient for particles >5μm
• automation	• requires gas (argon)

The primary limitation of the ICP technique is the need to dilute an oil sample with a solvent prior to analysis. This time-consuming step can lead to accuracy errors. Operation of the ICP also requires a relatively well-trained person with experience in basic laboratory procedures. Used oil samples are not ideally suited for the sample introduction system, and a certain amount of care and cleaning is routinely required. It is not the kind of instrument one would place on the factory floor. The particle size analysis efficiency is worse than the RDE technique and limits its ability to detect certain types of severe wear conditions

TOMORROW'S SPECTROMETERS FOR OIL ANALYSIS: Atomic emission spectrometers have become the primary technique for used oil wear metal and contaminant analysis. Our discussion regarding future improvements will thus be limited to this technique. The most logical approach is to review each of the main assemblies of a spectrometer system and discuss possible changes we might see in the future.

Excitation Source: Both the RDE and ICP technique provides acceptable analytical performance for used oil analysis. Wear metals and contaminants are usually in the oil samples in parts per million (ppm) concentrations; lower detection limits are not necessary. Accuracy is also acceptable for used oils but could be improved for the RDE technique when applied to quality control of additives. Most improvements will thus be in the cost, size and serviceability of future excitation sources.

All excitation sources are based on simple electronic circuitry using discreet components. Development trends tend to be towards solid state excitation circuits which result in cost, size and serviceability improvements. One of the side benefits will be the elimination of transformers that are bulky and extremely heavy.

Sample introduction is still one of the weak links of the RDE and ICP spectrometers. Pre-digestion and complex sample preparation procedure to improve particle size analysis capabilities have not been acceptable. Particle size limitations will be improved with simple ancillary systems such as the rotrode filter spectroscopy RFS method. Automation for sample introduction will also be a primary area of improvement. Automatic samplers and dilutors already exist for ICP spectrometers and work relatively well with many types of oils. The RDE technique has been a little more difficult to automate. However, several methods are in development and will be commercially available in the near future.

One of the biggest limitations to excitation source improvement is the requirement for instruments to meet and correlate to the existing JOAP data base. The JOAP data base has become a standard in the military, is extremely well documented and accepted and used by numerous worldwide military organizations. It is, however, based on instrument performance and referencing methods that are over thirty years old. Improvements that have been made to excitation sources often cannot be implemented if the instrument is to be used, and approved to analyze oil samples as an integral part of the JOAP.

Optical Systems: The optical systems of spectrometers have always been the primary factors in determining the size of an instrument. Improvements in gratings have made it possible to reduce size without sacrificing performance or dispersion. The early instruments with 3 meter (9 feet) focal curves were replaced with 2 meter, then 1 meter and today as small as 0.3 meter optics. Reduced requirements in focal curve size have made it possible to build much smaller instruments with better stability. The introduction of fiber optics has also led to further improvements in stability, complexity and size of optical systems.

The optical system is still the most expensive, the largest and the most delicate component of an instrument. However, rapidly developing solid-state detector technology in the form of charged coupled devices or CCD detectors may change all that. This technology provides devices, by a variety of processes, consisting of many small light sensitive semiconductors (pixels) which may be arranged in a single row (linear array) or as a rectangle or square (area array). CCD technology is widely applied in such commonplace and reliable devices as video cameras, copy and fax machines. Consequently, a robust industry exists to supply and improve such devices and they are available at favorable prices.

A linear array may be positioned at the focal surface of a spectrometer optic such that an entire emission spectrum may be measured pixel by pixel. Linear arrays are now commercially available having as many as 8000 pixels with pixel spacing of approximately 10 micrometers. Bearing in mind that in the present construction of a polychromator optic, exit slits placed in front of photomultiplier tubes are often 10 micrometers wide, never narrower and often wider, a CCD linear array can achieve the same selectivity as an exit slit, and does so for every location along the spectrum. This opens the possibility of simultaneous measurement everywhere along the emission spectrum which, in turn, means unlimited choice of elements within the spectral range of the optic. Further to freely choosing any element, two additional benefits now exist. Namely, simultaneous background correction is possible at any spectral position and an element may be measured at several emission wavelengths instead of one. These features are available in conventional polychromator optics, but only at substantial additional expense and without flexibility after construction, perhaps even compromising some wavelengths because of inevitable mechanical constraints.

By placing these miniaturized and high resolution linear arrays along the focal surface of spectrometer optics as described in the preceding paragraph, the possibility is created of designing a significantly smaller optic. Resolution may not be improved in all cases, but what exists may be adequate for routine oil analysis, especially if spectral interferences can be avoided by choosing analytical wavelengths with somewhat more freedom than photomultiplier tube detectors allow.

CCD technology has already been incorporated in several commercial ICP spectrometers, but not demonstrably with the cost/benefit that might ultimately derive. We should expect time to improve this situation.

At present, several problems hamper realization of full CCD potential.

1. Photometric sensitivity is not as good as with vacuum tube photomultipliers which have very low noise levels and excellent amplification.

2. CCD arrays require cooling (to -20 C and below) to improve noise reduction.

3. CCD semiconductors perform less well in the deep UV portion of the spectrum (which is probably not important for oil analysis - JOAP instruments operate at air path wavelengths, i.e., above 210 nm).

4. Fast computers and new software are required to interface the huge amounts of data generated by the many active pixels. However, the enormous increase in computing power that we have seen in the last few years has rendered it possible, even commonplace, to use analytical techniques that heretofore would have been considered impractical because of "number crunching" demands.

Readout Systems: The readout system of a spectrometer is used to control spectrometer operation and to collect the signals from the optics and to convert them to concentrations for the elements that have been detected in an oil sample. The readout system thus consists of hardware and software.

In the past, readout system hardware was complex and differed greatly among manufacturers. Today, except for one or two circuits, the personal computer is becoming more and more the only electronic readout system required for the spectrometer. It provides instrument control, data processing and troubleshooting. In the future, readout systems will become smaller and less expensive and in direct proportion with developments in the computer market.

The future will also bring improvements in software making instruments easier to operate, improving data bases, providing artificial intelligence and on-board instruction and maintenance manuals. Improvements in processing speed will be transparent as the integration time, up to 30 seconds, to collect the signal from the optics is still the limiting time factor.

As computerization was used more and more in spectrometers during the late 1980's, software was envisioned to be the panacea and solution to everything. Flato's quote [5] still applies today and most likely will apply in the future.

"*Those of us who were there (speaking of the late 1980's) can well remember the mistaken belief that software was easier to develop than hardware, and would take less time. This misunderstanding led to endless 36-hour workdays, many missed shipping schedules, and just as many financial problems.*"

CONCLUSION: There will be changes in tomorrow's oil analysis spectrometer, but they may not be apparent to the operator of the instrument. Changes will be in the internal components of the instruments, there may be additional capabilities and applications, but performance will probably improve little. Since performance for oil analysis with today's spectrometers is more than acceptable, the main benefit derived by the user from instruments in the future will be:

- smaller size and weight
- more capabilities
- easier operation
- additional applications
- lower cost

Changes to solid state excitation sources and CCD technology in optical systems will reduce the size and weight of instruments by at least a factor of two. More capabilities will be possible with solid state detector based electronics and instruments will no longer be limited to only those elements installed when they were purchased. Every spectral line will be available to the user and can be selected with software that is easy to use and operate. The availability of more elements and fewer compromises in the construction of the optics will open additional applications for oil analysis spectrometers. Fuel and coolant analysis is already possible but performance will be improved. Lower cost of instruments will also make them available to other markets and potential user that could not afford these instruments before.

Spectrometers for oil analysis have evolved from instruments that had to be used in environmentally controlled laboratories and by expert and well trained scientific personnel. Today they have become easier to operate, smaller in size, mobile, and more robust for use in non laboratory environments. The net effect has been an analytical capability that provides information immediately and at the site where it is actually needed. The primary advantage of tomorrow's spectrometer will be in its expanded capabilities and lower cost. They will open up additional markets, but most important, the lower cost will make the technology available to users who until then were economically unable to benefit from machine condition monitoring based on oil analysis.

REFERENCES:

[1] Slickers, K., 1993, Automatic Atomic-Emission Spectroscopy, p. 26-32, Brühlsche Universitätsdruckerei, Giessen, Germany.

[2] Sieber, J.R, Salmon, S.G., 1994, Elemental Analysis of Lubricating Oils and Greases, *Lubrication*, Vol. 80, No. 1, 1994.

[3] Lukas, M., 1993, Comparison of Spectrometric Techniques for the Analysis of Liquid Gas Turbine Fuels, *Transactions of the ASME*, Vol. 115, July 1993, p 620-627

[4] Rhine, W.E., Saba, C.S., Kaufmann, R.E., 1986, Metal Particle Detection Capabilities of Rotating Disc Emission Spectrometers , *Lubrication Engineering*, Vol. 42, No. 12, p 755.

[5] Flato, J.B., *Spectroscopy*, 10(8), 62-64, (1995)

AN INTRODUCTION TO THE CONCEPT OF PROFILES

Grahame Fogel CSI *Tribology Program Manager*
Diane Doll *Consultant CSI Tribology Program*

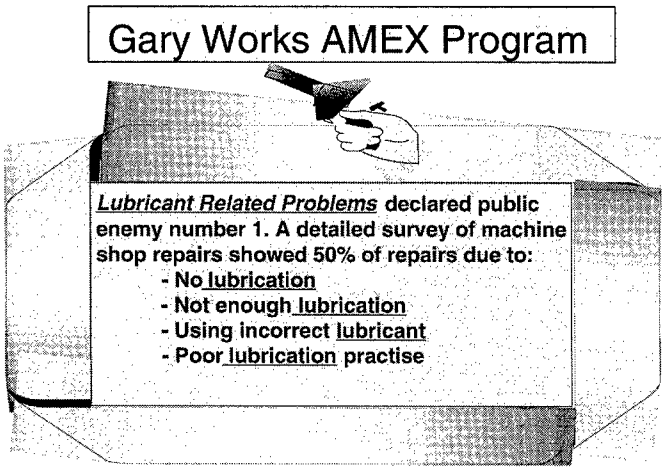
***Key Words**: Lubricant analysis, profiles, third generation oil analysis*

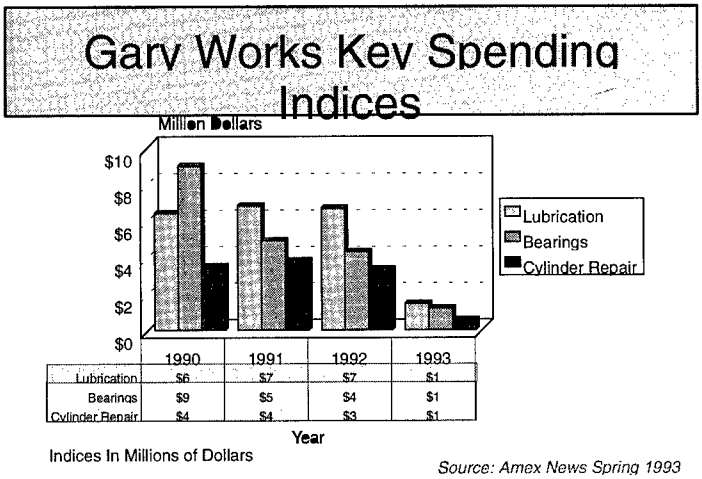
Abstract

There has been limited general acceptance of lubricant analysis as a predictive maintenance tool in industry as compared to say vibration monitoring. There is perhaps a perception within industry that Lubricant analysis has failed to deliver on some of the promises that it has made. However when one looks at so called benchmark programs that exhibit the practises of industry leaders, and measures the achievements of such programs there is no doubt that if applied correctly that lubricant analysis is an invaluable tool in the development of an advanced approach to maintenance. There are many issues which create a successful lubricant analysis program, one of which is the appropriate use of existing technologies to ensure optimum value in the implementation of a program. This paper introduces the concept of Profiles as a framework for the effective utilisation of the correct lubricant analysis technologies.

Examples of Benchmark Lubricant Analysis Programs

In a large integrated steel mill it was estimated that 50 % of machine shop repairs had a lubricant related origin. A comprehensive program was instituted with almost immediate dramatic results.





Source: Amex News Spring 1993

Other examples are an underground mechanised coal mine implemented a comprehensive on site lubricant analysis program to prevent unplanned downtime, and achieved a $2M return on investment in the first year. This was achieved by a combination of increased machinery availibility, decreased work interuptions, and the partial elimination of built in faults during refurbishment. The result was significant measurable increses in coal production.

The question that remains is why is there such a large disparity gap between so called benchmark programs and others?. There are many reasons for this, however one of the most significant is in the utilisation of existing lubricant analysis technologies. This paper will attempt to provide opportumities for ensuring that the appropriate technologies are used within a meaning ful conceptual framework which is referred to as a Lubricant Profile or Profile.

Historical Review Of Lubricant Analysis

Since lubricant related condition monitoring was first implemented just after the second world war it has grown through three distinct generations of development. Each has played a significant role in combining to provide a scientific insight into the workings of lubricated equipment.

First Generation Oil Analysis

Just after the second world war the US army began experimenting with traditional analytical methods for elemental analysis of oil in diesel engines. It was found that certain trends were observed relating to both the wear of the mechanical components together with information on the depletion of additives within the oil. This has led to the broadly applied and familiar field of spectrometric or SOAP (Spectrometric Oil Analytical Procedure) analysis.

SOAP analysis is fairly generally applied around the world and has been successful in the monitoring of internal combustion engines. It has been used with a lesser degree of success in the monitoring of industrial plant such as gearboxes, pumps etc.

A typical SOAP analysis profile would consist of the following components.

Spectrometric elemental analysis for:

- Wear Metals
- Additive Depletion
- Solids and Fluid Contamination

Additional tests were performed for Viscometry, TBN (total base number) or TAN (total acid number), water content, pentane insolubles, dispersancy, fuel dilution etc.

Traditional SOAP analysis suffered from two significant sets of criticisms.

1. The first was that due to the limitation of only seeing particles less than 8 microns in diameter, it was extremely limited in it's use in industrial applications which frequently in a healthy state produced particles with an average size of 30 microns.
2. Secondly a typical profile of SOAP analysis results may consist of in excess of 30 different parameters. Generation of huge amounts of data was confusing to the end user who was looking for simple answers.

Second Generation Oil Analysis

In the 1970's and 1980's there were a number of significant developments in instrumentation and analytical techniques for analysis of used oil. Amongst the most important of these were:

Ferrography (Wear Debris Analysis) :

Ferrographic or wear debris analysis consists of the physical analysis of wear debris particles, as apposed to the chemical based technique of SOAP analysis. The premise that this technique made was that differing wear mechanisms produced wear debris of differing physical appearance. By skilfully measuring the amount and physical morphology of the debris an accurate indication of the health of systems was obtained. The criticism of this technique was that the analysis of the actual debris was very skills dependent and subjective.

Particle Counting;

Particle Counting for many years was inaccessible for the routine analysis of used oil due to the cost of instrumentation and the sensitivity of the measuring technique. However in the last few years cheaper more user friendly particle counters have found their way onto the market. Particle counting has filled a hole in the oil analysis spectrum where in clean oil systems such as turbines, compressors and hydraulics the particle concentrations were so dilute that the other recognised methods were of limited value.

FTIR (Fourier Transform Infra Red) Spectrometry

FTIR Spectrometry has recently come of its own as a new applied technology in the field of oil analysis. This analytical technique has been developed around the use of the instrument in combination with computerised procedures for large volume data manipulation and storage. FTIR spectrometry allows accurate determination of the in service degradation of the lubricant. The technique is also able to give a measure of water content and diesel dilution together with a measure of soot in engines.

Other techniques and technologies which have come onto the market and are being utilised in specific applications are gas chromatography and various on line analysis methods.

Technology	Capabilities	Comments
Spectro Metals Analysis	**Wear Metals** **Metallo-Organic Additives Contaminants**	**Severe Particle Size limitations**
FTIR Spectrometry	**Lubricant Degradation Fluid Contamination (water) Good for QA & QC of lubricants**	**Becoming increasingly used as the technology of choice, As FTIR systems and associated data management systems improve.**
Wear Debris Analysis	**Diagnostic tool for wear related system deterioration. No particle size limitation. Excellent tool for root cause analysis**	**Limited proactive capability, mainly a diagnostic tool**
Particle Counting	**Measures solid contamination. Lead technology for monitoring clean oil systems. Laboratory based particle counting relies on accurate laser based detectors.**	**Measurement technologies, useful for setting target cleanliness levels. Also for specification of filtration, and quality control of unused lubricants.**
Viscometry	**Good indicator of fluid integrity. Can measure lubricant degradation or misapplication.**	

Third Generation Oil Analysis

In this paper we would like to present the concept of third generation oil analysis. As the concept of predictive/proactive maintenance becomes more accepted as a maintenance philosophy so too becomes the need for more accurate specific understanding of problems within mechanical systems. This means that no single analytical technique is better or more appropriate than another for solving a problem. However in combination the various technologies provide a powerful set of capabilities of deriving a holistic picture of machine health. There are two cornerstones to making third generation oil analysis work. These are firstly to select the appropriate ***Profile*** of tests in the required application and secondly to have a flexible sophisticated information processing system that turns the raw data into clearly understood maintenance information.

Within the development of third generation technologies we introduce the concept of ***Profiles*** (Lubricant Profiles)

Profiles

A survey of the lubricant analysis marketplace shows that the cost of analysis of a sample in the USA can vary from anywhere between $5- $85. Asked for pricing a typical vendor would offer $10 analysis, $15 analysis, $25 analysis or $40 analysis. The presumption that $25 analysis is an expanded $10 analysis may easily be made, however the basis of this premise is in many cases incorrect. For example the $25 analysis may be limited to analytical Ferrography where the $10 may be simple spectrometry. The real question which needs to be asked is what exactly should I be doing to cost effectively monitor my machines - and that is where the concept of Profiles is of assistance. The definition of a profile is:

A combination of tests which appropriately monitors the manifestation of failure for that system.
(In other words lets get away from treating everything as a variant of a diesel engine)

Typically in industrial plants there are a large variety of complex systems operating. The application of the profile concept will allow the user to move beyond a simple fault detection approach in order to provide a more comprehensive understanding of the dynamics within such systems. The use of profiles exploits the full capabilities of modern lube analysis technology. In practise this means understanding the operational dynamics within a system together with the failure modes and matching these to applicable tests which will detect and diagnose these conditions. Below I illustrate the selection of two typical profiles, one for a gearbox, and another for a hydraulic system.

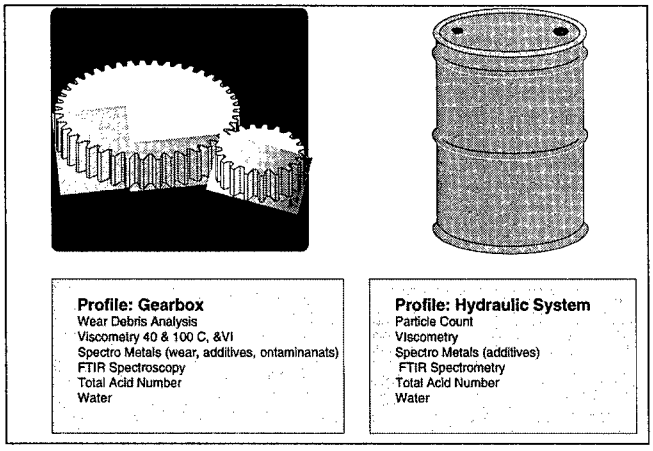


Table 1 Illustration of the use of Profiles for two differing industrial applications

Gearbox	Hydraulic System
Need to monitor: • Mechanical wear, which is normally seen in particles larger than the range of detection for spectrometers. Anti wear, anti oxidant and anti scuff additive depletion • Oil oxidation • Acid build up. • Water and other contaminants	Need to monitor • Solid contaminants by particle count • Water • Depletion of anti wear additive • Oil oxidation • Acid build up • Wear

Table 2 Monitoring Objectives for a gearbox profile compared to a hydraulic system

For example a gearbox in a large mechanical drive will require a differing set of tests (referred to as the Profile) to that of a turbine in a nuclear power station. It is therefore important for the condition monitoring practitioner to understand the differing damage and deterioration

mechanisms that occur within a particular system so as to allocate the correct profile in order to optimally benefit from a condition monitoring program.

PROFILE	TESTS	COMMENTS
Turbo Generator	Viscosity (40 C, 100 C and VI) Particle Count Elemental Spectrometry (additive depletion) Karl Fisher water FTIR Molecular Spectroscopy TAN	This profile would have a proactive bias. If wear is detected, inevitably there will have been an associated reduction in residual life of the system. We are monitoring to ensure elimination or at least very early detection of any incipient problem.
Gearbox	Wear Debris Analysis Elemental Spectrometry Viscometry (40, 100 and VI) FTIR Molecular Spectroscopy	Wear is a normal product of operation of most gearboxes. However the nature, extent and mechanism of this wear are vitally important to detect. Influence of lube functionality also monitored.
Standby Diesel Generator Engine	Viscosity (40 C, 100 C and VI) Particle Count Elemental Spectrometry (additive depletion) Karl Fisher water FTIR Molecular Spectroscopy TBN	The condition monitoring approach for such an application has to be very proactive ensuring that no conditions arise that would allow the initiation for a failure cycle.
Grease Lubricated Bearing	Wear particle extraction, followed by wear debris analysis and elemental spectrometry. FTIR Spectrometry	Predominantly a predictive profile which relies on wear debris analysis to define mechanism and extent of damage. FTIR spectrometry detect additive depletion or oxidation.
Hydraulic System with servo valves	Particle Count Elemental Spectrometry (additive depletion) Viscosity (40 C, 100 C And VI) FTIR Molecular Spectroscopy TAN	Predominantly a proactive profile dealing with the principal root cause of deterioration which is solid contamination.

Table 3 A set of typical machinery profiles

CASE STUDIES ILLUSTRATING THE USE OF PROFILES.

CASE STUDY 1 Turbine Lubricant

A turbo generator was sampled in order to review the health of the lubricant. The sample was analysed using a "clean" profile. The results were as follows:

Contaminants:

>2um	843
>5um	342
>15um	57
>25um	14
>50um	2
>100um	0
ISO 2	17
ISO 5	16
ISO 15	13
Silicon	0

Additives:

Magnesium	0
Calcium	0
Barium	0
Phosphorus	9
Zinc	18
Molybdenum	0

Wear Metals:

Iron	42
Chrome0	
Lead	1
Copper	69
Tin	0
Aluminium	1
Nickel	0

Viscosity:

cSt 40C30	
cSt 100C	5
Visc Index	104

Oil Condition:

PPM Water	0
Oxidation	0
Nitration	0

The diagnostic message read:
Excessive wear. In addition, the lubricant is slightly more contaminated than the recommended Target Cleanliness Limit of 17/15/13.

When looked at more closely, using detailed FT-IR spectroscopy, it was discovered that the lubricant was showing severe signs of additive depletion together with indications of base oil oxidation. The detailed conclusion of the analysis was:

Significant indication of lubricant degradation. This situation poses a threat to the reliability of this system. Immediate action is necessary to ensure reliability. Recommend system is drained and relubricated.

The standard FT-IR quantitative methods did not detect additive depletion because the published methods focused on the "wrong" wavelength for this proprietary additive. A well-trained Tribologist knows how often additive packages change and instinctively looks at the entire infrared spectrum instead of trusting a standard computer generated number. This specific new procedure would then be added to the ***Profile*** for this system.

By understanding the criticality of this system and going beyond "second generation" lubricant analysis, a major dollar loss due to system shutdown was prevented. In addition by modifying the Profile, new real value was added to the analysis set.

CASE STUDY 2 Rolling element bearings

A large rolling element bearing was monitored for 3 months using a "screening" (low cost) profile. The results were as follows:

		Month 1	Month 2	Month 3
Wear Metals:	Iron	51	102	36
	Chrome	0	0	0
	Lead	0	0	0
	Copper	2	2	1
	Tin	0	0	0
	Aluminium	0	0	0
	Nickel	0	0	0
	Silver	0	0	0
Additives:				
	Boron	0	0	0
	Sodium	7	10	6
	Magnesium	0	1	0
	Calcium	4	8	4
	Barium	0	2	0
	Phosphorus	10	16	0
	Zinc	3	5	1
	Molybdenum	0	0	0
	Titanium	0	0	0
Contaminants:				
	Silicon	3	1	0
	PPM Water	0	0	0
Oil Condition:				
	Viscosity (40C)	64	66	64
	Oxidation	0.00	0.02	0.01
	Nitration	0.00	0.01	0.01

Month 1 diagnoses read:
Sample has no abnormal indications.
Month 2 diagnoses read:
Increase in iron, will monitor in next sampling.

Month 3 diagnoses read:
High indications of wear . Wear debris analysis indicated wear in the form of fatigue platelets, and metallic oxide. A wear index of 13 was registered. In addition, there was a significant amount of fibre present suggesting perhaps the failure of the filter. We recommend that the unit is carefully analysed for possible problems (compare with your vibration data). Drain and relubricate unit and resubmit a sample after 100 hours of operation.

In month 3, the profile was changed to a more comprehensive industrial profile (which included wear debris analysis and wear index) after educating the client concerning the importance of proper monitoring of systems. If the profile had not been changed to monitor wear debris, the wear may have gone un-noticed as evidenced in the drop in iron content due to the particle size limitations of spectrometric analysis.

CASE STUDY 3 Electric Motors

An industrial plant interested in establishing a lubricant monitoring program sampled a motor bearing. The first analysis performed followed an "industrial" profile. The objective here was to get an overall health indication of the system. It was assumed the system would be too "dirty" for a clean (particle count) profile because the system had never been monitored before. Once the system was flushed and "cleaned up", a "clean" profile was used as a proactive approach to monitoring the newly replaced motor bearing. Results were as follows:

		Industrial Profile	Clean Profile
Wear Metals:	Iron	6	0
	Chrome0	0	
	Lead	79	0
	Copper	211	0
	Tin	142	0
	Aluminium	0	0
	Nickel	0	0
	Silver	0	0
Additives:			
	Boron	0	0
	Sodium	0	0
	Magnesium	0	0
	Calcium	3	5
	Barium	0	0
	Phosphorus	11	67
	Zinc	109	130
	Molybdenum	0	0
	Titanium	0	0
Contaminants:			
	Silicon	2	0
	PPM Water	0	0
Oil Condition:			
	cSt 40C	68	
	cSt 100C	8	8
	Viscosity index	97	97
	Oxidation	0.03	0.00
	Nitration	0.12	0.00

Wear Debris:		Contamination:	
Wear Index	24	>2um	1355
Ave Size	4	>5um	266
Max Size	1	>15um	32
Density	0	>25um	13
		>50um	4
		>100um	2
		ISO 2	18
		ISO 5	15
		ISO 15	12

By monitoring the system as a "clean" profile one adopts a more proactive approach, thereby eliminating a primary root cause of failure condition. The consequence of this is to exclude abrasive wear as a cause of system failure. This bearing is now maintained at a high level of cleanliness, where the slightest indication of wear, or contamination is detected very early on, and remedial action is immediately taken. It is assumed that this bearing will have infinate life, provided no external factors create a failure condition.

Lessons To Be Learned

As lubricant technology has matured we have a variety of advanced analytical tools which identify differing failure causes within systems. To ensure optimal value within a luricant analysis program it is important to :

1. Familiarise ourselves with the capabilities and deficiencies of the various technologies.
2. Recognise that the concept of ***Profiles*** provides a useful conceptual tool for devising appropriate tests for different systems.
3. For condition monitoring professionals to recognise the characteristics of a system in order to apply appropriate ***Profiles*** within the context of the analytical program.

THE MINI-LAB CONCEPT AS AN ALTERNATIVE TO CONVENTIONAL OIL ANALYSIS

Grahame Fogel *(Manager Tribology Program CSI)*

ABSTRACT

Analysis of used oil has proven itself to be a useful condition monitoring technique for mechanical equipment. Traditional oil analysis involves taking an oil sample and sending it away to a remote laboratory for analysis. This has many short comings such as lack of ownership of the program and long lead times for recovering results. Until recently installing lubricant analysis capabilities on site have been precluded as instrumentation has been cost prohibitive. However recent developments in both instrumentation and intelligent software allows successful and cost effective implementation of on site lubricant analysis capabilities.

Key Words: On site analysis, Mini lab, Lubricant testing, maintenance, condition monitoring.

1.0 INTRODUCTION

Lubricant analysis has been around as a maintenance tool for a number of years. Most plants know of it's availability and a number are using it in some form with varying success. However there has not been the general acceptance of the technology in industry as the has for example monitoring process control parameters. This is due to :

- The lube analysis business has been dominated by laboratories analysing internal combustion engines. This has led to industrial plant with complex and differing deterioration mechanisms being treated like "diesel engines" providing end users with limited success in industrial applications. Lubricant analysis in industrial applications has defaulted itself as a detection technology whereas end users require detailed knowledge of failure process and how to address these from a root cause perspective.

- One of the chief obstacles in achieving general acceptance of lubricant analysis as a valid technology for industrial plant is so called ***free oil analysis*** where the market place has forced lubricant vendors to provide such a service. Generally free oil analysis is characterised by very limited spectrometric analysis with limited diagnostic capabilities, which might be a totally inappropriate technology for say a steel rolling mill.

- There is often a disconnect between a plant and the laboratory analysing the sample. This is born out by frustration with turnaround times for results, lab diagnosticians poor awareness of industrial equipment, and a general break

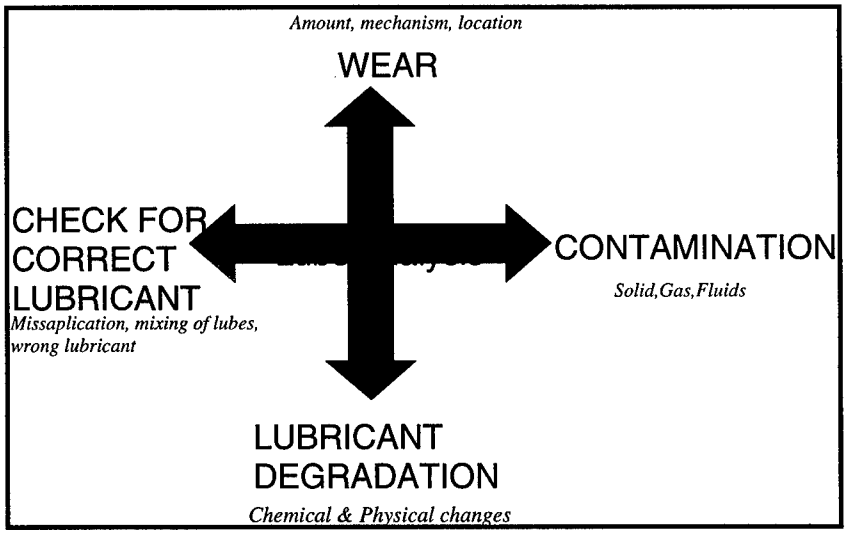
down of communication between what maintenance people require to make decisions and what lubricant analysis reports contain.

This paper argues that there is an opportunity of overcoming these frustrations and making real benefit from the technology by installing On Site Lube Analysis tools referred to as the Mini- Lab.

2.0 OVERVIEW OF LUBE ANALYSIS TECHNOLOGY

One conducts lubricant analysis for four reasons:

- To monitor mechanical surface damage (wear) of Componentry within a system.
- To check for both solid, fluid and gas contamination.
- To check for degradation (loss of function) within the lubricant.
- To check that the correctly specified lubricant is within the system.



2.1 Lube Analysis In Industrial Plants

There is a large variety of complex systems operating in industrial plants. In order to move beyond a simple fault detection approach and provide a more comprehensive understanding of the dynamics within such systems one has to adopt a more sophisticated approach which utilises the full capabilities of modern lube analysis technology. This means understanding the operational dynamics within a system together with the failure modes and matching these to applicable tests which will detect and diagnose these conditions. In order to accomplish this objective I would like to

introduce the concept of a Lubricant Analysis Profile. Which is essentially a collection of tests appropriate to specific systems. (Lets get away from treating everything as a diesel engine). Below I illustrate the selection of two typical profiles, one for a gearbox, and another for a hydraulic system.

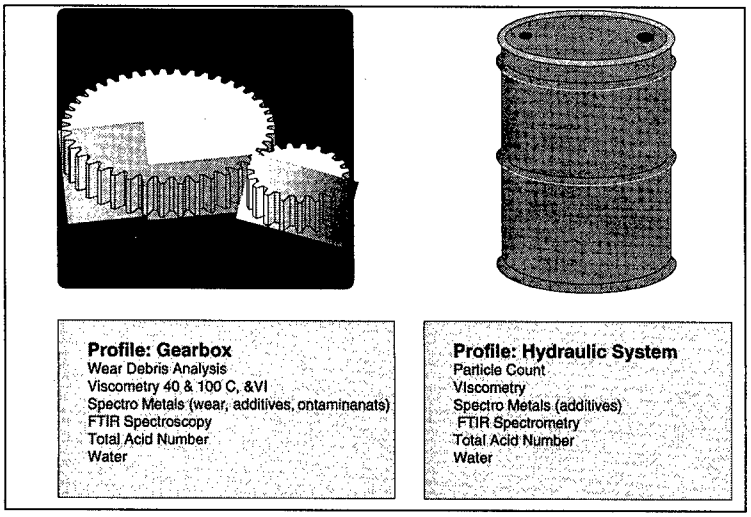


Illustration of the use of Profiles
for two differing industrial applications

Gearbox	Hydraulic System
Need to monitor: • Mechanical wear, which is normally seen in particles larger than the range of detection for spectrometers. Anti wear, anti oxidant and anti scuff additive depletion • Oil oxidation • Acid build up. • Water and other contaminants	Need to monitor • Solid contaminants by particle count • Water • Depletion of anti wear additive • Oil oxidation • Acid build up • Wear

Monitoring Objectives for a gearbox profile
compared to a hydraulic system

For example a gearbox in a large mechanical drive will require a differing set of tests (referred to as the monitoring profile) to that of a turbine in a nuclear power station. It is therefore important for the condition monitoring practitioner to understand the differing damage and deterioration mechanisms that occur within a particular system so as to

allocate the correct profile in order to optimally benefit from a condition monitoring program.

In order to understand how to set up Profiles lets us review the various capabilities of lube analysis technology.:

1. **Spectro Metals Analysis** providing elemental analysis (Fe, Cr, Zn, Si etc.) gives information on:
   - Wear Metals
   - Metallo Organic Additives
   - Contaminants

2. **Viscometry** indicates firstly whether one is using the correct grade of lubricant and secondly viscosity increases or decreases measure certain physical effects on the lubricant. Viscosity is an extremely important parameter to measure as it determines the thickness of the fluid film.

3. **FTIR Molecular Spectrometry** (Fourier Transform Infrared Spectrometry) measures the degradation of functionality within the lubricant.

4. **Wear Debris Analysis** (Ferrography) measures the extent, location and mechanisms of wear.

5. **Particle Counting** measures solid contamination within a system.

6. **TAN** (Total Acid Number), **TBN** (Total Base Number) measures acid build up, and reserve alkalinity depletion.

Technology	Capabilities	Comments
Spectro Metals Analysis	Wear Metals Metallo-Organic Additives Contaminants	Severe Particle Size limitations
FTIR Spectrometry	Lubricant Degradation Fluid Contamination (water) Good for QA & QC of lubricants	Becoming increasingly used as the technology of choice, As FTIR systems and associated data management systems improve.
Wear Debris Analysis	Diagnostic tool for wear related system deterioration. No particle size limitation. Excellent tool for root cause analysis	Limited proactive capability, mainly a diagnostic tool
Particle Counting	Measures solid contamination. Lead technology for monitoring clean oil systems.	Measurement technologies, lack absolute precision.
Viscometry	Good indicator of fluid integrity. Can measure lubricant degradation or misapplication.	

Summary table indicating the principle lab based fluid analysis technologies

3.0 THE MINI-LAB CONCEPT

This paper argues that as more emphasis is being placed on industrial productivity and the associated need to maximize machinery uptime, so the importance of the role that qualified maintenance decisions have in ensuring optimum plant availability. The quality of this decision making is going to be based on the availability of good reliable information. *This premise defines the role of the Mini-Lab.* It is the direct on site evaluation tool for system and lubricant evaluation whose purpose is to continuously supply current information in a form to enable such maintenance decision making.

What Can Be Measured On site?

Instrument	Capabilities	Comment
Oilview Analyzer 5100	Wear Ferrous Non Ferrous Wear Lube degradation Contamination	Sensitivity to 1 ppm Sensitivity to 100 ppm Excellent field tool for measuring oil degradation in the field. Good indicator of water. Can measure water in solution, dispersion and emmisable water.
Oilview Ferrous Wear Monitor	Measures amount of paramagnetic iron within the sample.	No particle size limitations, simple to use, excellent screening tool
Digital Viscometer	Measures absolute viscosity in the field.	Simple to use field test for viscosity. Verify correct lube. Detect fuel dilution.
Oilview Particle Counter	Absolute measure of solid contamination.	Has additional sensors that measure absolute colour and dielectric of the oil, giving indication of oil condition.
Wear Debris Analysis (ferrography)	From microscopic examination of the wear particles, determine the nature, extent and mechanism of wear	Excellent diagnostic tool, as it uniquely provides information on the mechanism of wear.

Summary table indicating instrument capabilities

3.1 Wear:

Wear is the primary mechanism by which industrial plant deteriorates, By observing the amount and mechanism of wear periodically one is able to monitor the deterioration of plant. Traditionally this has been done by SOAP Analysis (spectrometric oil analytical procedure).

As apposed to SOAP which is a chemical based analytical technique, there are a number of new instruments which are able to accurately measure wear on site. Typical

field instruments that ,measure wear are the CSI Oilview Ferrous Wear Monitor 51 FW and the Oilview Analyzer 5100, the Tribometrics and the Analex PQ90.

3.1.1 Oilview Ferrous Wear Monitor. 51FW

The Oilview 51FW measures the effect that an oil sample has on a magnetic field. In other words it gives a direct indication of the amount of Ferrous Wear within a sample. Operation is simple and involves no measurement of sample volumes or dilution. The sample bottle is simply set on the top of the Monitor which immediately gives an indication of the amount of paramagnetic ferrous debris within the sample. The assumption being that ferrous wear will give a significant indication of system deterioration. This is principally valid for industrial power transmission systems such as large gearboxes and heavy loaded rolling element bearings.

3.1.2 Oilview Analyzer 5100

The Oilview Analyzer 5100 is able to measure both Ferrous and Non-Ferrous wear. The sensitivity to Ferrous Wear has been determined to be very accurate in the range of 1ppm Iron and non ferrous wear less accurately to around 100ppm. The Analyzer uses a combination of measuring the electrical properties of the oil, the magnetic properties of the debris and the time based settling of suspended debris due to gravity. The combination of measurement technologies within the instrument allows not only determination of wear, but a measure of contaminants such as water (very accurately), and most significantly a field based indication of lubricant deterioration. The measure of lubricant deterioration is achieved by a comparison of the dielectric properties of a new oil which is stored in a database, as compared with the sample of oil which has been drawn from an operating system. Not only can this condition parameter give an indication of field degradation of the lubricant, it can be used as a quality control tool for the testing of variations within supply of new lubricants.

3.1.3 Wear Debris Analysis

Wear Debris Analysis is a physical technique which quantifies and characterises the amount of wear within a system. There are a limited number of wear mechanisms, each of which generates wear particles with a distinct morphological appearance. One can periodically sample, quantify and observe these wear particles in order to understand the extent and mechanism of wear within systems. Interpretation of the results from this process gives direct information on the type location and extent of damage within a system.

This has powerful and far ranging implications, for instead of being simply satisfied with a detection capability of measuring too much wear, one can now move into a diagnostic mode where the mechanism is identified, together with the extent of damage. This allows a more meaningful conclusion as to the level of threat that a particular system is operating under.

An additional important consideration where Wear Debris Analysis has significant advantages over SOAP techniques is that it does not suffer from the ever present problem of particle size limitations. Spectrometers used in oil analysis today have a restriction in that they do not measure particles over 8 microns in size. In power transmissions this is a major problem as the majority of damage mechanisms (not all) generate particles greater than 8 microns. In addition a general rule is that as damage progresses these particles increase in

size. SOAP analysis has failed to provide information on these situations, much to many end users frustrations. Wear Debris analysis has no such limitation, and besides being able to quantify the amount of wear, gives valuable information on mechanism, location, and extent of wear as well as the state of the lubricant and presence of contaminants.

Wear Debris Analysis or Ferrography has been presented has a high powered complex technology, where in essence it is the separation and measurement of wear particle characteristics under a suitable microscope. Even a limited capability for undertaking visual particle examination on site provides great insights into what is happening within a system. The viewing of the wear particles provides a direct understanding of damage modes and extents within systems. With a degree of training and experience, wear debris analysis becomes a powerful on site analysis tool especially appropriate in industrial power transmission systems.

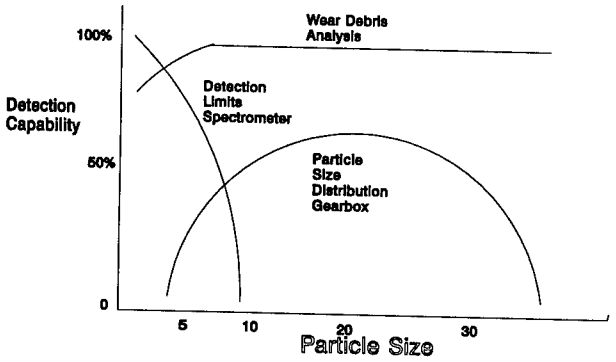


Fig. 2 Showing the particles size effects of SOAP vs. Wear Debris analysis.

3.2 Viscometry

The viscosity of a lubricant is the essential lubricating property of the fluid. It not only determines the lubrication regime but will also define essential lubrication conditions such as the fluid film thickness. Variations in viscosity are an indication of fluid degradation, or contamination, or just as meaningful, give the best indication of misapplication of a lubricant. A jump in viscosity from 220 Cst in a gear system to 100 Cst probably means that an incorrect lubricant has been applied to the system.

3.5 Contamination Control Through Particle Counting

Solid contamination especially in clean oil systems is the single most important contributor to wear, the primary root cause of failure. It is therefore simple common sense to maintain determined cleanliness levels within sensitive systems. Proactive contamination control has is becoming more and more well accepted as general engineering practise and consists of five essential principles. These are defined as:

Contamination Analysis the measurement and understanding of the amount type and composition of contaminants.

Contaminant Exclusion the identification and minimising of contaminant sources.

Contaminant Removal the practices and procedures for removing contaminant that can not be excluded, such as , residual manufacturing debris or generated contaminant within a fluid system.

Contaminant Tolerance ensuring that components can tolerate (not degrade performance beyond allowable limits) the level of contaminant that is not excluded or removed.

Systems Approach establishing a systematic approach that ensures that the system Target Cleanliness Level (TCL) is achieved and maintained.

A proactive contamination control program is not a condition monitoring program where one is trying to detect unhealthy conditions. Essentially one is dealing with a root cause which initiates ill health within machinery, thereby eliminating initiation of the failure cycle. This provides enormous benefits in terms of reliability and life extension such as indicated in the BHRA published results shown below.

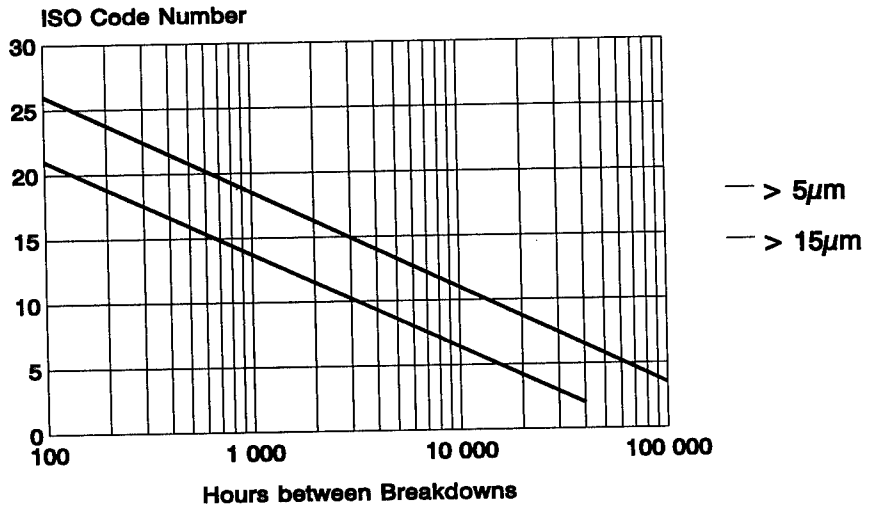


Fig. 1 BHRA Study of Effect of Cleanliness on Life Extension of Hydraulic Pumps

A proactive contamination control program by it's nature involves extensive testing. Relying on a remote laboratory is firstly to time consuming and secondly will result in testing costs mounting. Having an accurate particle counting capability on site as part of the testing capability provides an effective solution.

The Oilview Particle Counter 51PC is a state of the art laser based particle counter with the highest degree of accuracy which is important in contamination control. In addition it has two additional sensors that measure the fluid condition, these are a sensor that measures the electrical properties and the fluid colour. This allows information on solid contamination and fluid condition in a single pass.

3.6 Mini Lab Profiles

Similar to the concept of using Profiles for effective industrial lubricant analysis, the mini lab can be configured in terms of what would be most effective in monitoring the range of systems operating within a plant.

For example a precision machining shop would be running predominantly clean systems which require a "proactive approach". The Mini Lab Profile which would be configured for this application is as follows:

Instrument	Capability
Particle Counter 51PC	Accurate measure of solid contaminants, together with indication of lubricant condition.
Analyzer 5100	Measures, water contamination and lubricant degradation
Digital Viscometer 51DV	Verify viscosity changes and use of correct lubricant.

Minilab Clean System Profile

For comparison a Mini Lab profile for an Industrial Site which predominantly monitored gear trains would have an instrumentation profile as follows:

Instrument	Capability
Ferrous Wear Monitor	Detection of Ferrous wear.
Analyzer 5100	Detection of ferrous and non ferrous wear, together with water contamination and lubricant degradation
Wear Debris Analysis (Ferrography)	Detailed examination of particles in order to understand nature, mechanism and extent of wear.
Digital Viscometer 51DV	Verify viscosity changes and use of correct lubricant.

Minilab Industrial System Profile

4.0 DOES THE FLUID ANALYSIS LABORATORY STILL HAVE A ROLE ?

The concept of the Mini-Lab is specifically designed to be able to be simply operated by maintenance practitioners (as apposed to chemists) to provide simple detection and diagnosis capabilities on-site (in the plant work environment).

The fluid analysis laboratory (where the chemists work) has an integral role in this scenario by providing advanced back up to the mini lab on an as needed basis. In order

for the fluid analysis lab to be able to do this it has to redefine it's purpose from that of handling large volumes of samples efficiently to keep costs down, to being the detailed diagnostic backstop with all the associated expertise and capabilities for the Mini-Lab contributing meaningful solutions to real problems. (No more diagnosing gearbox problems, with an engine analysis mind set)

An example of this relationship is indicated as follows: A Mini-Lab at a plant detects a contamination problem in a high speed compressor. The condition is detected on site using the particle counter. The role and function of the Fluid Analysis Laboratory would at this point be to identify the exact nature of the contaminant, define the threat that it poses to the specific componentry within the system and then make recommendations as to a set of activities which will address the source of the contaminant together with the Tribological management strategy to contain the situation, by addressing the root cause. This may be far removed from existing practises within lubricant analysis laboratories at this point in time. However what is achieved is an effective solution, rather than a symptomatic evaluation. It is up to users of oil analysis services to force this as a competitive aspect of purchasing such services. This will provide a realistic and meaningful solution to those plants serious about their condition monitoring programs.

5.0 SUMMARY OF BENEFITS OF ON SITE LUBRICANT ANALYSIS CAPABILITIES.

Setting up in house on site capabilities for the routine analysis of lubricants has many advantages over contracting out oil analysis. A summary of these is as follows:

- The end user has total ownership of the program. Success or failures is determined by a defined effort, with no external influence.
- There is immediate access to results, no waiting for external lab turnaround. In addition samples don't go "*missing*".
- There is ownership of the test process together with the results. With this site specific expertise around the plants own machines can be built up.
- One can set up appropriate test profiles per compartment.
- Over time costs are significantly reduced. In addition sampling and testing frequencies are now not cost constrained. (you can test us much as you want with no increased cost consideration)
- When a problem is tested, one can immediately resample, and much more closely manage the situation. One can sample daily, hourly etc. with immediate access to the results.
- There is a relatively low skills requirement to set up, and run such a program.
- An on site facility can easily be integrated into existing programs such as vibration, thermal imaging etc.
- One can electronically link into an external testing lab where samples requiring additional testing can be routed.

6.0 SOME ADDITIONAL SIDE BENEFITS OF HAVING A MINI LAB ON SITE

Beyond the main concern to prevent unexpected failures and machine life extension which most users of such systems have, there are a number of side benefits that have been experienced. These side benefits are easily overlooked but all contribute to instigating better use of machinery and reducing the running costs for users. Some of the hidden benefits which have been experienced are:

° regular sampling ensures that there is lubricant in the system
° the oil is in a state whereby it will serve the required function
° reduction or elimination of unexpected failures allows better use of manpower
° regular monitoring of mature machines, where redundancy was considered has allowed considerable life extension.
° allowing slightly degraded machines to be run until they can be conveniently rectified, whereas previously this machine may have been immediately taken out of service with resultant loss of availability.
° due to the reduction in failures, the high costs of transport, storage and handling of spares and sub assemblies is greatly reduced. In the mining industry this figure is estimated to be as high as 22% of the life cycle costs of machines
° by sampling a unit immediately prior or after entering service an indication of the quality and workmanship during maintenance is obtained. At one site this has led to an improved standard of maintenance both internally and by external contractors.

7.0 CONCLUSIONS

Several methods exist for the monitoring of machine health. Present experience indicates the simple method of extracting oil samples periodically and monitoring the rate of wear, together with measuring contamination and fluid degradation are particularly appropriate to industrial applications. The ability to conduct testing on site is shown to be both a relatively simple and cost effective option, together with providing a number of benefits which would not have accrued by contracting the samples to a Lube analysis laboratory.

A set of instrumentation has been described which is both simple to use and measures the primary deterioration mechanisms within industrial systems. These instruments make up what is known as the Mini-Lab. The role of the Mini-Lab is one of problem identification (detection) with limited diagnosis, where it is suggested that the role and function of the advanced fluids analysis lab has a role of more advanced problem definition and diagnosis. This combined with the capability of providing recommendations as to eliminating the root cause of the problem.

References:

1 Moreton, G and Yardley, E D-The Use Of Wear Debris Analysis to Monitor Gear Transmissions Underground. Proc. of the Conf. on Condition Monitoring, 1984, p491
2. Lester M F and Brooks A - Debris Testing Progress and Further Developments. Proc. of the Conf. on Condition Monitoring 1987
3 Fogel A G and Wright G.J. -Case Study: The Application of wear Debris Analysis for Monitoring Underground Continuous Coal Miners. Conf. Proc. of " Tribology 86 "
4 Fogel A G and Wright G.J. -Refining Wear Debris Monitoring as a useful Applied Technology in Maintenance Management. Conf. Proc. of " Tribology 88 "
5 Fogel A G. The Impact Of Condition Monitoring In The Next Decade.
SA Mech Engineer Nov/Dec 1989.
6 Fogel A G and du Toit A.- Exploiting Failure for Management Advantage. Proc. of Tribology 90.
7 Information - Bridging the Gap Between Condition Monitoring and Predictive Maintenance. G Fogel and T Hayzen 1990 Maintenance Management Convention (SA Institute of mechanical Engineers)
8 Condition Based Maintenance Using Oil Debris Analysis. G Fogel National Productivity Conference 1990

9.Third Generation Oil Analysis G Fogel and F Lourens SA Mechanical Engineer September 1992.
10.Condition Monitoring - Where to Now? G Fogel and T de Sousa Mechanical Technology May 1993
11 Wear Debris Atlas Published CMS International Aug 1993
12.Proactive Contamination Control G Fogel Lubrication Engineering Notes 1993
13 SOAP, Wear Debris Analysis, Particle Counting, how do we really measure contamination .G Fogel SAIT Workshop on contamination Control

LUBRICANT CONDITION MONITORING #2

Considerations in Applying Molecular Analysis of Different Fluids in a Condition Monitoring Program

Jay R. Powell
Bio-Rad Laboratories, Digilab Division
237 Putnam Avenue
Cambridge, MA 02139

Abstract: Molecular analysis of oils by Fourier Transform Infrared (FT-IR) spectroscopy is becoming commonplace in condition monitoring programs. Providing a fast, direct measurement on species and properties of interest, molecular spectroscopy can supplement or replace other measurement techniques which are time consuming, or provide only an indirect measurement of the properties of interest. Molecular spectroscopy also offers other capabilities, such as calibration to classical physical tests, or the identification of an unknown fluid by searching against a spectral library of known fluids. Here, we will review some of the promises and pitfalls in molecular analysis for a condition monitoring program, with emphases on the general capabilities of molecular spectroscopy, and the amount of useful information (to a condition monitoring program) which can be generated when progressively less information is known about the machine being monitored.

Key Words: Condition monitoring; oil analysis; molecular spectroscopy; FT-IR, infrared; physical property test.

INTRODUCTION: Molecular analysis of new and used lubricating oils, hydraulic fluids, and fuels by Fourier Transform Infrared (FT-IR) spectroscopy is rapidly becoming commonplace in condition monitoring programs. Providing a fast, direct measurement on species and properties of interest, such as water, soot loading, coolant, oxidation, etc., FT-IR spectroscopy can supplement or replace other measurement techniques which are time consuming and generate additional wastes (such as Karl-Fisher titration for water and pentane insolubles for soot/dirt/carbon loading), or provide only an indirect measurement of the properties of interest (such as TAN/TBN/NN for oil breakdown and contamination). Infrared spectroscopy also offers other capabilities, such as calibration to classical physical tests (such as kinematic viscosity and TBN) by Principle Component Regression and Partial Least Squares (PCR/PLS), or the identification of an unknown fluid by searching against a spectral library of known fluids.

When applying molecular analysis in a condition monitoring program, the primary point to remember is that the information generated from the FT-IR spectrometer (or any other technique,

for that matter) is to be used as a reliable, consistent source of information on the health of a system, not as an isolated exercise in pure analytical chemistry and instrumental analysis. One unfortunate misapplication of FT-IR spectroscopy in condition monitoring has been the use of spectral search (also called scan and search) to identify unknowns, and to attempt to use the search and identification results for condition monitoring purposes. As will be shown, this application will produce useful information only as long as nothing is wrong! Another desire in condition monitoring is to have the analyzer operate as much like a "black box" as possible, without the need for adjustments or modifications for the analysis of different fluid classes. An example of some instruments that approaches this ideal are the Arc and ICP spectrometers used for wear metal analysis. As these instruments can produce results for wear metal levels without knowledge or adjustment of the fluid characteristics, another unfortunate assumption is that any spectrometer, such as an FT-IR spectrometer, produces results in the same manner. However, even the Arc and ICP spectrometers do show a sample dependent response, and must be taken into account even more so when applying molecular analysis for condition monitoring. Finally, as many established condition monitoring programs which use the traditional physical property tests implement FT-IR spectroscopy, one "roadblock" they encounter is the desire for the results to be expressed in units and formats the maintenance personal are familiar, either for compatibility with their current expert system database or for simple "comfort". One way infrared spectroscopy can be quickly implemented in these programs is to use the power of Principle Component Regression / Partial Least Squares analysis (PCR/PLS), which can predict the physical response based on the infrared spectrum. PCR/PLS has been successfully used to generate predicted vacuum pentane insolubles, Total Base Number, and viscosity, in several condition monitoring programs. This allows the programs to begin implementation of the "newer" technology, reducing per-sample analysis costs, without forcing those who receive and act on the laboratory results into a sudden reeducation program.

AUTOMATIC SAMPLE IDENTIFICATION: Infrared spectroscopy has long been used in analytical chemistry for the identification of unknown compounds. As the infrared spectrum of each molecular species is unique, the identity of an unknown compound can be determined by comparison to a library of known compounds. These libraries were first distributed in large reference books, and later as compressed, computer searchable collections. The use of infrared spectroscopy in conjunction with easy-to-use search software and these library collections for unknown identification is one of the largest reasons FT-IR spectroscopy is so widely accepted in analytical laboratories.

As such capabilities are found in just about every modern infrared spectrometer, one application of infrared spectroscopy with search and identification has been to reduce the amount of paperwork and data entry in a condition monitoring program. In this application, it is assumed that the spectrometer can identify the type fluid from the infrared spectrum of the sample. From this identification, the system can then automatically determine the appropriate infrared analysis techniques to apply, and the condition alert or alarm limits, without the operator needing to keep track of the source of the sample and the characteristics of the machine. As noted above however, the objective is to provide useful information on the condition of a mechanical system, not a chemical analysis of submitted unknowns. Figure 1 shows the infrared spectrum of an unidentified fluid (top), and a small "library" of known fluids. While infrared spectroscopy can easily identify

the fluid as a typical petroleum based crankcase lubricant, the real question to be answered is: "is this machine operating within normal limits?"

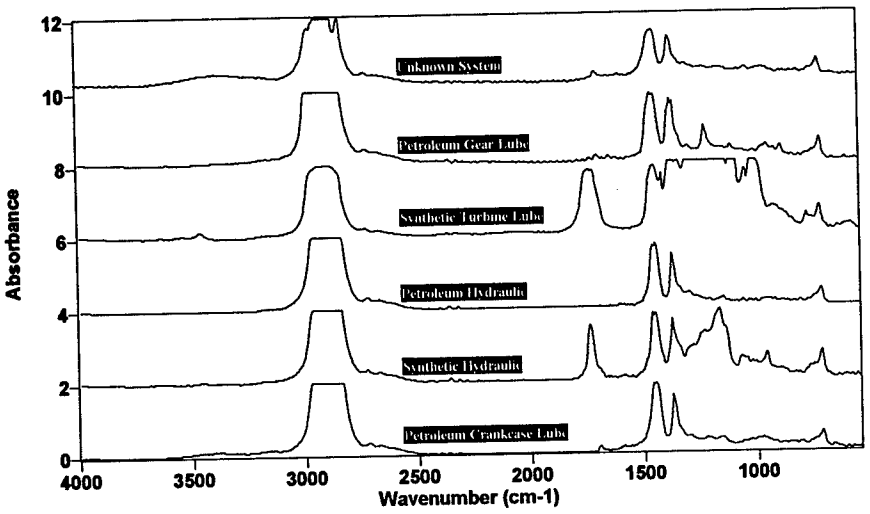


Figure 1. Does the fluid from the "Unknown System" (top) indicate that the machine being monitored is functioning within limits?

In this case, the correct answer can not be determined. While the spectrometer and software can identify the fluid as most similar to a petroleum crankcase lubricant, the analysis information would only be useful if the sample actually came from an engine crankcase. The obvious failure of the search and identification approach is the inability to alert the operators when the wrong fluid is used in a machine. While this may be an uncommon mistake in the real world, even one occurrence may place the component in unacceptable risk, not to mention the total system, productivity, associated property, and lives. A more common example is the case where an engine crankcase oil sample is correctly determined to be a petroleum crankcase, and the soot or dirt level is determined by the absorbance at 2000 cm^{-1} (Figure 2). Once again, the question is not "what is the soot/dirt/carbon level", but "is this machine operating within normal and acceptable soot/dirt/carbon levels?" Once again, without knowing if this sample came from a diesel engine, gasoline engine, or a natural gas engine, the answer can not be determined. Diagnosis of the condition of a machine can not be made when the requirements and characteristics of the machine are unknown.

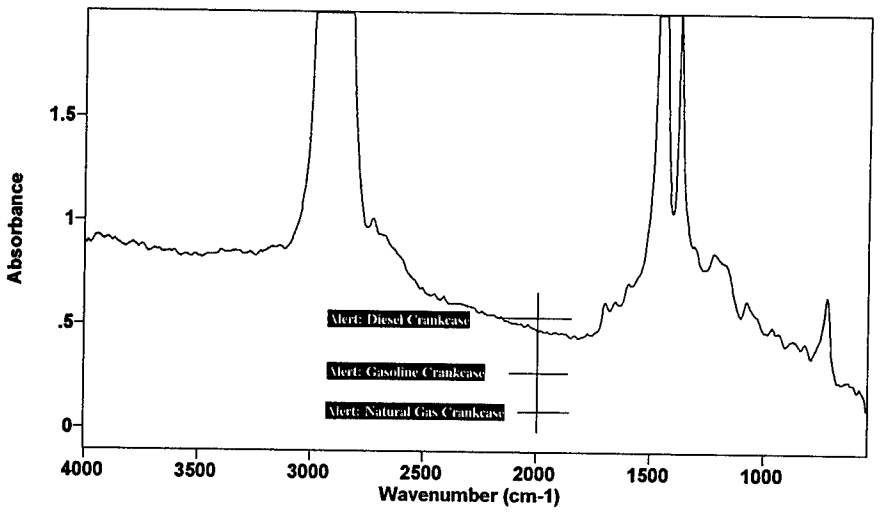


Figure 2. Does this engine have an excessive soot loading?

THE "BLACK BOX" APPROACH: The ideal analytical instrument operates without any user input or interaction, and produces accurate, informative results irrespective of the conditions of use, knowledge of the sample, source of the sample, chemistry of the system, or information from other sources. Part of this has been visited above. One example of some instruments that approach this ideal are the Arc and ICP atomic emission spectrometers used for wear metal analysis. While at first it may seem that these instruments produce informative results irrespective of the sample, in fact these instruments used in condition monitoring programs have been typically optimized for the analysis of the lubricants. Thus, much of the information and knowledge of the samples analyzed has already been built into the hardware and software of these instruments.

Not even atomic emission spectrometers provide accurate measurements of wear metals in lubricants. Because of the different energy characteristics of the excitation sources in arc and ICP spectrometers, different emission lines may be used for some elements. Thus, measurement of a given metal may be performed using different emission lines with different relative intensities. Arc and ICP spectrometers correct this variation, as well as producing results in concentration units, by calibration with organometallic compounds dissolved in a base oil matrix. Even with calibration, the characteristics of sample introduction to these different instruments will produce different results. The nebulizer and plasma on the ICP spectrometers are unable to generate a signal from particles less than 3 microns in size, as compared to emission signals which can be generated from particles 10 microns and possible larger from arc spectrometers. [1] Thus, for the same lubricant sample with wear metal particles the two different types of spectrometers will

produce different results for wear metal levels. These two different instruments are useful in condition monitoring programs because most condition monitoring programs trend the wear metal levels from on type of instrument, or in essence, the spectrometers are always measuring one consistent, well defined matrix: a high temperature plasma.

Molecular spectroscopy can also achieve this "black box" level, if as noted in the last sentence above, a single, well defined matrix is measured by the FT-IR spectrometer. However, many condition monitoring programs perform analysis on a wide range of fluids from different systems, and thus different samples constitute a different matrix measured by the spectrometer. Previous work has demonstrated the differences in response in the infrared absorbance of water in two different types of lubricants: petroleum based crankcase lubricants and synthetic turbine lubricants. [2] Figure 3 shows the difference in the hydrogen bonded OH (water) absorbance region for the same amount of water in a typical petroleum based crankcase lubricant and

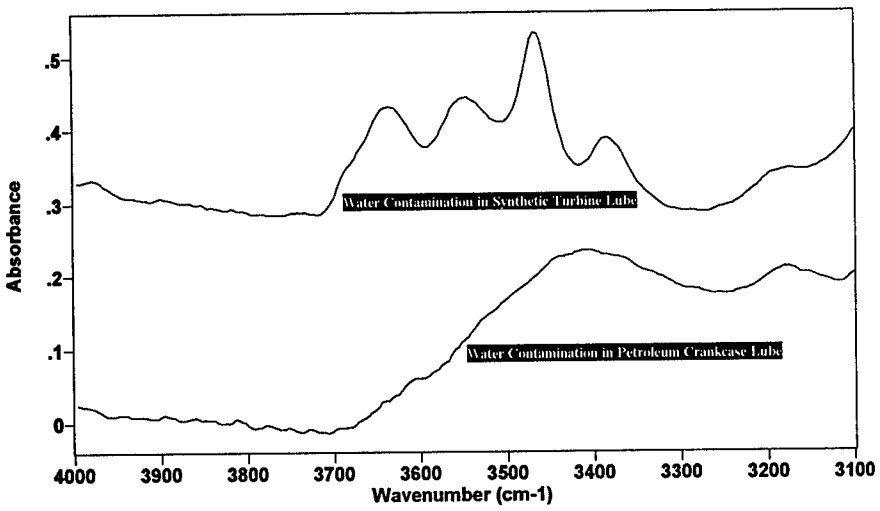


Figure 3. Water contamination at 2500 ppm in petroleum based lube (bottom) and a polyol ester based lube (top).

a polyol ester based synthetic turbine lubricant. As can be seen, there are significant differences in the water response between the petroleum based lubricant and the synthetic based lubricant, mostly due to differences in hydrogen bonding between the water and the lubricant and additive package. Assuming that water will give the same response independent of the matrix in which it exists clearly will lead to erroneous results.

A common technique for observing individual components in a strongly absorbing matrix is to use computer spectral subtraction, which produces spectral results that are similar to using a pair of matched cells on an old scanning double-beam spectrophotometer. Spectral subtraction is then used to remove the spectral response of the base oil, with the assumption that this mathematical technique removes the interactions from the base matrix (the oil) with the sample response (component of interest), and thus the same quantitative analysis methods and calibrations can be used independent of the class of lubricant. [3] Figure 4 shows the danger of

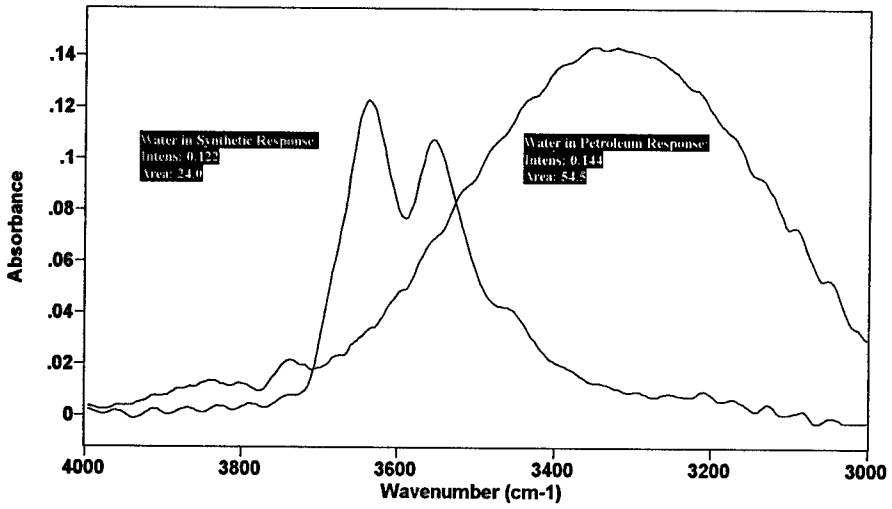


Figure 4. New oil subtracted spectra, each with 1000 ppm water.

this assumption. Both lubricants, a typical petroleum crankcase lubricant and a polyol ester synthetic (Mil-L-23699) were spiked with 1000 ppm of water, and then the new reference oil subtracted. Classical infrared quantitative analysis relates either the intensity of the band or the area under the band to the component concentration. As can be seen on the figure, a calibration based on an intensity / concentration correlation for water will result in a nearly 20% error, and a calibration based on the area under the curve using water in petroleum standards will be in error by 50%! While it is possible to set up instrument such that it produces a number no matter what type of sample it is presented, one should be wary of such an analysis technique without further investigation.

ADVANCED QUANTITATIVE ANALYSIS: While molecular analysis using infrared spectroscopy offers great promise in condition monitoring programs, there is always reluctance to employ new analytical methods because of either the familiarity of the old methods, machinery supplier guidelines and recommendations, or because the various maintenance rules and

guidelines, either human interpreted or as dictated by an expert system, require results from the old methods. Many of these tests for molecular species of interest are based on what is generally referred to as physical property tests, where a physical measurement of the lubricant is related to a machine condition or recommended action. These tests usually involve "wet" chemical sample preparation and analysis, such as measuring the soot or carbon loading by diluting a sample of used oil with pentane, filtering or centrifuging the solids out, weighing the separated solids, and then calculating the percent solids in the oil. While infrared spectroscopy can measure the amount of soot based on the amount of baseline offset and tilt due to scattering of the suspended particles, one limitation is that such a measurement is affected by both the concentration of the suspended soot and the average particle size and distribution, making a classical infrared correlation to percent insolubles difficult. [4]

Besides the classical infrared quantitative analysis tools, relating the intensity or area of a single band to the desired property, statisticians have created a wide range of advanced interpretation and analysis tools such as Principle Component Regression and Partial Least Squares (PCR/PLS) analysis. Rather than "reducing" a single band in the infrared spectrum to a single number based on the intensity or area, PCR/PLS analysis tools use the entire "profile" of the infrared spectrum, and with a sufficient training set of spectra, can often create surprising correlations between the infrared response and a desired physical property. One example of this was recently encountered at CSX laboratories, where the infrared soot measurements from a smaller set of EMD engines

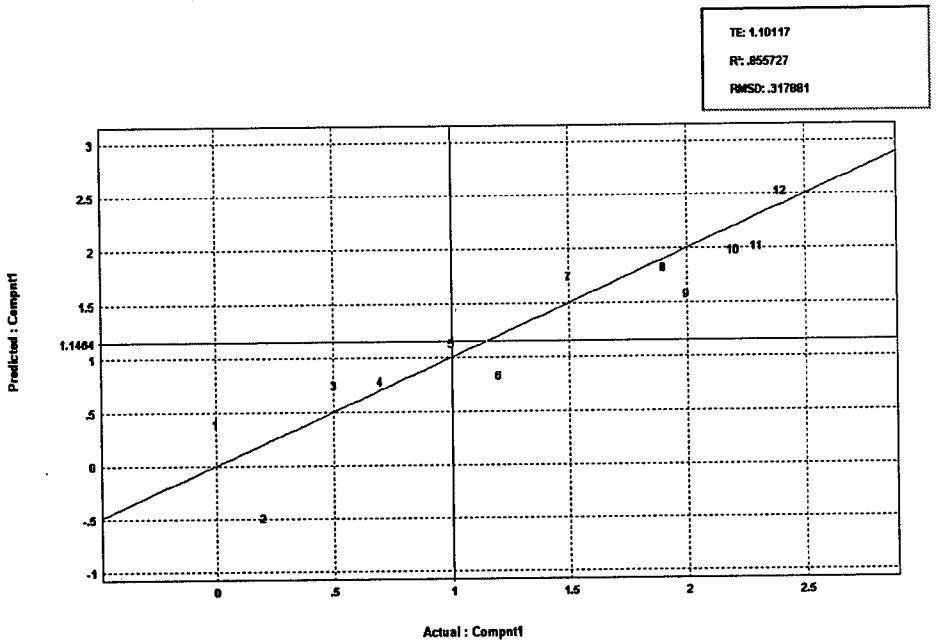


Figure 5. PCR/PLS correlation for vacuum pentane insolubles.

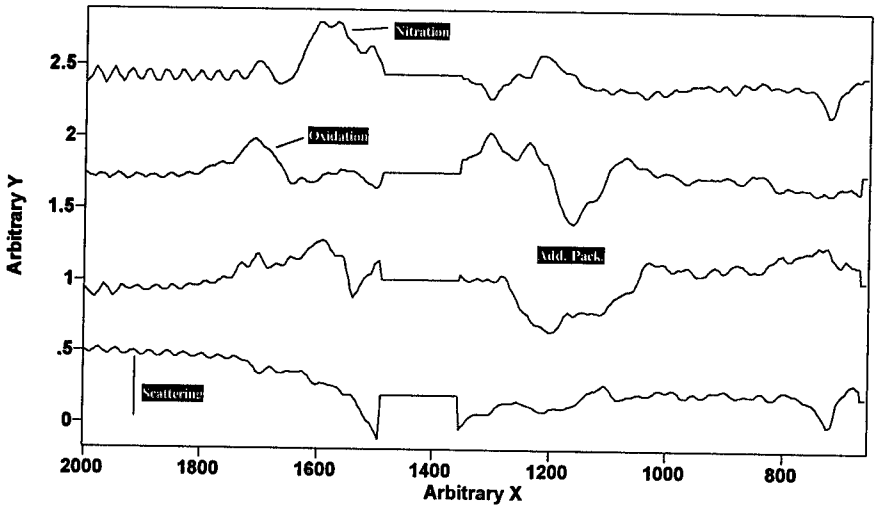


Figure 6. Principle component spectral factors influencing vacuum pentane insolubles measurement.

did not correlate to the infrared soot measurements from the GE engines. This too was traced back to average particle size and distribution differences, as the EMD and GE engines used different filters with different pore sizes. The PCR/PLS analysis tools were able to generate an acceptable correlation to their vacuum pentane insolubles test (Vac. P.I.), to within +/- 0.5% (figure 5), as compared to the wet chemical method reproducibility of +/- 0.2 % across the range. While of lower precision, the savings in manpower and solvent usage / disposal more than justified the change. The PCR/PLS also allows an examination of some of the molecular level components and interactions which are related to the physical property of interest. In figure 6, some of the principle component spectra generated from a PCR/PLS analysis of used samples from 62 EMD engines, covering the range of Vac. P.I. readings, are displayed. From inspection of the spectra, some molecular level correlations and influences can be determined. For example, the first component (bottom spectrum in the figure) shows the classical baseline offset and tilt due to the scattering from the soot particles. However, the PCR/PLS analysis uses not only this factor, but some of the other factors plotted above. The second spectra (factors) from the bottom show a correlation between the additive package and the vacuum pentane insolubles measurement, while the third and fourth spectra (factors) show a correlation between the oxidation and nitration to the Vac. P.I. response. Both of these correlations become obvious when one reflects that the Vac. P.I. levels will increase as the additive package components, especially the detergents and dispersents, decrease, and the Vac. P.I. levels will increase with the age of the oil charge, which will also show an increase in the oxidation and nitration due to age and use. The PCR/PLS

correlation uses the above calculated factors, along several other factors not displayed, to produce a predicted Vac. P.I. sufficient for the maintenance needs.

While generating a correlation to the pentane insolubles was shown to be feasible, other classical wet chemical techniques may give one pause. One example of this is the Total Base Number (TBN) titration, which is used to measure the amount of acid species in an oil, with the assumption that the acid species are due to oxidative breakdown of the oil. The sample is titrated to an end point, and the actual measurement is expressed as milliequivalents of potassium hydroxide per gram of sample. While acid generated species from oil breakdown are measured, there are a considerable number of possible internal or external species which can also give rise to acid species. In fact, while ASTM method D 5372 notes the use of this titration test, it also notes that "the use of infrared analysis may serve as a more reliable method for detection of oxidation". [5] In this case however, the laboratory and data system management needs dictated

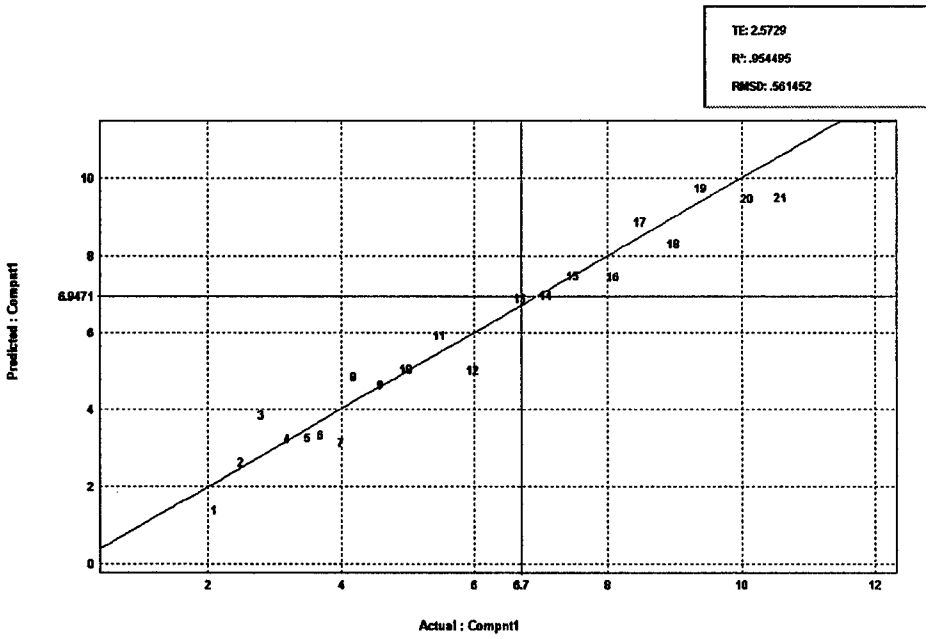


Figure 6. PCR/PLS correlation to TBN

that the new system continue to produce TBN measurements for ensured compatibility with current procedures. Here, PCR/PLS analysis once again produced an acceptable correlation with the desired physical property (Figure 6).

In both these cases however, note that the typical fluids analyzed represented a very narrow range of all possible lubricants which may be encountered in a large condition monitoring program. All the above sites used a relatively restricted range of lubricants, typically specifying both base oil characteristics and a particular additive package blended at a specific level, and thus the number

of variances that the PCR/PLS analysis tools had to account were relatively limited. In an independent, commercial condition monitoring laboratory, where a much larger range of petroleum based lubricants are used, the size of the training set must be increased to account for the additional variances. While PCR/PLS analysis offers the promise of correlating the infrared spectrum to many physical properties of interest, anyone embarking on this path should be aware that these tools can only accurately predict characteristics in an unknown if it has "seen" that type of unknown previously. As in the above example of applying the calibration of water in a petroleum based lubricant to the analysis of a synthetic lubricant will lead to errors, similar misapplication of advanced analysis tools will produce similar misinformation.

REFERENCES:

1. Toms, L., "Machinery Oil Analysis: Methods, Automation, and Benefits", Larry A. Toms, Pensacola, FL, 1995.

2. Toms, A., "A Preliminary Report on the Evaluation of FT-IR for Lubricant Condition and Contamination Determination in Support of Machinery Condition Monitoring. I. Synthetic Lubricants", Proceedings of Condition Monitoring '94, ed. M. Jones, Pineridge Press, Swansea, UK, 1994.

3. Garry, C., Bowman, J.C., Leimer, B.L., "Analysis of Synthetic Lubricants by Fourier Transform Infrared Spectroscopy", Proceedings of the JOAP International Condition Monitoring Conference, ed. A. Toms, JOAP-TSC, Pensacola, FL, 1992

4. Garry, M.C., "Applied Interpretation of FT-IR Oil Analysis Results for Improving Predictive Maintenance Programs", Proceedings of the JOAP International Condition Monitoring Conference, ed. M. Squalls, JOAP-TSC, Pensacola, FL, 1994

5. "D 5372-93 Standard Guide for Evaluation of Hydrocarbon Heat Transfer Fluids", Annual Book of ASTM Standards, V 5.03, ASTM, Philadelphia, PA, 1995.

Use of FT-IR Spectrometry as a Replacement for Physical Property Testing of Railway Lubricants

David E.G. Crutcher
255 Montee des Trente
Mt. St. Hilaire, Quebec
CANADA J3H 2R7

Roger Gervais
Bio-Rad Laboratories (Canada) Ltd.
5671 McAdam Road.
Mississuaga, ONT
CANADA L4Z 1N9

Larry A. Toms
Larry Toms Technical Services
5018 High Point Drive
Pensacola, FL 32505

Abstract: The railway industry has long used physical property tests to determine the condition and suitability of lubricating oils for continued use in locomotive diesel engines and air compressors. This paper describes the adaptation and implementation of FT-IR spectral analysis as a means of determining lubricant physical and chemical parameters at Canadian National Rail (CNR). Specifically the paper reviews the application of traditional FT-IR measurements for fuel dilution, water contamination, insolubles loading, and oil oxidation, and describes the implementation of new measurement parameters for viscosity and total base number. A study of atomic emission and FT-IR data from over 8000 locomotive samples provides summary data on the reliability of the FT-IR for routine diesel engine oil analysis.

Key Words: Condition Monitoring, Oil Analysis, Railway, FT-IR, Infrared, Physical Property Test.

INTRODUCTION: The North American railway industry was one of the first to use oil property measurements to support engine oil condition monitoring. Routine property tests reveal the degree of contamination and degradation of the lubricant, and these data can be related to common failure mechanisms to determine the risk of damage and any required action. Canadian National Rail (CNR) operates a fleet of 1700+ locomotives. Oil analysis services are provided internally by

5 laboratories. The traditional tests used by CNR for diesel engine oil analysis were kinematic viscosity @ 100°C, total base number (TBN), percent pentane insolubles and percent water. In addition, atomic emission spectroscopy is performed to determine the level of wear metals and other data, such as coolant inhibitor contamination, dirt and silicas, etc.

In recent years, CNR has upgraded oil analysis technology in order to maintain high levels of data integrity, repeatability and reproducibility, and reduce costs. By mid 1993, CNR had state-of-the-art methods, data management and expert systems, and felt the next phase of improvement would be to find a replacement for the physical properties test methods. These methods were time consuming, potentially hazardous, subject to error, and required the operation of individual instruments for each parameter measured. While traditional property tests are available in automated forms, anticipated capital costs and labor savings did not justify renewal of these test methods. However, if a single instrument could be used to determine all property data, over $700,000 per year would be saved at CNR in laboratory labor alone.

Recent published work [1,2,3,4,5] presented significantly improved FT-IR application software, and the new FT-IR systems showed promise as a single-instrument-replacement for the traditional oil condition tests. In addition to measuring oil contaminants, the new FT-IR system provided data about oil degradation (oxidation, sulfation, nitration, etc.).

CNR FT-IR DEVELOPMENT: Before considering FT-IR practical, the CNR Mechanical Department specified six parameters to be determined from each sample analysis. These are viscosity, percent fuel dilution, percent insolubles, percent water contamination, TBN, and oxidation level. Obtaining reliable FT-IR absorbance data for each parameter was only part of the task. CNR mechanical personnel were familiar with oil data in traditional units of measurement and wanted the absorbance data reported in the same units, and required the conversion of oil contaminant parameter data (fuel, water & insolubles) and the synthesis of viscosity and TBN from the FT-IR absorbance data. Moreover, a means to extract viscosity and TBN data was not commercially available and would require the development of suitable analysis and calibration methods. Oxidation was the only new parameter in the CNR program and is reported in integrated absorbance units.

All commercial FT-IR spectrometers offer a variety of infrared ranges, sensitivities, sample introduction methods and other features. After detailed review of the major FT-IR manufacturers instruments and capabilities, the Bio-Rad FTS-155 was selected in April 1994. Each instrument was equipped with a 100um flow transmission cell, an autosampler and a calibration matrix to convert absorbance data into the traditional units. To avoid the problems involved with spectral subtraction and the requirement for up-to-date reference oils, CNR chose a data trending methodology [1,2] where the area under the infrared curve for each oil parameter is measured and reported. The parameter trends are then calculated using prior measurements and engine usage data. Using this simple trending method, in conjunction with statistically based level and trend limits for each engine class, the need for operator interpretation of sample spectra was eliminated. In addition, the conversion of the spectra into properties data permitted interface of the FT-IR to the existing data management system without significant modification. This greatly reduced implementation costs.

CNR Engine Oils: The engine oils used at CNR include multi-grade SAE 20W40 and single-grade SAE 40 blended with Chevron Oronite Division 2990 additive package at 17 TBN level. The oil is supplied by two different manufacturers and is produced by two different refining processes. The differences in the oils required the preparation of standards from both manufacturers to achieve good calibration reliability.

Calibration Standards: Calibration standards for each measured parameter were provided by the following organizations:

oxidation - Chevron Chemical Company, Oronite Additives Division,
10, 13, 27, 35, 64, 116 - ABS (integrated absorbance units)

viscosity - original oil manufacturers,
13.0 to 16.5 cSt @ 100°C in 0.5 cSt increments

TBN - original oil manufacturers,
1, 3, 5, 7, 10, 12, 15, 17, 20 total base numbers

%insolubles - CANAC International,
0.5% to 5.0% in 0.%5 increments

%water - CANAC International, and;
0.05% to 0.5% in 0.05% increments

%fuel - CANAC International.
1.0% to 10.0% in 1.0% increments

The remainder of the paper will describe the extraction of oil condition parameters, the areas and baselines selected and the calibration methods used to convert the data into traditional reporting units.

Viscosity: At CNR, oil viscosity was measured by the ASTM D445 method, utilizing a calibrated glass capillary viscometer immersed in a heated bath at 100°C . Although multiple samples could be analyzed simultaneously, the method was labor intensive. In addition, experience has shown that the measurement is only reliable when the oil is relatively free of contamination such as free water. CNR required the FT-IR method to indicate viscosity changes due to fuel dilution and viscous shear. This was accomplished by analysis of changes in oil chemistry using PCR/PLS methods to integrate several measurements over the range 2500 to 650 cm^{-1}.

The viscosity calibration was established from prepared samples of 17 TBN 40 and 17 TBN 20W40 oils over the required viscosity ranges, with additional standards using both viscosity and fuel dilution prepared standards and 60-80 samples of used oil. The viscosity of the used samples was measured by the ASTM D445 method and the fuel concentration was calculated from the viscosity data. This calibration matrix provided the best correlation between FT-IR and ASTM D445 results.

Figure 1 shows the frequency distribution obtained from 4446 routine samples (one year of operation) from 175 EMD locomotive engines. Over 75% of the samples range from 14 to 15 cSt--the normal operating range of the engine oils used at CNR. Since the locomotive fleet ranges

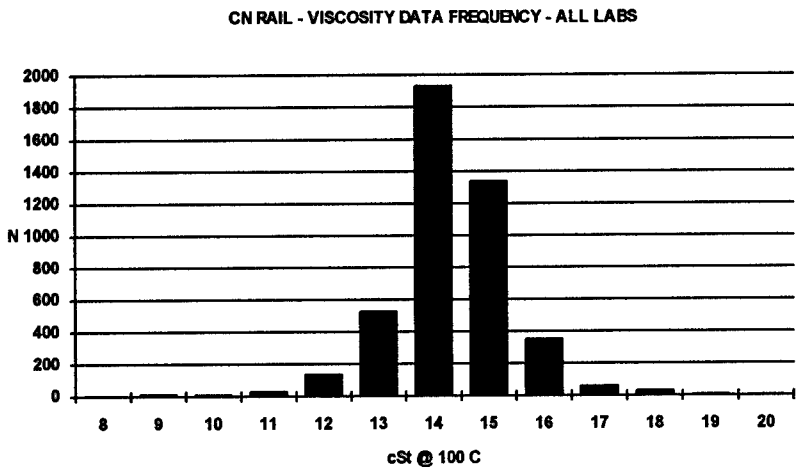


Figure 1. Frequency distribution for FT-IR predicted viscosity obtained 175 EMD engines.

across the country, the sample population includes measurements from all instruments. The low variability shown, thus indicates the high reproducibility in the FT-IR viscosity measurement.

Fuel Contamination: At CNR, fuel dilution was traditionally calculated from the ASTM D445 viscosity measurements and suffered from any problems inherent with that method. It was assumed that any decrease in viscosity was due to fuel contamination, although viscous shear and/or thermal cracking could result in the same symptoms.

CNR required the FT-IR to indicate the concentration of fuel contamination and report the data in percent. This was accomplished using PCR/PLS analysis over the range 850 to 760 cm^{-1}. The percent fuel calibration was established from prepared samples of used (uncontaminated) 17 TBN 40 and 17 TBN 20W40 oils spiked to cover the required range of fuel measurement. Figure 2 shows a classic example of a severe fuel dilution problem in engine CNR 5406. The viscosity and %fuel trends clearly show the sensitivity of the FT-IR measurements to the presence of fuel in the oil.

While the calibration model and frequency distribution data indicate a high correlation between viscosity generated and FT-IR generated fuel contamination data, in practice, the FT-IR failed to reliably indicate fuel contamination at laboratories located in Western Canada. The fuel measurement problem was found to be the result of a lower aromatic content in the fuels used in the Western Canada region. While the aromatic peak (810 cm^{-1}) of the fuel was measured accurately by the FT-IR, the level did not correlate with the actual fuel contamination as indicated by oil viscosity decrease. Since the current FT-IR method uses the 810 cm^{-1} peak for fuel quantification, the reliability problem was attributed to the variability in the aromatic content of the fuels used by CNR. As the spectra in Figure 7 shows, the 810 cm^{-1} peak of Western diesel fuel measures substantially lower then the peak from Eastern diesel fuel.

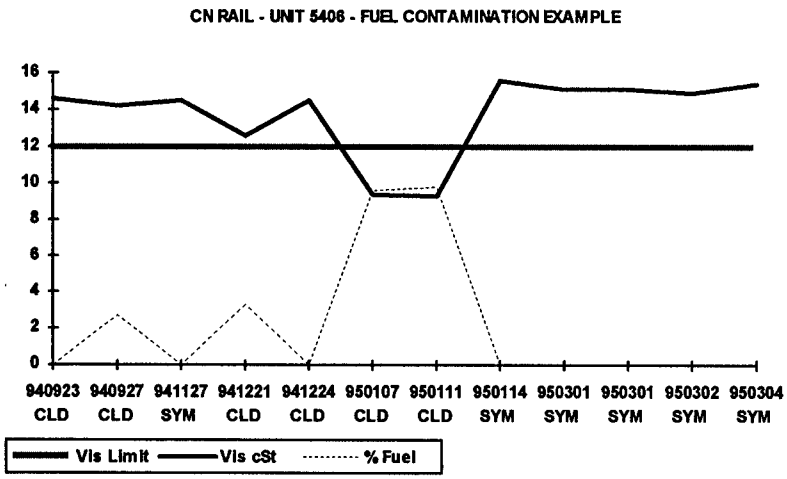


Figure 2. Unit 5606 fuel contamination. X axis in compressed date (yymmdd), with three character location code. Y axis in cSt @ 100C.

To compensate, the CNR FT-IR system at the Edmonton laboratory calculates the percent of fuel contamination from the FT-IR predicted viscosity data. The viscosity was not similarly affected by fuel aromatic concentration, and remained reliable. The FT-IR fuel contamination measurement is compared with the viscosity predicted fuel result and the larger value is reported. While lab-to-lab trending of fuel data was affected, it was determined that this was the best way to preserve engine components from fuel related damage modes.

Water Contamination: CNR laboratories traditionally used the crackle (hot plate) test and the presence of high levels of sodium and boron as the primary indicators of water in engine oils. Both methods suffer problems--the crackle test is non-quantifiable, insensitive and operator subjective, and any water ingress into the engine lubrication system may not be the result of a coolant leak, and thus, not show high sodium/boron levels.

As the most common contaminant found in diesel engine lubrication systems, water is also a very strong infrared absorber, and perhaps the easiest oil contaminant to detect. The hydrogen bonded O-H stretch vibrations (symmetric and asymmetric) of liquid water are found over the range of 3700 to 3100 cm^{-1} in petroleum based lubricants. [1] A general baseline over the range of 4000 to 2000 cm^{-1} will correct for the most common interferences, e.g. soot, insolubles, dirt loading, etc.

CNR required the FT-IR water measurement to indicate changes in concentration calibrated in percent. The calibration was accomplished using PCR/PLS analysis over the range 3700 to 3150 cm^{-1}. Prepared samples of 17 TBN 40 and 17 TBN 20W40 used (uncontaminated) oils were spiked with water to cover the required measurement ranges. Figure 3 shows the frequency distribution for the FT-IR percent water measurement for 175 engines over one year of operation.. As expected, most samples were free from serious water contamination. Figure 4

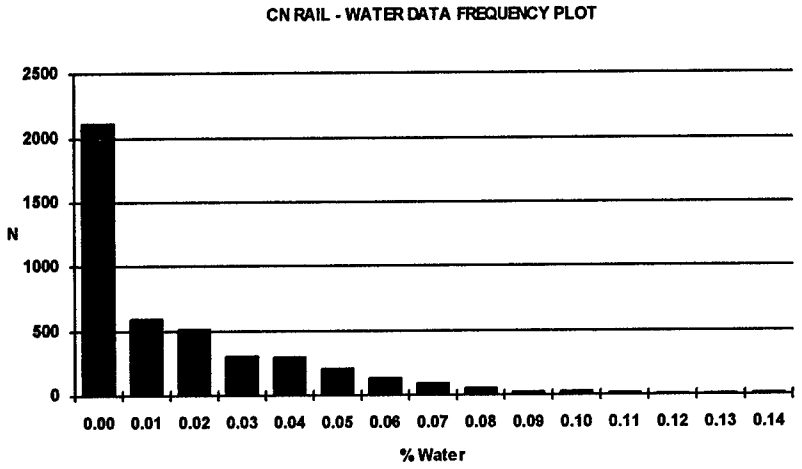


Figure 3. Frequency distribution of the FT-IR water measurement from 175 engines over one year of operation.

shows a classic example of two severe water leaks in engine CNR 5405. The percent water data is trended with the sodium and boron indicators of the coolant corrosion inhibitor. Notice the high correlation in the data trends.

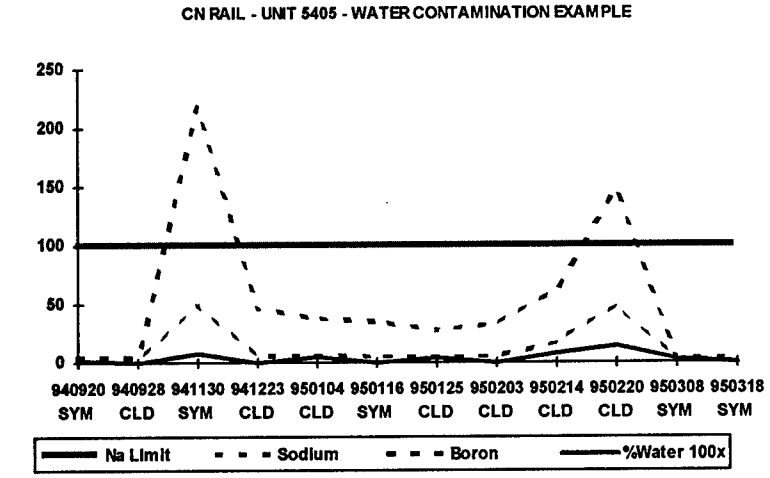


Figure 4. Unit 5405 water contamination. X axis in compressed date (yymmdd), with three character location code. Y axis in ppm for elements, % x 100 for water.

Overall, the FT-IR water measurements correlated well with other measurement methods such as emission spectroscopy (Na, B). In addition to coolant contamination, the FT-IR detected water which may resulted from other contamination sources. In these samples, the atomic emission data showed no change in Na or B levels. It is probable, these contamination occurrences were the result of condensation or fresh water ingress, and would have been missed prior to the implementation of FT-IR. While the Na and B trend data were routinely correlated to engine damage progress, the crackle test was only used as a go/no-go test. The FT-IR water contamination data can be trended over time and correlated to damage mode progress in a manner similar to the atomic emission data.

Insolubles: The FT-IR measurement of insolubles, such as soot, are not easily correlated to physical separation test methods, such as ASTM D-893 pentane and toluene insolubles [6]. The poor correlation is due to differences in particle size distribution, and differences in readings from the various ASTM methods, such as D-893, D-4055 and their many derivatives.

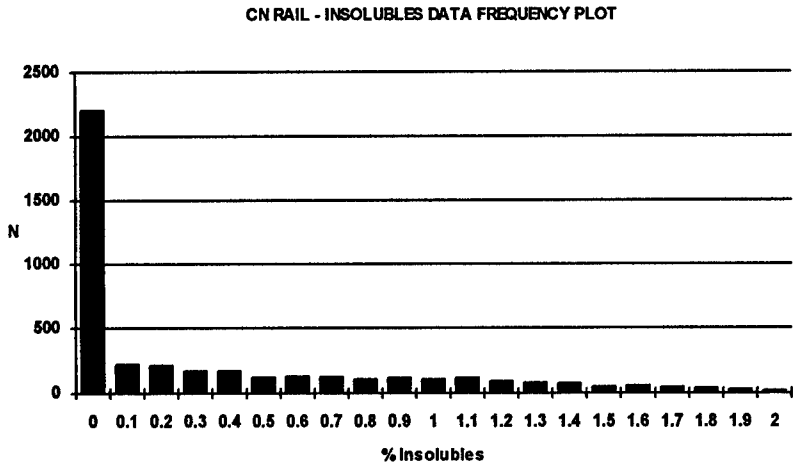


Figure 5. Frequency distribution of the FT-IR insolubles measurement from 175 engines over one year of operation.

CNR required the FT-IR analysis to correlate to the standard method developed by the Locomotive Maintenance Officers Association (LMOA). The LMOA method utilizes absorbance measurements from a UV-Vis-NIR spectrophotometer. The standard was established by a round robin analysis of sooted samples (0.2 to 6.0%) at a number of North American railway laboratories. Correlation of FT-IR generated percent insolubles with the LMOA standard was achieved by a PCR/PLS calibration over the range 3950 to 3700 cm^{-1} from prepared samples of 17 TBN 40 and 17 TBN 20W40 oils. In practice, the FT-IR analysis of used oil samples correlated very well to the LMOA method:

For values < 4%, Repeatability is 0.5% and Reproducibility is 1.3%
For values > 4%, Repeatability is 0.75 % and Reproducibility is 2.0%

Figure 5 shows the frequency distribution for the FT-IR percent insolubles measurement. As expected, most samples were free from serious contamination

Oil Oxidation/TBN: In diesel engine lubrication systems, the oil is exposed to high temperatures and stresses in the presence of oxygen, resulting in the formation of partially oxidized compounds. Antioxidants are added to inhibit this process by serving as free radical scavengers. Oxidation by-products are generally acidic in nature and usually result in a depletion of the reserve alkalinity of the oil. The alkaline depletion has been traditionally monitored by measurement of the total base number (TBN). In keeping with past practice, CNR required both TBN and oxidation level be monitored.

Oxidation by-products in petroleum lubricants absorb infrared energy due to the C=O stretch from by-products such as ketones, esters, carboxylic acids, carbonates, aldehydes, anhydrides, and amides. A measurement region of 1760 to 1660 cm^{-1} was determined to be acceptable for the oils in use and a calibration curve was generated from prepared samples of 17 TBN 40 weight oil covering the range of 0 to 100+ integrated absorbance units. In practice the levels of oil oxidation remained low, due in part to the high oil consumption and frequent make-up oil, or periodic oil changes as are recommended by the individual engine manufacturers.

Obtaining reliable TBN data correlating to current standard methods was more difficult, in part due to the variability of results from current ASTM methods. The ASTM D4739 test was determined to be the most applicable for analysis of used railway engine oils. A TBN calibration method was accomplished by a matrix of 10 prepared samples. The analysis is performed over the range 1500 to 1420 cm^{-1}.

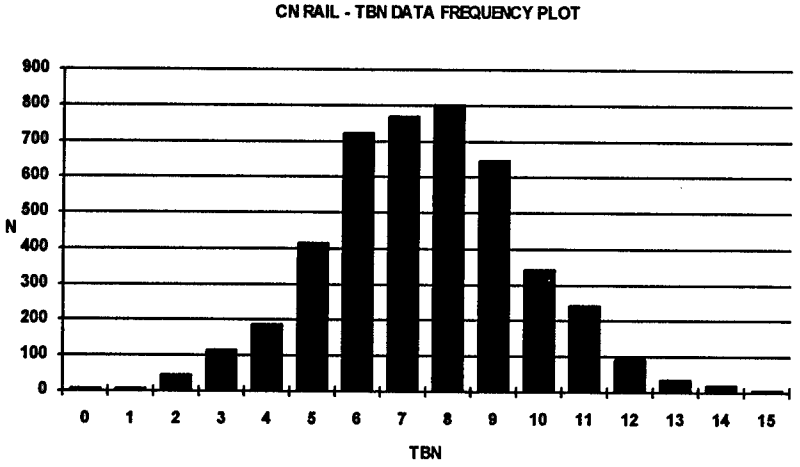


Figure 6. Frequency distribution of the FT-IR TBN measurement from 175 engines over one year of operation.

Figure 6 shows the TBN data frequency plot and indicates a relatively broad normal range spanning several base numbers. The FT-IR TBN area includes the CH_2 scissor band, a strong absorbance band, and it's variability may be causing some of the variability observed. Overall, TBN showed a general correlation with oxidation data although high oxidation occurrences were not always followed by a reduction in TBN. Make-up oil additions are not reported and the effect of infusions of fresh oil could not be determined. In any case, variability in the TBN data did not impair the oil condition assessment reliability and further effort to improve the measurement was not pursued.

BENEFITS: During a 12 month period over which FT-IR analysis was used in conjunction with wear metal analysis, the CNR monitoring program detected 119 cases of fuel dilution and 325 cases of water contamination. This was based on routine sampling from 1600 locomotives sampled every 200 to 250 hours of operation, or about once every eight days. Analyzing 80,000 samples / year with a total program cost of $1,200,000.00 translates to roughly $15.00 / sample. Historical data from CNR indicates that a fuel contamination problem which is allowed to progress to failure will cost an average of $17,000 to repair, and water contamination which is allowed to progress to failure will cost an average of $30,000 to repair. Assuming that the fuel and water contamination problems were not caught prior to failure, the 444 problems detected above could of incurred in excess of $11.7M in repair costs to CNR. Based on the laboratory operating costs and cost avoidance estimates, CNR is able to achieve roughly a 10:1 return on investment in their oil analysis condition monitoring program. An extended analysis of the financial costs and benefits in the CNR program has already been presented [7].

CONCLUSIONS: FT-IR analysis provides CNR with a fast, easy to operate method to determine engine oil properties data and is used to complement atomic (wear metal) analysis for a complete two-instrument oil condition monitoring program. The experience at CNR indicates the flexibility of FT-IR analysis software allows easy creation and implementation of new measurement parameters and was easy to interface to the existing computer infrastructure.

Viscosity measurement by FT-IR is practical. Although anyone attempting to duplicate this method should consider that, CNR maintains tight lubricant specifications and, while oils may be blended by different manufacturers, a single additive package is specified. TBN measurement by FT-IR also appears to be practical, although the above consideration also applies.

Contamination measurements met or exceeded the physical property tests replaced with the exception of fuel dilution measurements which appear to be dependent on the level of aromatics in the contaminating fuel. This problem was corrected by using the viscosity as an alternative fuel dilution indicator.

Calibration of the FT-IR to generate oil condition data in standard engineering units is practical, however, each instrument must be calibrated individually and may require re-calibration should significant maintenance alter an instrument's measurement profile. The major problem with using calibration matrices is the general lack of off-the-shelf standard fluids. Standards' prepared by individual FT-IR users may lack global consistency and this can lead to poor correlation between

different users' instruments. Poor correlation among users will entrust the new FT-IR method to the old controversies surrounding the lack of correlation among the various physical tests which are variously in or out of vogue. However at CNR, the calibration matrices developed met all goals established for the program. In addition, other authors have presented an alternative to calibration [1,2,3]. This methodology suggests the development of new alarms and limits based on the FT-IR absorbance data rather than converting the absorbances to engineering units so the data can be used with existing alarms. The choice seems to lay with the user.

REFERENCES

1. Toms, A., "A Preliminary Report on the Evaluation of FT-IR for Lubricant Condition and Contamination Determination in Support of Machinery Condition Monitoring. I. Synthetic Lubricants", Proceedings of Condition Monitoring'94, ed. M. Jones, Pineridge Press, Swansea, UK, 1994.

2. Toms, A. " JOAP-TSC Fourier Transform Infrared (FT-IR) Spectroscopy Evaluation", Proceedings of JOAP International Condition Monitoring Conference, ed. M. Squalls, JOAP-TSC, Pensacola, Florida, 1994.

3. Toms, A. " FT-IR for the Joint Oil Analysis Program - Uses, Advantages and Benefits", Proceedings of JOAP International Condition Monitoring Conference, ed. M. Squalls, JOAP-TSC, Pensacola, Florida, 1994.

4. Garry, M., "Fourier Transform Infrared Spectroscopy and its Role in Lubricant Analysis", Proceedings of JOAP International Condition Monitoring Conference, ed. A. Toms, JOAP-TSC, Pensacola, Florida, 1992.

5. Powell, J., "The Development and Application of Infrared Spectroscopy in Lubricant Condition Monitoring", Proceedings of JOAP International Condition Monitoring Conference, ed. M. Squalls, JOAP-TSC, Pensacola, Florida, 1994.

6. Desjardins J. and Seifert W., " Test Results Obtained by LEM, TGA and IR for Measuring Soot in Diesel Oil", Proceedings of Condition Monitoring'94, ed. M. Jones, Pineridge Press, Swansea, UK, 1994.

7. Crutcher, D. and Toms, L., "Fuel and Water Contamination Monitoring of Diesel Engine Crankcase Lubricants at CN", P/PM Technology, Minden, NV, (in press).

PARTICULATE/WEAR DEBRIS ANALYSIS #1

AN ALTERNATIVE CODING SYSTEM DEFINING THE TOTAL AND SEVERITY OF WEAR

M.H.Jones and J.Alberich
University of Wales Swansea

Abstract: A ferrous debris monitor is described which is capable of measuring the concentration of ferrous wear debris suspended in a lubricant and the severity of wear associated with particle size of this suspended debris. A coding system is proposed : PQ index(total wear):initial TDPQ slope(large particles):final TDPQ slope(small particles). Correlation with existing measurements is detailed.

Keywords: Ferrography; Ferrous Debris Monitor; Spectroscopy; Time Dependent Particle Quantifier;

Introduction: Users of spectrographic oil analysis have, for some considerable time, been aware of the limitation of this technique in relation to the accuracy of this method when analysing samples which contain particles larger than 5um. Emission, absorption and inductively coupled plasma spectrometers are each limited in their ability to analyse these large particles.

Direct-read ferrography, Wear Particle Analyser (Tribometrics), the ferrous wear debris monitor (PQ90) and the use of magnetic plugs have addressed this problem. These techniques are however, restricted to measuring ferrous wear debris. Most equipment, however, which generate large particles, gearboxes for example, produce ferrous wear debris and it is this material which is monitored.

The ferrous debris monitor (PQ90), which is a sensitive magnetometer, measures the quantity of ferrous wear debris in a sample without the disadvantages of time and dilution requirements associated with the other techniques. The unbalanced condition is measured and displayed as the 'PQ Index'. This index is measured in arbitrary units which may be correlated with the Dl and Ds ferrographic measurements. These measurement units, like all concentration measurements obtained by spectroscopy, should not be considered absolute measurements due to particle size sensitivity of each technique.

Correlation of the PQ Index with other measurements: The introduction of a new technique requires an assessment of the correlation of the measurement with measurements obtained by techniques already in use. The following results were prepared by Wearcheck, South Africa.

There was good correlation between the PQ Index and Iron measured by ICP for samples taken from diesel engines and transmissions. Both these types of component generate a higher percentage of small particles, typically less than 10um: correlation would therefore be expected. The case of drive trains (final drive axle units) indicates less correlation between the PQ Index and ICP measurement of iron than for either engine or transmission units. In practice, the drive train units generate a greater population of larger particles than for the other components.

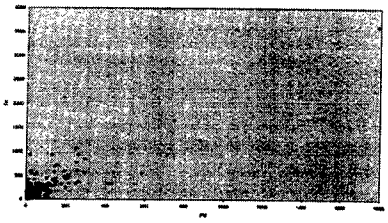


Figure 1 Correlation : 961 Engine Samples

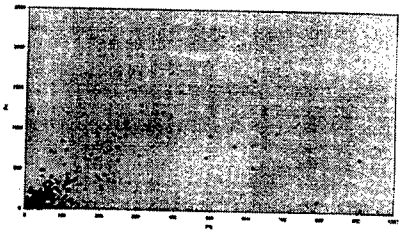


Figure 2 Correlation : 762 Transmission Samples

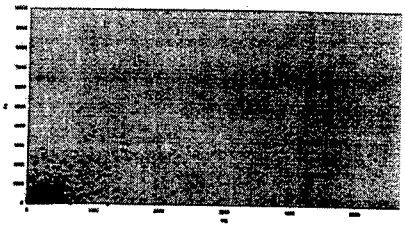
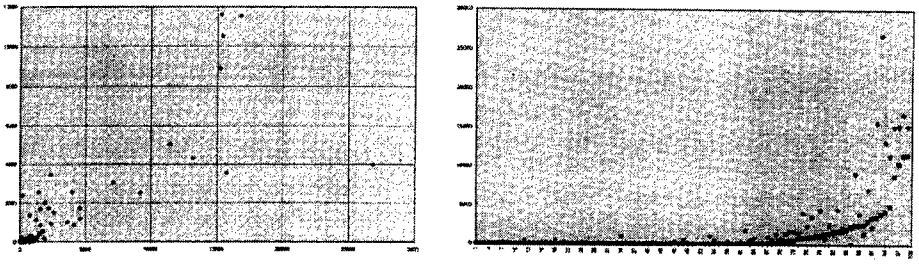


Figure 3 Correlation : 3310 Final Drive Damples

Direct - read ferrography measures the large and small particles by magnetic separation. The summation of Dl + Ds may be considered a measurement of the total wear debris separated from a fixed volume of sample. The PQ Index may also be considered a similar measurement. The volume of the sample is not however critical when using the PQ90 instrument.

The following figures show the correlation between Dl + Ds and the PQ Index for over 200 samples. These samples were taken from diesel engines, transmissions, drive units and hydraulic components.The correlation in Figure 5 has been drawn as a function of increasing values of the PQ Index. Both figures show good correlation between the two techniques.



Figures 4 and 5 Correlation between PQ Index and Dl+Ds Ferrography

The Case for a New Coding System: The development of the bottle method measures the sedimentation rate of particles in a sample of fluid which may be related to the severity of wear that has occurred in the sampled machine. A 20 to 30ml plastic bottle provides an adequate volume of sample for the ferrous debris monitor. As particles settle, the effective particle concentration will increase in the region subjected to the measuring field. The TDPQ method measures the magnetic out of balance at 10 second time intervals for a total of 100 seconds.

Previous reported results (1) of samples of oil with controlled concentrations and proportions of various size ferrous particles show, in a number of cases, a higher initial rate increase of the PQ index. The rate of increase of the PQ index decreases with time. Relating these rate increases to the particle size distribution suggests a possible method for measuring the severity of wear occurring within a machine.

Figures 6 and 7 show the higher initial rate of change of the PQ index. In these examples particles larger than 100um appear to have a higher rate of sedimentation, as perhaps would be expected, than the smaller particles. Particles less than 50um in size and in concentrations of 1000 and 500 ppm continue to settle over a 100 second interval.

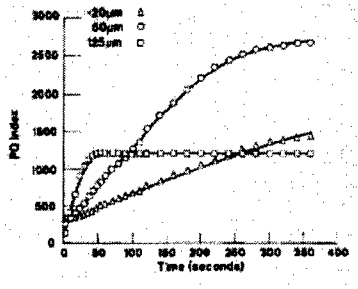


Figure 6 PQ Index as a function of time

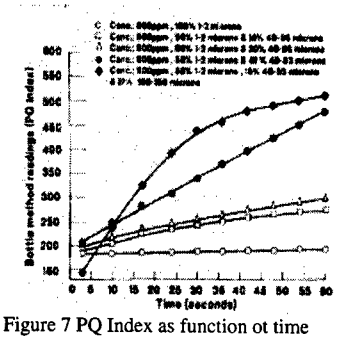


Figure 7 PQ Index as function ot time

This last observation has provided a possible source of a measurement of severity of wear. The combination of PQ Index, initial TDPQ slope and final TDPQ slope may offer a coding system suitable for measuring the total and severity of wear occurring in lubricated equipment. The next section illustrates the practical implication of these observations.

Application of the TDPQ to Field Samples: A series of 21 samples were analysed from earth moving equipment. Components included diesel engines, transmissions, hydraulics and final drive axle units. These units were selected at random from part of a typical daily sample batch received at the laboratory.

Diesel Engine Units: Figure 8 shows results which are typical of diesel engine units. The PQ index for an engine is usually less than 30 and the index remains constant with time. From the previous discussion these results would be expected to be typical of a sample with only small particles present. Figure 9 shows the Rotary Particle Deposit (RPD) associated with one of these samples. The particles deposited are all less than 10um and, as expected, when compared with spectrographic analysis show good correlation. For example the samples which provide PQ indices of 30 and less have iron concentrations less than 60 ppm. The sample with a PQ index of 56-70 measured 174 ppm iron by spectroscopy.

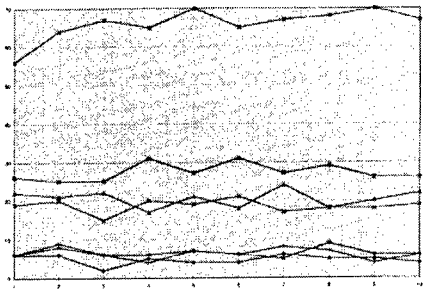


Figure 8 TDPQ readings for Diesel Engine samples

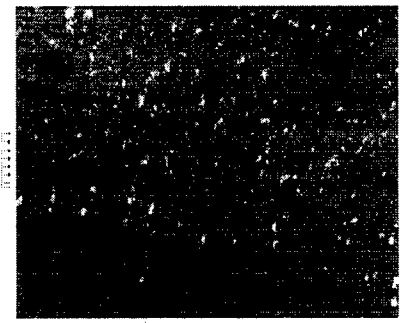


Figure 9 RPD typical particles of a diesel engine

Transmission Units: Figure 10 shows a summary of the TDPQ results for two of the transmission units tested. These graphs show a similar trend to that obtained for diesel engines i.e. a constant PQ index for the test duration. Figure 11 shows the particles that were deposited by the RPD. It will noted that the particles are all less than 10um and would therefore all be analysed by spectrographic methods.

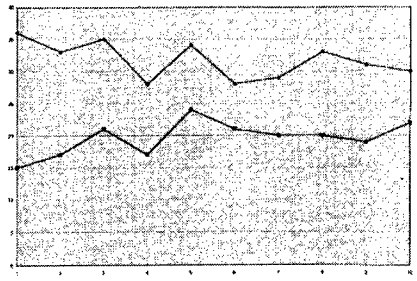


Figure 10 TDPQ for transmission samples



Figure 11 RPD for transmission sample

Final Drive Axle Units: Figure 12 summarises the TDPQ results obtained for nine final drive axle units. Each of the units has a higher initial TDPQ slope than that obtained for the latter stages. One example suggests an anomaly, i.e.a much higher TDPQ slope although all the samples had spectrographic results for iron of between 350 and 600 ppm. Figure 13 through 16 show the reason for this anomaly. The sample which had the high TDPQ slope also had a high density of larger ferrous wear particles. These particles varied in size from 30 to 300 um. These larger particles would not be expected to be analysed by spectroscopy which has the size limitation for analysis.

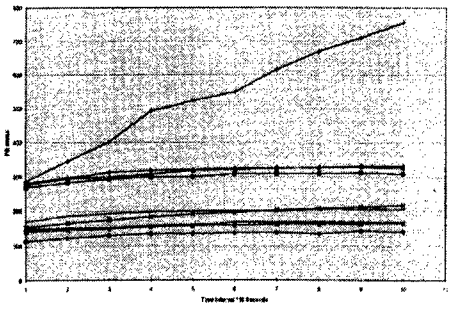


Figure 12 TDPQ for Final Drive Axle samples

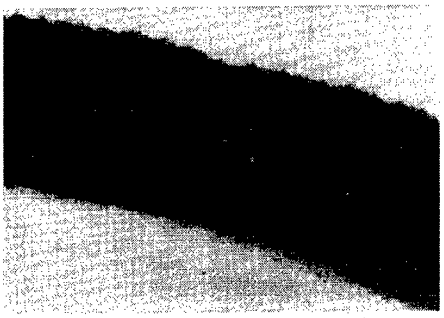


Figure 13 RPD inner ring Final Drive Axle TDPQ slope average for Drive Train

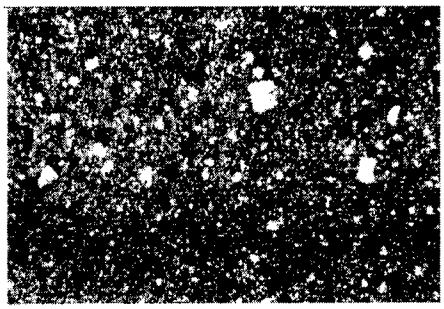


Figure 14 Low level of particles larger than 50 μm

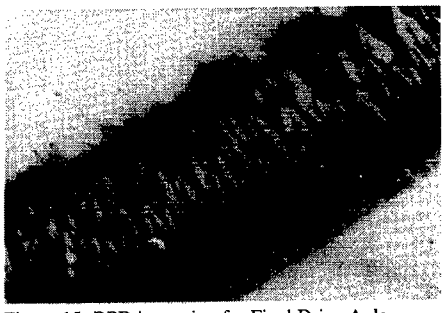


Figure 15 RPD inner ring for Final Drive Axle sample. TDPQ slope higher than other samples

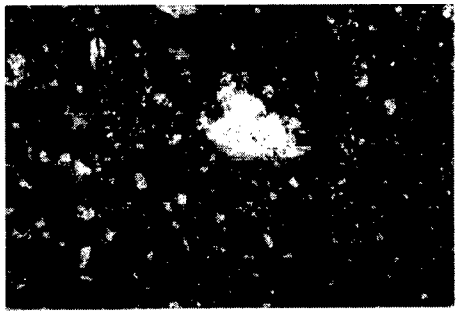


Figure 16 Numerous particles larger than 100μm

The following example further illustrates the effect of large particles on the spectroscopy and TDPQ measurements. A low, acceptable spectroscopy reading of 22ppm of iron was measured whilst the PQ index was 582 and 430 Dl and 113 Ds abnormal readings for this transmission unit. The TDPQ initial slope was higher than average whereas the final TDPQ slope had fallen to a low level. This result, from the above observations, would point to a proportion of large particles (greater than 100um) in a suspension of particles less than 10um. The following SEM micrographs show the supporting evidence.

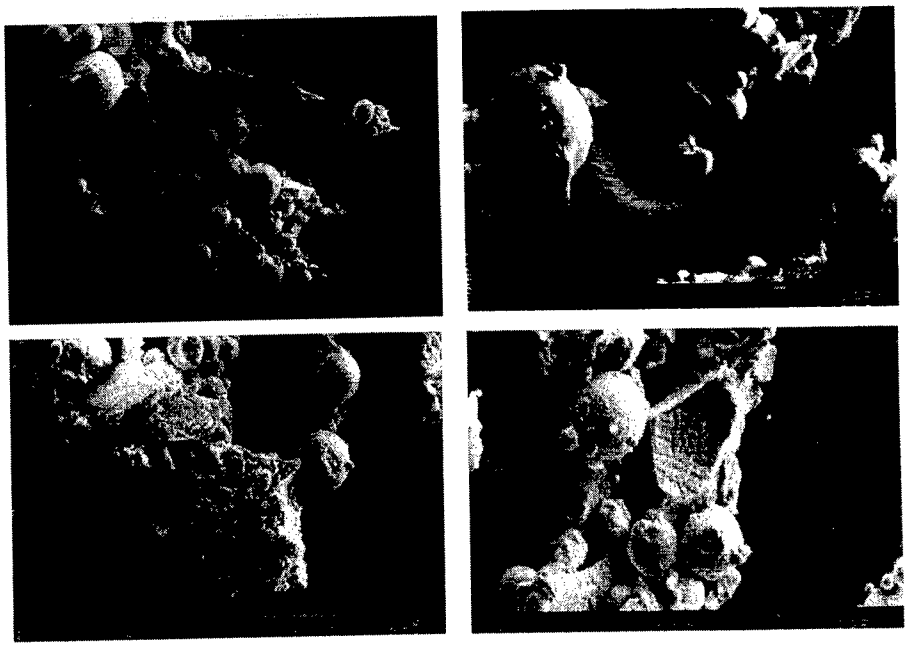


Figure 17 Stereo Electron Micrographs of an RPD of Final Drive sample with a high TDPQ initial slope

A proposal is therefore make to consider a coding system which includes PQ Index : initial TDPQ slope : final TDPQ slope.

The PQ Index is a measurement of total level of wear : initial TDPQ slope is a measurement of large particles : final TDPQ slope a measurement of the proportion of small particles.

Conclusions: The Time Dependent Particle Quantifier (TDPQ) offers both a portable and laboratory based instrument capable of measuring the total amount and severity of wear occurring in mechanical equipment. This instrument, in its portable mode, would provide a simple site screening technique capable of selecting fluid samples which should be sent to a central laboratory for further tests.

Acknowledgements: The help of colleagues at Swansea Tribology Services Ltd.,Analex and Wearcheck South Africa is gratefully acknowledged.

References

1. Jones M.H. and Massoudi A.R.; The TDPQ ; a solution to the analysis of large wear particles. Insight, vol 37 No.8 August 1995.

DEVELOPMENT IN WEAR DEBRIS MORPHOLOGICAL ANALYSIS AT RAF EARLY FAILURE DETECTION CENTRES

Brian J. Roylance, Lewis M. Jones and Anthony R. Luxmoore
Department of Mechanical Engineering
University of Wales, Swansea, U.K.

A. Killingray and S. Harris
RAF

D. Hodges
Defence Research Agency

Abstract: The RAF's Early Failure Detection Centres (EFDC) carry out routine inspection and analysis of magnetic drain plug samples taken from aircraft while in service. The captured debris is assessed in which morphological analysis is an important integral activity. The development of computerised procedures to identify and categorise the debris has led to the implementation at specific EFDC's of new, improved methods for detecting and diagnosing the wear condition of critical components. In this paper, the background to the development and application of these procedures is described in which the scope for predicting the future course of the wear and the prospects of introducing automatic analysis procedures are discussed.

Key Words: Computerised Procedures; Morphological Analysis; Wear Debris

Debris Analysis at RAF EFDC's: The RAF's Early Failure Detection Centres (EFDC) carry out routine inspection and analysis of samples taken from aircraft while in service. Magnetic Drain Plugs (MDP) are removed from aircraft for analysis in which the appearance of the captured debris is assessed and related to the wear condition of critical components. Techniques have been developed at Swansea for analysing particle morphology using computerised image processing and analysing methods to obtain detailed information about the wear condition of critical rubbing parts.

Figure 1 shows a schematic outline of the present scope for performing early failure detection and the current techniques in use at an EFDC. Clearly, a number of different monitoring technique are employed in various combinations to effect the most appropriate information for a wide range of operational conditions and aircraft types (1-2). Not all the techniques shown are utilised in every instance. However, the use of MDP units is common to all categories and this necessitates that the quantity of debris captured by a plug is measured (3-4). This is presently undertaken using the Debris Tester wear monitoring unit (5) which is then logged against 'Operating Hours'. Specific criteria are sometimes used to determine whether to recommend that the system being

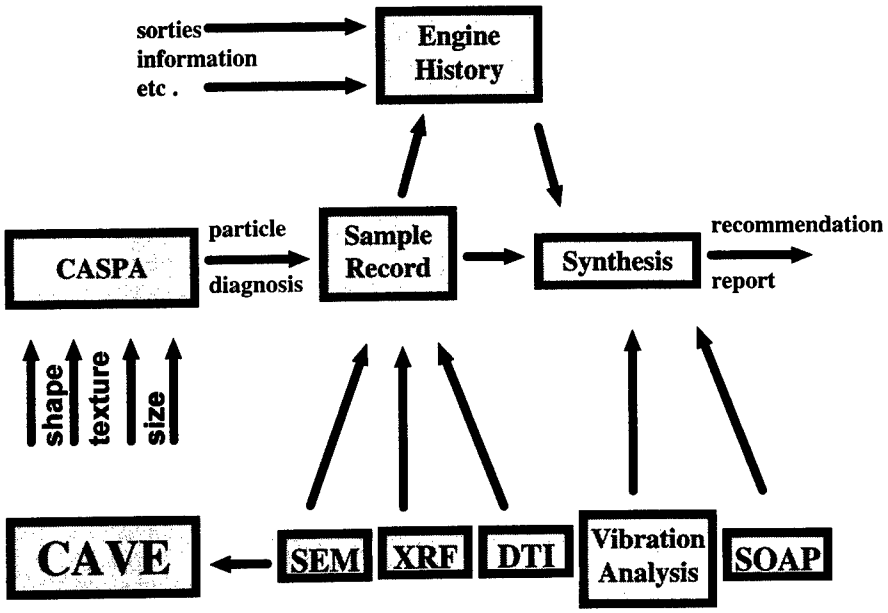


Figure 1 **The Integration of Wear Debris Analysis with Oil and Other Analyses and Data at RAF Early Failure Detection Centre**

monitored warrants closer inspection, or more frequent sampling, or even that it be immediately withdrawn from service. In most instances, where, based on the criteria, follow-up action is to be recommended, there is also a requirement that the debris be visually inspected and analysed to determine its morphological characteristics as a means of determining more precisely the source and type of wear occurring.

This is either undertaken by inspecting the debris directly on the plug itself, (large debris, typically of the order of a mm in linear dimension) or more usually, by placing a plug 'wipe' under an optical microscope and viewing it at a specific magnification, typically X50 overall magnification. Guidelines are provided to aid particle identification, but these are not universally applicable to every EFDC situation and are currently in the process of undergoing revision by the RAF.

The need to develop computer-aided procedures to assist in the identification of wear debris and performing reliable diagnosis has come to the fore in recent years because of the increasing complexity of modern machinery and the increased turnover of staff in many organisations. This

often means that there is insufficient time available to develop the skills required to carry out this type of analysis to the level demanded in a highly specialised area involving a subjective type of activity. Consequently, this had led increasingly to the need to introduce procedures which assist in formalising the decision-making processes in order to provide systematic, and thus more objective, ways to perform the analytical tasks. This is something for which computers are ideally suited and is the basis of the developments at Swansea (6-7).

The main task in hand has been to adapt the original programs, (which are relevant to oil sampling and associated ferrographic wear particle analysis) to the needs of the EFDC methods of working. This is with particular reference to the different terminology used to describe magnetic drain plug debris and also, the fact that widely differing procedures and requirements are involved.

The purpose of this paper is to describe the background and procedures developed which has led to their implementation at EFDC's and to discuss the implications of the need to predict the future course of the wear in an engine once it has been detected and diagnosed. The scope for introducing automatic procedures for performing morphological analysis of debris is discussed.

Wear Debris Analysis: Debris transported from the wear site by the lubricant carry with them important information about the condition of the wear state of critical components in the machinery. The first, and most important information concerns the quantity of debris generated as a function (usually) of operating hours. It provides an excellent opportunity to 'trend' monitor the extent of the wear and the rate at which it is progressing. It also has the advantage that the debris quantity, or concentration, is quantitative and can usually be obtained quickly, simply and cheaply (5).

The RAF currently use the Debris Tester for this purpose (5) in which the debris concentration is plotted on a basis of wear debris concentration count per unit time versus operating (flying) hours. Based on experience, threshhold limits are preset for each engine type which permit further action to be taken when the limits are exceeded. The technique is thus used as a 'front-line tool' to detect the onset of active wear in a potentially deteriorating component wear situation.

Whereas most of the debris generated by the wear processes involved will be ferrous free metal, for certain aircraft types, e.g. the Tornado, it is important also to establish the composition of some of the non-ferrous debris generated so that a distinction can be made between 'active' and 'benign' wear. X-ray Fluorescence (EDXRF) and also, Scanning Electron Microscope (SEM) techniques are utilised for this purpose, as and when necessary.

Morphological analysis requires that the debris be viewed through a microscope (optical mainly, but sometimes scanning electron). The particular features, or attributes of interest are: size, (plain dimension and also the thickness), outline shape and edge detail, surface features and colour. The type of analysis usually employed is generally time-consuming, often tedious and invariably subjective. However, it is essential to carry out this type of analysis if we are to properly determine the type and cause of the wear and also clarify its severity .

The methods devised to perform morphological analysis are several and highly variable, although there have been attempts made to specify good practise which utilise either a wear particle atlas, (8) or attempt to lay down a set of rules for performing the analysis (0). There are now considerable opportunities available to exploit modern computerised procedures in association with artificial intelligence techniques, including the use of neural nets and image processing and analysis (10).

To adapt and apply these methods to the analysis of MDP samples, two separate, though related procedures have been developed in parallel, as follows:

- Wear Particle Analysis
- Computer-Aided Analysis of Morphology

Figure 2 indicates schematically the processes involved and how they interface with one another. In devising procedures for ultimate use at EFDC's, the following points must be borne in mind:

- the procedures devised must be clearly defined, easy to follow and quickly executed
- they must be compatible with other facilities and procedures already in operation at an EFDC
- they must be helpful and reliable in assisting personnel arrive at a correct diagnosis of the problem.

Application to EFDC Operations:

Wear Particle Atlas: The purpose of developing a series of annotated images is two-fold: it provides a reference document for debris associated with specific engine types which can be accessed by operators in the course of performing routine analyses of samples. It can also be used for tutorial sessions during training of inexperienced personnel.

The process used in the development was Hyper Text Mark up Language (HTML) to generate hyper-linked text and images comprising scaled images of debris, galleries of 'thumbnail' images with explanations of components and other relevant information, (including engine history, drawings of key components, etc.).

This method of displaying information permits the operator to switch rapidly from one source of information to another and is entirely 'mouse-operated'. Up-dates on information or images can be inserted easily with appropriate commentary to aid future analysis. Electronic transmission, e.g., from one EFDC to another, is easily effected either through normal (telephone) communication channels, or 'secure' channels if preferred.

The principal advantage of using this method is that the atlas can be specifically dedicated to the requirements of a particular engine type or EFDC operation, and the scope for building a 'library' of particle images is unlimited without the need to generate 'paper' records.

Figure 3 shows a sample 'page' extracted from the atlas.

Computer-Aided Morphological Analysis: The purpose in utilising a computer-aided system for analysing particle morphology is to establish a systematic procedure to arrive at a reliable identification of a specific particle which is linked to a capability to document the information and prepare a report defining the results of the overall analysis of the sample. When linked to other information about the sample it should also be possible to formulate an accurate diagnosis of the wear condition and associated fault in the system.

The present development is an extension of a previous program, CASPA (Computer-Aided Systematic Particle Analysis) which was established to assist in the analysis of wear debris extracted from oil samples using the ferrography method (6). The new version is based on the original version but is dedicated to the analysis of magnetic drain plug debris and is configured in 'Windows' format. It can be run on a 486-type, personal computer in conjunction with an optical microscope in which the user describes the appearance of a particle in response to a series of questions, or 'prompts', leading to a diagnosis of the particle type. Documentation procedures follow and the results of debris concentration measurements (e.g, Debris Tester) are tabulated and plotted and cross-referenced to the morphological analysis.

Quantitative analysis of particle morphology is undertaken by means of image processing. One such application, known as CAVE (Computer-Aided Vision Engineering) was developed previously to perform quantitative analysis of particle size and shape (7). A neural net procedure has since been added to define a particle's surface features in relation to the associated wear mode (11) and applied to different controlled wear test situations (12).

A Windows version of CAVE has been developed for the analysis of magnetic plug samples. However, there are some unresolved difficulties with its application due to problems with the procedures currently used at EFDC locations to deposit the samples prior to analysis. This has the effect of causing particle overlap and piling up of particles. The quality of incident light illumination of the particles when deposited on opaque substrates is also highly variable. In combination, the two effects result in inconsistent results and further work is required to establish procedures which will permit reliable analysis. This is regarded as being important in relation to the prospects for establishing automatic procedures necessary to speed up morphological analysis and render it a less tedious and subjective activity.

The implementation of these procedures has commenced on a trial basis at EFDC locations responsible for two entirely different types of aircraft: the Tornado, powered by the Rolls Royce RB199 engine, and the Hercules C130, powered by the Allison engine. In the latter, the main requirement is to introduce appropriate procedures for trend monitoring of Debris Tester results linked to periodic particle morphological analysis in which the particles are almost entirely comprised of large, > 100 μm ferrous free metal particles. In the case of the former, the debris is invariably much smaller, < 50 μm in which there is a need to identify particles of different composition with a corresponding requirement to carry out morphological analysis on a more routine basis.

Concluding Remarks: In performing condition monitoring of wearing components using wear debris analysis procedures, the principal requirements are satisfied (ideally) when a fault is

detected and correctly diagnosed with no corresponding loss of life, equipment or operational time. However, the other major requirement is to be able to determine the optimum period in which to carry out remedial repairs or replacement of wearing parts. While detection and diagnosis is intended to guarantee that there will be no unforeseen failure, they do not guarantee that repairs, or other interruptions to operation, are not undertaken prematurely. In the absence of any certainty as to when a component will malfunction or cause a deterioration in performance efficiency, a tendency to err on the side of caution is to be expected even if it is known to add considerably to the overall long term cost of the operation.

The procedures now being introduced are necessary to ensure that the present effectiveness of the EFDC operation is maintained and improved in the future against a background of possible reductions in available resources. However, it is recognised that if there was greater confidence in predicting the course of a deteriorating wear situation there would be considerable potential savings without the accompanying increased risk of unscheduled failure. This is most likely to be achieved when there is increased scope for automated analysis procedures coupled to improve co-ordination of condition monitoring data with maintenance histories and operational data.

References

1. Eucams : The Story; Pub. Future Systems (Air) 42, Ministry of Defence (Procurement Executive), January 1994, ISBN 0 9522854 0 1.

2. The RAF Spectrometric Oil Analysis/Centralised Oil Analysis Programme; Proc. JOAP Conference, Pensacola 1994.

3. Identification of Particle Debris, Ch. 4, Annex B, AP119A-20006-1, Ministry of Defence.

4. Hunter; Engine Failure Prediction Techniques, Aircraft Eng., 45, 1984, 4-13.

5. Hunt; Handbook of Wear Debris Analysis and Particle Detection in Liquids, Elsevier 1993.

6. Roylance, I.A. Albidewi, A.L. Price and A.R. Luxmoore; "The development of a computer-aided systematic particle analysis procedure - CASPA"; Lubr. Eng., 48, 12 (1992), 940-946.

7. Roylance, I.A. Albidewi, M.S. Laghari, A.R. Luxmoore and F. Deravi; "Computer-Aided Vision Engineering (CAVE), Quantification of Wear Particle Morphology", Lubr. Eng., 50, 2 (1994), 111-116

8. Anderson; Wear Particle Atlas (*revised*), Naval Army Eng., Centre Report No. NAEC-92, 163, 1982.

9. Anderson and J. Stecki; A Diagnostic Key for Wear Debris Obtained from Oil and Grease Samples from Operating Machinery, Proc. 3rd Int. Conf. on Condition Monitoring 1990, Paper E2 - BHR Group Ltd, London.

10. Roylance; Developments in Wear Particle Analysis Using Computerised Procedures; Proc. I.Mech.E., Aerotech '92 Conf., C428/15/216.

11. Laghari, I.A. Albidewi, A.K. Muhamad and B.J. Roylance; Wear Particle Texture Classification Using Artificial Neural Networks; Proc. 3rd Nordic Transputer Conf. Denmark 1993.

12. Roylance and S. Raadnui; The Morphological Attributes of Wear Particles - Their Role in Identifying Wear Mechanisms; Wear, 175 (1994) 115-121.

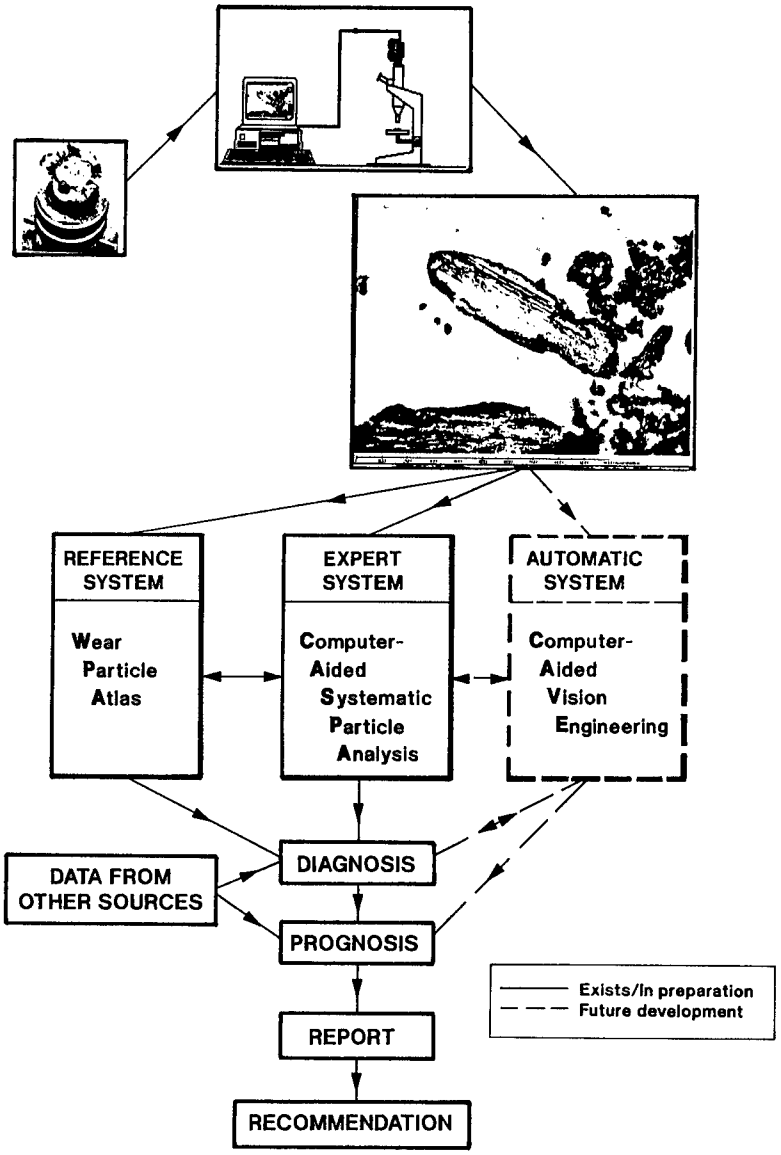
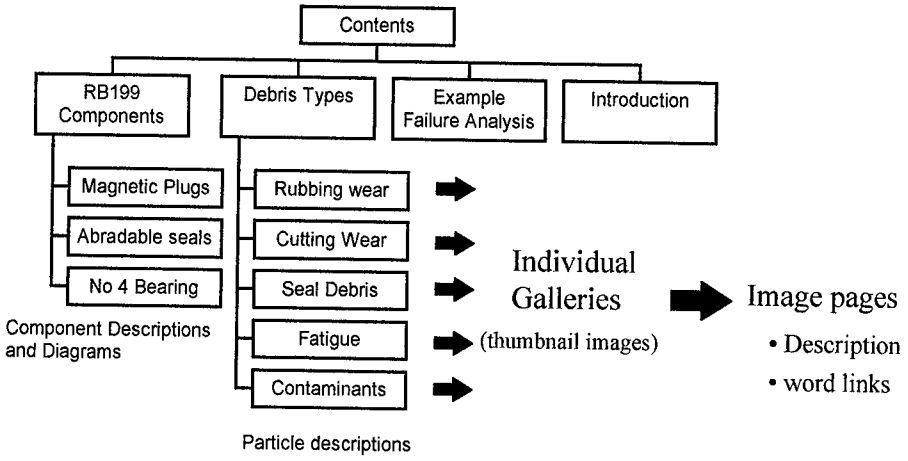


FIGURE 2 PROCESSES INVOLVED IN PERFORMING MORPHOLOGICAL ANALYSIS OF MAGNETIC DRAIN PLUG SAMPLES

RB199 Wear Particle Atlas Structure



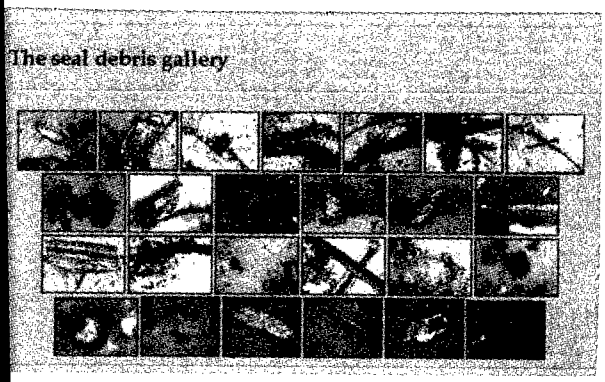




Figure 3 - Sample 'pages' extracted from the Wear Particle Atlas.

Accompanying text including a particle description and word links to other photographs within the Atlas.

ROYAL AIR FORCE

EVALUATION OF VARIOUS METHODS TO DETECT METALLIC WEAR PARTICLES IN LUBE OIL FILTER DEBRIS

Gary C. Fisher
Defence Research Establishment Atlantic
P.O. Box 1012
Dartmouth, NS Canada B2Y 3Z7

Abstract: Studies were conducted to evaluate methods of detecting metallic wear particles in lube oil filter debris. The methods studied were low temperature oxidation of the organic constituents in the debris, separation of ferromagnetic particles by passage through a magnetic field, and counting/sizing of metallic particles in a fluid stream using a commercially available detector. Laboratory-prepared simulations of wear debris and actual filter debris samples from operational military aircraft systems were used in the evaluation. The relative ability of each method to detect metallic wear particles is discussed, along with the effects of each method on subsequent debris examination.

Keywords: Wear particles; filter debris; filter debris analysis; magnetic separation; oxidation; particle counting

INTRODUCTION: Historically, condition monitoring of military systems in Canada has relied on Spectrometric Oil Analysis (SOA) and Ferrography. [1, 2] However, increased usage of fine lube oil filtration (10 μm or less) has lessened the efficacy of these techniques due to a reduction of wear-related debris in the lube oil.[3] One technique that has been touted as a potential replacement or supplement for these methods is Filter Debris Analysis (FDA). [3, 4] FDA involves removal and examination of particulate debris trapped by in-line lube oil filters. Wear particle quantity, size, and morphology are used to interpret system wear condition. As the predominant wear modes vary among systems, the success of any FDA program depends on proper definition of what constitutes "normal" versus "abnormal" wear for the particular system under study.

While it is relatively straightforward for most systems to define "abnormal" wear particle sizes and morphologies, definition of "abnormal" wear particle quantity has not been simple. Filter debris contains many particles, such as oil oxidation products, fibers, atmospheric pollutants, etc., which, while they may be of diagnostic significance, tend to obscure the detection and quantitation of metallic wear particles. [3, 4] Traditionally, wear particle quantities have been estimated by visual or microscopic examination, a method requiring significant operator expertise. In the military environment this expertise has proved difficult to achieve and maintain at the field level and widespread implementation of FDA has been correspondingly hindered. Therefore development of an accurate method by which the quantity of metallic wear particles in

a debris sample can be measured by field condition monitoring technicians has become a topic of considerable research activity. [5 - 7]

Based on the results of some preliminary work [8], three methods were chosen for evaluation as potential field methods for measuring wear particle quantity. These were: removal of organic materials by low temperature oxidation, physical separation of magnetic particles, and counting and sizing of ferromagnetic and conductive particles using a commercially available detector.

Discussion of these methods will proceed in two parts. Firstly, the ability of each method in detecting metallic wear particles will be discussed. Secondly, each method will be evaluated in terms of: effect of the method on wear particle appearance (destructive nature of method), specificity of the method for detecting metallic particles, method simplicity, cost, and analysis duration. It should be noted that this work focuses on the detection capabilities of these methods. Complete evaluation of their quantitative capabilities must await further work.

PROCEDURES AND EQUIPMENT: Each method was evaluated using both old lube oil filter debris samples from CH124 main transmissions and T-58 gas turbines, and simulated lube oil filter debris. Simulated debris samples were prepared from crushed activated charcoal (sieved to an average particle size of less than 75 μm), carbon steel and copper filings, magnetite, hematite and sodium chloride. Metal filings were sieved so as to consist of two particle sizes; average size greater than 75 μm and average size less than 75 μm.

Optical microscopy work was accomplished using a Zeiss Axioplan reflectance polarized light microscope using a 5X 0.15NA strain-free objective. Scanning electron microscopy (SEM) was done utilizing an ISI DS-130. Energy dispersive x-ray analyses (EDXA) were completed with a Princeton Gamma-Tech IMIX system using a light element detector mounted on the SEM. All SEM/EDXA work was done at an accelerating voltage of 20KV, a working distance of 30 mm, a specimen tilt of +20°, and a take-off angle of 38°.

DETECTION CAPABILITIES: Each method will first be briefly described followed by a discussion of the techniques used to evaluate detection capabilities.

Low Temperature Oxidation: The inorganic content of filter debris has been shown to consist of less than 30% of total debris weight. [3] It has been suggested that if the organic contribution to total debris weight could be removed, then direct correlation of debris weight to system wear condition may be possible. [3, 8] It was attempted here to remove the organic contribution by oxidation at an elevated temperature. Such a process is complicated by undesirable oxidation of metal surfaces and, if temperatures are too high, by melting of some metal particles. A temperature of 300°C was chosen as a compromise. It was thought to be sufficient to oxidize most organics but low enough to prevent melting of lead and other low-melting metals.

To determine the effect of exposure to 300°C on metal surfaces, samples of carbon steel and copper filings were placed in an oven at temperature for 8 hours. Portions were removed after 1, 3, 6, and 8 hours for examination. Figure 1 indicates that the level of oxygen detected on the copper surfaces by SEM/EDXA increases with exposure time. Oxide formation can be been seen in Figure 1 to correspond to a "pebbling" of the metal surface. These results were mimicked by

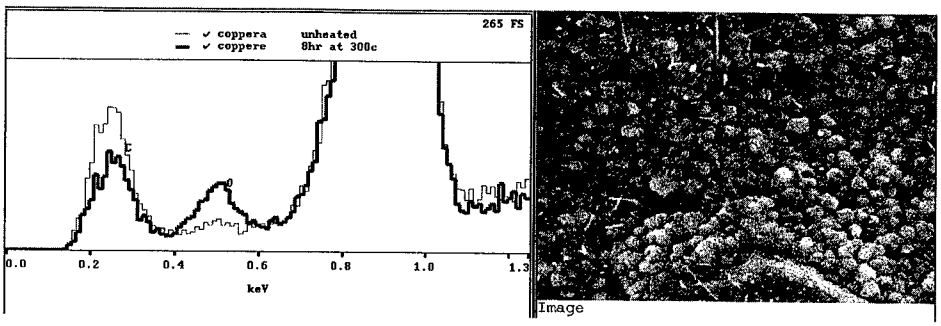


Figure 1: EDXA spectra of copper filings. Oxygen content is seen to increase in sample heated for 8 hours at 300°C (thick line) from untreated sample (thin line). Large peak is due to copper L lines. Image shows "pebbled" oxide structure that forms on heated copper surfaces.

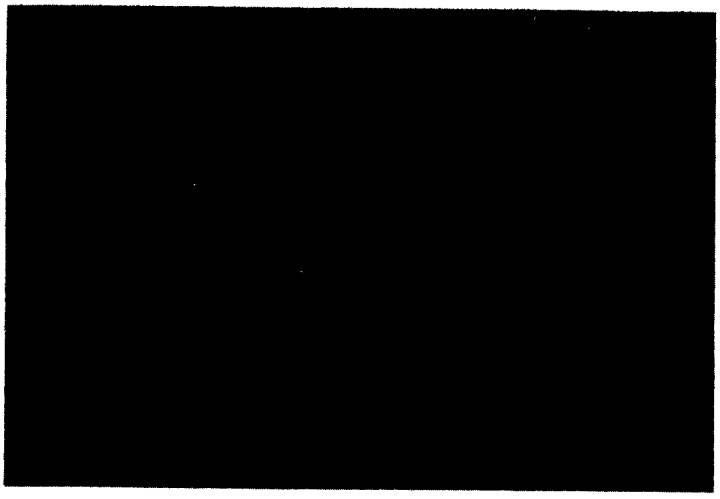


Figure 2: Copper filing after exposure to 300°C for 8 hours.

the steel surfaces, although to a lesser extent. It should be noted that the oxygen peak shown in the x-ray spectra is, even after 8 hours, a minor peak and would not lead to chemical misidentification of the particle.

Exposure to the heat also visibly darkened the metal surfaces. However, this was not pronounced and would not preclude identification of the particle as metallic, as can be seen in Figure 2.

Figure 3 shows cumulative weight losses for two different CH124 main transmission debris

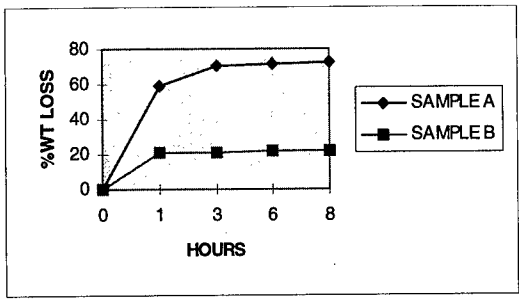


Figure 3: Cumulative weight loss curves for two CH124 main gearbox filter debris samples exposed to 300°C.

samples. Sample A was a typical filter debris sample in that it contained relatively few visible metal particles. Sample B was from a transmission which had experienced significant component wear. It should be noted that in both cases the majority of weight loss due to organic oxidation occurred within the first hour of heat exposure.

While both samples A and B experienced significant weight loss as result of the oxidation, little qualitative difference was visible in the appearance of the debris as a result. Metal particles were no easier to observe than in the untreated samples. Many non-oxidizable materials, presumably inorganic in nature, were still obscuring detection of the metal wear particles.

Magnetic Separation: In virtually every mechanical system the critical load-bearing components are made from steel. Therefore, quantitation of the ferromagnetic particles in a filter debris sample should be a good indicator of component wear. It is true that some nonferrous metals are used in most mechanical systems, but it is the author's experience that virtually every serious component wear situation warranting unscheduled and significant remedial action has involved wear of one or more steel components. Therefore an apparatus for separation of ferromagnetic particles from filter debris was assembled. This apparatus consisted of a shielded permanent magnet device having a magnetic field centered along a 10 cm long, 2 cm deep notch. By turning a switch the magnet could be shielded, thus essentially "turning off" the magnetic field. Field strength was 0.15T along the apex of the notch, falling to 0.025T at the roots. The separation apparatus consisted of a glass funnel suspended 30 cm above the magnet device. 0.25" I.D. Tygon™ tubing was run from the funnel, along the magnetic notch and then into a reservoir.

The apparatus was designed for separation of magnetic particles from filter debris. Generally, debris is removed from on-line lube oil filters by extracting entrapped particulates from the filter into a solvent by some physical means, such as ultrasonic agitation. At this point the solvent/debris mixture would be passed through the device with the magnetic field "on".

Magnetic particles would be held in the field and non-magnetic particles washed into the

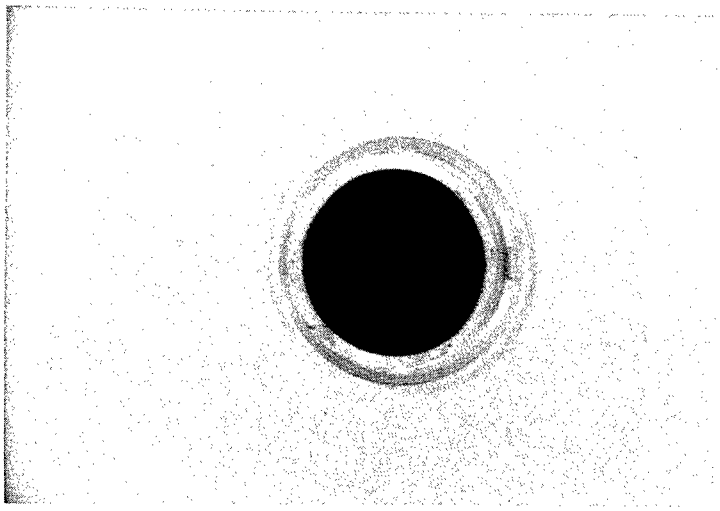


Figure 4: T-58 gas turbine filter debris before magnetic particle separation.

reservoir. Afterwards, the field would be turned "off" and the magnetic particles washed into the additional solvent into a second reservoir. In this way, the filter debris would be partitioned into two samples: magnetic and non-magnetic debris.

Various weights of steel filings were dispersed in petroleum ether and percent recoveries determined using the separation apparatus. To determine if size of ferromagnetic particle affected recovery, two sizes of steel filings (greater than and less than 75 μm) were used. Essentially quantitative recoveries (average of 92%) were achieved using only one pass of the mixtures through the magnetic field. Size of steel particles had no apparent affect on recovery rate. Quantitative recoveries were also achieved using fine magnetite powder (particle diameters less than 20 μm), but three passes through the magnetic field were required. It was therefore decided to employ three passes for all remaining tests.

Recovery rates using laboratory prepared filter debris samples often exceeded 100%. This was due to entrapment of nonferromagnetic material in the plug of magnetic particles held in the magnetic field and due to solvent absorption by the finely divided activated charcoal. Subsequent SEM/EDXA of the nonferromagnetic debris did not indicate the presence of ferrous species, suggesting quantitative recovery of ferromagnetic material was achieved by the separation apparatus.

The effect of the separation apparatus on old debris samples was evaluated by scraping debris from existing CH124 main gearbox and turbine samples and then dispersing the weighed debris in petroleum ether. As an example, compare Figure 4, which shows the original appearance of turbine filter debris, with Figure 5, which shows the ferromagnetic particles recovered from that debris sample. A considerable reduction in particulate numbers is evident. Quantitative analysis

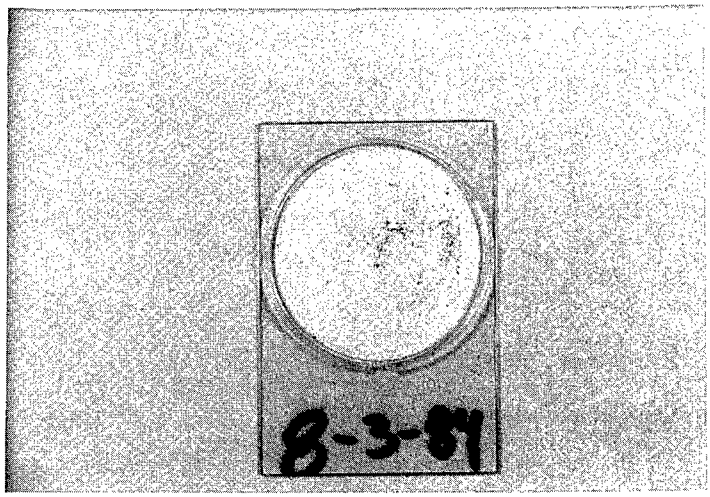


Figure 5: Magnetic particles separated from T-58 gas turbine debris shown in Figure 4.

of the ferromagnetic debris, gravimetrically or otherwise, would obviously be simpler than in the original sample. Although the relative amounts of recovered ferromagnetic particles varied, similar separation results were obtained for all operational debris samples evaluated.

Wear Particle Counting/Sizing: A commercially available wear particle measurement system, called MetalSCAN™, was employed as a means of counting metallic particles. The system is an inductive device which measures disturbances to an alternating magnetic field caused by the passage of a ferromagnetic or electrically conductive particle. If the passing particle is ferromagnetic, magnetic field strength is enhanced as a result of displacement of low magnetic permeability solvent. The magnitude of the disturbance is proportional to the volume of the particle. Nonferromagnetic conductive particles are detected by the formation of eddy currents within the particle which act to reduce the field strength. In this case, the magnitude of the field disturbance is proportional to the surface area of the passing particle.

The MetalSCAN™ system sizes particles by making a few assumptions about their composition. Ferromagnetic particles are assumed to be carbon steel. Nonferromagnetic particles are assumed to be aluminum. Using these assumptions, size detection limits can be determined. The manufacturer reports limits of 125 μm for ferromagnetic particles and 400 μm for nonferromagnetic conductive particles. [8] The instrument then counts and reports the number of particles detected within specific size ranges.

It should be noted that if the composition of the passing particle is different from the assumed composition, erroneous values will be reported for particle size. It should also be noted that the

size calculation further assumes spherical particles. As the detection limit is actually dependent on particle volume or surface area, the detection limit is considerably larger for wear particles, which tend to be relatively flat plates.

For the MetalSCAN™ system to be effective, metallic particles must individually traverse the detector. Should two particles pass simultaneously (or at least within the time-response period of the detector) a single response will be recorded. If both particles were ferromagnetic or nonferromagnetic conductive, the overall response will be additive. If dissimilar particles traverse simultaneously, the response will be more complex.

The apparatus consisted of a fixed displacement vane pump driven by a 0.018 Hp AC motor at 1440/1730 rpm which draws solvent (in this case distilled water) from a stainless steel reservoir through 5/8" ID Tygon™ tubing. Water was pumped out the discharge side through 3/16" ID tubing to the MetalSCAN™ detector and emptied into a beaker. Pump and tubing sizes were selected so as to provide a fluid velocity of 9 ft/sec, as is required by the detector.

The MetalSCAN™ detector software was configured to calculate the total weight of detected particles. This was done by determining the weight of a particle of a size at the midpoint of each range, using the composition assumptions described earlier. The total number of particles detected within each range could then be multiplied by this weight and the weights for each range summed to give total weight. Preliminary testing in this manner indicated a linear response for up to 30 mg of carbon steel particles between 250 and 300 μm in size dispersed in water. It was found during this testing that a water volume of 3.5 L was required to sufficiently disperse the particles to permit discrete counting by the detector. This volume was used for all subsequent tests.

To evaluate the efficacy of the detector, 12 CH124 main gearbox debris samples were selected. These samples were graded, using current FDA practices, into 5 severity categories, with 1 being the least severe. The categories were based on the total amount and size of metal wear particles observable with a stereoscope. Using current guidelines, samples having a severity of 3 or higher were considered significant. Each sample was then dispersed ultrasonically into a small volume of isopropanol and diluted to 3.5 L with distilled water. Each diluted sample was then examined using the apparatus described above.

Table I shows the 12 samples arranged by severity level, sample number and MetalSCAN™ results. The MetalSCAN™ results are arranged in descending order of the total weight of detected ferromagnetic and nonferromagnetic particles with the original sample number maintained for comparison. Some of the tests (Sample #1, for example) deemed acceptable by the stereoscopic examination contained among the highest levels recorded by the MetalSCAN™.

Given the past success rate of the visual/stereoscopic method, these results suggested shortcomings in the particle counting method. However, two points must be remembered. First, the stereoscopic approach counts only metallic particles that were visible within the debris while some nonmetallics, particularly ferrous oxides, would be counted in the MetalSCAN™ results. Second, it is possible, as has been discussed earlier, that low numbers of wear particles could be

TABLE I: COMPARISON OF VISUAL RATING WITH METALSCAN™ RESULTS

VISUAL RATING	SAMPLE NUMBER	MAGNETIC METALSCAN SAMPLE # (MG)	CONDUCTIVE METALSCAN SAMPLE # (MG)
5	12	1 (14.4)	5 (31.4)
	11	5 (6.7)	1 (11.1)
4	10	6 (4.3)	6 (8.2)
	9	3 (3.7)	9 (3.6)
	8	7 (2.1)	3 (3.4)
3	7	2 (1.4)	7 (1.8)
	6	9 (1.2)	11 (1.4)
	5	12 (1.1)	12 (1.3)
2	4	11 (1.1)	2 (0.6)
	3	10 (0.8)	10 (0.2)
1	2	8 (0.2)	8 (0.2)
	1	4 (0.1)	4 (0)

erroneously reported by the visual method due to obscuration of metallic particles by nonmetallic particles.

Figure 6, which shows Sample #1 prior to any treatment, indicates that this filter debris sample did in fact contain a significant amount of nonmetallic particles that may have obscured metallic particle detection by the stereoscopic method. SEM/EDXA examination of the magnetic particles separated from this debris indicated that the majority were ferrous oxides, which would not have been counted by the visual method.

As a further check, the debris samples utilized in this testing were magnetically separated using methods similar to those described earlier. The weight of magnetic debris recovered in each test was quite similar to the ferromagnetic MetalSCAN™ results, further supporting the validity of the MetalSCAN™ results. It is true that ferromagnetic particle counting will detect metal oxides (which, strictly speaking, aren't metallic wear particles), but it would seem beneficial to have such particles detected as this will provide some measure of system component corrosion.

METHOD EVALUATION: As stated earlier, the criteria used to rank the methods for further field testing were: destructive nature of the technique, specificity, simplicity, cost, and analysis duration. Each criterion will be discussed separately and then the overall evaluation summarized.



Figure 6: Sample #1 used in MetalSCAN™ evaluation, prior to any treatment.

The magnetic separation and wear particle counting methods are non-destructive and do not affect wear particle appearance or the overall composition of the debris. The oxidation method is destructive as organic materials, which could be of diagnostic significance, are oxidized. Also, the color of wear particles can be altered. Specifically, copper alloys tend to blacken while steels take on straw or bluish hues, depending on alloy type. It should be noted that this change does not affect identification of the particles as "metallic". Further, compositional analysis by SEM/EDXA is not significantly affected.

None of the methods are entirely specific for metallic particles. Both the magnetic separation and particle counting methods record magnetic ferrous oxides as wear particles. However, this is not necessarily a disadvantage as it provides a measure of system corrosion. The particle counting method also detects nonferrous metals and therefore provides a greater degree of specificity than magnetic separation. The oxidation method is not very specific. Any inorganic material, and even some temperature resistant organics, will be detected and counted as metals.

The oxidation and magnetic separation methods would be quite simple to use. Both would require preparation of filter patches using current techniques and simple correlation of patch weight with system component condition. The particle counting is somewhat more complicated in that analyst familiarization with the detector's operating system is necessary.

Cost for implementing any of the methods would be relatively minor. Start-up costs for both the oxidation and magnetic separation methods should be less than $1K. Start-up costs for the

counting method will be somewhat higher, but should not exceed $10K. Operational costs for all three methods should not be significantly greater than that required for current FDA practices.

In terms of analysis duration, both the counting and magnetic separation methods should be only slightly greater than current practices, as both require only passage of the debris/solvent mixture through a detection apparatus prior to patch preparation. The oxidation method would add approximately 2 hours to the analysis time given a 1 hour duration in the oven and 1 hour cooling before weighing.

CONCLUSIONS: Separation of wear particles by passage of a debris/solvent mixture through a strong magnetic field and counting/sizing of metallic particles using a commercial detector both hold promise as field methods for measuring wear particle quantity in lube oil filter debris. Field trials utilizing operational equipment should begin to determine the quantitative capabilities of these techniques. The oxidation method is not recommended for further evaluation due to increased analysis time requirements, poor specificity and destruction of potentially significant organic constituents.

REFERENCES:

1. D.E. Veinot, "X-ray Fluorescence Spectrometric Analysis of Wear Metals in Used Lubricating Oils", DREA Technical Memorandum 80/J, December 1980.

2. H.P. Dominique & C.A. Waggoner, "Experience with Ferrography for the Wear and Condition Diagnosis of CF Aircraft Propulsion Machinery", DREP Technical Memorandum 84-10, November 1984.

3. P.V. Madhavan, "Monitoring Aircraft Power Plant and Transmission Systems Via Diagnostic Filters", Proceedings of the 43rd Meeting of the Mechanical Failures Prevention Group, 1991, pp. 44 - 47

4. G.C. Fisher & D.E. Veinot, "Wear Debris Examination as a Condition Monitoring Technique for the Sikorsky Sea King Helicopter Main Gearbox", 1989 Condition Monitoring and Preventative Maintenance Conference Proceedings, STLE Special Publication SP-27, 1989, pp. 119 - 131.

5. K.K. Yeung et al, "Development of Computer-Aided Image Analysis for Filter Debris Analysis", Lubrication Engineering Vol.50, No.4, April 1994, pp. 293 - 299.

6. P.R. Roberge et al, "Developing an Expert System Assistant for Filter Debris Analysis", Lubrication Engineering Vol.50, No.9, September 1994, pp. 678 - 683.

7. G.F. Fisher, "Quantitative Filter Debris Analysis (QFDA) as a Means of Monitoring Wear in Oil-Lubricated Systems", M.Eng. Thesis, Royal Military College, Kingston, Ont. Canada, 1991.

8. F.W.M. Dopplinger, "Investigation of Quantitative Screening Methods for Filter Debris Analysis", DREA Contractor Report CR/95/464, August 1995.

CONDITION MONITORING OF AIRCRAFT BY QUANTITATIVE FILTER DEBRIS ANALYSIS (QFDA)

L.G.I. Bennett and Sarah Swanson
Department of Chemistry and Chemical Engineering
Royal Military College of Canada - Collège militaire royal du Canada
Kingston, Ontario, Canada K7K 5L0

F.D. Koza and J.S. Poland
Department of Chemistry, Queen's University
Kingston, Ontario, Canada K7L 3N6

Abstract: Various condition monitoring techniques are used collectively to monitor the health of aircraft engines and transmissions, a concept known as Integrated Health Monitoring (IHM). A well-established quantitative technique is Aircraft Oil Analysis (AOA) in which spectroscopic techniques such as Rotating Disk Electrode Atomic Emission Spectroscopy (RDE-AES) is employed to analyse periodic oil samples for wear debris. Usually, no sample preparation is undertaken as the oil sample containing both dissolved and suspended metallic wear debris is analysed directly. AOA works well for oil-lubricated systems with relatively coarse filtration that allows circulation of the debris and its subsequent abrasive contact with moving components. To avoid this secondary wear, finer filtration is employed on new and older aircraft. Less wear debris, and thus information, is available in the oil. A technique that quantitatively analyses the wear debris caught on the filter has been developed and is termed Quantitative Filter Debris Analysis (QFDA).

Actual oil filters from Challenger ALF 502L-2C and Hornet F404 engines were obtained in sequence, when possible, prepared with the developed procedure and analysed with AOA instrumentation. With sufficient results, both normal and abnormal levels of wear rates have emerged, as has been recorded and used for AOA. Moreover, trending of the data for sequential samples has demonstrated the capability of QFDA for condition monitoring.

Key Words: ALF 502L-2C engines; condition monitoring; F404 engines; integrated health monitoring; quantitative filter debris analysis

INTRODUCTION: Integrated Health Monitoring (IHM) is the collection, analysis and application of many types of data related to the life usage and the mechanical and aerodynamic health of an aircraft's engine and drive train which will assist in its operation, maintenance, management, design, safety and logistics [1]. Condition Monitoring (CM) techniques are used to repair or replace components depending on their analysed condition rather than on a fixed schedule that the manufacturer recommends to prevent failure.

The overall concept of IHM uses CM techniques, such as wear debris analysis, to monitor the health of aircraft engines and related components. By monitoring the condition of aircraft engines, maintenance personnel may be able to predict engine failure before it occurs and reduce costly repairs.

The nature and quantity of wear debris (material, size, shape and concentration) are indications of the condition of various machine components which are in contact with the lubricant. The

higher the concentration of metal particles and the larger the particles being generated, the more severe the wear [2]. The identity of a particular component as well as the specific wear processes can be determined by Filter Debris Analysis (FDA) and the level of wear (and thus engine health) can be determined by Quantitative Filter Debris Analysis (QFDA).

Currently, a tendency towards finer oil filtration (10 microns or less) in aircraft engines restricts the applicability of the present oil analysis methods in the Canadian Forces (CF) [3]. Although fine filtration removes the risk of secondary damage caused by wear debris, there is now insufficient debris remaining in the oil for valid analysis. However, the wear debris may be made available for analysis, both qualitative (FDA) and quantitative (QFDA), by cleaning the filter of debris and preparing it for an analytical technique.

SAMPLE PREPARATION: Known metal samples were used to develop and test the sample preparation procedure [4] and then actual oil filters were analysed. After the engine filters were received from the various CF Bases, they were cleaned ultrasonically with a solvent and the wear debris was then filtered. (At this step, the morphology of the wear particles may be inspected qualitatively by FDA). The filter patch was then dissolved by acid digestion. At this second step, the debris could be analysed quantitatively by Atomic Absorption Spectroscopy (AAS) or Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES). However, in order to determine if the Rotating Disk Electrode Atomic Emission Spectroscopy (RDE-AES) available for AOA could be employed, a third step was developed. To prepare the solid sample for quantitative analysis by RDE-AES, it must be converted into an oil-based form. All of the oil filters were cleaned of debris (step 1) and the debris was digested with acid (2), converted into an oil matrix (3), and then analysed by RDE-AES.

(1) Ultrasonic Cleaning: Engine oil filters from the Hornet and Challenger aircraft were cleaned by a combination of ultrasonic agitation and filter backwashing in Petrosol solvent with low pressure (3 psi) pulsating air. This cleaning technique [3] was developed to ensure that the metallic wear particles enmeshed in the engine filter are dislodged. After the cleaning period, the solvent containing wear debris was vacuum filtered through pre-weighed 47 mm diameter 0.45 μm pore size cellulose acetate membranes (now termed filter patches). After drying, the filter patches were weighed, the difference representing the mass of the engine wear debris collected. A cleaning time of 30 minutes was chosen for a recovery of debris between about 60 to 80 % (by mass) [4].

(2) Sample Dissolution: Filter patches were then dissolved in acid at high temperature through use of a microwave oven. A combination of three acids, HNO_3, HCl and H_2SO_4, was added to the samples in Teflon digestion vessels and placed in a microwave oven at 80% power for 5 minutes [4]. The filter patches were completely dissolved and some were analysed by AAS.

(3) Organic Conversion: The conversion method involves the use of a chelating agent. The chelate molecule readily attracts metal ions, making the ion soluble in oil. An excess of the chelating agents is used to ensure that the conversion of metal ions from aqueous to organic phase proceeds as desired. The chelating solution chosen was 1% 8-hydroxyquinoxine (oxine) in base oil [4].

A double extraction technique was developed and proceeded as follows. First, a buffer (2M ammonia and 3.17M ammonium bromide) was added and the reaction allowed to proceed to completion. Then 4.00M KOH and n-butylamine were added to bring the solution to a pH of 8.6. This solution and oxine in base oil were shaken and left long enough for the two phases to separate. The aqueous phase was collected for the second extraction and the oil was collected and set aside. To the aqueous solution, readjusted to a pH of 8.6, oxine in base oil was added, shaken and left to stand until the two phases separated. The second organic phase was added to

the first and shaken so as to combine the two extraction products. The final aqueous phase was discarded. The organic phase containing most of the dissolved wear debris was then ready for analysis by RDE-AES.

The chemical preparation took approximately 4-5 hours for twelve samples (the capacity of the microwave oven). Calibration and standardization took approximately 4 hours. The actual analysis with the RDE-AES required less than 15 minutes for the twelve samples.

The metals which are extracted adequately include Fe, Ag, Al, Cu, Mg, Ni, Pb, Ti, Cd, Mn, V and Zn. (This order of elements is the order that the Baird FAS 2C and MOA analysis is read and is also the order from major to minor wear constituents). With the exception of magnesium, silver and titanium, all the metals are extracted with percent recoveries of >80%. The remaining metals suffer from various difficulties from a different detector response to standards and samples by the RDE-AES (Cr), relatively large reagent blanks (Na, Si, Sn) and poor extraction (Ba, B, Mo). Note that as long as these recoveries, although different, are consistent, trending is valid.

ENGINE FILTER DEBRIS RESULTS: Engine filters were received from Challenger, CF188 and USN F/A-18 aircraft, for a total of 121 filter debris patches. After the sample preparation described above, these were analysed by a Baird FAS-2C RDE-AES instrument. Wear rates are used to represent the wear debris data since oil filters are not always removed at consistent time intervals. The values from RDE-AES were adjusted in the following manner:

$$\text{ppm metal} = \frac{[(\text{measured ppm value}) \times (10.0 \text{ mL oil})]}{\left[\frac{\text{debris weight in g } [/2(\text{for half sample})]}{0.835 \text{ g/mL density of oil}}\right]}$$

This calculation takes into account the different amounts of debris from the various filters, the density of the oil, and that only half of the dissolved sample would have been used in the conversion to the organic sample. This latter calculation was included for those samples in which the other half of the dissolved sample was diluted and analysed by AAS in order to evaluate the organic conversion efficiency.

Wear rates (μg metal/h) were then calculated from the concentration results (ppm metal) to depict a time dependent picture of the condition of the aircraft engine as follows:

$$\text{ppm metal} \times \text{amount of prepared sample} = \mu\text{g metal}$$
$$\mu\text{g metal} / \text{interval between filter change} = \mu\text{g/h}$$

This calculation for wear rates gives an average value of mass of metal collected by the filter during the previous interval that the filter was in the system.

For the engines from the CF188 and USN F/A-18, the wear rate results for one of the twelve elements, iron, are shown in Figures 1 to 2 and reveal a definite pattern of grouped normal wear with few high values. To clearly show this, the average (dotted line) of all the wear rates as well as two standard deviations (2σ) from that mean (upper solid line) are plotted. The upper solid line marks a 95% confidence limit that all the results below this limit can be considered normal wear. Values above this line are considered abnormal wear. Rather than plot all of the elements which behaved in a similar manner, the grouping of the results are indicated in Table I, where the mean and two standard deviations are given for all twelve elements. Since consecutive samples were obtained for the Challenger aircraft, wear rate results for four metals (Fe, Al, Cu and Ni) are plotted consecutively for various engines in Figures 3 to 6.

Table I - Mean and Two Standard Deviations from Mean in μg/h for CF188, F/A-18 and Challenger Aircraft Engines

Elements	CF188		F/A-18		Challenger	
	Mean	2σ	Mean	2σ	Mean	2σ
Fe	0.4747	1.6813	6.11	13.27	4.1	12.7
Ag	0.0805	0.2543	2.51	13.28	2.2	6.6
Al	0.0259	0.0537	0.6	0.87	2.6	9.4
Cu	0.1128	0.4883	0.37	0.6	0.4	1.0
Mg	0.0032	0.0067	0.1	0.21	0.2	0.9
Ni	0.6270	2.4289	7.71	18.31	0.5	1.9
Pb	0.0451	0.0651	1.33	4.44	0.4	1.2
Ti	0.0068	0.0326	0.18	0.48	0.0	0.1
Cd	0.0119	0.0415	0.18	0.64	0.0	0.1
Mn	0.0055	0.0184	0.07	0.13	0.0	0.1
V	0.0056	0.085	0.07	0.12	0.1	0.2
Zn	0.0363	0.0893	0.48	0.96	0.3	0.7

CF188 Hornet F404 Engine Filter Debris Results: Twenty six CF188 engine filter debris samples were analysed for the 12 elements and, as an example, the wear rate for iron is shown in Figure 1, and the mean and 2σ are given in Table I for all the elements. The results gave trends similar to those found by Fisher [5] by INAA. The dominant wear constituents of the CF188 wear debris were found to be Fe, Ag, Cu, Ni and Zn. The elements Al, Mg, Ti, Pb, Cd, Mn and V were present in lower quantities.

Wear rates for engine serial #'s 376028 and 376269 are well outside the 95% confidence limit. From the records received from BFC Bagotville, these aircraft had known failures at the time the samples were taken. Engines 376028 suffered from a radial drive shaft failure (elements outside the 2σ limit are Fe, Ag, Cu, Ni and Mn) and engine 376269 was inspected because a metal chip was found in the metal screen plug. Since an analysis was not conducted at BFC Bagotville on the metal chip, it is uncertain that the elements found outside the limit (Ag, Al, Ni, Ti and Cd) were those found in the metal chip. The wear rates are outside the limit for two other engines 376268 (Pb) and 376022 (Cd and Zn). For these two aircraft, no major maintenance was found to have been conducted. Notice that the two engines with failures had five elements out of limit whereas engines 376268 and 376022 have only one or two elements out of limit. Possibly, only one element above the 95% confidence limit is an indication of a potential problem, but more than one element indicates a definite problem with the engine. This premise could be confirmed by a more extensive data base.

United States Navy F/A-18 Hornet F404 Engine Filter Debris Results: Thirty eight USN F/A-18 engine filter debris samples provided by the Joint Oil Analysis Program - Technical Support Center (JOAP-TSC) were analysed. The patch made at the JOAP laboratory (after cleaning for 5 minutes) [6] was added to the patch made at RMC (after cleaning for 30 minutes) and their masses added together. As an example, the wear rate results for iron are shown in Figure 2 and the mean and 2σ are given in Table I for all the elements. The results gave trends similar to those found by Fisher [5] by INAA. The dominant wear constituents of the USN F/A-18 wear debris were found to be Fe, Ag and Ni. The elements Al, Cu, Mg, Ti, Pb, Cd, Mn, V and Zn were present in lower quantities.

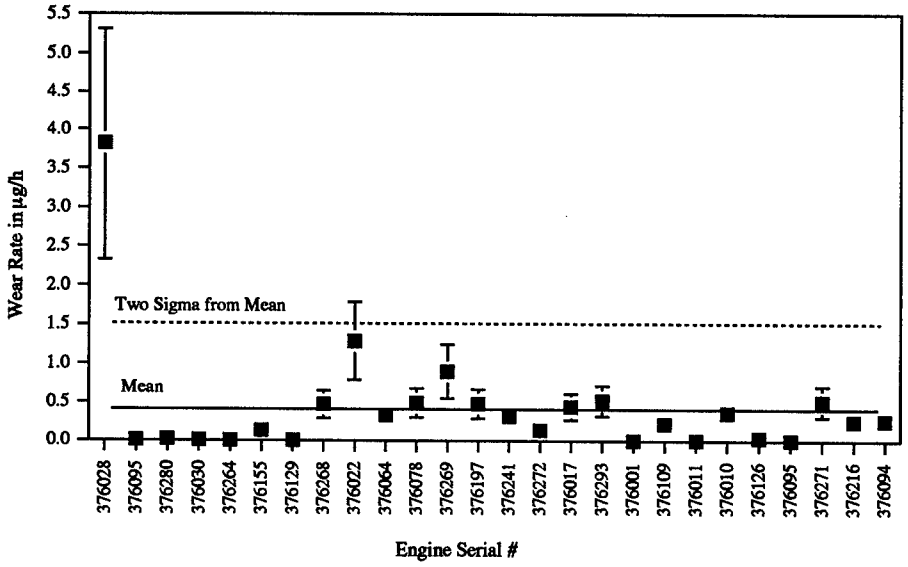


Figure 1 - Wear Rate of Iron from CF188 Engine Filters from April 94 to Dec 94

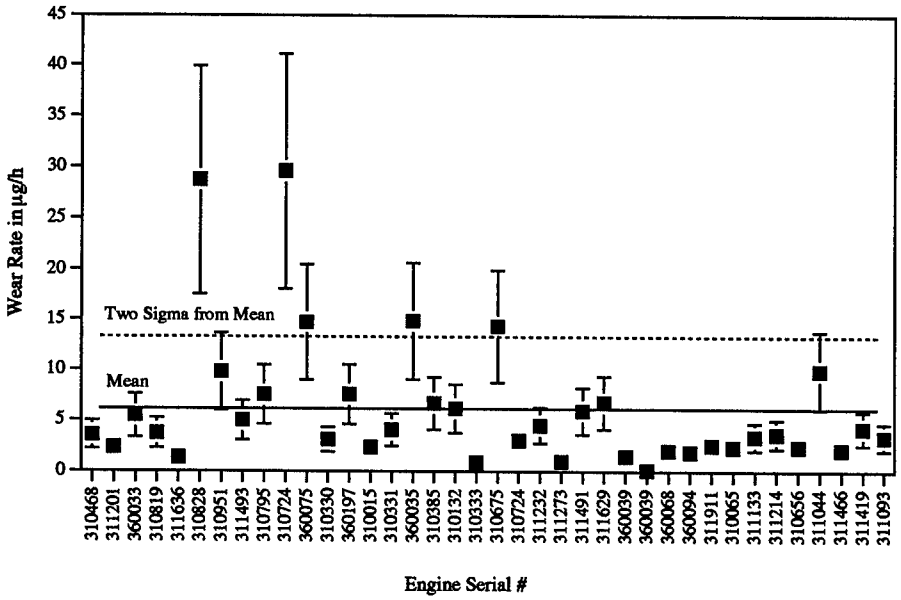


Figure 2 - Wear Rate of Iron from USN F/A-18 Engine Filters

Wear rates for engine 310828 are outside the 95% confidence limit for Fe, Al, Mg, Pb, Ti, Cd, Mn and Zn. Engine 310724 has wear rates outside the limit for Fe, Ni, Cd, Mn and V; engine 360075 for silver; engine 360035 for aluminum; and engine 311498 for lead. Confirmation of any difficulties with these engines is being requested through JOAP-TSC [6]. As no multiple engine samples were received, no engine trending can be shown.

Note that the wear rates for the USN F/A-18 aircraft are greater (as much as ten times) than those for the same engine type in the Canadian CF188. The US aircraft may be flown under more stressful performance conditions than the Canadian equivalent and the maintenance may be different.

CC144 Challenger ALF 502L-2C Engine Filter Debris Results: Fifty seven Challenger engine filter debris samples were analysed for 12 elements and the mean and 2σ are given in Table I. The dominant wear constituents of the Challenger wear debris were found to be Fe, Ag, Al, Cu, Mg and Ni. The elements Pb, Ti, Cd, Mn, V and Zn were present in lower quantities.

Since Challenger aircraft 604 and 605 have consecutive samples for a one to two year time period, trending can be demonstrated (Figures 3 to 6). The wear rate results for aircraft 604, engines 3061 and 3014 (replaced by 3038) are shown in Figures 3 and 4; the results for aircraft 605, engines 3189 and 3016 (replaced by 3163) are shown in Figures 5 and 6. All of the wear rate results show a similar pattern of grouped normal wear with few high values.

For aircraft 604, engine 3061 for 25 Sep 94 (Figure 3), the Fe and Al results are above their 2σ (39 μg/h and 33 μg/h, respectively) as are the other elements except Pb and Ti. In Figure 4, for engines 3014 and 3038, normal wear rates occur with all of the data points within the 95% confidence limit (see Table I for mean and 2σ values). However, engine 3014 has results for Ag, Cd, Mn and Zn above the 2σ limit. Engine 3014 was replaced by 3038 for time-expired maintenance.

In Figures 5 and 6, the wear rates for the left (3189) and right hand engines (3016, 3163) on aircraft 605 are plotted consecutively. In Figure 5, for aircraft 605, engine 3189 on 16 Jan 95 the wear rate for iron is above the 2σ limit at 27 μg/h (as are Al, Mg, Mn and V). The first three data points in Figure 6 are the wear rates of engine 3016 which was replaced by 3163. Apparently, engine 3016 was experiencing problems; however, the wear rates were normal for that engine. The last three points in Figure 6 were sampled at the same time (15 Oct 94, 20 Nov 94 and 16 Jan 95) as those in Figure 5. The wear rates for all elements for 16 Jan 95 (last data points) indicate an increase, although still within limits, except for Ag, Cd and V. Two explanations are possible, with the first being abnormal wear. However, aircraft 605 and 604 have been sent to a contractor for periodic maintenance so that any malfunction will be repaired. The second possibility and the more likely, since both engines are affected, is that the aircraft performed a flight profile that increased the loading on the engines which in turn increased the amount of wear. In this case, engine damage could only by concluded if that one engine did not have reduced wear rates at the next filter change.

DISCUSSION: Aircraft engines with fine filtration in their oil-lubricated systems cannot be monitored by the conventional AOA method. Most of the wear debris has been removed from the oil system and is in the oil filter. Previous work confirms that examining the wear debris from the oil filter does allow condition monitoring to be accomplished [5, 7]. Several analytical techniques are available for QFDA at various stages after the sample preparation. After the cleaning step (1), INAA could be employed, or after the dissolution step (2), AAS or ICP-AES could be used.

However, in order for QFDA to be employed in the field where the routine determination of the condition of the aircraft equipment in question is critical to safety and maintenance costs, the

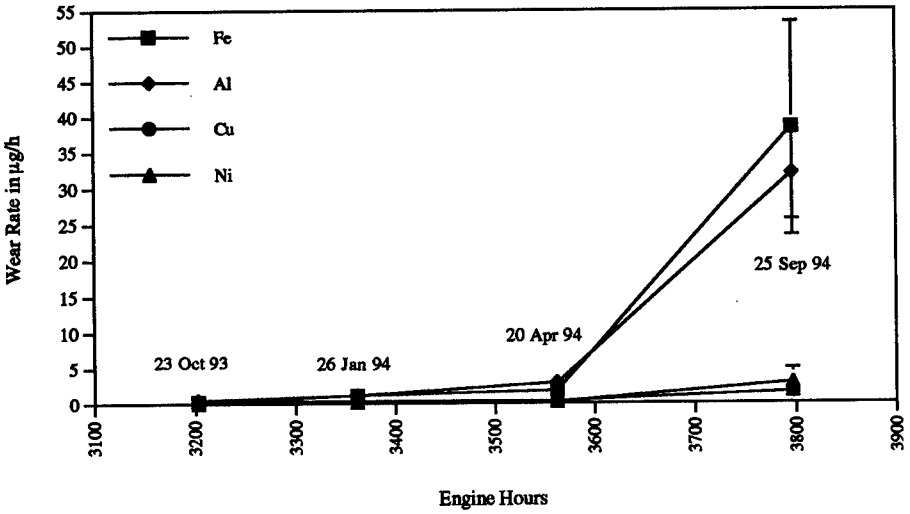


Figure 3 - Wear Rates from Challenger Engine Filters Aircraft 604 Position #1 Engine 3061 from Oct 93 to Sep 94

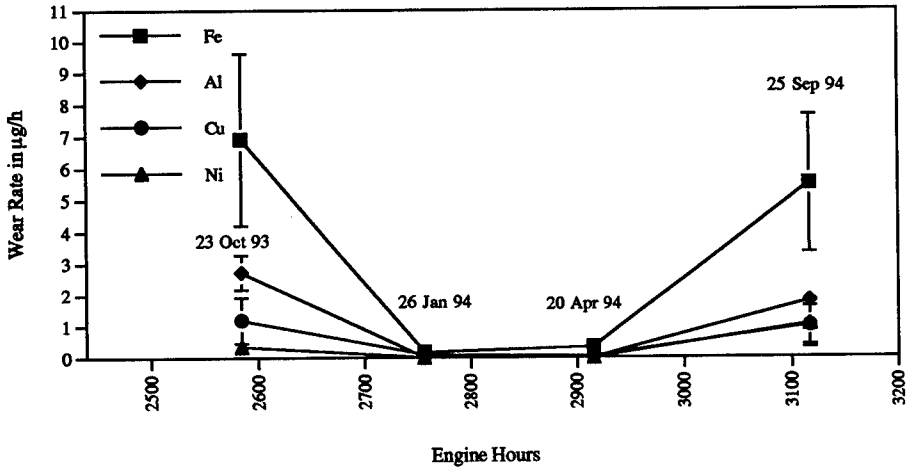


Figure 4 - Wear Rates from Challenger Engine Filters Aircraft 604 Position #2 Engine 3014 & 3038 from Oct 93 to Sep 94

Note: Engine 3014 (first three data points) was replaced by Engine 3038 (last three data points)

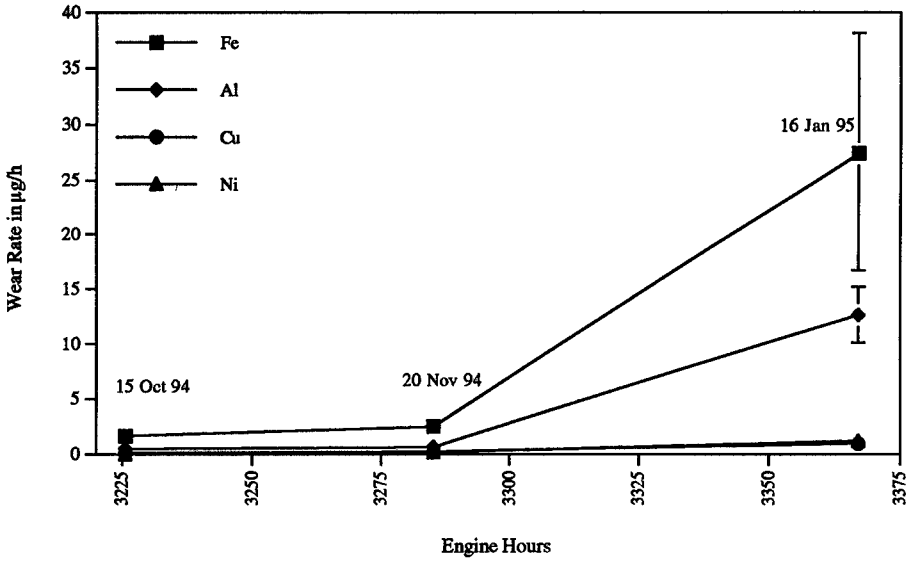


Figure 5 - Wear Rates from Challenger Engine Filters
Aircraft 605 Position #1 Engine 3189 from Oct 94 to Jan 95

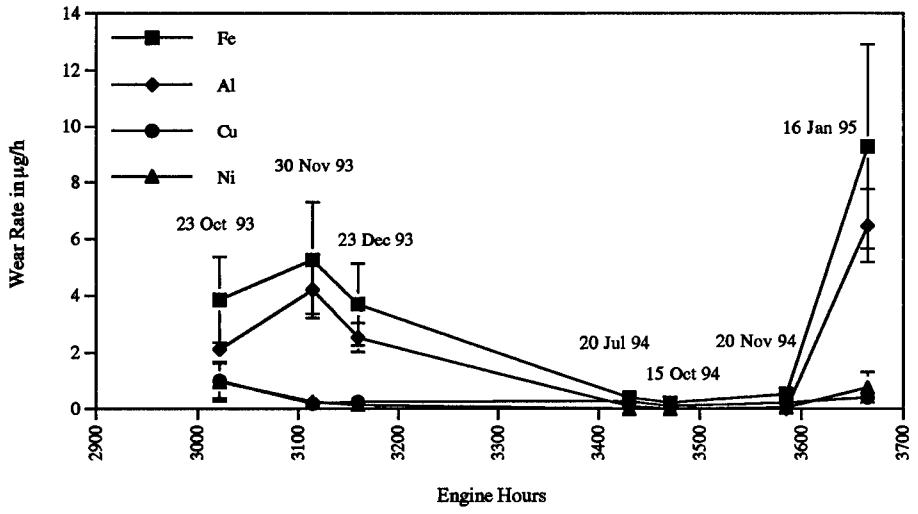


Figure 6 - Wear Rates from Challenger Engine Filters
Aircraft 605 Position #2 Engine 3016 & 3163 from Oct 93 to Jan 95

Note: Engine 3016 (first three data points) was replaced by Engine 3163 (last four data points)

technique has to be usable by aircraft technicians on RDE-AES instruments already acquired by the CF. Thus, the sample was prepared with three steps of cleaning, dissolution and conversion. The first two steps are relatively simple and fast in comparison to the third step which was carried out in order for the sample to be oil based. Even though each of these steps were kept as simple as possible, sample preparation requires four to five hours for batches of twelve.

The cleaning and dissolution steps are fairly straightforward and are not sensitive to small changes in procedure. The organic conversion step is the most time consuming. It is very sensitive to pH changes in the solution and very small changes in pH ($\pm$ 0.5) can reduce the % recovery to zero. Of the three sample preparation steps, the conversion to an organic matrix contributed the most to the sample loss. This conversion loss was estimated by determining the % recovery for the elements of interest in known samples made from AAS standards [4].

However, in spite of these apparent difficulties, the wear rates of the actual filter debris samples showed that most samples were normal and a few were beyond the 95% confidence limit. This coincided with what was found previously with INAA [5]. Of the four CF188 samples with one or more elements out-of-limit, only the two with multiple elements were associated with actual failures (radial drive shaft failure and chip indication on metal screen plug). None of the normal samples were found to be failures. It appears that multiple out-of-limit results are required to indicate an engine failure.

Since only a small number of engine filter debris samples were received to date, the conclusion made above cannot be considered comprehensive, but it does point out that the technique is able to indicate failures. As more samples are received, a greater confidence in wear rate trending will be possible. As with any condition monitoring technique, the ability to determine whether the technique is pinpointing actual starts of failure is dependent upon the records and history kept on the corresponding components. The number of samples in this study are not sufficient to ensure that all points above the 2σ line are indicative of abnormal wear, but, with each additional sample and knowledge of its maintenance history, an acceptable limit for abnormal wear can be determined.

CONCLUSIONS: With the advent of fine filtration on new and old aircraft engines and related components in the CF, the ability to monitor effectively the condition of those components by current practices, such as AOA, has been decreased. The potential for QFDA to monitor wear debris from the oil-lubricated components has been shown. Filter debris samples from the Challenger, the CF188 Hornet and the USN F/A-18 aircraft engines were analysed to indicate the levels of normal wear rates for twelve elements.

The calculated wear rate results for the Hornet and Challenger aircraft revealed that two distinct levels of wear can be indicated as normal and abnormal wear. By using a normal distribution, the abnormal wear rate level was set at the 95% confidence limit. Usually there was a wide gap observed between the normal and abnormal wear rate regions so that the difficulties with the % recoveries used in the conversion step will not affect the determination of abnormal wear rates. The wear rates for the elements showed moderate scatter about a standard mean. The 95% confidence limit was successful in determining abnormal wear in two CF188 engines that experienced failure. The other samples were generally well below this limit. Overall, the F/A-18 aircraft engines experienced higher wear rates than the same type of engine on the CF188 Hornet. With more samples and a different mean, these might be considered normal wear. This research demonstrated that trending the wear rates of wear elements by the proposed technique can provide a means of monitoring the condition of aircraft engines.

RECOMMENDATIONS: QFDA investigations should be continued for the ALF 502L-2C and F404 engines in order to continue building a data base. This technique could be expanded to other candidate engines and related components. A project has been started to involve unit

personnel in the evaluation and implementation of this technique in the AOA laboratories. The sample preparation procedure efficiency, especially the % recovery for the conversion step, should be determined by using actual samples and by comparing other techniques (i.e., INAA and AAS). Other analytical techniques such as ICP-AES, which is an appropriate method of analysing trace metal content in acid solutions, could be studied.

References

1. Morin, J., "Canadian Air Force Condition Monitoring Program", Joint Oil Analysis Program - International Condition Monitoring Conference, Department of Defense: Florida, 1992.
2. Flanagan, I.M. et al, Wear-debris detection and analysis techniques for lubricant-based condition monitoring, Journal of Physics E Scientific Instruments, Vol 21, 1011-1016, 1988.
3. Waggoner, C.A., Application of Diagnostic Wear Debris Analysis to CF188 Aircraft Engines, DREP: Esquimalt, 1988.
4. Swanson, S.A., Quantitative Filter Debris Analysis: Sample Preparation for Analysis by Rotating Disk Electrode Atomic Emission Spectrophotometry, M. Eng. Thesis, Royal Military College, Kingston, 1995.
5. Fisher, G.F., Quantitative Filter Debris Analysis (QFDA) as a Means of Monitoring Wear in Oil-Lubricated Systems, M.Eng. Thesis, Royal Military College, Kingston, 1991.
6. Humphries, G., personal communication, Nov, 1994.
7. Harper, W.A., The Determination of Wear Elements in Engine Lubricating oils and Filters by Neutron Activation Analysis, M.Eng. Thesis, Queen's University, Kingston, 1988.

Acknowledgments

This research was supported financially by the Head/Air Vehicle Research Section 3, the Directorate of Technical Airworthiness 3 of the National Defence Headquarters, and the Principal, Royal Military College of Canada - Collège militarie royal du Canada. The participation of BFC Bagotville, CFB Cold Lake and CFB Shearwater in sending oil filters and records is also gratefully acknowledged.

ON-LINE CONDITION DIAGNOSIS

FLEXIBLE EXPERT SYSTEM FOR AUTOMATED ON-LINE DIAGNOSIS OF TOOL CONDITION

Erkki Jantunen
Technical Research Centre of Finland, Manufacturing Technology,
P.O.Box 1702, FIN-02044 VTT, Finland

Harri Jokinen
Technical Research Centre of Finland, Manufacturing Technology,
P.O.Box 1702, FIN-02044 VTT, Finland

Robert Milne
Intelligent Applications Limited,
1 Michaelson Square, Livingston, W. Lothian, Scotland EH54 7DP

Abstract: An increasing number of Flexible Manufacturing Systems (FMS) have been installed in Europe during the past few years. The general experience is that the availability of the FMS is not as high as was originally expected and especially their unmanned use during three shifts has not been successful. One of the major problems is the deterioration or failure of the tools. To develop a system to address this, a wide cutting test and analysis program for tool wear was performed. The test program covered both shank end and end mills together with twist drills and tread taps. For monitoring the tool wear a number of monitoring methods such as vibration, acoustic emission, sound, spindle power and current, axial force, torque were tested. The relations between the analysed signals and tool wear form a basis for the diagnosis rules that are used in an diagnostic expert system module. An expert system for automated on-line diagnosis of tool wear of different types of tools was built using a new approach. In this approach the faults are described in a fault tree database and the corresponding features of condition monitoring signals together with the machine status information are described in a symptom tree database. Using a rule synthesiser program the information gathered in the databases is automatically converted to expert system code.

Key Words: Artificial intelligence, Condition monitoring, Expert systems, Flexible manufacturing systems, Tool wear monitoring

INTRODUCTION. An increasing number of Flexible Manufacturing Systems (FMS) have been installed in Europe during the past few years. A general experience is that the availability of the installed FMS is not as high as was originally expected, and especially the unmanned use has not

been successful [1]. A major problem is the condition of the tool. One of the most important reasons for this is today's existing real-time tool condition monitoring techniques do not cover the wide range of different machining situations and machining parameters that normally take place in practice.

There is a need to group and synchronize sensor signals together to avoid poor correlation between a single signal source and the measured event. However, it is obvious that the commercially available monitoring systems are exploiting just a limited number of the capabilities of modern sensor and analyzing techniques [2]. In this survey several sensors were installed and comprehensive laboratory tests done before the concluding sensor validation [3]. The aim was to accomplish the requirements of condition monitoring at critical points of the machine tool and in the cutting processes. The validation was performed according to the following criteria: sensitivity of the sensor to the measured event, correlation between the signal and measured event, the amount of deviation and universality. The information received from multiple sensors was analyzed with many different methods. The relations between the analyzed signals and wear form a basis for the diagnosis rules that can be used in an AI system.

The diagnosis of the condition monitoring signals has to be done using an expert system since the goal is to be able to use FM-systems unmanned in three shifts. One big problem in reaching this goal is how much the FM-systems differ from each other. From this it follows that an expert system should be easy to configure for each task. Unfortunately, the behavior of measuring signals is a function of the machine tool and tool type. In this study a lot of emphasis has been put on creating a generic solution, and a new approach to configuring an expert system. The basic idea is to use databases for defining the varying information, and to use specific software to write the expert system code automatically. The user only defines the necessary data with the aid of modern tools for database handling and the computer translates that information into a working expert system code.

TEST ARRANGEMENT. A small horizontal-type machining center with an 11 kW main motor power was used in the cutting tests for tool and machine tool condition monitoring. The tests concentrated on tool wear, tool breakage and collision monitoring. Cutting tests were performed in order to create situations where a measurable event was present due to tool wear or failure. In this part of the cutting tests, different types of cutting tools were used to cover a wide range of different cutting methods. The tools investigated in the tests were shank end mill (diameter 6 and 10 mm, HSS), end mill (diameter 50 mm with carbide inserts), twist drill (diameter 3.3, 5.0, 6.8, 8.5 and 10.2 mm, HSS) and thread tap (M4, M6, M8 and M12, HSS). The tool-monitoring tests were carried out by using different kinds of measuring arrangements. The main configuration of the measuring arrangement can be seen in Figure 1. A more detailed description of the measuring arrangement is in reference [3].

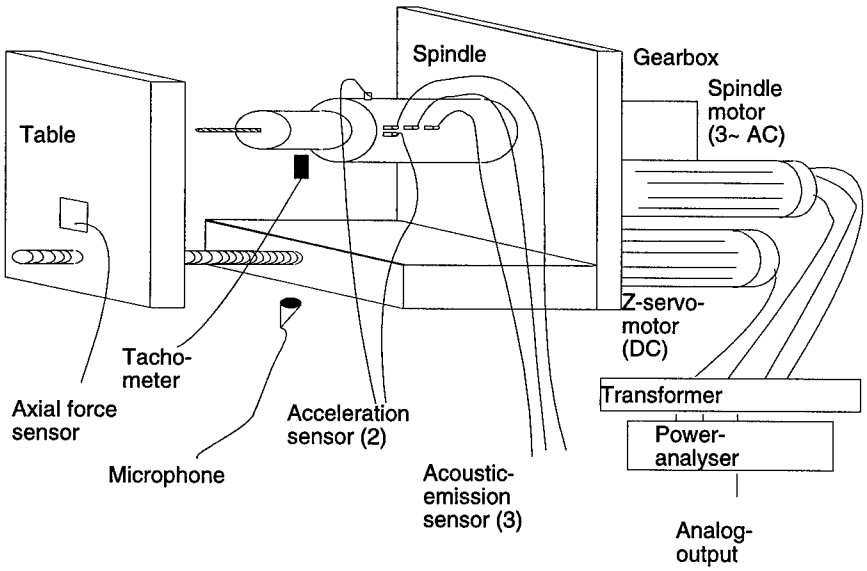


Figure 1. Measuring arrangement

EXPERT SYSTEM. With the increasing complexity of NC machine tools and flexible manufacturing systems, and with the growth in diversity of systems, it is increasingly difficult for maintenance personnel to assess the problems rapidly. Long delays in identifying the precise problem greatly increases downtime and causes even bigger secondary problems in plants. This is especially true for FM systems where the material flow from a machine tool to another is performed quickly with a minimum number of buffers.

To automatically identify the condition of the machine tools and cutting process two types of diagnostics are needed. A reactive diagnosis is needed to identify the cause of a current problem and a predictive diagnosis is needed where ever it is possible to anticipate the need for a maintenance action based on the condition monitoring of the machine tool or cutting process.

The development of a diagnostic expert system is based on diagnostic rules which are derived from the results of the condition monitoring tests. From the results of the analysis it became apparent that rules would become rather complex if the system were completely generic, i.e. suitable for a number of different types of FMS environments, and also that there would be a need to divide the tasks within the system so that responses would be fast enough to perform the diagnosis in the case of collision and tool breakage also.

Principles of the chosen approach. In order to make an expert system flexible and suitable for a wide range of FM systems, a new approach for defining and modifying the rules was developed.

The basic idea is to use the fault tree database definition program for defining the faults, and describe corresponding condition monitoring tools (symptoms) using the symptom tree database definition program. After that, the user starts a rule synthesiser program that translates the contents of the fault and the symptom databases into expert system rule code for the computer performing the monitoring task. In principle the translation program simply takes one page from the symptom tree at a time and writes a module to the expert system code from that. The procedure is shown in Figure 2.

It is considered that there are several advantages to this approach: It is not necessary to write enormous amounts of expert system code manually. It is very easy to make changes or add more information and, especially, it is possible to configure the program for a specific FM-system. Apparently, there are also certain disadvantages to this approach: The amount of code in the final expert system will be considerable since it is not possible to use the sophisticated features of an expert programming package. Instead of a sophisticated system the expert system programming module always writes rather simple modules for each condition defined in the symptom tree.

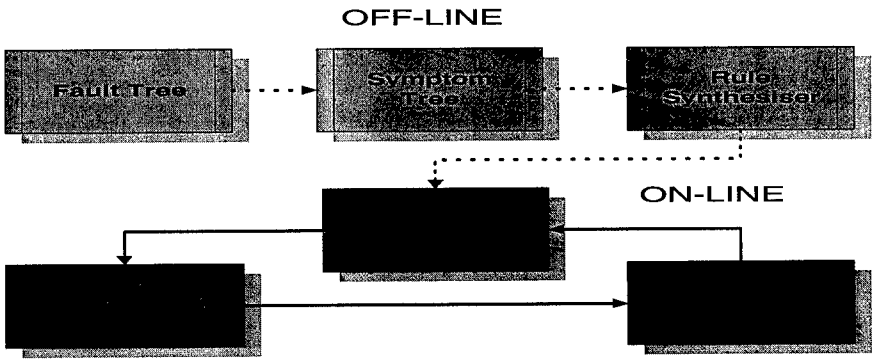


Figure 2. Principles of the new approach to expert system rule generation.

Fault tree. In the fault tree database the machine tool is defined with the chains of subcomponents. The number of subcomponent levels is limited to five. For the lowest level of subcomponents where a principal fault can take place, all the possible faults are described. The fault tree database program has all the typical search and editing functions of a normal database program. The structure/window of the fault tree is shown in an example in Figure **3.**

Symptom tree. Following the definition of all the relevant faults with the fault tree database, the next step in building an expert system with this new approach is to define all the symptoms related to the faults together with information about the machining process. As shown in Figure 4 the following input is defined in the symptom database: fault tree component chain identification, fault to be processed, tool identification, the status information of the machine tool, machining information and condition monitoring method information (symptom). The symptom is defined with signal, general, analysis and limit value information (Figure 4).

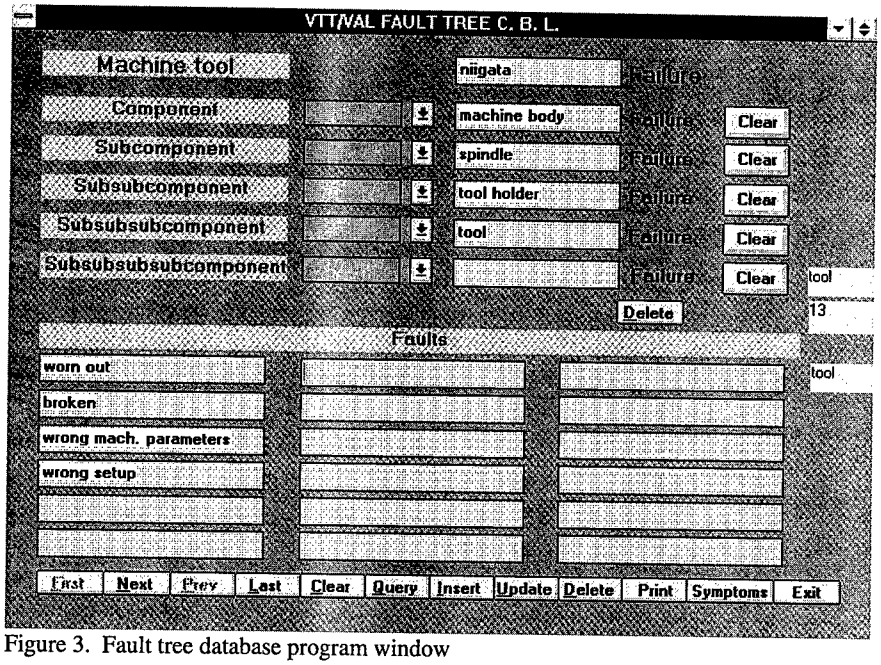


Figure 3. Fault tree database program window

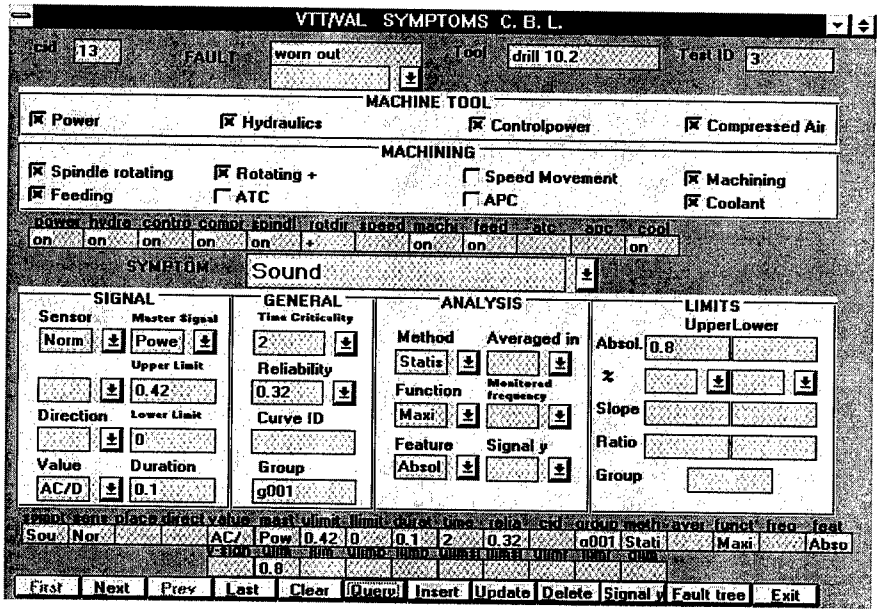


Figure 4. Symptom tree database program window

SIGNAL PROCESSING. For the automatic analysis of the huge amount of data to be gathered an interface was created using the Visual Basic programming environment for Windows. The system gathers the data with a data acquisition board (16 A/D channels, Keithley) and the necessary calculation procedures are defined using a collection of subroutines (VTX) for this AD card family.

Data acquisition. The statistical analysis is based on the acquired data. Attached to the acquisition board, a sample and hold board is used to synchronize dynamic signals. The measured signals are analyzed with a number of different methods in the time domain and in the frequency domain. In the case of dynamic signal analysis, only so-called cursor values are gathered to minimize the amount of information to be stored in the databases. These actions are also controlled by the interface. The flowchart of data acquisition is shown in Figure 5.

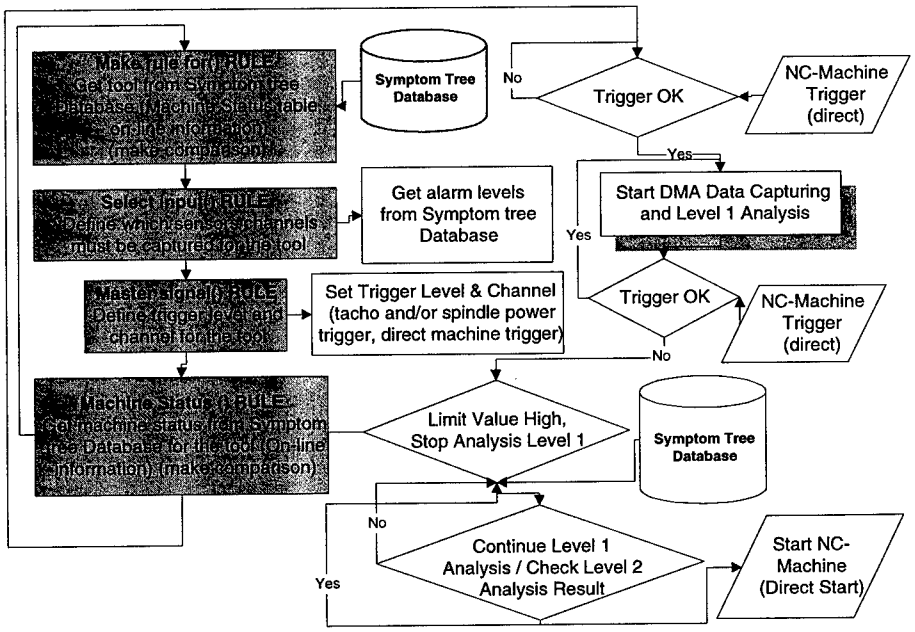


Figure 5. Level 1 data acquisition

Signal analysis. Depending on the measured events, the data to be analyzed is first cleaned of irrelevant signals, e.g., rapid movement during drilling, that has not been recorded during the actual machining process. Data measured and recorded simultaneously from the sensors is studied by calculating a number of statistical parameters: arithmetic mean, root mean square (RMS), mean deviation, standard deviation, skewness, kurtosis, maximum and minimum.

In the case of dynamic signals containing frequency information, Fast Fourier Transformation (FFT) techniques are employed [4]. The sample and hold function of the data acquisition board is used to get data from four channels simultaneously. It is possible to perform a FFT with both time and frequency domain averaging. Different kinds of analysis functions such as spectrum, cross-spectrum, frequency response, coherence, coherent output power, autocorrelation, crosscorrelation, cepstrum, liftered spectrum, 1/3 octave spectrum, 1/1 octave are available.

Regression analysis. The results of the statistical analysis and FFT analyses are further analyzed using regression analysis techniques. Different regression functions were tested to find the highest correlation between measured tool wear and analyzed measurement signals. The sets of data points are approximated as closely as possible with the four smoothing functions [3]: first, second and third order polynomials and one logarithmic function based on a simplified mathematical definition of wear [5] using the least square principle.

The degree of the fit is much higher for cursor values of the FFT functions than for the statistical parameters. This result is logical, since the idea of the FFT analysis is to separate meaningful information from noise. However, it takes time to carry out the FFT analysis, which makes it impossible to use the FFT for collision and tool breakage monitoring but enables tool wear monitoring. The goodness of fit varies with the tool-type. Drilling and shank end milling are the easiest to monitor. Functions describing how much two signals are related to each other, show a rather high goodness of fit (3). The use of at least two signals for tool wear monitoring coincides with the findings of the statistical analysis, since the best methods for monitoring purposes vary between the tool-types.

Simulation module. This work is focused on problems of automatically identifying the condition of the cutting process. The simulation program module is used as a tool to see how the expert system reacts in different kinds of situations with different kinds of limit values for the chosen regression models. The program fits a selected regression curve to the existing data. The fitted curves are the same as those used in the regression analyses. After each time the curve fitting has been done, the program checks to see whether the tool worn-out limit has been reached, giving a warning when the machining process has to be terminated by the expert system. The advantage of using the simulation module together with the regression functions is that with this procedure the amount of measuring data to be stored in the databases is greatly reduced and the method is not too sensitive for sudden changes in the measuring signals i.e. it is possible to avoid most false alarms.

Rule Synthesiser. Knowing what signals to process, how to process them and what features or thresholds to look for after processing, requires considerable knowledge about how tools deteriorate and the signal processing capabilities that are available. This 'expert' knowledge was gained during the initial experimentation of the project. A key facet of the knowledge is the processing to perform changes from tool type to tool type. Like traditional expert systems, a mechanism is needed to allow the expert to naturally specify his knowledge of how to process

the signals for a given tool. However, the system cannot use one fixed knowledge base, since the best way to analyze the signals will vary from tool type to tool type. Hence a more flexible approach to building the expert system is needed, since the 'expert system element' will also change from tool type to tool type.

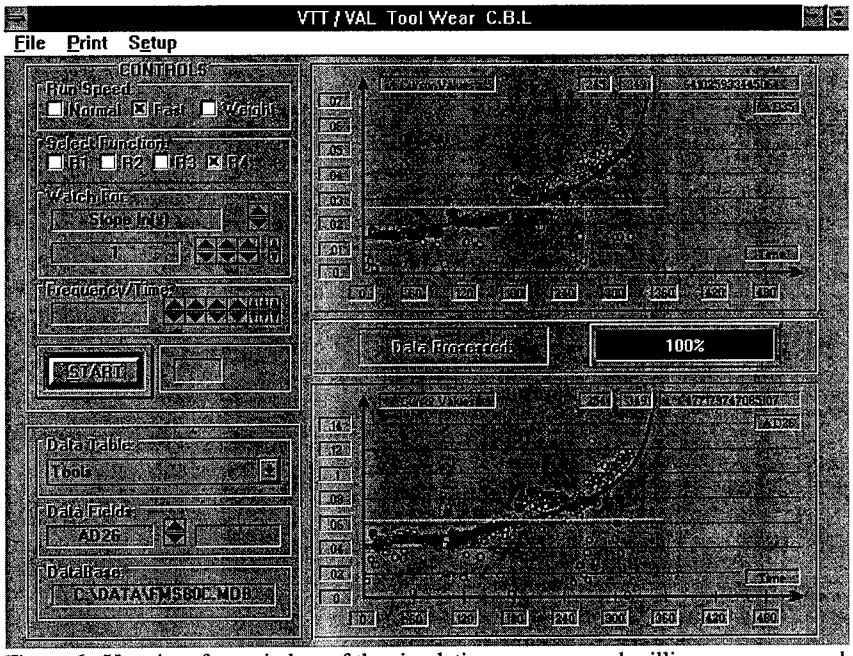


Figure 6. User-interface window of the simulation program, end milling, upper curve: horizontal vibration, mean deviation, lower curve: sound, root mean square

The information in the symptom tree database can be viewed as a specification for the knowledge based rules. Each entry specifies what a rule should look like; what signals to examine, what processing to perform, which features to extract and what thresholds to use. Using a traditional expert system environment would be too complex for experts in this domain to specify all this information. Also, hidden from the expert is the fact that each 'rule' has several parts, corresponding to the above steps. The database front end provides an user interface that is natural for the expert to use and hides this underlying complexity. Our system uses a 'rule synthesiser'. This takes the specification of a rule that is contained in the symptom tree database and automatically generates the computer programs needed to implement these 'rules'. Figure 7 shows the steps and database interactions that result from each rule. Contrasting this with the simple layout of Figure 4, illustrates the gain that results from the use of the rule synthesiser.

The rule synthesiser works by processing each rule specification in the symptom tree database, breaking each rule into several function calls. It also builds the links between these function calls so that the data can progress through the steps of acquisition, signal processing, feature extraction

and testing against the specified limits. In addition, it automatically combines rules into groups, for example, all the rules to detect worn out 10.2 mm drills. This process hence implements one of the key ideas of expert systems - let the expert specify the knowledge in a natural way and have the system do all the hard work.

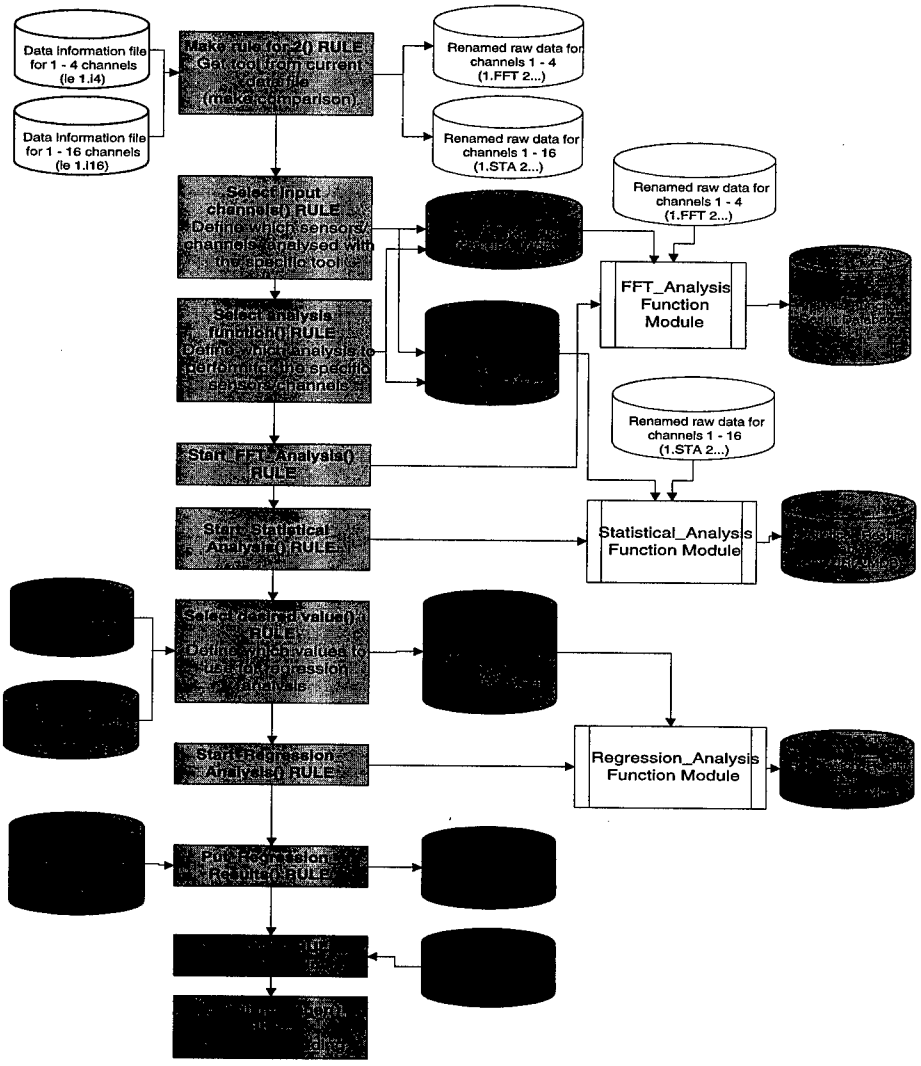


Figure 7. Level 2 rules and data model

CONCLUSION. It has been shown that there exists a great potential to improve the utilization rate of a machine tool by an advanced condition monitoring system using modern sensor and signal-processing techniques. A comprehensive cutting test procedure has been carried out with different tools and measurement arrangements. The recorded signal information has been processed in several ways, both in the time and the frequency domain. The effectiveness of the best sensors and analysis methods have been verified in the prediction of the remaining lifetime of a tool in use. The relations between the analyzed signals and tool wear form a basis for the diagnosis rules that are used in a diagnostic expert system. A new approach to development of a rule-based expert system is reported. The approach makes it easy to configure the expert system for different types of FM-systems with different types of tools. The solution is based on the adoption of fault and symptom tree databases with sophisticated user interfaces for the definition of the relevant fault types together with the corresponding monitoring methods. A rule synthesiser is used to take the specification of a rule and produce the detailed expansion to control the various sub systems of the condition monitoring system. The hides from the expert the underlying complexities of the task and lets him specify knowledge in a way that is natural to him. Without this capability, the system would be too complex to be used by the people who have the appropriate knowledge, and hence the value of the whole system would be greatly reduced. This work has combined extensive laboratory testing with the implementation of a state of the art signal processing environment and an easy to use way to specify the knowledge about how to interpret the data that is collected. Such a system is not only practical, but is an essential part of the how automation can be used to increase the utilization of flexible manufacturing systems.

ACKNOWLEDGMENTS. The results of the tool wear tests are used in the development of an integrated condition and machining process monitoring system for flexible manufacturing systems and for stand-alone NC-machine tools (FMS Maint System, 1993-1996, EUREKA MAINE EU 744 project). The work of the project team, D. Deasy, J. Hinkkanen, M. Karhu, H. Komulainen, P. Kontkanen, A. Poikonen, D. Smith, R. Vannela and K. Vähä-Pietilä is gratefully acknowledged.

REFERENCES.

1. Lakso, T. The Influence of FMS-techonology on the Efficiency of NC-controlled Machine Tools. Tampere University of Technology. Publications 50. Tampere 1988. 106 p.
2. Kluft, W. Techniken zur Werkzeugüberwachung auf der Basis von Kraft- und Verformugsmessung. Sonderdruck aus der Vortagsreihe des VDI-Seminars "Überwachung und Diagnose von Maschinen und Prozessen " von 3 -4-Juni 1985 in Aachen, Germany. 14 p.
3. Jantunen, E. & Jokinen, H. Automated on-line diagnosis of cutting tool condition. 5th International Conference of Flexible Automation & Intelligent Manufacturing, FAIM 95, June 28-30, 1995, Stuttgart, Germany
4. Randall, R.B. Application of B&K equipment to frequency analysis. In Brüel & Kjær. Nærum, Denmark. 239 p.
5. Jantunen, E. & Poikonen, A. Dynamics in monitoring gear faults. Proceedings of the 47th meeting of the mechanical failures prevention group. Virginia Beach, Virginia, USA, April 13-15 1993. p. 279-288.

ADVANCES IN OPTICAL OIL DEBRIS MONITORING TECHNOLOGY

J. Reintjes, R. Mahon*, M. D. Duncan, L. L. Tankersley†, A. Schultz**, V. C. Chen**, P. L. Howard***, S. Raghavan††, and N. Gupta††

Laser Physics Branch, US Naval Research Laboratory
Washington, DC 20375

Abstract: The status of two optical debris monitoring programs is described. The optical debris monitors are directed at developing on line technology for identifying type and severity of faults in machinery through measurement of size, shape and morphology of debris particles in real time. Operational characteristics of the monitors in two different size ranges is described.

Key Words: Bearings; early warning; catastrophic failure; gears; hydraulic fluid; real-time; shape classification; wear debris

Introduction: We have previously described real-time, imaging optical debris monitors. [1,2] These allow real time or near real time detection of ferrous, non-ferrous and non metallic debris in lubricating oil or other fluids in the general size range above about 5 micrometers. The optical debris monitor provides the capability for determining both the type of fault and its severity through combined analysis of particle size, rate of production and particle morphology. The general form of the optical debris monitor is illustrated in Fig. 1. It consists of a laser illuminator, an image detector and an image processor. Because of limitations associated with optical resolution and depth of field we have found it convenient to break the size of the detected particles into two ranges - 1) larger than about 50 μm and 2) larger than about 5 μm.

The large particle detector, referred to as LASERNET, is appropriate for detecting failure related debris in machinery lubricating oil in real time. One application of the monitor in this form is in helicopters to provide improved early warning of faults before they advance to catastrophic failures. Because the relatively large particles are trapped in the filter, the LASERNET detector will typically be located in the scavenge oil line ahead of the filter. The particles must then be detected on their only pass through the viewing cell. Typical operating conditions (flow speeds of 10 m/sec, viewing apertures of ~ 1 cm), along with the requirement to illuminate all of the oil column to maximize the detection efficiency, requires that the laser pulse duration be of the order of 1-2 μsec, the repetition rate be of the order of 0.5 - 1 kHz and that the image detector have approximately 1000 x 1000 elements and be capable of framing at 1 kHz. The images must be scanned in real time to determine the presence and characteristics of debris particles. Besides the high data rates the greatest challenge is to distinguish debris particles from air bubbles. In our system air bubbles are identified on the basis of their round shape, while debris particles larger than 50 μm will have irregular shapes.

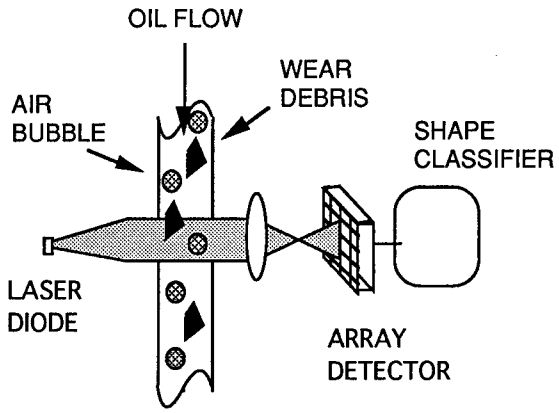


Fig. 1. A schematic diagram of the optical oil debris monitor, showing the illuminating laser diode, the array photodetector and the shape classifier. Objects suspended in the oil flow are imaged in transmission onto the array detector, appearing as dark shadows against a bright background.

LASERNET AVDS SYSTEM: We have implemented the LASERNET detector for test on the T700 engine and the main gearbox of an SH-60 power train in the test cell at the Naval Air Warfare Center Aircraft Division at Trenton, NJ under the US Navy's Air Vehicle Diagnostic System (AVDS) Advanced Technology Demonstration program. The test system is illustrated schematically in Fig. 2. The LASERNET system consists of a laser illuminator, a flow cell adapter, a high speed camera and a high speed image processing system. For the tests at Trenton, additional components for test validation have been included. These consist of a video camera for visualizing the flow conditions, an auxiliary high speed memory and a standard chip detector.

The illuminator is a commercial laser diode operating at 830 nm. The laser is coupled to a single mode fiber and the output beam expands to the desired final diameter and is collimated with an output lens. The final diameter is chosen to be larger than the illuminated area of the LASERNET viewing cell to ensure uniformity of illumination for the image detector. Typical center to edge variation in intensity is 0.8. The center of the viewing cell is imaged onto a high-speed camera from Silicon Mountain Design. This camera is capable of framing at speeds up to 1 kHz with up to 8 bits per pixel. In order to accommodate the speed requirements of the image analysis in the LASERNET system we have implemented the camera with 1 bit for the current set of tests. The high speed camera thus provides black and white images for analysis. Some control over the details of the image can be obtained by adjusting the bright-dark threshold of the camera and the peak laser intensity.

Each existing engine or gearbox requires its own adapter and fittings for retrofit. For the engine tests at the Trenton test cell we have chosen to attach the LASERNET system to the engine auxiliary gearbox manifold at the chip detector port as shown in Fig. 3. Since the T700 engine has cast internal oil channels the engine adapter has a re-entrant design. The oil flows out of the

engine through a center column, into the LASERNET chamber and then returns to the engine through an outer annulus. The flow pattern at the engine is similar to that provided by the existing chip detector.

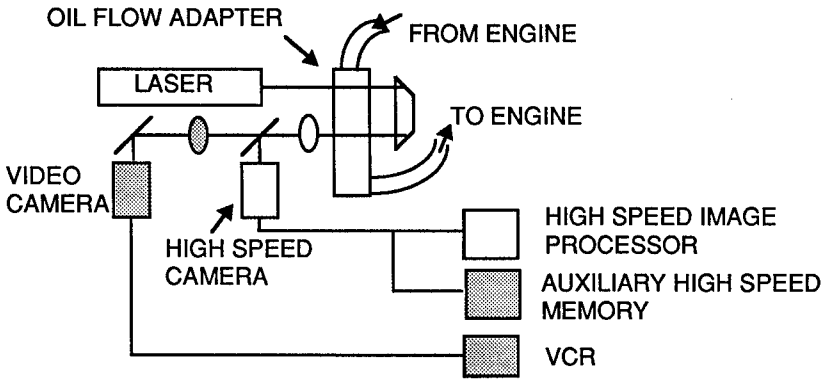


Figure 2. Schematic Diagram of LASERNET test system for T700 engine at the Trenton test cell. Items shown shaded are for test validation and not part of basic debris monitor.

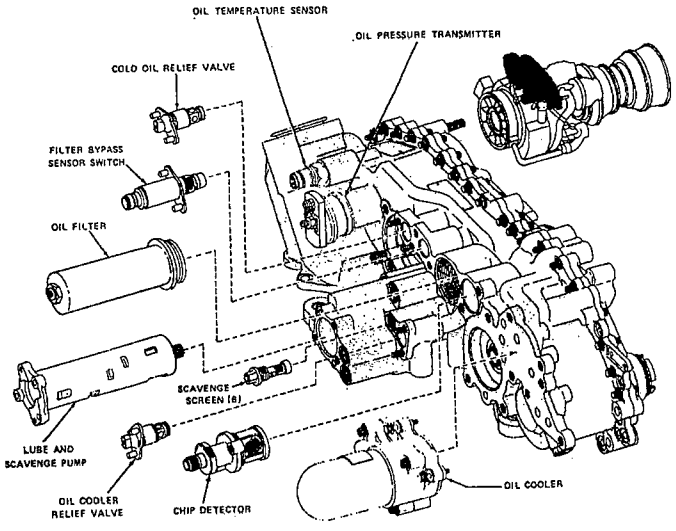


Figure 3. Illustration of the T700 auxiliary gearbox manifold showing location of chip detector port where LASERNET is connected.

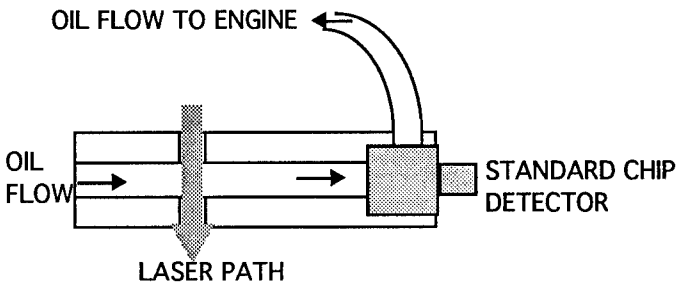


Figure 4. Schematic diagram of a portion of the LASERNET oil flow adapter showing the location of the standard chip detector after the laser viewing port for performance evaluation.

A portion of the LASERNET flow system is shown in Fig. 4. The viewing port is a rectangular section with flat windows for illumination and imaging. Typical dimensions of the viewing chamber are 0.5 - 1 cm in all three dimensions with specific values chosen to optimize debris detection and analysis performance levels, while minimizing illumination depth and required optical beam area. The standard T700 chip detector is located just after the viewing area for verification of LASERNET performance.

The image processing is broken into four stages shown schematically in Fig. 5. The raw image data is stored in a frame buffer and processed in a field programmable gate array to identify addresses of bright-dark transitions. These pixel addresses are transferred to a dec α digital processor for identification of individual objects. The objects are then subjected to a series of tests to determine shape. Roundness is first tested on the basis of variation of the length of normalized radial spokes. Objects that are determined to be round are identified as bubbles and discarded. Overlapping objects and objects at the edge are handled separately. Objects at the edge are tested for roundness of the partial circle contained in the image. Overlapping bubbles are identified by testing for the roundness of the partial circles of two potential objects lying on a line joining the centers. If the two objects are identified as partial circles, the entire objects is identified as overlapping bubbles and discarded. Any remaining particles are identified as debris particles and are analyzed for shape. The shape analyzer extracts a feature vector as described later and the particles are then classified according to wear source.

The auxiliary high speed memory is operated as a continuously streaming FIFO which can be interrupted by a response at the chip detector. In this way, image frames associated with debris detected by the chip detector can be retained for visual inspection and for evaluation of the real time processor performance. This comparison will establish the relative performance of the LASERNET system and the chip detector, allowing comparison not only of debris detection but also of false alarms by both systems. A second high speed memory system is contained within the real time image processor and retains images of frames in which debris particles are detected for visual inspection. Long term behavior can be compared to debris collected in the engine filter.

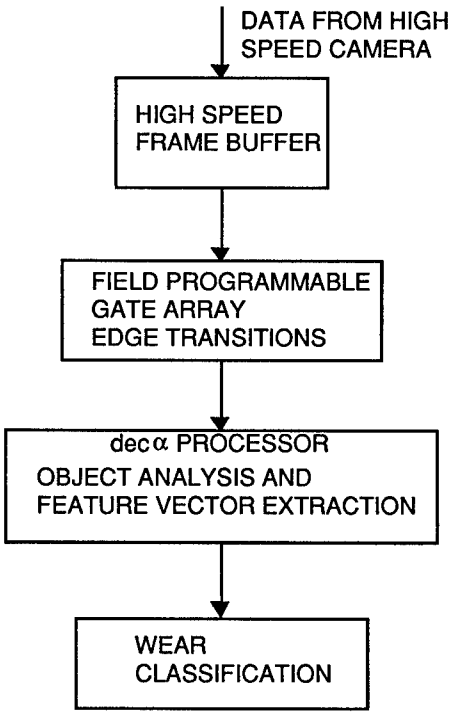


Figure 5. Block diagram of image processor for LASERNET

Long term tests on the T700, which is expected to be a relatively clean system, are planned to test the false alarm performance of the LASERNET system. Provisions are also being made to attach the LASERNET system to the SH-60 main gearbox, into which seeded faults will be introduced. This system will verify the ability of LASERNET to detect and classify debris particles.

LASERNET FINES: The fine particle detector, LASERNET FINES, is designed to detect and classify particles in the size range above 5 μm. It is appropriate for detection of debris particles associated with normal and early wear in lubricating systems. It is also capable of detecting particulate contamination in hydraulic and fuel systems, and in other fluids. Again the goal is to identify type and severity of fault through determination of particle size distribution and morphology. However, since it is sufficient to determine the representative concentration of particles in the size range of LASERNET FINES, rather than counting each and every particle, sampling in time and space is acceptable. This allows a reduction in the amount of fluid sampled and slower data detection and processing rates that are compatible with standard tv formats, reducing the special demands on the imaging system. The depth-of-field requirements for detection of the smaller particles requires use of cells that are of the order of 100-200 μm thick. The increased spatial resolution required for detection of the smaller particles is obtained with

magnification of the image of the order of 10x. With the particle concentrations that are typically encountered, particle distributions from different frames must be combined to obtain statistically significant results. Total detection time and image processing rates can be traded within an overall system. The LASERNET fines system can be configured as a batch processor, a full autonomous in line processor or a hybrid in which the sampler is located in line, but the fluid sample is ported to the unit.

We have currently configured the LASERNET Fines unit as a benchtop batch processor. In the present unit the illuminator is a commercial laser diode operating at 830 nm with 100 mW peak power. The fluid is gravity fed through a cell approximately 200 μm thick. In our current configuration the image is obtained with a Cohu full field tv camera using 4x magnification. The images are captured and processed on a Power Macintosh 8100 computer.

The image processing code scans the frame for objects. It determines the largest dimension of each object and then develops a record of particle number for various size ranges. Once finished with a frame, the program acquires another frame and repeats the analysis routine, after which it combines the results with the results of the previous frame. The process continues, providing a cumulative record of particle numbers in various size ranges, which can be chosen to coincide with the current ISO size ranges. The program can be set to accumulate data for a set number of frames, a maximum number of particles in a given size range or a set counting time.

The image processing code for the LASERNET FINES is different from that for LASERNET in several respects. First, round objects cannot be discarded from consideration because there are wear sources that can produce round debris particles in this size range, such as particles from fatigue cracks. Secondly, more than one particle is expected in each frame. As a result, the program does not test objects for roundness and discard round objects, but rather counts and analyzes all objects in the frame. However, because of the sample handling methods, it is anticipated that air bubbles will be much less of a problem. In the event that they remain, side illumination can be used to give the bubbles a characteristic highlight as discussed in previous papers.

TABLE I.

PARTICLE DISTRIBUTIONS FOR CALIBRATION SAMPLE OF MIL-H-5606 HYDRAULIC FLUID

PARTICLE SIZE RANGE (μm)	% PARTICLES Pratt&Whitney	% PARTICLES LASERNET FINES
5-15	87.0	80.2 ± 3
15-25	10.0	15.9 ± 2
25-50	2.7	3.5 ± .95
50-100	.29	.25 ± .27
> 100	.01	.08 ± .14

The LASERNET FINES system has been exercised with samples of MIL-H-5606 hydraulic fluid provided by Pratt&Whitney. The particle distribution in different size ranges for a calibration sample containing A/C Fine Test Dust is shown in Table I, along with the distribution determined by Patt&Whitney with a commercial particle counter. The results for LASERNET

FINES should not be expected to reproduce the results of the commercial counter since the commercial unit determines spherical diameter, while the LASERNET system determines the largest dimension of the particles. Non-spherical particles, which our images indicate are quite prevalent in the calibration sample, will thus be classed differently by the two instruments. Nevertheless, the size distributions can be seen to be quite similar.

PARTICLE FEATURE VECTORS: One of the main goals of this program is to obtain a classification of the particles according to wear source based on an analysis of particle morphology. Several authors have discussed this approach using various characteristics of the particle shape, including aspect ratio, kurtosis of the distribution of angles in the edge contour and variations of curvature in the edge contour.[3-6] We have analyzed images of particles obtained from the Wear Particle Atlas from the University of Swansea, and from bearing test stands. Some of the characteristics that we have investigated are:

aspect ratio, defined as particle area/ (maximum dimension)^2
external compactness, defined as average diameter/(diameter of maximum circumscribed circle)
bending energy, defined as the normalized second moment of curvature

An example of the separation of particles into various wear classes using these shape characteristics is shown in Figure 6 for a limited data base. Further investigation of particle classification are underway for a more extensive data base.

*Jaycor, Vienna, VA
†Dept. of Physics, US Naval Academy, Annapolis, MD
**Naval Research Laboratory, Code 5362, Washington, DC 20375
***P. L. Howard Enterprises, 1212 Clearbrook Rd., West Chester, PA
††LNK Corp., 6811 Kenilworth Ave., Suite 306, Riverdale, MD 20737

References

1. J. Reintjes, R. Mahon, M. D. Duncan, L. L. Tankersley, A. Schultz, V. C. Chen, D. J. Kover, P. L. Howard, M. Chamberlain, Srini Raghavan, and Naresh Gupta, "Optical Debris Monitoring", JOAP Annual Meeting, Pensacola FLA, November 1994
2. J. Reintjes, R. Mahon, M. D. Duncan, L. L. Tankersley, A. Schultz, V. C. Chen, D. J. Kover, P. L. Howard, M. Chamberlain, Srini Raghavan, and Naresh Gupta "Optical Oil Debris Monitor", in "Life Extension of Aging Machinery and Structures", H. C. Pusey and S. Pusey, eds. pp. 57-66, 1994
3. A. Albidewi., A. R. Luxmore, B. J. Roylance, and G. Wang, "Determination of Particle Shape by Image Analysis-the Basis for Developing an Expert System," in "Condition Monitoring '91," M. H. Jones, J. Guttenberger and H. Brenneke, eds., Pineridge Press, Swansea, UK, 1991, p. 411
4. B. J. Roylance and S. Raadnui, "The morphological attributes of wear particles - their role in identifying wear mechanisms", Wear **175**, 115 (1994).
5. B. J. Roylance, I. A. Albidewi, M. S. Laghari, A. R. Luxmore and F. Deravi, "Computer-Aided Vision Engineering (CAVE) - Quantification of Wear Particle Morphology", Lubr. Eng. **50**, 111 (1993)
6. J. J. Hamalainen and P. Enwald " Inspection of wear particles in oils by using a fuzzy classifier", SPIE vol 2249 "Automated 3D and 2D Vision", 390 (1994).

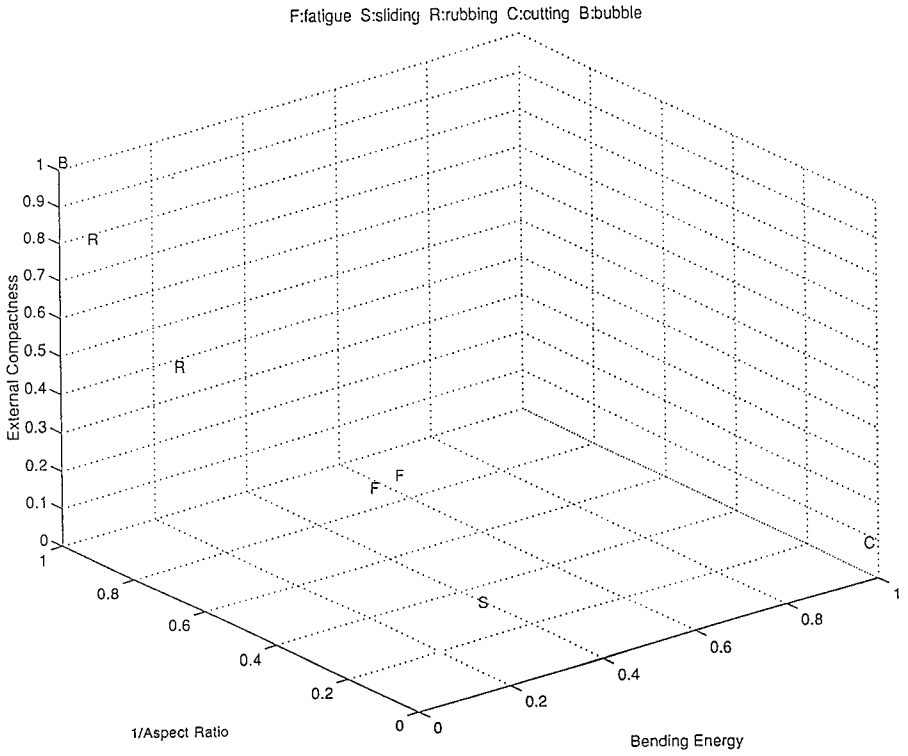


Figure 6. Classification of particles according to wear type using external compactness, aspect ratio and bending energy as shape features.

A NON-REDUNDANT SENSOR VALIDATION SCHEME FOR TRANSIENT AND STEADY-STATE CONDITION MONITORING

Michael J. Roemer

Stress Technology Incorporated
1800 Brighton-Henrietta Town Line Rd.
Rochester, New York 14623
Phone: (716) 424-2010
Fax: (716) 272-7201

Abstract: This paper presents a sensor validation scheme capable of detecting failed sensor hardware without sensor redundancy and during non-steady state monitoring conditions. The technical approach utilizes neural networks and fuzzy logic to accomplish the desired goal. Neural networks are used to recognize the non-linear, inter-relationships between the different types of sensors used in a transient or steady-state measurement environment. Fuzzy logic is used to pre- and post-process the measurement data in order to determine general characteristics about the state of the process being monitored. Different types of neural network architectures were developed and tested to determine their suitability to solving this problem. The feasibility of the method was proven through computer simulation utilizing gas turbine engine data as input to the validation system.

Key Words: Sensor Validation, Condition Monitoring, Neural Networks, Fuzzy Logic, Artificial Intelligence, Transient Conditions, Non-Redundant Sensors, Health Monitoring.

Introduction: Integrated sensor systems play a major role in the rapidly expanding area of on-line diagnostics and condition monitoring of all types of industrial, commercial and military equipment. After all, without accurate and reliable information on the equipment being monitored, it is impossible to diagnose the machines current condition or "health" in order to make informed maintenance and safety decisions.

Numerous sensor validation and recovery systems have been developed and tested over the years to separate failed sensor hardware from "real" equipment malfunctions [1-6]. In particular since 1980, when a ground test of the Space Shuttle main engine experienced erroneous combustion chamber pressure measurements that were used in the closed-loop thrust level control algorithm [7]. In this case, the failed pressure sensor led to running the engine in a severely abnormal operating condition, and nearly self-destroyed the engine. Following that incident, sensor validation and recovery research has focused primarily on utilizing sensor redundancy and knowledge-based systems that operate well under steady-state conditions [8-11]. Today, with the increased application of neural networks to solve non-

linear, pattern recognition problems, non-redundant equipment sensor patterns can be "learned" by dedicated neural networks to detect and isolate sensor failures. The additional advantage of utilizing neural networks is reducing the dependency on redundant sensors and steady-state operating conditions.

Algorithms designed to perform non-redundant sensor validation in transient monitoring environments are based on the principal that the various sensors used in a particular application are non-linearly related over a particular speed/power range. For example, if a machine is increasing in speed and power, then the temperatures, pressures, etc. are related to each other in some non-linear sense. Neural networks would appear to be a good tool for solving this type of problem because of their non-linear, pattern recognition and classification qualities. In addition, fuzzy logic was chosen as the best tool for deciding if the speed/power range of the equipment is increasing, decreasing, or in a steady-state condition and whether the sensor confidence level output of the neural networks depicts a good, bad, or marginal sensor condition.

Neural Networks for Non-Linear Pattern Recognition: Neural networks are systems of elemental processing units connected in a way analogous to how neurons are connected in the brain. Like the brain, neural networks exhibit learning and associative memory skills. A neural network is trained to perform a task by showing it examples of an input it will receive, paired with the output it is to deliver. The network learns the associations between these pairs of input examples and corresponding outcomes, and is able not only to reproduce these associations, but also to generalize these relationships for inputs that it has not encountered before. Neural nets are therefore capable of intelligent interpolation and therefore make them particularly well suited for this type of application.

The artificial neural network can be viewed as a collection of elemental processing units massively interconnected among themselves. Some of the processing units, sometimes called nodes, communicate with the outside environment. We distinguish between the different types of processing units with the following nomenclature:

1.) Input Nodes: Receive signal from the environment

2.) Output Nodes: Send signals to the environment

3.) Hidden Nodes: No direct contact with the environment

The processing unit or node is the component within neural networks where the computations are carried out. The input signal come from either the environment or other processing units, and form an input vector containing all the inputs. Figure 1 is an illustration of one processing unit or node.

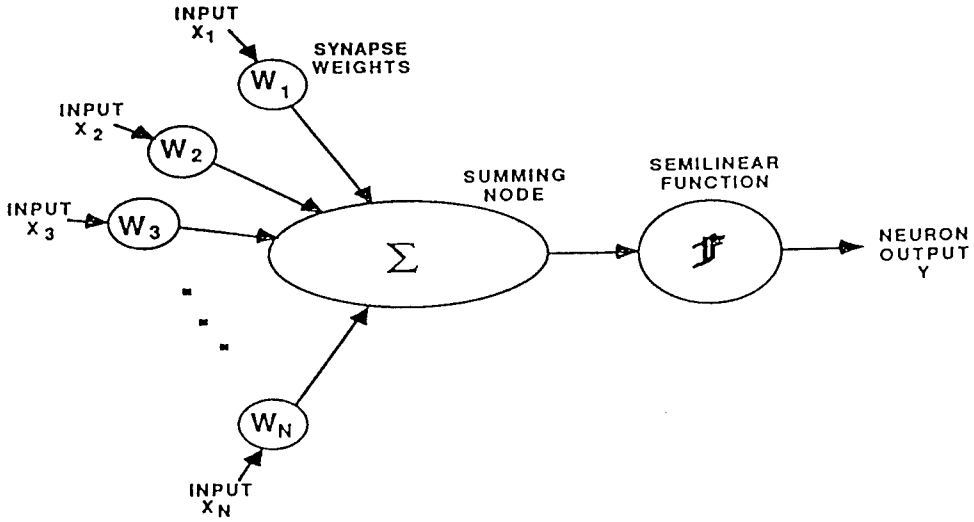


Figure 1 Neural Network Processing Unit

Figure 1 also shows weights corresponding to each input. These weights are used to compute the output value of the processing unit. This computation is performed by taking the product of each input value x_i and its corresponding weight w_i. These products are then summed together and "passed" through a sigmoidal activation function to determine its final output activation level. Other types of activation function can be used, but the sigmoid is the most commonly used function.

When we talk about neural network's abilities to learn cause and effect relationships, we are really discussing the supervised learning procedure. Supervised learning involves the task of teaching the network associated input/output pairs. The network is presented with data that shows what response (output) should be generated by a given stimulus (input). The network then self adjusts its internal parameters in order to represent this underlying relationship between the inputs and outputs. This is the basis for a neural network's ability to generate appropriate outputs for all other inputs of a similar category, even if these inputs have never been previously encountered.

Supervised learning is commonly implemented using a Generalized Delta Rule network architecture with backpropagation of error. During this procedure, the network architecture is specified in terms of the number of input and output nodes, as well as hidden layer nodes. The training set is then used to specify what target outputs should result from an input pattern, and the network automatically learns the set of parameters (weights and thresholds) that will generate this desired output. In this learning procedure, the network learns a single set of network parameters that satisfies all the training input/output pairs. The learning is not perfect, but is optimum on the basis of the least mean square error. In the consulting mode, the network is able to generalize and generate appropriate output patterns for any input pattern

applied to the network. This attribute is the principal advantage to utilizing neural networks in condition monitoring applications.

Fuzzy Logic for Approximate Reasoning: Fuzzy logic is a programming tool that is capable of incorporating imprecise or ambiguous information into algorithmic expressions. However, contrary to its name "fuzzy", the mathematics involved are based on precise and rigorous calculations with respect to fuzzy sets or membership functions. The four basic processes required to develop fuzzy logic systems are fuzzification, rulebase development, inference, and defuzzification. The fuzzification process begins with the development of membership functions which relate linguistic variables like "cool", "hot", and "cold" to particular numerical ranges used in the "fuzzy" calculations. For instance, "cool" might have a membership value of 1.0 (the highest degree of membership) for 60 degrees F, a membership of 0.6 for 50 degrees F, a membership of 0.25 for 40 degrees F, and a membership of 0.0 (no degree of membership) for 30 degrees F.

The rulebase development is typical of any if/then rule set implemented in standard expert systems, except the rules now incorporate the "fuzzy" linguistic variables that have membership functions associated with them. An example of a rule would be; If *temperature* is cool, Then *velocity* is medium. In this rule, *temperature* is the *input variable* and *velocity* is the *output variable*, both or which have membership functions associated with them that include cool and medium respectively.

The strategies for "inferring" conclusions/decisions from cause-and-effect relationships provided by the rulebase and membership functions (knowledge base) are often called fuzzy inference techniques. Some inference techniques include; Product-Sum, Max-Min, and Min-Sum. The first expression of the inference technique name refers to the method for scaling the membership function variables. The second expression refers to the technique for combining the scaled membership function variables. A more complete description of the inference techniques is given in Reference [12]. The final process of calculating a single value from the scaling and combining of the variables described in the membership functions is called defuzzification. Techniques such as Centroid, Max-height, and Max-moment are used to determine the value that best represents the outcome of the fuzzy rule evaluations.

Sensor Validation System Architecture: The sensor validation system architecture involves the integration of the neural networks, fuzzy logic, and miscellaneous arithmetic and logic operations. A block diagram of the basic system architecture is given in Figure 2. The speed/power sensor data is first accepted by two parallel fuzzy logic modules. The first module determines the state of the speed/power condition (i.e increasing, decreasing, or steady-state) and the second verifies the validity of speed/power sensor itself. The output of the speed/power condition module triggers a particular neural network module that was specifically trained to know the sensor relationships for either increasing, decreasing or steady power output. Only one neural network module is triggered at a time, depending on the outcome of the prior fuzzy logic decisions. The sensor confidence values predicted by the

neural networks are trended over time and passed through another fuzzy logic module to interpret the results. These extra steps are used to ensure that false alarms do not occur.

For the gas turbine engine application discussed in this paper, there are four primary performance related sensors that are monitored during turbine operation. These sensors include; fuel flow (Wf), HP compressor delivery pressure (P3), LP compressor delivery temperature, and Jet Pipe temperature (T6). The outputs of the neural networks yield a confidence factor associated with the probability of a failed sensor. A confidence factor near one represents proper sensor operation, while a confidence factor near zero indicates a faulty sensor mode. A fuzzy logic module is used at the output of these neural networks to decide whether the sensor is good, bad, or somewhere in between. For instance, if a hard decision was utilized to alert the crew when a sensor confidence factor reached a level less than 0.80, false alarms would likely occur even though a sensor confidence factor of 0.78 might still indicate a properly working sensor.

The same reasoning applies for using fuzzy logic as a pre-processor for the neural network in terms of determining the speed direction. When examining the difference in speed change based on several different time differences, speed changes approximately near zero would indicate a steady-state operating condition. By implementing fuzzy logic, the "approximately near zero" term can be accounted for in algorithmic expressions.

Neural Network Architecture and Training: Two different neural network architectures were examined for this application. Both networks utilized a multi-layered, feed-forward architecture with five input nodes and four output nodes. The first network contained one hidden layer with 13 nodes and the second used 2 hidden layers with 10 and 5 nodes respectively. Figure 3 is an illustration of the neural network with only one hidden layer and 13 nodes.

Determining the "optimal" number of hidden layers and nodes for each network is a non-trivial task and depends on many factors, some of which include; number of input/output nodes, quantity and accuracy of training data, complexity of problem, and resulting network generalization performance. The "standard" feed-forward architectures used for this problem were picked due to the large quantity of training data available and the resulting network generalization performance required.

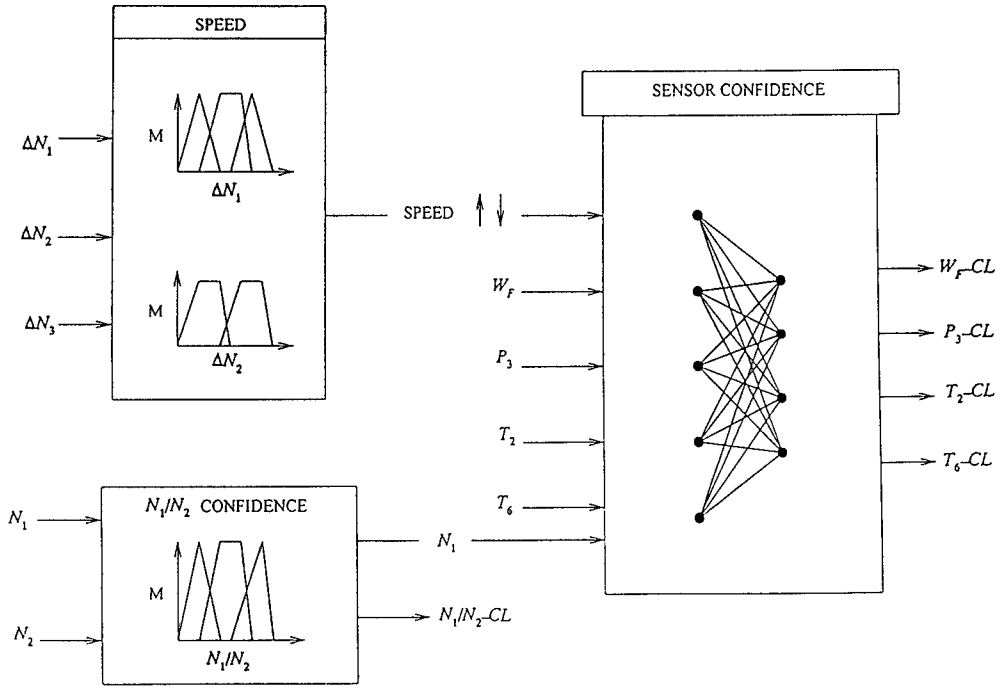


Figure 2 Sensor Validation System Architecture

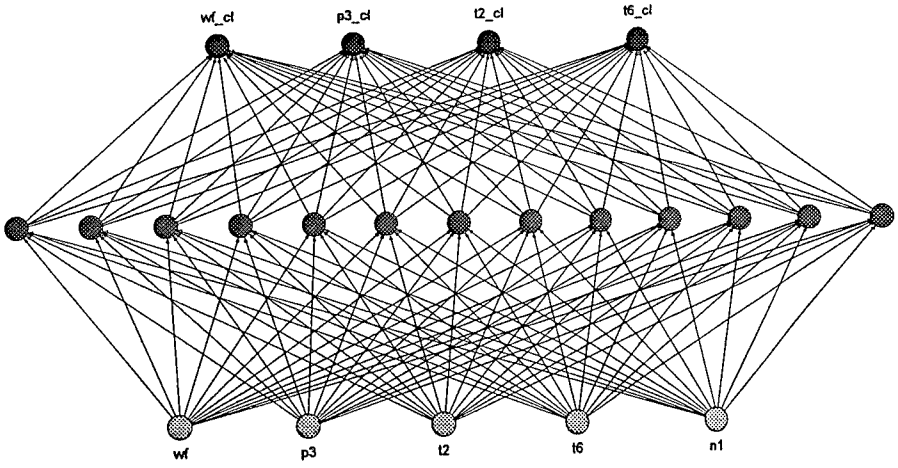


Figure 3 Neural Network Architecture

The simulated gas turbine engine data used to train the neural network architectures is given in Figure 4. The data set represents sensed fuel flow, pressure, and temperature readings during a start-up condition. Simultaneously measured data staying within the illustrated confidence limits for each sensor would represent properly operating sensors. Data going outside these limits would indicate a failure mode associated with the particular sensor. For training purposes, any measurement within the confidence limits of each sensor for a particular engine speed would indicate a sensor confidence level of 1.0 (highest confidence). As a sensor measurement moves outside the confidence limits, the neural network output confidence level decreases from 1.0 towards 0.0 indicating the graduating sensor failure mode. Each network architecture was subjected to the same training data set consisting of 300 input/output pairs.

Training the sensor validation networks was accomplished with a supervised learning procedure. Each of the 300 training pairs or patterns used during the training process consisted of 5 sensor input signals and it's corresponding set of 4 outputs sensor confidence factors. The input and output training data was normalized to values between 0 and 1. An error-back-propagation algorithm was used to minimize the mean-squared error between the actual network output and the target values set by the training set. Training parameters such as the learning rate, gain of the activation function, and momentum coefficient were adapted during the training session to aid in minimizing the error. A final RMS error associated with all training pairs was reduced to 0.199.

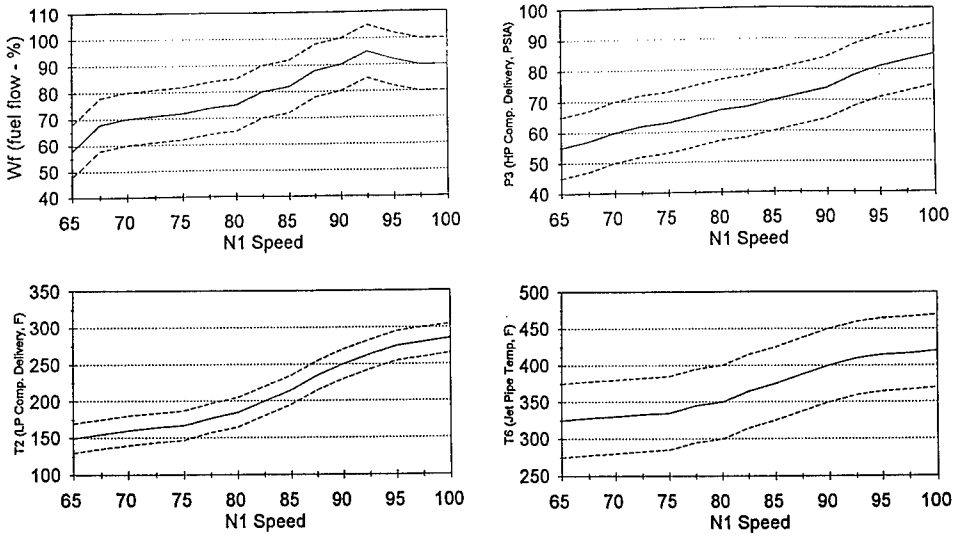


Figure 4 Gas Turbine Measurement Data

Neural Network Results: A computer generated data file simulating normal and faulty sensor measurements was developed to test the accuracy of the two neural network architectures. Figure 5 is an illustration of a small section of that file including a range of fuel flow measurement data. The data represented by the + signs are all within the confidence limits of normal operating sensor patterns. In this case, sensor confidence levels predicted by the neural network should all be close to one. The data indicated by X's and O's are outside the confidence limits and therefore indicate worsening sensor operation. The X's are just outside the confidence limits and should predict sensor confidences between zero and one. The O's are significantly outside the sensor confidence band and should predict sensor confidence levels close to zero. The results from this data file are given in Table 1 below. Note, the other sensor measurements including temperatures, pressure, and speed were all within the confidence limits shown in Figure 4.

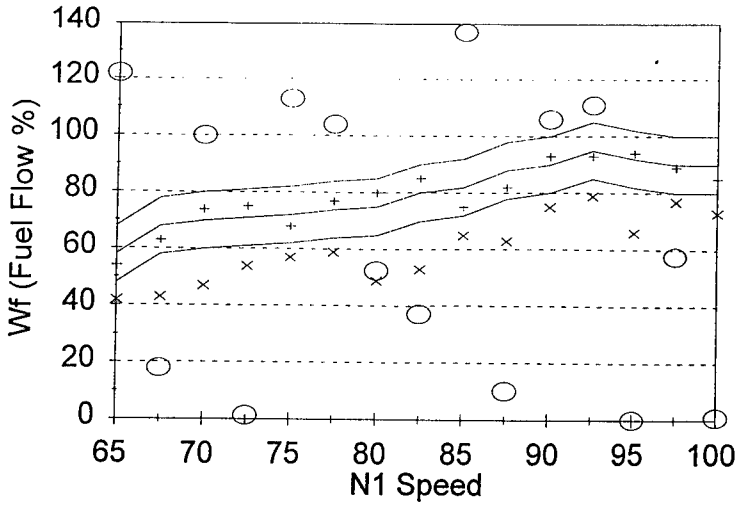


Figure 5 Simulated Fuel Flow Test Data

Table 1 Neural Network Results

Case #	"+"	"x"	"o"
1	0.9197	0.2711	0.0755
2	0.9990	0.1674	0.0543
3	0.9990	0.2625	0.0923
4	0.9861	0.7777	0.0500
5	0.9990	0.7700	0.0923
6	0.9990	0.7852	0.1061
7	0.9756	0.2378	0.4670
8	0.9112	0.1263	0.0380
9	0.9734	0.8133	0.1389
10	0.9723	0.3508	0.1736
11	0.8831	0.9244	0.4835
12	0.9259	0.9549	0.4616
13	0.8556	0.6185	0.2459
14	0.9345	0.9057	0.0980
15	0.9057	0.7111	0.1866

Note: The network output confidence levels for the other sensors were all above 0.975 for "+" test cases, above 0.927 for the "x" test cases, and above 0.946 for the "o" test cases.

Several test cases similar to the results described in Table 1 were conducted for the other sensor measurements, all yielding similar results. Although the trained network yielded good results in terms of accuracy and generalization capabilities, overtraining was a concern that was monitored carefully. Initially, 600 training patterns were used to train the network with output error similar to the 300 training pattern case. The resultant trained network had much worse generalizing capabilities than the network trained with only 300 patterns.

The network architecture with two hidden layers was trained and tested with the same data as used for the previous single hidden layer network. In this case, the output RMS error was only reduced to 0.289 and the network generalization capability degraded. The worsened generalization capability can be explained by the additional degrees of freedom that were introduced by the additional nodes (neurons) in the hidden layers. The higher network RMS error is most likely due to finding a "local" minimum associated with the gradient decent BPE algorithm. In theory, the error should have been reduced to at least the level of the previous single hidden layer network.

Conclusions: A sensor validation scheme was presented for checking sensor hardware operation without the need for sensor redundancy and during transient equipment monitoring conditions. Fuzzy logic and neural networks were applied to meet the non-linear and non-exact nature of this problem. Fuzzy logic modules were used to determine the transient operating condition of the machine and to interpret the neural network sensor validation network outputs in terms of normal, faulty or marginal sensor operation. Neural networks were developed and trained for increasing, decreasing, and steady-state operating conditions. These trained networks accepted the non-redundant sensory data and interpreted the non-linear relationships between them in order to recognize when a sensor reading did not match the patterns used in the training process. The output of the networks included the sensor confidence factors which ranged from zero (faulty sensor) to one (normal sensor). An output fuzzy logic module interpreted the sensor confidence values to determine when the equipment operator should be notified of a sensor hardware problem. Trending the sensor confidence factors over time is an additional step used to ensure accurate diagnosis of failed sensor hardware.

Results of the sensor validation scheme when subjected to computer simulated data representing gas turbine engine sensor hardware measurements was encouraging. The neural network never predicted sensor failures when measurements stayed within the trained sensor confidence bands. Also, sensor confidence levels were always predicted to be less than 0.5 if a sensor measurement drifted more than 50% outside the sensor confidence bands.

References

1. Walker, N. D. and Wyatt-Mair, G. F., "Sensor Signal Validation Using Analytical Redundancy for an Aluminum Cold Rolling Mill", **Control Engineering Practice**, Vol. 3, No. 6, June 1995, pp.753-760.

2. McAvoy, T. J., "Sensor Data Analysis Using Autoassociative Neural Nets", **World Congress on Neural Networks**, Vol. 1, 1994, pp.161-166.

3. Holbert, K. E., Heger, A. S., and Alang-Rashid, N. K., "Redundant Sensor Validation by using Fuzzy Logic", **Nuclear Science and Engineering**, Vol. 118, No. 1, Sept. 1994, pp.54-64.

4. Harrison, P. R., and Harrison, P. A., "Validating an Embedded Intelligent Sensor Control System", **IEEE Expert**, Vol. 9, No. 3, June 1994, pp. 49-53.

5. Ahmadi-Echendu, J. E., and Hengjun, Zhu, "Detecting Changes in the Condition of Process Instruments", **IEEE Transactions on Instrumentation and Measurement**, Vol. 43, No. 2, April 1994, pp.355-358.

6. Lee, S. C., "Sensor Value Validation Based on Systematic Exploration of the Sensor Redundancy for Fault Diagnosis", **IEEE Transactions on Systems, Man and Cybernetics**, Vol. 24, No. 4, April 1994, pp. 594-605.

7. Aguilar-Crespo, J. A., de Pablo, E., and Alaman, X., "A Fuzzy Logic Approach for Sensor Validation in Real-Time Expert Systems", **Advanced Methods in Artificial Intelligence**, 1993, pp.330-337.

8. Uhrig, R. E., "Use of Neural Networks in the Analysis of Complex Systems", **Third Workshop on Neural Networks - WNN92**, 1993, pp.16-24.

9. Upadhyaya, B. R., Eryurek, E., and Mathai, G., "Neural Networks for Sensor Validation and Plant Monitoring", **Proceedings of the 1990 International Fast Reactor Safety Meeting**, Vol. 3, 1990, pp. 349-358.

10. Lee, S. C., Ratliff, M., Peterson, C., and Lollar, L. F., "AMPERES: A Real-Time Fault Monitoring and Diagnosis Knowledge-Based System for Space Power Systems", **Applied Artificial Intelligence**, Vol. 3, No. 5, July 1991, pp. 281-308.

11. Boger, Z., "Applications of Neural Network Techniques for Fault Detection", **Nuclear Societies of Israel Transactions**, Vol. 16, 1990, pp. 217-222.

12. Ray, A., and Luck, R., "An introduction to sensor signal validation in redundant measurement systems", **IEEE Control Systems Magazine**, Vol. 11, No. 2, Feb. 1991, pp. 44-49.

MACHINERY DIAGNOSTICS & PROGNOSTICS #3

A MODEL-BASED FAILURE DETECTION, ISOLATION AND RECOVERY SYSTEM

Amit Misra and Janos Sztipanovits
Department of Electrical and Computer Engineering
Vanderbilt University
Box 1824, Station B
Nashville, TN 37235
email : misra@vuse.vanderbilt.edu

Abstract: The Failure Detection, Isolation and Recovery (FDIR) in the International Space Station Alpha (ISSA) requires timely monitoring and diagnosis of failures so that recovery actions can be employed to safeguard the mission and the life of crew. Using traditional methods for representation of domain knowledge and for diagnosis proves to be ineffectual because of the scale, complexity and dynamics of ISSA. Model-based approach for representing systems and for diagnosis is an attractive and feasible solution. We have developed and field tested a model-based real-time robust monitoring and diagnostic system for ISSA and other aerospace systems. The system is represented using hierarchical and multiple-aspect models, which include representation of functional structure as well as the physical component assemblies. A discretized model of the failures and their effects is represented using timed failure propagation graphs. The monitoring mechanism is modeled by using a discretized sensor space, with mechanisms for sensor validation. The diagnostic reasoning applies structural and temporal constraints for the generation and validation of fault hypotheses using the "predictor-corrector" principle. The diagnosis is generated in real-time amid an evolving alarm scenario, and uses progressive deepening control strategy. The robust diagnostic system has been tested and demonstrated using ISSA models obtained from the Boeing Company.

Key Words: Diagnostics; fault models; hierarchical models; model-based systems; multiple aspects; program synthesis; sensor failure

INTRODUCTION: An increasingly competitive aerospace market requires requires computer integrated Failure Detection, Isolation and Recovery (FDIR) systems to perform complex and sophisticated analyses that are capable of providing real-time embedded vehicle health management. Simultaneously, a general trend in the evolution of complex, large-scale, computer integrated systems is the rapidly increasing use of design-time models in system operation. The goal is to synthesize vehicle health management systems that are formally and automatically derived from the integrated model sets developed during the vehicle design, development, build, test, and verification.

In this paper we describe a model-integrated approach to synthesis of FDIR system for aerospace vehicles. The model-integrated approach is based upon the MultiGraph Architec-

ture (MGA) [15]. The primary specifications for FDIR come from the International Space Station Alpha (ISSA) program requirements, though FDIR tools and formalisms described here could be applied (with appropriate modifications) to task of health management for most large-scale, complex, computer integrated system. The FDIR system consists of modeling formalisms and a health management system, synthesized from the models of the artifact.

FDIR MODELING PARADIGM: It is evident that the practical use of a model-integrated system is limited by the "goodness" of the models themselves, which in turn is influenced by the formalism used for modeling. Thus, one must develop modeling formalisms that capture the essence of the system being modeled *and* the FDIR requirements. One must also recognize the fact that there is a critical need for software technology which makes high-performance computing and communication capabilities accessible for end-users. Systems engineers need domain-specific modeling and analysis environments that support the building and verification of vehicle fault models, provide interfaces to engineering databases and systems engineering tools, and allow the synthesis of FDIR systems that are consistent with the vehicle models. Further, there typically are some mature engineering disciplines underlying the design and systems engineering. Thus, the modeling paradigms are not "negotiable": systems engineers need to be supported by rich, domain specific concepts, relations, and composition principles routinely used in the field.

The main challenges in using a model-based approach for the FDIR in large-scale heterogeneous dynamic systems are the following :

1. The size of the systems in terms of the number of components, the complexity of physical processes and their interactions can be large. In providing models for system-wide diagnosis, *scaleability* of the modeling technique becomes a major issue.

2. Design of such systems involves different engineering disciplines with different focus and tools. In the FDIR modeling four such disciplines are identified – signal, fluid, electrical, mechanical.

3. The source of failures may be outside of the system boundary. Propagating effects of *external disturbances* must be traced.

4. The primary goal of diagnosis in critical systems is the prevention of the occurrence of critical failures. Prediction of the propagation of discrepancies requires not only the spatial but also the *temporal isolation of fault events.* For this purpose, steady-state models are often useless, because processes may only slowly converge to steady-state and because steady-state models do not capture the dynamics of fault propagation.

5. Fault diagnosis is based on the observation of the behavior of the plant during a fault incidence. Consequently, the models to be used in fault diagnosis should capture the dynamic behavior of processes when it is out of the normal operation range. Needless to say, *modeling uncertainties* in these regions are even more significant than in the normal operation range.

6. The FDIR system must be able to reason about *abrupt* faults and input disturbances, which means that assumptions about the system (valid only during normal operation) become invalid and unusable.

7. Faults propagate through the system. That is, the effects of a fault rarely tend to be localized unless specific measures are taken. The goal of FDIR is to contain and rectify the faults locally. Thus the propagating effects of faults must be modeled.

The first two challenges listed above address the issues common to all complex, heterogeneous, large-scale system, more or less independently from the application. In other words, these issues are not specific to FDIR, but arise in control, simulation and many other applications, and relate to the formalisms used for overall organization of *system models.* The rest of the challenges listed above arise out the FDIR task specifically. These issues are addressed in the formalisms used for *fault models*, which are a subset of system models. We will first give an overview of the organization of system models, followed by a description of the fault models.

Hierarchical, Multiple Aspect, Discipline Oriented System Models: Because our goal is to model engineering systems, the modeling technique should utilize the well known engineering techniques to manage complexity.

One of the primary model structuring method is focusing on selected types of interactions; i.e. to model a system from *multiple aspects.* Different modeling aspects use different concepts (e.g. the physical structure is defined in terms of assemblies and sub-assemblies, while the functional structure of a temperature control system is defined in terms of material and energy flows). Each aspect may simultaneously be sub-divided into *views*, that contain discipline oriented information. The models are typically organized into *decomposition hierarchies* controlling the level of details shown. On each level, the system is modeled as an aggregate of connected sub-systems. The type of the connections are determined by the modeling aspect and view. The subsystems are connected through an *interface*, which defines their boundaries and separates the internal and external environments. The decomposition hierarchies and the connected set of subsystems on each level constitute the *structural model* of the system. Each subsystem can also be characterized in terms of the relationships among its input/output quantities. These models are called *behavioral models.*

For purposes of FDIR, the system models are broken down into two primary hierarchies – the physical assembly, which models the components in the system, and the functional decomposition. The physical models and the functional models are both described in terms of their structure and behavior. Separation of the functional and physical structure is an essential difference from the primarily component-oriented modeling in model-based diagnostic systems (e.g. [1, 2, 3]). Our rationale for this separation is the following :

1. There are many examples for multi-function components that are involved in the implementation of several functionalities in the same time. Well known examples are computers and energy distribution systems.

Table 1: Physical Model Aspects

Aspect	Concept(s) Modeled	Model Elements
Assembly Aspect	Component assemblies and energy and material flows.	Sub-component and input and output flows.
State Transitions Aspect	Operation states and State Transition Machine	Component states, sub-states, state transitions, local, input and output events.
Alarm Generation Aspect	Sensors	Alarms that the sensor generates, and sensor attributes like cost and time to use, probability of false alarm, etc.
Component Faults Aspect	Faulty behavior	Failure modes and failure rates of components.

2. Assignments among physical components and functionalities are not always static. Physical redundancy and the use of multiple-function components are frequently used to achieve fault tolerance in critical systems.

Both the physical and functional models have many different aspects. In this paper we present only a very brief description of the different modeling aspects of physical and functional models, given in Table 1 and Table 2. For a more detailed description, the reader is referred to [14].

Fault Modeling: Model-based diagnostic systems work with a model (a suitable representation) of the system. The level of detail in the models can be kept at the level required by the FDIR requirements. These diagnostic systems interpret the observed discrepancies in the context of the system model. There are primarily two kinds of models that have traditionally been used for diagnosis – *functional models* and *fault models.*

Functional models (also called behavioral models) describe the "correct" behavior of the system, i.e., how the system is supposed to behave when no faults are present. The level of abstraction in the functional models can vary from system to system, depending on the application – from quantitative models (also called *analytical* models) using state-space representation to qualitative models.

Quantitative functional models (analytical models) use Ordinary Differential Equations (ODE), state-space or similar representations (examples of such systems can be found in [11, 12, 13], among others), to diagnose anomalies. However, the usefulness of analytical models is limited to small, stable sub-systems only, which have a well defined and simple domain theory, as opposed to the large-scale, complex systems addressed in this paper.

Table 2: Functional Model Aspects

Aspect	Concept(s) Modeled	Model Elements
Structure Aspect	Functional structure and energy and material flows	Sub-functionalities and input and output in the four disciplines (signal, fluid, electrical and mechanical).
State Transitions Aspect	Operation states and State Transition Machine	Functionality states, sub-states, state transitions, local, input and output events.
Failure Propagations Aspects	Failure interactions	Component failure modes, functional discrepancies and timed failure propagations.
Failure Observation Aspect	Fault monitoring	Alarms and sensor states.
Implementation Aspect	Relationship between functionalities and components in the system	The physical components that fulfill the functionality and the redundancy between the components.

To address the complexity in most engineering systems, some researchers have used *qualitative* functional models. Qualitative functional models divide the process variable space into "ranges of interest" and use qualitative physics to generate the behavior of a system. They have met with varying degrees of success in analyzing and predicting the complete and correct behavior. The functionality can be described using just input/output relationships as in [4], using a mathematical description, or using a set of connected components and causal sequences which give a description of how the system behaves [1, 2, 3].

Using functional models to diagnose faults has its own problems, the foremost being the accuracy and validity of models, particularly if faults are present. Further, while the models might be good for identifying the presence of a malfunction (using simulation or analytical methods), they are not necessarily helpful in diagnosing, i.e., locating the faulty component. This is because using functional models can lead to an explosion in the size of the diagnostic search space and hence the number of possible hypotheses, thereby rendering diagnosis intractable. The large diagnostic search spaces arise out of the attempt to reason about *abnormal* behavior of a system using models that describe the behavior under *normal* conditions. Except in limited cases, such attempts have not been successful.

Fault models (fault trees, cause-consequence diagrams, diagnostic dictionaries etc.), as opposed to functional models, describe system behavior when faults are present. These models use qualitative representation of faults, discrepancies and their interactions. This is done by discretizing the failure space of the systems in terms of the failure modes of components,

functional discrepancies, alarms etc. Such fault analysis of systems is standard practice in systems engineering (e.g., FMEA, fault trees, etc.) and has been used to diagnose faults. Since our goal is to develop a modeling environment which is based on the concepts and relations used by systems engineers, a fault model representation is better suited for our purpose.

Fault models help in diagnosis by reducing the diagnostic search space. Hypothesis generation is straight-forward - just consider all the failure modes that could have caused the discrepancies. Diagnosing with a single fault assumption is simple. Diagnosing with multiple faults and/or sensor failure assumption can possibly result in a large number of combinations of faults to be examined. In this case, some reasonable heuristics can be used which are derived from the structure of the system.

Fault models using diagnostic dictionaries (the kind used in [7, 8] etc.), provide a simple mapping form faults to effects. The effects of a fault, once all the propagations have taken place and the system has reached a steady state, are listed. Thus, the temporal relationships between faults and the dynamics are lost, making this representation less attractive for FDIR task.

The temporal relationships and the dynamics are captured in the fault models using directed graphs (as in [5, 6]). In the research described in this paper, we use a similar representation. This is done in the following manner (for a more detailed description, see [14, 9]) :

1. Discretize the physical and functional failure space to model only the plausible fault states, called *failure modes* and *discrepancies*, respectively.
2. Discretize the observation space to correspond to the discretized failure space, specifying the discrete *alarms* and *sensor states*. Describe the observation of failures using alarms and sensor states.
3. Specify component and functional boundaries and the input and output failure interfaces.
4. Describe the interactions between failures in terms of *timed failure propagations*, which capture the dynamics of system behavior when it is out of the normal operation range. The uncertainty in dynamics is expressed by using a *propagation interval*. The failure propagations can describe the interactions of failures within a sub-system or between sub-systems.

The above method of modeling faults and their interactions address the challenges of FDIR task outlined earlier. The use of these models for real-time robust diagnostics during system operation is briefly described in the next section.

REAL-TIME ROBUST DIAGNOSTICS: An embedded robust diagnostic system was developed, which is synthesized from the hierarchical fault models. The goal here was to develop diagnostic software which doesn't have a pre-defined structure, but instead, the

structure of the diagnostic system is derived from the structure of the system, as captured in the models. Thus, the overall diagnostic system consists of many monitoring and diagnostic sub-systems, as shown in Figure 1.

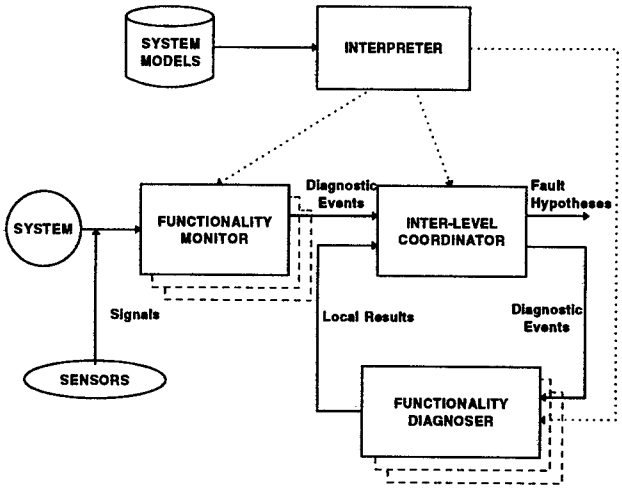


Figure 1: Block Diagram of Robust Diagnostic System

The diagnostic system structure is determined by the functional hiererachy in the models. For each functionality, a monitoring sub-system, the Functionality Monitor (FM), and a diagnostic sub-system, the Functionality Diagnoser (FD), is generated. The interfaces of these sub-systems are determined by the interfaces in the models. An FM receives the sensor signals pertinent to the functionality it represents (as specified in the model of the functionality). If and alarm condition exists, or if the sensor signatures change, the FM generates "diagnostic events" and sends them to the Inter-Level Coordinator (ILC). The FDs are also generated according to the functional breakdown, and there is one FD for each functionality. The interfaces of the FDs (incoming and outgoing failures and diagnostic events) are determined by the respective functionality models.

The ILC (1) receives the diagnostic events from the FMs, (2) routes the events to the proper FDs, (3) controls and guides diagnostic search (4) receives the *local* diagnostic results from FDs (5) combines the local diagnostic results and generates the fault hypotheses.

The prominent features of the diagnostic system are :

- Diagnosis of multiple faults (no assumption of single or multiple points of failures).
- Identification of observation errors.

- Robustness against a large number of sensor failures and graceful degradation as the number of sensor failures increase.
- It is event-driven and uses incremental non-monotonic reasoning.
- It predicts future events and uses the predictor-corrector principle to revise its hypotheses.
- Restricts the diagnostic search to the relevant parts of the functional hierarchy.
- Identifies loss of model validity in case of large faults and restricts its search to those parts of the hierarchy where the model of the system seems to be valid.
- Uses algorithms of polynomial complexity.

For details of the robust diagnostic system and the algorithms used, please see [9, 10].

CONCLUSION: A model-integrated approach to FDIR of complex, large-scale systems was presented. Although the primary motivation for this research came from the FDIR requirements for ISSA, the approach used here could be used for a variety of engineering systems, since it provides a solution approach for FDIR modeling and embedded health management for any complex, large-scale engineering system. The modeling formalisms are derived from standard engineering practices and domain specific concepts and relations, thus making it more accessible to systems engineer. The structural and functional organization of models makes the complexity and the scale of the systems easier to tackle. The feasibility of the model-integrated approach for using design information to formally and automatically derive embedded health management systems is demonstrated by the real-time robust diagnostic system synthesized from the models.

References

[1] E. Scarl *et al.*, "Diagnosis and Sensor Validation through Knowledge of Structure and Function," *IEEE Trans. Syst., Man and Cybernetics*, vol. SMC-17, no. 3, May./June 1987, pp. 360-368.

[2] J. de Kleer and B. C. Williams, "Diagnosing Multiple Failures," *Artificial Intelligence*, vol. 32, 1987.

[3] R. Davis, "Diagnostic Reasoning Based on Structure and Behavior," *Artificial Intelligence*, vol. 24, 1984.

[4] B. Kuipers, "Qualitative Simulation as Causal Explanation," *IEEE Trans. Syst., Man and Cybernetics*, vol. SMC-17, no. 3, May./June 1987, pp. 432-444.

[5] N. H. Narayanan and N. Vishwanadham, "A Methodology for Knowledge Acquisition and Reasoning in Failure Analysis of Systems," *IEEE Trans. Syst., Man and Cybernetics*, vol. SMC-17, no. 2, Mar./Apr. 1987, pp. 274-288.

[6] Padalkar, S., Karsai, G., Biegl, C., Sztipanovits, J., Okuda, K., Miyasaka, N.: "Real-Time Fault Diagnostics," *IEEE Expert,* pp. 75-85, June, 1991.

[7] S. J. Chang, F. DiCesare and G. Goldbogen, *Evaluation of Diagnosability of Failure Knowledge in Manufacturing Systems*, Proceedings, 1990 IEEE International Conference on Robotics and Automation, Vol 1, pp. 696-701.

[8] Pattipati, Krishna R. and Alexandridis, Mark G., "Application of Heuristic Search and Information Theory to Sequential Fault Diagnosis," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 4, July/August 1990, pp. 872-887.

[9] Misra, A., "Sensor-Based Diagnosis of Dynamical Systems," Ph.D. Dissertation, Vanderbilt University, May 1994.

[10] Misra, A., Sztipanovits, J., Carnes, R.: "Robust Diagnostic System: Structural Redundancy Approach," in Proc. of the SPIE's International Symposium on Knowledge-Based Artificial Intelligence Systems in Aerospace and Industry, Orlando, FL, April 5-6, 1994.

[11] R. Patton, P. Frank and R. Clark, "Fault Diagnosis in Dynamic Systems: Theory and Application," Prentice Hall International (UK), 1989.

[12] Shogo Tanaka, "Diagnosability of Systems and Optimal Sensor Location," Chapter 5 in the book *Fault Diagnosis in Dynamic Systems: Theory and Application,* Prentice Hall International (UK), 1989, pp. 21-45.

[13] Andow, P. K., and F. P. Lees, "Process Computer Alarm Analysis: Outline of a Method Based on List Processing," *Trans. Inst. Chem. Eng.*, 53, 195 (1975).

[14] Misra A., J. Sztipanovits, Carnes, J. R., "Modeling Paradigm for Failure Detection, Isolation and Recovery," Technical Report #95-001, Measurement and Control Systems Laboratory, Department of Electrical Engineering, Vanderbilt University, Jan. 1995.

[15] Sztipanovits, Janos *et al.*, "MULTIGRAPH: An Architecture for Model-Integrated Computing," submitted to ICECCS 95, Int'l Conf. on Eng. of Complex Systems, Nov. 6-10, 1995.

SOLID-STATE ACTIVE SENSING FOR IN-SITU HEALTH MONITORING

Craig A. Rogers and Frederic Lalande
Center for Intelligent Material Systems and Structures
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0261

Abstract: Currently, there exists a need for sensors which can perform on-line health monitoring of structures. Such sensors will lead to utilization of a greater fraction of useful life, greater reliability, and reduced maintenance and inspection cost. An on-line health monitoring sensor system must meet three requirements: i) it must be small, non-intrusive, and must offer the possibility of being located in inaccessible remote areas of the structure and, as yet, it must be able to transmit information easily to a central processor; ii) it must be as sensitive as conventional non-destructive evaluation (NDE) techniques, i.e., it must be able to detect minor cracks; and iii) it must be able to monitor a certain minimum area of the structure as opposed to the point measurements offered by presently available NDE equipment. Our recent research efforts indicate that the piezoelectric (PZT) materials used as sensory actuators may meet these requirements. The method being investigated relies on tracking the high-frequency (typically > 50 kHz) point impedance of the structure through a surface-bonded PZT to qualitatively identify damage. At such high frequencies, the technique is comparable in sensitivity to sophisticated traditional NDE techniques, and is capable of qualitatively detecting incipient-type damage by looking at changes in structural impedance. As yet, it can be implemented in a remote sensing scenario. Several case studies involving built-up complex structures will be presented to demonstrate the potential benefits of the technique.

Key Words: Damage detection; health monitoring; impedance approach; nondestructive evaluation (NDE); smart structures

INTRODUCTION: A new qualitative non-destructive evaluation (NDE) sensor system to perform in-situ on-line monitoring of critical sections of structures and to provide a warning in the event of damage is the focus of this paper. How this new method differs from conventional NDE techniques like ultrasonics and X-rays, is in its ability to monitor a relatively larger area with minimal intrusion. While these conventional techniques can pin-point microscopic damage, they require ready access to the area as well as relatively large pieces of equipment. Such access, in some cases, is simply not available and in other cases can mean taking the structure out of service. Taking the structure out of service in itself can cause major disruption; in addition, the process is usually very labor intensive and costly.

Therefore, the need to develop a sensor system which can be permanently and non-intrusively installed in the critical sections of the structure to monitor their integrity. How does the new method overcome the limitations of the conventional NDE equipment and strain gage sensors? The new method employs high frequencies (typically > 50 kHz) at which there is a measurable change in structural point impedance even for minor damage such as small internal cracks, surface cracks, and loose connections. As yet, it can be implemented via small non-intrusive surface-

bonded transducers for remote on-line monitoring. With regard to strain-gages and such local strain measuring techniques: first, the new method does not need a calibrated load on the structure since the actuator-sensor itself provides the excitation; and second, it can monitor a relatively large area. The extent of the area the sensor system can monitor is certainly much larger than a strain gage; but, the exact sensing range is not exactly known.

The proposed sensing system and methodology, in principle, utilizes the changes in the vibration/impedance signature (due to damage) to qualitatively identify damage. But, it is different from modal analysis-based methods. The fundamental difference lies in the frequency used to interrogate the structure, i.e., excite and sense the resulting vibration magnitude and phase. The frequencies used in this technique are much higher than those typically used in modal analysis-based methods. To sense incipient-type damage which does not result in any measurable change in the structure's global stiffness properties, it is necessary for the wave-length of excitation to be smaller than the characteristic length of the damage to be detected [1]. Finding damage when it is at an incipient level and before the global structural integrity is compromised, is most useful. This is because it can provide us with a warning before actual failure occurs.

The proposed technique described in this paper utilizes small PZT patches to provide high-frequency excitation, typically in the high kHz range, to the structure being monitored. At such high frequencies, the response is dominated by local modes and incipient damage like small cracks, loose connections, and delaminations, produces measurable changes in the vibration characteristics. The high frequencies also limit the actuation/sensing area. The effect of excitation frequency, geometry, material properties, structural joints, etc., on the size of the sensing/actuation area is still under investigation, but it is our observation that the sensing area, as a minimum, extends to the boundaries of the solid member to which the PZT is bonded. This limited sensing area helps to isolate the effect of damage on the signature from other far-field changes in mass-loading, stiffness and boundary conditions. The insensitivity to far-field boundary conditions comes at the cost of a limited sensing area. Therefore, this technique will be most useful in identifying and tracking damage in those areas of structures where high structural integrity must be assured at all times.

ELECTROMECHANICAL PRINCIPLES: Piezoelectricity describes the phenomenon of the generation of an electric charge in a material when subjected to a mechanical stress (direct effect), and conversely, a mechanical strain in response to an applied electric field (converse effect). Contrary to most applications where the piezoelectric material is either used as a pure actuator or as a pure sensor, this signature acquisition method utilizes the direct as well as the converse effect allowing for simultaneous actuation and sensing (e.g., collocated actuator/sensor) of the structural response. A PZT bonded or attached to a structure, and driven by a fixed alternating electric field, excites and induces vibrations in the structure. Meanwhile, the resultant vibrational response, which is characteristic of the particular structure, modulates the current flowing through the PZT. This modulation is representative of the degree of mechanical interaction between the PZT actuator-sensor and the structure at different frequencies. In electrical terms, the variation in the current modulates the electrical admittance, which is defined as the ratio of the resulting current to its energizing voltage. The admittance signature then provides the same information, and serves the same purpose as the conventionally known transfer function. The size of the PZT patch-sensor required to get a good vibration signature is typically small (less than a 0.5 sq. in., .01 in. thick). This allows for non-intrusive installation.

Mechanical impedance (the reciprocal of impedance) is defined as the ratio of the applied force and the resulting velocity. Analogous electrical impedance is the ratio of the applied voltage and the resulting current. Electromechanical transducer materials such as piezoelectrics provide a means for coupling the mechanical and electrical impedance. This allows for the extraction of vital mechanical impedance information from pure electrical impedance measurements. The electrical impedance measurements are greatly facilitated by use of a commercially available impedance analyzer. In the electrical impedance *vs.* frequency measurements, the structural resonance will show as sharp peaks above the base-line electrical capacitive impedance. Since these peaks correspond to specific structural resonances, they constitute a unique signature of the structure's dynamic behavior. Therefore, any change in the impedance is considered an indication of structural integrity variation.

PROOF OF CONCEPT APPLICATIONS: To date, the new health monitoring method has been applied to a wide variety of situations in the laboratory. A brief outline of the successful applications done at CIMSS is presented.

Aircraft Fuselage [2]: The tail section of the Piper Model 601P airplane was selected to test the NDE technique on complex structures. The goal was to monitor the integrity of the two main brackets which connect the rear fuselage to the vertical tail. PZT patches were mounted on the fuselage side of the vertical tail support brackets, each within one inch of the two securing bolts (Fig. 1).

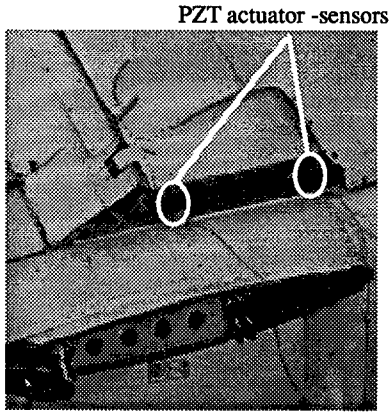


Figure 1. *The integrity of the two main brackets which connect the rear fuselage to the vertical tail of a Piper Model 601P airplane was monitored.*

The experimental results for the bolted aircraft joint are presented below. The two graphs in Fig. 2 show the real admittance measurements of the PZT actuator/sensor with local damage and global damage, respectively. The change in the electrical admittance measurements with a small defect (smallest possible

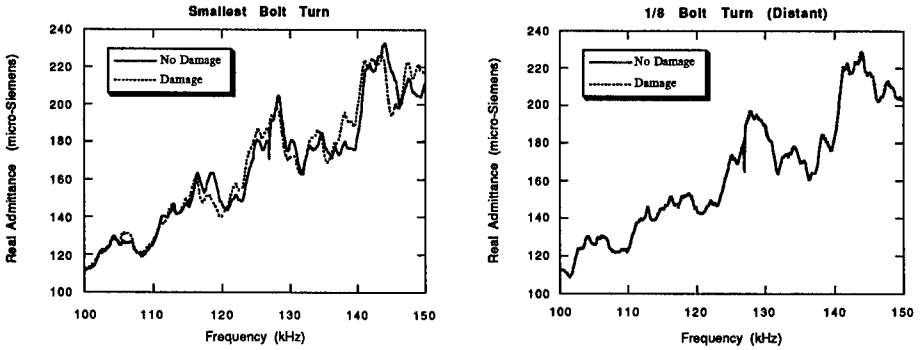


Figure 2. *Graph on the real electrical admittance of the PZT actuator-sensor. Note the large variations in the values of the local bolt, while the distant bolt values remain unchanged.*

turn of the connecting bolt) is significant while the measurements are unaffected by distant damage. The principal conclusions drawn from this work were that: i) the technique, because of the high-frequency, is very sensitive to minor change; ii) the actuation-sensing area is limited to a small area. The localization of the sensing area provides a practical means of utilizing vibration/impedance measurements to monitor critical sections of structures.

A scalar damage index used to simplify the interpretation of the impedance changes is shown in Fig. 3 for the aircraft bolted joint, with five different types of damage, both local and far-field. The damage index is defined as the sum of the differences of the real admittance change, squared at each frequency step. This damage index is then normalized to 100% with respect to the local smallest bolt turn to which all other damage is then relatively compared. A larger damage index than 100% means more structural damage than the reference smallest bolt turn. The principal advantage of this damage index is the facilitation of the result interpretation. Based on the damage index, damage can be reported to the operator automatically when a threshold value is reached, in a green/red light form.

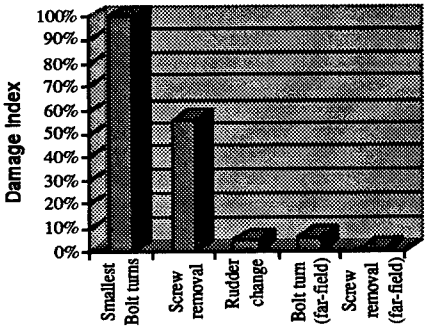


Figure 3. *Damage index of a bolted joint between the fuselage and the vertical tail of an aircraft subjected to various local and far-field alterations.*

Composite Repair Patch [3]: Critical to the success of composite repair of metallic aircraft structures is the integrity of the bond between the base aluminum panel and the reinforcing high-strength composite patch. Monitoring of the repair is equally important to ensure the integrity of the composite patch throughout the service life of the structure. The new NDE method was applied to test coupons repaired with a composite patch to monitor the growth of the crack under fatigue loading (Fig. 4). The detection of debonds in the composite repair patches was also investigated.

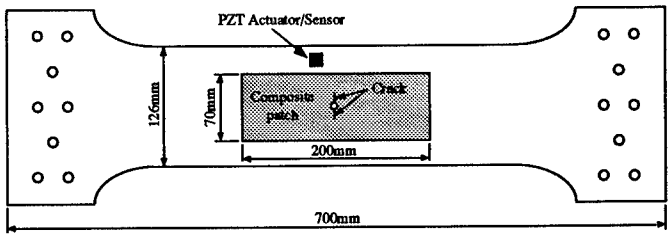


Figure 4. *An aluminum dog-bone specimen with a small crack was repaired with a bonded graphite/epoxy composite patch.*

The electrical measurements of the dog-bone specimen before and after the crack growth are presented in Fig. 5. A clear shift in the resonant frequencies can be observed, indicating a change in the structural properties of the specimen, i.e., crack growth. Similar tests performed on dog-bone specimens without composite repair patches showed similar variations in the admittances measurements. Also, very minor debonds between the composite patch and the metal base structure were qualitatively detected.

Deck Truss Bridge [4]: To this point, the research has involved relatively light structures. In this next proof-of-concept demonstrator, a massive one-quarter model of a steel deck truss bridge joint that experienced in-service failure is investigated. The size of the demonstrator is 72 inches tall, 41 inches wide, and over 12 inches deep (Fig. 6). The entire structure weighs more than 500 pounds and is considered representative of a typical high-strength civil engineering steel structure. Three piezoceramic incipient damage actuator/sensors were bonded at critical locations on the structure. As for the aircraft structure, local and global damage due to loose connections and structural damage was investigated.

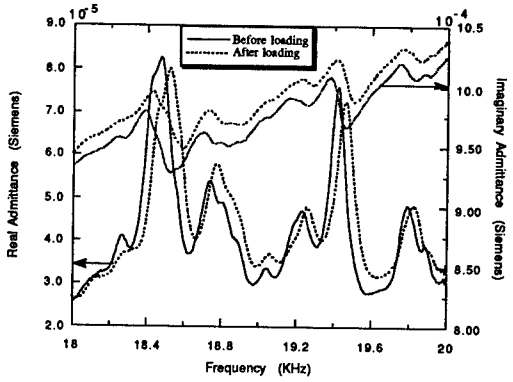


Figure 5. *Effect of crack growth on the admittance measurements of the dog-bone specimen with a graphite/epoxy repair patch.*

In Fig. 7, the admittance measurements for the healthy structure and the damaged structure (one loosened local bolt) are shown. It is to be noted that loosening one of the six bolts constituting the connection between the two structural members does not constitute a change in the global stiffness of the structure; in fact, it is quite comparable to incipient-type damage. In the experiments, the incipient damage manifested itself as a vertical shift of the admittance, which differs from the previous cases were a non-uniform horizontal admittance change was observed. This different behavior was attributed to the high stiffness of the structure. Nevertheless, the nondestructive evaluation technique demonstrated the capacity of detecting incipient damage once again. The results also indicated that the technique has the potential for quantifying damage: the effect of two loose bolts is almost double that of a single bolt. The localization phenomena of the NDE technique was once again observed.

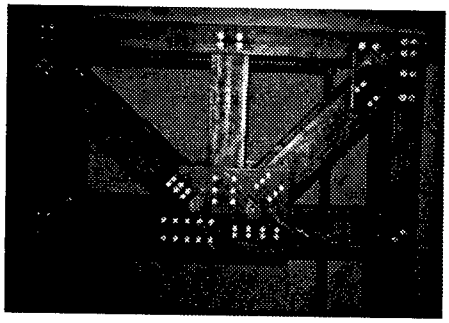


Figure 6. *A quarter-scale model of a bridge joint was used to study the NDE technique in massive structures.*

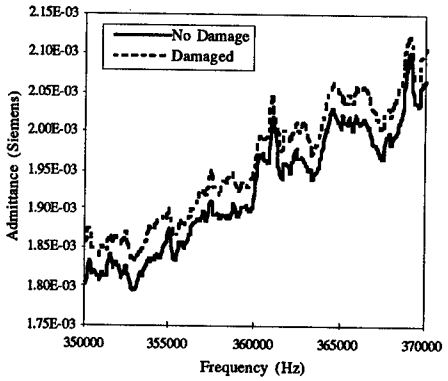


Figure 7. *The loosening of a local bolt has the effect of shifting the admittance measurements.*

Precision Parts [5]: The application of the impedance-based technique was then extended to high- precision parts. Gears were chosen as complex precision parts for the experimental procedure because of their high tolerances, high quality, and broad use. The goal was to show that incipient damage in the gear teeth, which are an extension of the base structure, can be monitored. The most common types of damage in gears, i.e., abrasive wear and bending fatigue cracks, were successfully detected. The impedance measurements before and after damage were converted into a scalar damage. Also, quality inspection was successfully demonstrated using the impedance-based technique.

In Figure 8, the damage metric of two frequency ranges for the adjacent (0°) and distant (90°) cracks is shown to be dependent on the frequency range. A greater structural activity in a frequency range, which reflects in an impedance reading consisting of many peaks and valleys, will cause the damage metric to increase because more variations in the impedance are present. A dynamically active frequency range should be picked for increased sensitivity of the impedance-based health monitoring technique. Also, as mentioned earlier, the damage metric is larger for adjacent cracks to the PZT actuator/sensor due to the localized effect. The damage metric for the abrasive wear is also shown in Figure 8. The damage created by the wear of the gear teeth is less important than the damage created by the crack.

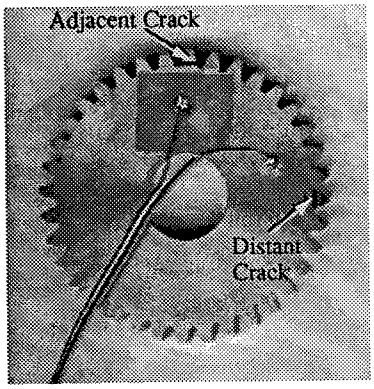


Figure 8. *Detection of bending fatigue cracks and abrasive wear in high-precision gears was investigated.*

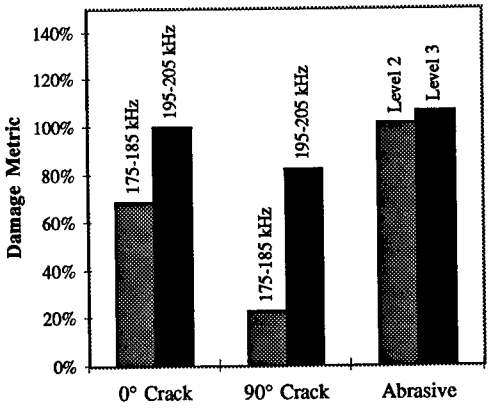


Figure 9. *The damage metric of all studied cases shows the frequency dependency of the health monitoring technique.*

Three-Bay Aluminum Truss [6]: Finally, the health monitoring technique was implemented on a three-bay aluminum truss. The purpose of this experimentation was the on-line implementation of a system using PZT actuator/sensors at multiple critical locations. Small PZT's (approximately 8 x 8 x 0.2 mm) were bonded to the eight nodes of the middle-bay. The impedance signature from each node was sequentially acquired and controlled by a personal computer. Damage was simulated by loosening one of the member's connection with the nodal Derlin-ball. Within the personal computer controller, the damage metric was computed and the health status of the structure was displayed in a green/red light fashion. Thus, through this investigation, the possibilty of implementing the technique on-line was confirmed.

ADVANTAGES OF THE NDE TECHNIQUE: The impedance-based health monitoring technique has been shown to be very efficient for the detection of incipient damage in complex structures. In summary, this new qualitative NDE technique has the following principal advantages:

- The technique is not based on any model, and thus can be easily applied to complex structures;
- The technique uses non-intrusive and small-size actuators to monitor inaccessible locations;
- The technique is unaffected by changes in boundary conditions, loading, or operational vibrations;
- The technique can be implemented for on-line health monitoring;
- The continuous monitoring provides a better assessment of the current status of the structure, which can eliminate scheduled base inspections;
- The added weight of the actuator/sensor is negligible.

CONCLUSION: A high-frequency impedance-based health monitoring technique is described. The method utilizes the changes in the electrical impedance of a surface-bonded PZT actuator-sensor, acquired through an impedance analyzer, as a qualitative indicator of damage. The five proof-of-concept applications showed the effectiveness of the new method in real-life complex structures. This technique, because of the high frequency, is very sensitive to minor damage, and the actuation-sensing area is limited to a small area. The localization of the sensing area provides a practical means of utilizing vibration/impedance measurements to monitor critical sections of structures.

ACKNOWLEDGMENTS: The authors gratefully acknowledge the support of the Army Research Office - University Research Initiative Program, Grant No. DAAL03-92-G-0181, Dr. Gary Anderson, Program Manager.

REFERENCES:

[1] Stokes J. P. and Cloud, G. L., "The Application of Interferometric Techniques to the Nondestructive Inspection of Fiber-reinforced Materials," *Experimental Mechanics*, 1993, Vol. 33, pp. 314-319.

[2] Chaudhry, Z., Sun, F., and Rogers, C. A., "Localized Health Monitoring of Aircraft *via* Piezoelectric Actuator/Sensor Patches," *Proceedings, SPIE 1995 North American Conference on Smart Structures and Materials*, San Diego, CA, 26 Feb. - 3 March, 1995, pp. .

[3] Chaudhry, Z., Lalande, F., Ganino, A., Rogers, C. A., and Chung, J., "Monitoring the Integrity of Composite Patch Structural Repair *via* Piezoelectric Actuators/Sensors," *Proceedings, AIAA/ASME/ASCE/AHS/ASC 36th Structures, Structural Dynamics, and Materials Conference*, New Orleans, LA, April 11-14, 1995, pp. 2243-2248.

[4] Ayres, J. Chaudhry, Z., and Rogers, C.A., "Localized Health Monitoring of Civil Infrastructure via Piezoelectric Actuator/Sensor Patches," *Proceedings, SPIE 1995 North American Conference on Smart Structures and Materials*, San Diego, CA, 25-29 February 1996, in press.

[5] Lalande, F., Childs, B., Chaudhry, Z., and Rogers, C.A., "Utilization of High-Frequency Vibration Pattern Recognition for NDE of Complex Precision Parts, ," *Proceedings, SPIE 1995 North American Conference on Smart Structures and Materials*, San Diego, CA, 25-29 February 1996, in press.

[6] Chaudhry, Z., Sun, F., and Rogers, C.A., "Health Monitoring of Space Structures Using Impedance Measurements," *Proceedings, 5th International Conference on Adaptive Structures*, Sendai, Japan, 5-7 December 1994, pp. .

ARTIFICIAL KNEE JOINT CONTROLLER CHARACTERISATION AND CONDITION DIAGNOSIS

C.K. Mechefske and J.T. Davenport
Department of Mechanical Engineering
Victoria University of Technology
P.O. Box 14428
Melbourne, Victoria, Australia, 8001

Abstract: The development of the hydraulic artificial knee joint controller has significantly improved the quality of life for many people who have suffered above the knee amputations. While these devices are extremely reliable, lasting up to 3 years before servicing is required, gradual deterioration does take place. Minor deterioration of the knee controller performance can be compensated for using adjustments on the device, however, changing performance has largely been left to the user to detect qualitatively during use. Given the wide range of operating conditions that the controller may regularly be used within and the continuous usage of the device, it is not surprising that gradual deterioration of the performance is difficult to detect. Sudden failures of hydraulic knee controller units are known to occur and can result in oil leakage and loss of support, bringing about embarrassing and dangerous situations for the user. At present there is no simple quantitative test which allows for early detection of performance deterioration. The work reported on in this paper was carried out to establish clear performance characteristics for the controller. Various controller settings were used for simulated normal operating condition tests and full flexion relaxation tests. All the operating and test parameters, including performance test results are presented for a new and known to be significantly deteriorated controller. It was found that the time for the unit to return unassisted to its equilibrium position from full displacement (full flexion), varied greatly between the new and old units, suggesting a simple quantitative method for the early detection of controller deterioration.

Key Words: Artificial knee joint controller; condition characterisation; diagnostics; prostheses; quantitative test method.

INTRODUCTION: The development of the hydraulic artificial knee joint controller has significantly improved the quality of life for many people who have suffered above the knee amputations. This device has allowed for the improved functionality of leg prostheses. The ability to adjust the controller allows the user to pick settings which optimise the controller response for different user activities. The user can select settings which improve the ease with which he/she is able to climb or descend stairs, walk on a grade, jog, ride a bicycle, etc.

Given the wide range of operating conditions that the controller may regularly be used within and the continuous usage of the device, it is not surprising that gradual deterioration of the performance is difficult to detect. In general, the gradual change during regular use is too slow

for the user to notice until the performance has deteriorated significantly. Even then some users accept this performance because they do not remember the original "feel" of the device.

With an expected operational life of up to 3 years between services it would be advantageous for users to be able to assess the remaining life of their knee controller. Knowing, quantitatively, the degree of performance deterioration which has taken place would also allow users to better plan their activities, including when to have a replacement unit on hand. It could also be useful as an early warning of the increased probability of sudden and complete failure of the controller unit as performance deterioration advances.

Practitioners in the field also have difficulty determining the exact state of deterioration of the controller because they may have little or no experience with that particular unit. They are also hampered by the fact that there exist no published performance guidelines to allow comparisons against known conditions or standards.

The work reported on in this paper was carried out to establish clear performance characteristics for the controller. Various controller settings were used for simulated normal operating condition tests and full flexion relaxation tests. All the operating and test parameters, including performance test results are presented for a new and known to be significantly deteriorated controller. It is expected that tests at these two extremes of operational performance will set limits within which controllers of unknown condition can be compared. Future work will involve further testing under a variety of operating conditions and the endurance testing of several controllers from new condition through to fully deteriorated condition. This will allow complete performance characteristics to be compiled and the establishment of objective controller testing procedures

BACKGROUND: Prosthetic knees provide three functions; support during stance phase, a smooth and controlled swing phase, and unrestricted flexion for sitting, kneeling, and related activities. Hydraulic knee controllers regulate the flexion and extension motion of the artificial knee joint to achieve the above functions in an optimum manner. Hydraulic knee controllers consist of a cylinder and piston attached to the thigh section of the prosthesis above the knee and to the lower leg (shank) section below the knee (see Figure 1).

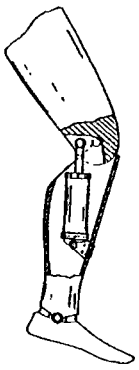


Figure 1: Artificial leg and hydraulic knee controller unit.

Knee flexion forces the piston down into the cylinder. This in turn forces fluid through a bypass channel at the bottom of the cylinder. The fluid travels upward within and around the sides of the cylinder, through a port at the top of the cylinder, and back into the central cylinder above the piston [1]. The resistance to motion provided by the liquid depends on the liquid viscosity and temperature. Silicone oil is used in most prosthetic hydraulic units because it minimises viscosity changes with temperature, thus minimising stiffness in cold weather and looseness in hot weather. The size of the oil passages and the number of passages open to allow flow also governs the rate of flexion and extension under a given load [2][3].

The hydraulic knee assists the extension process by the use of a spring. The spring is compressed during knee flexion. Energy stored in the spring during flexion is then used to push the shank into full extension during the forward swing of the leg when walking. A full description of the hydraulic knee controller operation is given elsewhere [4].

The hydraulic units can be reconditioned several times throughout their useful life, with each recondition providing approximately two million cycles or, conservatively, two years of use. The units can last as long as 20 years with the appropriate maintenance.

The Monash Rehabilitation Technology Research Unit (REHAB Tech) located at the Caulfield General Medical Centre in Melbourne, Australia suggests that prosthetic hydraulic knee units exhibit sudden catastrophic failure on a regular basis [5]. Catastrophic failure in this case means the complete failure of the unit to adequately resist flexion and retard extension and thereby provide acceptable functionality. Two of the leading manufactures of hydraulic knee controller units were contacted regarding the possibility of detecting the gradual deterioration of the performance of the units in a quantitative manner [6][7]. One of the manufacturers indicated that the user should be able to feel when the performance of the unit has deteriorated and would then return the unit for service. In their experience the units did not fail catastrophically unless used beyond their recommended useful life. The other manufacturer declined to comment.

The units that were tested were CATECH units [3]. Both units have been reconditioned approximately seven times. The 'as new' unit was recently reconditioned, while the 'old' unit came from a patient complaining of poor performance. Clinical examination of the unit failed to reveal any obvious dysfunction of the unit.

PERFORMANCE CHARACTERISATION: The action of the hydraulic knee controller unit can be modelled as a simple spring and damper system. The unit obtains its resistance to motion by the use of a viscous fluid. This allows control over the rate of flexion and extension of the knee joint. The spring provides extension assistance to reduce the amount of exertion required by the user. The behaviour of the unit may then be characterised using the rate of displacement of the piston either in flexion or in extension when acted on by a known force. The force required to move the piston at a known rate of displacement may also be used to characterise the controller unit performance. This was taken into account when setting up the experiments to assess the performance characteristics of the hydraulic knee controller.

TESTING PROCEDURE: Two types of tests were designed and conducted on an 'as new' hydraulic knee controller and on an 'old' unit. These tests are fully described in this section. The test results and a discussion of the results follows in the next section.

Flexion Relaxation Tests: The flexion relaxation tests involved compressing the piston of the hydraulic knee controller unit a known distance as would take place during knee flexion. The piston was then released and forced by the internal spring to return to the fully extended position. The displacement versus time plot generated can then be used as an indication of the controller unit's condition.

For these tests the controller unit was clamped in a vertical position. A displacement transducer was then attached. The transducer sent a voltage to a computer via an analogue-to-digital converter where the data was captured using in a data acquisition program called DTV. A program was written in DTV to allow for the appropriate sampling to be obtained. The hydraulic knee unit piston was depressed fully (flexion) and released. Data acquisition took place during the relaxation of the unit (extension of the piston). This procedure was then repeated for various settings of the unit's extension and flexion adjustments on both the 'as new' and 'old' units.

Simulated Step Tests: In the simulated step tests the hydraulic knee controller units were mounted in an Instron 8501 dynamic testing machine. The knee controller was then forced through repeated flexion and extension cycles simulating the motion induced during a typical walking cycle. The simulated step cycle is shown in Figure 2.

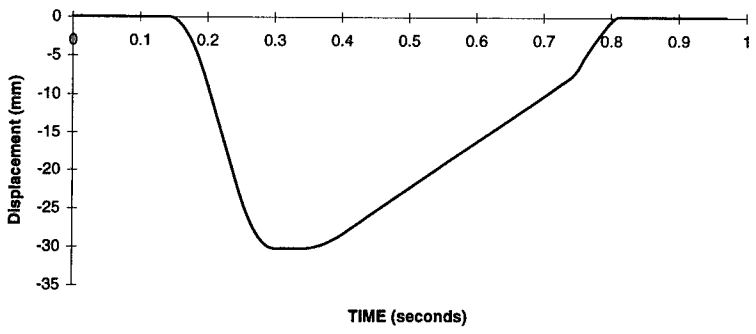


Figure 2: Typical displacement cycle of controller unit.

As can be seen from the figure, a typical step cycle can be divided into four phases. The first phase of the step is the flexion of the controller unit (knee flexion). This is a relatively short duration, high rate of displacement compression of the unit piston into the cylinder. The flexion phase is followed by a short period of time where there is no or little displacement of the piston.

The third phase is the extension of the controller unit (knee extension). This phase is of longer duration than phase one (lower rate of displacement) and ends just before the piston is fully extended. The final phase is a short duration, high rate of displacement extension of the piston to full extension. This portion of the step simulation is meant to mimic the user's flick of the lower leg into hyper-extension just before placing weight on the leg to begin a new walking cycle (see detailed description of hydraulic knee controller unit operation [4]).

The data stored by the Instron testing machine was the force required to obtain the simulated walking motion and the corresponding time values. The test on the 'as new' and 'old' unit were again performed at various settings of the flexion and extension adjustments available for each device.

RESULTS AND DISCUSSION: Due to the large number of tests conducted and the limited space available for the presentation of results, this section contains only the results from some of the tests conducted.

Flexion Relaxation Test: The results of the flexion relaxation tests are shown in Figures 3 to 6. While tests were conducted on both the 'as new' and 'old' units at most of the flexion and extension adjustment settings, only the extreme settings are shown here. The results at these settings show most dramatically the difference in performance between the two units.

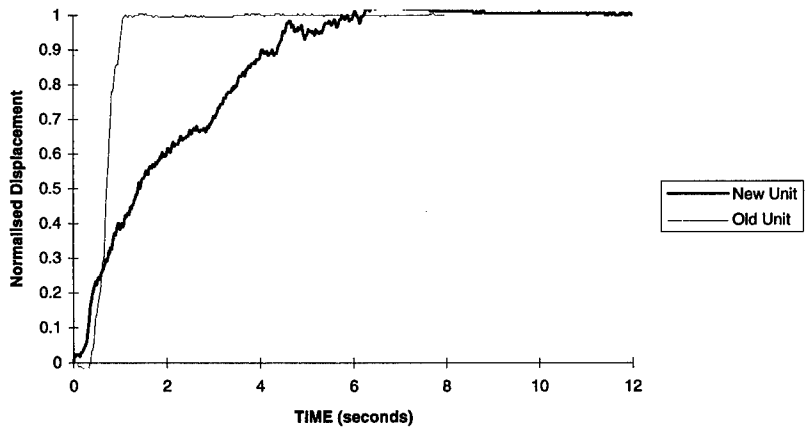


Figure 3: Flexion relaxation test at minimum flexion and minimum extension setting.

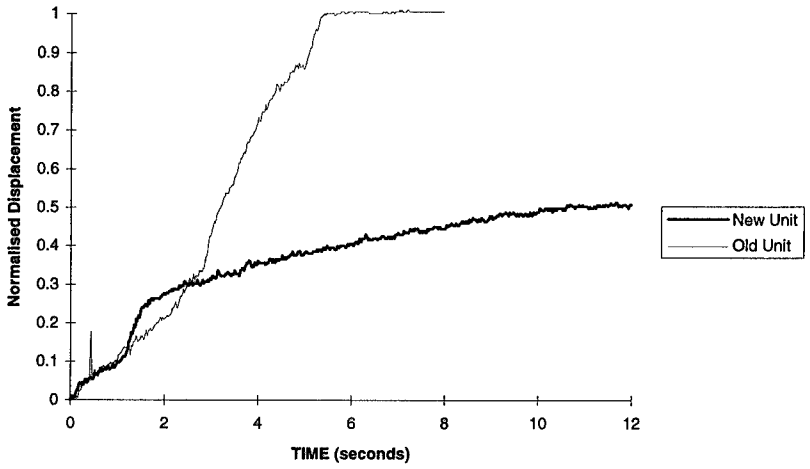


Figure 4: Flexion relaxation test at minimum flexion and maximum extension setting

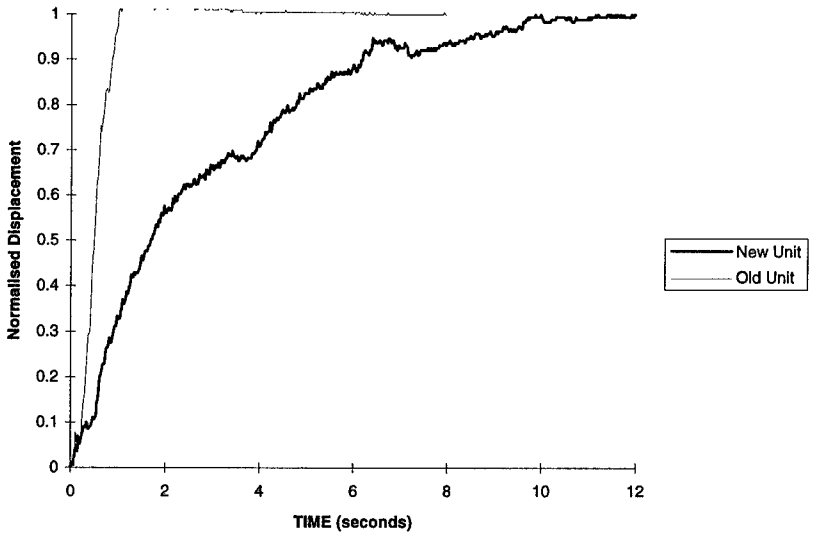


Figure 5: Flexion relaxation test at maximum flexion and minimum extension setting

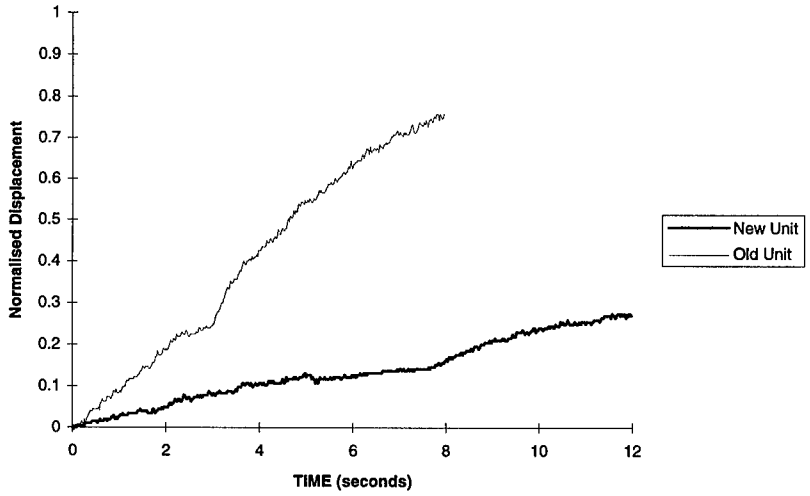


Figure 6: Flexion relaxation test at maximum flexion and maximum extension setting

In these figures it is clear that the 'as new' unit takes much longer to reach full extension at each of the different adjustment combinations shown. This result could be translated easily into a quick and simple quantitative diagnostic test for hydraulic knee controllers of unknown condition. By testing a greater population of units representing various known states of performance deterioration and other designs and manufacture's models, a series of tables or charts could be developed which would allow the performance of a controller to be quantified. It should be possible to design such a test unit for use in a clinical setting or even at the home of the artificial knee user

Simulated Step Tests: A representative sample of the results of the simulated step tests is presented in Figure 7. This figure shows the force exerted by the Instron dynamic testing machine when moving the knee controller units through the typical displacement cycle shown in Figure 2. In this plot the 'as new' unit is compared to the 'old' unit. Both units were set on minimum flexion and minimum extension. Only one plot is shown because of the similarity between plots at all the different settings.

The first point of note is that during the first test on the 'as new' unit a loud crack was heard when the unit's piston was moved through the typical step cycle. This was interpreted as catastrophic failure of the unit. As can be seen from the plot of force required to move the 'old' unit through the step cycle, the force is significantly greater than that required for the 'as new' unit. Hence, in this set of tests the 'as new' unit represents a failed unit while the 'old' unit represents a deteriorated but still functional unit. Further evidence of the failure appeared later in the test when oil began to leak from the 'as new' unit.

Two further points should be noted. The first is that there is obviously a significant difference in the force required to move the controller pistons in these two tests. Measuring the force required to move the unit piston through a prescribed displacement could, therefore, be another method of distinguishing the degree of unit functional deterioration. The second point to be made is that catastrophic failures do occur despite the manufacturers assurances that they do not.

A long term endurance test was planned for both the units but because of the failure of the 'as new' unit, it was not completed.

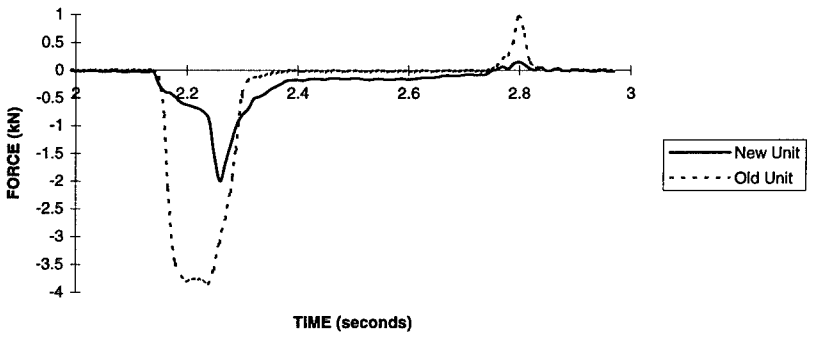


Figure 7: Simulated step test at minimum flexion and minimum extension setting

CONCLUDING COMMENTS: Hydraulic knee controllers representing 'as new' and 'old' conditions were tested using a flexion relaxation test. Knee controller units representing completely failed and deteriorated but still functional conditions were tested using a simulated step test. Both tests compared controllers which were of identical design but in significantly different operating conditions over the full range of their flexion and extension settings.

It is clear from the results presented that differences in performance can be measured and potentially used to distinguish the degree of performance deterioration of hydraulic knee controllers. Two simple methods for accomplishing this task were investigated. Noting the rate of displacement of the unit's piston under a constant load or measuring the force required to move the piston at a set rate of displacement both give clear indications of being useful tools.

The flexion-relaxation test indicates the best method for determining in a simple manner how deteriorated the performance of the hydraulic unit has become. If this test was to be designed for clinical use, it would require only a simple timing device or data acquisition program to obtain the performance data. A comparison of this data to standards representing knee controllers of known condition could then be made by the clinician or automatically using a computer based classification program. If the values were close to critical the user could then receive the appropriate servicing, if not the approximate remaining useful life of the unit could be estimated. In this way servicing could take place at opportune times, minimising lifestyle disruptions for the user and maximising the overall useful life of the unit by helping to prevent excessive damage to

units through over use. If the test and testing device were simple and inexpensive, the user would not require to go any where for testing. Instead, the user could simply test the unit and check the results in a manual at home.

Future work on this project will involve the design of a simple and accurate knee controller testing device for clinical or home use. Further testing will also be conducted on more units to check the statistical variability between units. Gradual deterioration tests should also be conducted to determine the rate of deterioration under different types and intensities of usage.

Further development work may be to design and place a built-in monitoring system into or adjacent to the knee controller unit which would be able to continually monitor the condition of the unit and give a warning if the unit operation had deteriorated excessively. This could be achieved by setting up small transducers in the knee with a microchip capable of comparing the data received from the transducers to that stored in the memory as recommended values. This work may also lead to the design of an active controller. Such a device could possibly sense the type and intensity of use taking place, estimate the amount of unit deterioration having already occurred and adjust the flexion and extension settings automatically for the optimum performance.

ACKNOWLEDGMENTS: The authors wish to acknowledge the contributions of Mr. Bill Contoyannis of Monash Rehabilitation Technology Research Unit in regard to providing artificial knee controller units for testing and valuable technical advice, and Mr Vincent Rouillard of Victoria University of Technology for his technical assistance during testing.

REFERENCES:

1. Rubin, R.M.D, Clinical Prosthetics and Orthotics, Volume 7, Academy of Orthotists and Prosthetists, 1983.

2. Mauck, H., Stance Control for Above-Knee Artificial Legs - Design Considerations in the S-N-S Knee. Mauck Laboratories Inc., Dayton, Ohio, USA, 1968.

3. Swing Phase Hydraulic Knee Controls, Catech Inc., Dayton Ohio, USA, 1994

4. Davenport, J.T., Honours Bachelor of Engineering Fourth Year Thesis. Department of Mechanical Engineering, Victoria University of Technology, Melbourne, Australia, 1995.

5. Contoyannis, B. Manager, Monash Rehabilitation Technology Unit, Caulfield General Medical Centre, 260-294 Kooyong Road, Caulfield, Victoria, Australia. Personal communication, April 1995.

6. Phillips, W. President and CEO, Maulk Laboratories Inc., Dayton Ohio, USA. Personal communication, April 1995.

7. Wiggins, L. President, Catech Inc., Dayton, Ohio, USA. Personal communication, April 1995.

Comprehensive Machinery Monitoring with FFPI Sensors

Robert X. Perez
RXP Consulting
6733 Deerwood
Corpus Christi, TX 78413

Robert A. Atkins and Chung E. Lee
FFPI Industries
909 Industrial Blvd.
Bryan, TX 77803

Henry F. Taylor
Department of Electrical Engineering
Texas A&M University
College Station, TX 77843-3128

Abstract: This paper introduces the benefits of applying the optical fiber Fabry-Perot interferometer (ffpi) pressure sensor as a means of detecting flow instabilities in centrifugal pumps and compressors. Actual pump test data will be presented, demonstrating how harmful hydraulic instabilities, such as cavitation and low flow instabilities, can be readily detected using this new sensor. The authors will explain how centrifugal compressors can be protected from surge by sensing flow reversals with an ffpi differential pressure sensor. Also covered is how the ffpi pressure sensors can be integrated into a complete ffpi monitoring system capable of sensing bearing housing vibration, shaft vibration, and shaft rotary speed.

Key Words: accelerometer; cavitation; centrifugal compressors; centrifugal pumps; fiber optics; Fabry-Perot interferometer; pressure sensor; surge

Introduction: Today, more than ever, it is vital that we design and operate our process equipment so that catastrophic failures and product releases are extremely rare events. This means that, in addition to purchasing well designed and constructed equipment, we must monitor their condition to ensure they remain healthy during their operational lives. A vibration monitoring program is a common means of protecting mechanical equipment from catastrophic failures. For critical pumps, such as those handling highly flammable or toxic fluids, prudent operation also requires users to monitor pressure pulsations as a means of ensuring proper hydraulic operation and preventing flow related mechanical failures. For centrifugal compressors, dynamic flow fluctuations should be sensed as a

means of protecting them from the dangers of surge. The authors will discuss how it is now possible and practical to monitor both the mechanical condition and hydraulic or aerodynamic condition of high performance centrifugal equipment.

Centrifugal Pump Characteristics: Centrifugal pumps are designed to produce a fairly constant differential pressure at a given flowrate. Idealized head-flow curves give users the illusion that centrifugal pumps generate only static pressure. However, in reality, this class of pump generates a dynamic pressure component along with the static component. This dynamic pressure component is an excellent indicator of how the pump is being operated. Many have tried to use vibration information to detect flow related problems; but this evaluation method is imprecise when it comes to assessing the severity of the flow problem. Dynamic pressure, on the other hand, can be converted to force by knowing the peak to peak pressure magnitude and the impeller's projected area.

A small dynamic pressure component (<10% of the static differential pressure) means there is plenty of net positive suction head available (NPSHA) and that the pump is operating close to its best efficiency point (BEP); but a large dynamic pressure component means the pump is hydraulically unstable [1]. The primary danger of excessive dynamic pressures is that they lead to large dynamic forces causing premature bearing failures and shaft deflections that can lead to wear ring contact and seal failures [2],[3].

Until recently, a continuous-duty dynamic pressure monitoring system was impractical because standard pressure transducers were limited to processes operating below about 300 ^{O}F. With the recent development of the optical fiber Fabry-Perot interferometer (ffpi) pressure sensor, pumps can now be monitored in services operating at much higher temperatures [4]. In the past year, ffpi pressure sensors have been used successfully in harsh process environments with temperatures exceeding 700 ^{O}F.

Pressure Sensor Design: The ffpi sensing element, which is the basis for the pressure transducer, consists of two internal mirrors separated by a length L of single mode optical fiber, as illustrated in Fig. 1. Each mirror is produced by vacuum deposition of a thin film of the dielectric material TiO_2 on the cleaved end of a fused silica (SiO_2) fiber. Electric arc fusion splicing is used to integrate the mirrors, which each have a reflectance of about 5%, into a continuous length of the fiber. For the pressure sensor, L is about 1 cm.

The next step in making a pressure sensor is to embed the ffpi along the axis of an aluminum alloy rod by a casting process. After machining the cast rod to the desired dimensions, it is inserted into a stainless steel housing with a thin (0.5 mm) lower wall. A nut at the top of the housing is torqued to produce a slight compression of the aluminum rod. The sensor is then mounted in a threaded port in the pump inlet or outlet line, as in Fig. 2.

To monitor the sensor, light is coupled into the fiber and a portion of the optical power reflected from the ffpi is converted to an electrical signal by a photodetector. The amplitude of the reflected power, as determined by coherent interference of light reflected from the two mirrors, is very sensitive to small changes in L. The pressure sensor is designed such that the fluid pressure produces a slight strain (of the order of 10 μstrain, or 0.1 μm change in L) in the fiber, leading to a large fractional change in the reflected optical power [4],[5].

System Design: A pressure measurement system developed at FFPI Industries in which one laser provides optical power for up to 24 sensors is illustrated in Fig. 3. Light from a semiconductor laser diode (LD) modulated with a sawtooth waveform is coupled into a single mode fiber and through an optical isolator to prevent feedback into the laser. The laser light is split by a star coupler to provide optical power to each sensor. A portion of the reflected light from each sensor is routed through a directional coupler to a PIN photodiode, which converts the raw optical signal to an electrical signal. The two unprocessed sensor signals are digitized and a microprocessor computes the pressure. The final stage of the signal processor contains a digital-to-analog converter, providing analog pressure vs. time for each sensor. The dynamic pressure response of each sensor is calibrated against a conventional pressure sensor.

Test Setup and Description: To test the ffpi pressure sensor, we chose to monitor a hydraulic performance and net positive suction head (NPSH) test in a pump manufacturer's test facility. In this way, we were able to investigate the effects of low flowrates and actual cavitation. As an added benefit, the owner of the pump asked for a suppression-type NPSH test, which allowed us to assess the effect of falling NPSH on dynamic pressure.

The pump tested was a 6 x 8 x 11 double-suction, overhung process pump rated at 125 hp and 3600 rpm and designed to pump light hydrocarbon liquid in a refinery. The pump was instrumented with an ffpi pressure sensor on the discharge piping and one on the suction piping. At the time, it was not known if dynamic pressure pulsations would be more pronounced on the suction or discharge of the pump. The optical fiber cables from each transducer were then connected to the FFPI signal conditioning unit through optical fiber connectors. The converted electrical signals exiting the signal conditioning unit were connected to a PC for viewing and storage in digital form.

It is important to note that the test pump had a suction specific speed of 11,800 and BEP flow of about 1688 gpm. During the performance test, the pump was operated at flows of 300, 600, 900, 1200, 1600, 1800, and 2000 gpm, while the suction and discharge pressures were recorded. During the NPSH suppression test, the pump flows were held at 300, 1000, 1600, and 1800 gpm, while the effects of NPSH on differential pressure were recorded.

Historically, the onset of cavitation has been detected by a loss of head. The Hydraulic Institute defines NPSH to be the point where a 3% loss of pump differential pressure is observed. In our NPSH test, the flow was held constant while the NPSH was reduced by pulling a vacuum in the test stand suction tank. Once the NPSHA equaled the NPSH required (NPSHR), a drop in discharge pressure was observed. During the 1000 gpm suppression test, for example, the NPSHA was varied from 32 ft down to about 8.4 ft. Since, by testing, the 3% drop in discharge pressure for this flow was found to be about 9.2 ft, we can say the ratio of NPSHA/NPSHR was varied from 3.48 to .91.

Test Results and Conclusions: Several key observations were made from the resulting dynamic pressure waveforms recorded during the pump test. There seemed to be three major categories of pressure pulsations that arose during the test. First, when there was plenty of NPSH available, i.e. NPSHA/NPSHR > 2, there were only rare signs of cavitation. Typically, components of vane pass frequencies and lower were seen. At flows of 50% of BEP flow and less, dynamic pressure pulsation rose to about 30 psi (peak to peak). These types of pulsations are expected during off-design operation due to internal recirculation and inefficient flow distribution in the impeller and cutwaters.

The second category of pressure pulsation seen was of classical cavitation. At all flows, when the NPSHA equaled the NPSHR, high frequency pressure spikes were observed. These pressure spikes were clear signs of vapor bubble implosions in the impeller suction. The magnitude and frequency of occurrence increased at lower values of NPSHA and at lower flows.

The third category of pressure pulsation observed was that of pressure surging. This phenomenon resulted in pulsation frequencies of about 5 Hz and was only detected at flow rates less than 60% of BEP and when the NPSHA equaled the NPSHR (see Fig. 4). At the onset of surging, dynamic pressure amplitudes, at times, exceeded 40 psi (peak to peak). At these lower flows, as the NPSHA fell below the NPSHR, dynamic pressure pulsation became erratic, random, and destructively large (> 80 psi pk-pk), as seen in Fig. 5.

As a result of this testing we can draw the following conclusions:

* Dynamic pressure pulsations increase dramatically whenever a pump is operated at flow significantly below its BEP flow.
* Dynamic pressure pulsations increase dramatically whenever the NPSHA drops to/or below the pump's NPSHR.
* The combination of low flow and NPSHA can lead to excessive and potentially destructive pulsation levels.
* The ffpi pressure sensor has the sensitivity and dynamic response required to sense hydraulic phenomena typically seen during events of cavitation and hydraulic instability.

Monitoring Centrifugal Compressors: Similar to centrifugal pumps, centrifugal compressors are designed to produce a constant discharge pressure for a given flow. The absence of significant pressure pulsations and a predictable differential pressure at a given flow make this class of compressor popular in the chemical process industry. Fig. 6 shows a typical performance map for a centrifugal compressor, where the abscissa is flowrate and the ordinate is discharge pressure. The relationship between discharge pressure and flow is defined by the N curves. N_1, N_2, N_3, and so forth represent the expected compressor performance at different compressor speeds. N_1 represents performance at the highest operating compressor rpm and N_4 represents performance at the slowest operating rpm.

Operation in the region defined by the surge control line and the end of each performance curve will usually ensure safe performance. However, operation to the left of the surge line can result in severe flow surging. Surging is the condition where rapid pulsations in flow occur. At higher discharge pressures, compressors eventually reach a point where they can no longer generate pressure and experience stall. Once the compressor stalls, the compressor flow actually reverses and the discharge pressure drops. (The time for the flow to go from forward flow to reverse flow has been measured at .05 seconds. [6]) As a result of the lower discharge pressure, the compressor is able to reestablish forward flow and does so until the discharge pressure reaches the surge point. This cycle of reverse flow and forward flow repeats itself time and time again until discharge pressure is lowered, suction pressure in raised, or the gas density increases.

The frequency and amplitude of these flow and pressure fluctuations very widely between different classes of centrifugal compressors. Some of the factors affecting these fluctuations are compressor design, i.e. radial flow, axial flow, horsepower rating, and the piping system design. In installations where fluid horsepower is high and the piping and vessel designs are such that fluid energy is readily stored, the potential for destructive flow and pressure pulsations during surging is great.

Probably the most disastrous consequence of surge is its effect on the compressor thrust bearing. Usually, this bearing carries a constant load; but if surge is experienced, the thrust load fluctuates wildly. These rapid and violent changes in thrust load and position have been known to result in thrust bearing failures--and soon after massive compressor rotor damage occurs.

Another consequence of surge dreaded in axial compressors is that of gas reheating. Surging in axial compressors may only occur in a few of the stages. If surge occurs in the later compression stages, the gas is reheated with each surge cycle. This results in a rapid rise in gas temperature and can reach the point where the mechanical properties of the blading degrade--often leading to catastrophic blade failures.

Detecting Surge in Centrifugal Compressors: To avoid surge, it is essential to measure the right process variable. Two of the best process variables to measure are

compressor differential pressure, ΔP_c, and the differential pressure across the gas flow orifice, ΔP_o. Fig. 7 and 8 show typical plots of these variable at the time of a surge event. Notice how quickly these events occur and that ΔP_o represents a much greater change on a percentage basis. Also notice that after the initial flow reversal a lower frequency flow and pressure surge begins. These sample surge plots suggests that monitoring the flow orifice differential pressure is an effective means of monitoring flow stability.

To sense the rapid pressure fluctuation generated by surge events, a high frequency pressure sensor is required. An ffpi differential pressure sensor possesses the ability to sense high frequency (up to 2 khz) pressure changes, while being immune to high process temperatures. This makes it ideal for high compression ratio applications, where high discharge temperatures can be encountered. A typical ffpi sensor installation is shown in Fig. 9. Sensing ΔP_o across the inlet flow orifice, as shown, will allow a user to immediately sense a surge event and act quickly upon this information before compressor damage can occur.

Vibration Monitoring Options: To augment ffpi dynamic pressure sensors for monitoring centrifugal pumps and compressors, users can utilize a full array of ffpi machinery monitoring sensors, which will also be immune to electromagnetic interference and high temperatures. This array includes the Fabry-Perot interferometer based accelerometer, bearing defect sensor, proximity probe, and speed sensor.

Fig. 10 shows a typical monitoring scheme for a centrifugal pumps with antifriction bearings [7]. Notice the dynamic pressure transducer in the pump's discharge line for measuring cavitation and flow instability. ffpi accelerometers, used to detect rotor-related vibration, are mounted on the bearings housings. A speed sensing ffpi probe, capable of measuring speed and sensing the direction of rotation, is mounted near the pump's key way. And finally, an ffpi bearing defect probe is mounted near the pump's thrust bearing to sense bearing damage.

Fig. 11 shows a typical monitoring scheme for a centrifugal compressor with sleeve bearings. First, you will notice the ffpi differential pressure sensor mounted across the inlet flow orifice. The purpose of this sensor is to detect the occurrence of surge. A single ffpi pressure sensor can be used at the compressor discharge or between stages to sense pressure pulsations due to impeller damage or off-design operation. Next, there are ffpi proximity probes near the hydrodynamic bearings for sensing shaft vibration, along with a speed sensor. Finally, there is an ffpi proximity sensor at the thrust bearing to ensure that thrust position is maintained.

The configurations shown in Fig. 10 and 11 represent the most comprehensive monitoring schemes available for turbomachinery. They allow users to accurately assess pump and compressor performance--mechanical as well as hydraulic or aerodynamic.

The Future of FFPI Sensors: Ongoing development of fiber optic sensor technology at FFPI Industries and Texas A&M is directed towards establishing long-term

durability, improving sensor capability, and reducing system cost. Durability tests of in-cylinder pressure sensors at operating temperatures in the 200°C - 300°C range in reciprocating engines are continuing. New sensor designs which can extend operating temperatures to 500°C - 800°C range and improve the sensitivities by 1 to 2 orders of magnitude are being investigated. The replacement of laser diodes (LD) in the present signal conditioning units with less expensive light emitting diodes (LED) shows promise as a cost-saving measure.

The widespread use of fiber optic sensors in industrial monitoring and control should become a reality within the next decade. We can envision networks of tens to hundreds of point sensors connected to computers which process the raw optical signals, store parametric data, and implement feedback algorithms in the control of equipment and processes. Fiber optic sensor networks will eliminate electromagnetic pickup problems so common with conventional electrical sensors, and they will enhance safety by making it possible to physically isolate all electrical cables and electronic component from volatile materials. It is anticipated that technology will become affordable in the years ahead as the application of multiplexing techniques makes it possible to operate an increasing number of sensors and control loops from a single signal conditioning unit.

References:

1. Florjancic, S. and Frei, A., "Dynamic Loading on Pumps--Causes for Vibrations," *Proceedings from the 10th International Pump Users Symposium*, Sponsored by Texas A&M Turbomachinery Laboratory, Houston Texas, 1993.

2. Nelson, W. E., "Pump Curves Can Be Deceptive," *NPRA Refinery and Petrochemical Plant Maintenance Conference*, San Antonio, Texas, January, 1980.

3. Dufour, J., Nelson, W., *Centrifugal Pump Sourcebook*, McGraw-Hill, 1992.

4. Aktins, R. A., et al, "Fiber-optic Pressure Sensors for Internal Combustion Engines," *Journal of Applied Optics* Magazine, March, 1994.

5. Lee, C. E., et al, "Metal-embedded Fiber-optic Fabry-Perot Sensor," Optics Letters, 1991.

6. Staroselsky, N., Ladin, L., "Improved Surge Control for Centrifugal Compressors," *Chemical Engineering* Magazine, May, 1979.

7. Mitchell, J. S., *An Introduction to Machinery Analysis and Monitoring*, Pennwell, Tulsa, Oklahoma, 1981.

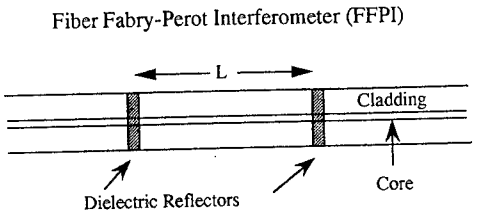


Fig 1 Fiber Fabry-Perot Interferometer

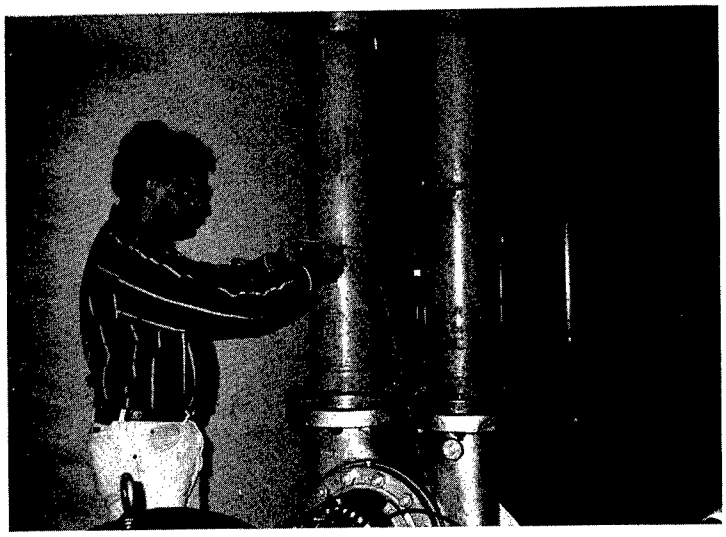


Fig 2 Photo of author standing next to pump

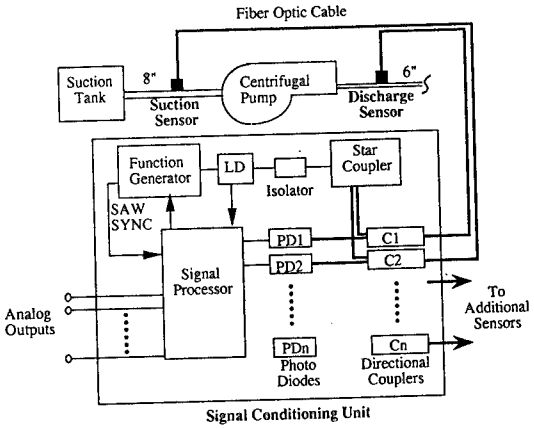


Fig 3 Schematic of pump test arrangement

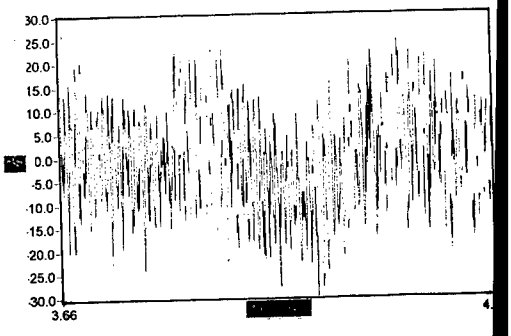


Fig 4 Pressure pulsation at 1000gpm and NRSHA< NRSHR

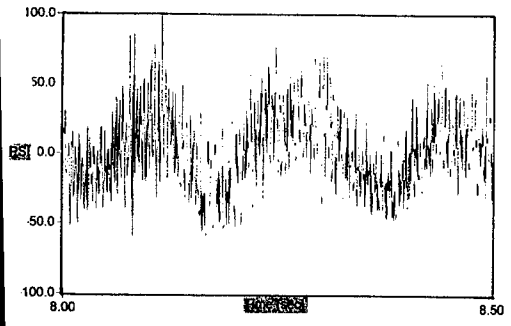


Fig 5 Pressure pulsation at 300gpm and NRSHA<NRSHR

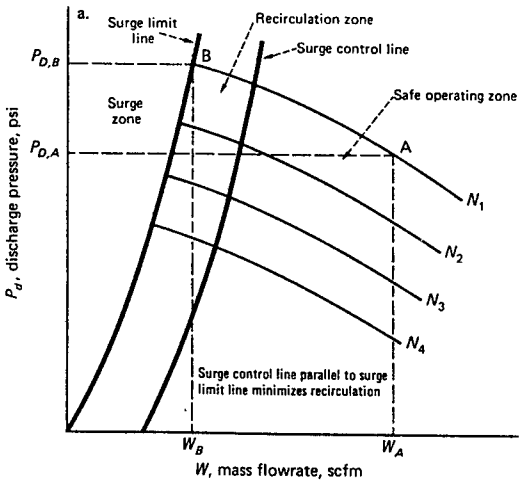


Fig 6 Characteristic curves and surge control lines define regions of operation for compressor

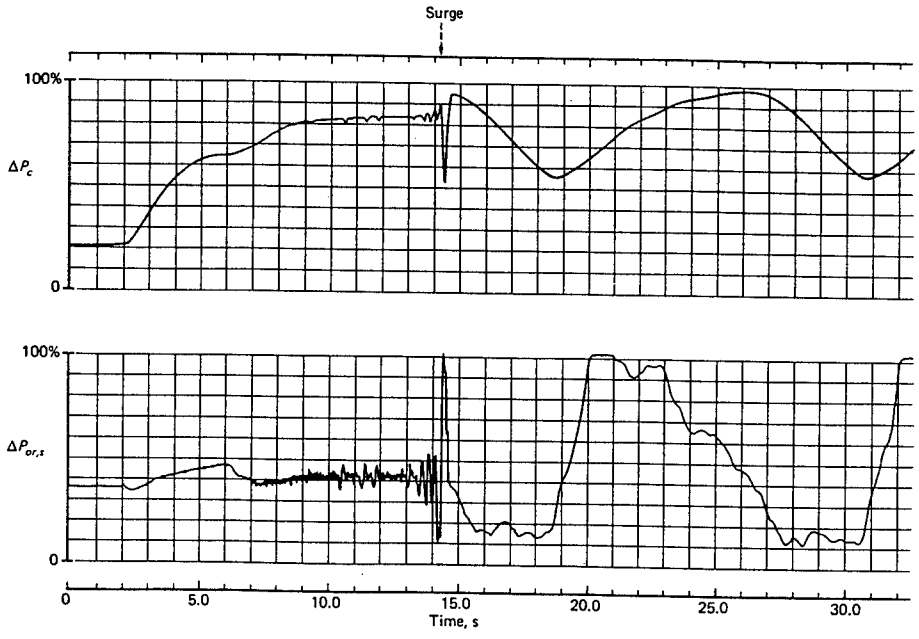


Fig 7,8 Flow drops precipitously before surge cycles begin then reverses quickly

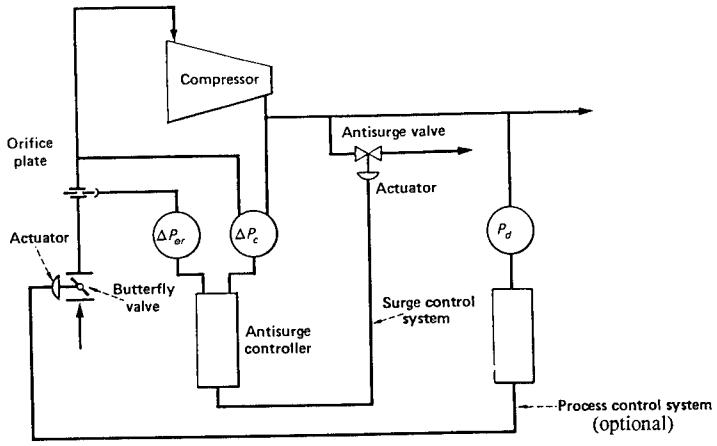
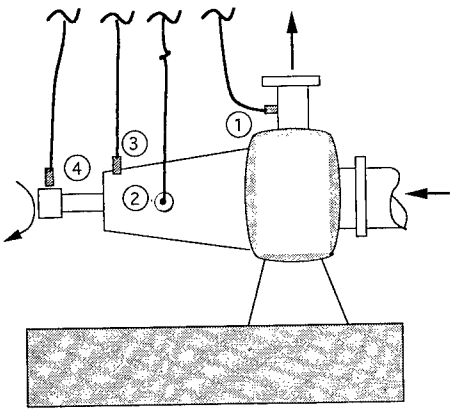
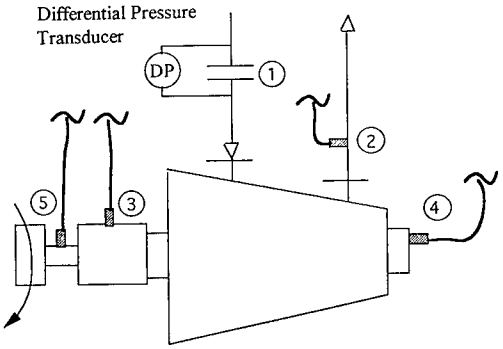


Fig 9 Basic schemes for surge protection



1. Dynamic pressure sensor 2. Accelerometer
3. Bearing defect sensor 4. Speed sensor

Fig 10 Typical pump monitoring arrangement



1. Surge sensor 2. Dynamic pressure sensor
3. Shaft vibration probe 4. Thrust probe
5. Speed sensor

Fig 11 Typical compressor monitoring arrangement

CLASSIFICATION OF COMPOSITE DEFECTS USING THE SIGNATURE CLASSIFICATION DEVELOPMENT SYSTEM

Jeffrey S. Lin
The Johns Hopkins University Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20723-6099

Lawrence M. Brown and Carol A. Lebowitz
Naval Surface Warfare Center
Carderock Division
Annapolis, Maryland 21402-5067

Abstract: The Johns Hopkins University Applied Physics Laboratory and the Carderock Division of the Naval Surface Warfare Center are developing a Signature Classification Development System (SCDS) to transfer classification technology to nondestructive evaluation (NDE) field equipment. SCDS is a personal-computer-based software tool-kit for developing classification algorithms. It includes support for digital signal processing, gating of the signatures, generation of feature vectors, and classification of vectors using artificial neural networks. SCDS successfully classifies ultrasonic signatures from defects in thick section, graphite/epoxy composite test panels. Seven test panels were fabricated with programmed defects embedded one-eighth or halfway into the panel. Six of the panels contain defects representing delaminations, porosity, and contaminations, and one panel serves as a reference standard. Ultrasonic signatures were recorded from the test panels using an ultrasonic C-scan system. SCDS was used to process the signatures and generate feature vectors for input to the artificial neural networks. The classifier achieved a 94% accuracy for one defect, and perfect accuracy for two other defects.

Keywords: Composites, defect classification, artificial neural networks, ultrasonics

INTRODUCTION: In 1966, manual ultrasonics was introduced into the U.S. Navy as an NDE inspection technique for welded structures [1]. Recent studies show that computer-assisted ultrasonics offers several benefits over conventional manual ultrasonics, including: an increased inspection speed [2], a more repeatable and reproducible inspection, less operator dependency, better evidence of weld coverage, the potential for improved consistency of length measurement, and an automatic, hard-copy of the inspection [3]. As a result, considerable effort is now being directed towards developing automated ultrasonic systems that size [4,5] and classify defects, and apply acceptance criteria [6].

Prior to the introduction of artificial neural networks (ANNs), classification algorithms were mainly limited to rule-based and statistical techniques [7]. Current research [8,9] demonstrates that ANNs can out-perform traditional techniques at classifying ultrasonic defect-signatures. Correct classification rates greater than 95% have been achieved using simulated weld-defects in laboratory studies [10,11], and about 90% for real weld-defects [2].

The U.S. Navy recognizes the recent advancements in computer-assisted ultrasonics, and the potential to produce reliable classification algorithms based on ANNs. To transfer this emerging technology to NDE field equipment, the U.S. Navy is developing the Signature Classification Development System. This system is a fully interactive, personal-computer-based package providing software tools necessary to build ANN classifiers for 2-dimensional NDE signatures. Automatic conversion of the developed classification algorithm into source code will permit integration of the classifier into computer-based NDE field equipment, increasing both the reliability and repeatability of inspection results.

Development of SCDS is an on-going effort, and this paper briefly presents the capabilities of the current version of SCDS, Version 1.02, where the targeted system application is ultrasonic-signature classification. SCDS is demonstrated by classifying ultrasonic signatures from thick-section, graphite/epoxy test panels.

SIGNATURE CLASSIFICATION DEVELOPMENT SYSTEM: SCDS is a software tool-kit designed to assist the research engineer in developing algorithms for classifying signatures. The system is fully programmable and graphical in nature, and allows the user to view interactively the processing and classification results on NDE signatures. The current version of SCDS assembles the tools needed to: select signatures for analysis, assign classification categories, perform digital signal processing (DSP), define gates or subsets of the signal, perform feature vector generation, and train artificial neural networks. Future versions of SCDS will include statistical classifiers, and a source-code generator to transition the classification algorithms into run-time applications.

The prototype, SCDS Version 1.02, runs on an IBM® PS/2® or AT-compatible computer operating under Microsoft® Windows™ 3.1 or above. SCDS Version 1.02 contains an Ultrasonic Library for the analysis of ultrasonic signatures, but will support analysis of any signature that is a function of a single variable (e.g. amplitude as a function of time). The main benefits of SCDS are the abilities to process NDE signatures interactively, and to develop ANN classifiers from a single, self-contained software product.

SCDS Version 1.02 has the following capabilities:

selection of data files and signatures,
assignment of classification categories,
random assignment of signatures to training, testing, and validation sets,
selection of DSP functions, templates, and gates,
selection and calculation of signature feature parameters,
batch mode processing of the data and feature generation,

classification of feature vectors via artificial neural networks,
graphical review of classification results.

SCDS User Interface: The SCDS user-interface is developed for Microsoft® Windows™. All displays in the SCDS graphical user-interface are developed using Microsoft® Visual Basic™, an event-driven high-level language for Microsoft® Windows™ that supports dynamic-link libraries [12]. An example window of the user-interface is shown in Figure 1. To increase the speed of processing, some process control and computations are performed in dynamic-link-library routines written in Borland® C++, Version 4.02.

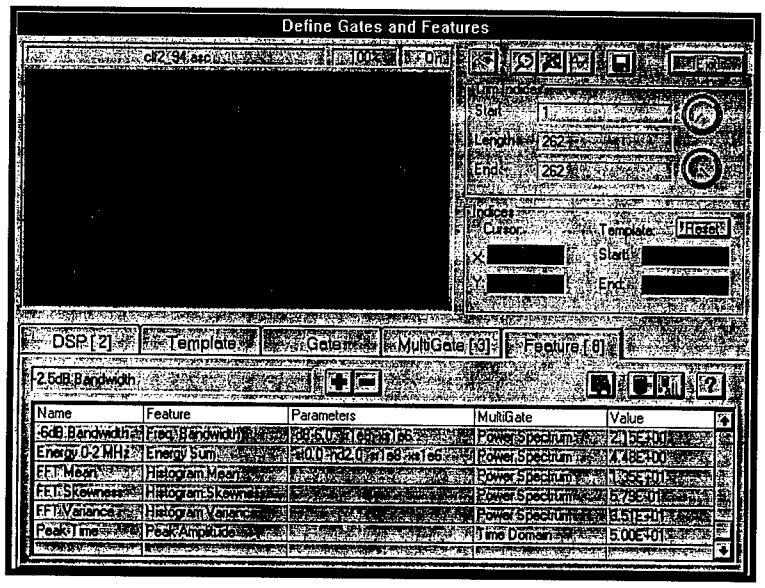


Figure 1. SCDS Define Gates and Features Window

SCDS Data-Processing Module: The SCDS data-processing module assists the user in processing the signatures and generating input feature-vectors for the ANN. The processing module supports digital signal processing, templates to locate discontinuities in the signatures, definition of gates and grouping of gates into multigates, and generation of features based on the gates and multigates. Processing routines are accessed via links to the SCDS functional libraries. On-line help provides a detailed description of each available SCDS library function and optional script-control parameters for the functions.

SCDS extracts features from gated, or sub-selected, portions of the NDE signature. A signature gate can be a fixed gate, with a specific starting and ending index, or can be a variable gate, where templates are used to locate the discontinuities in the signature and define the edges of the gate. Pre-

and post-processing of the signature can extract the discontinuities and transform the signature prior to calculation of the feature parameters. Gates may be defined in different signal domains, for example, time domain, frequency domain, or discrete-wavelet domain. If desired, gates can be combined into multigates so that unwanted portions of the signal can be removed.

The SCDS *Define Gates and Features* window is shown in Figure 1, where an ultrasonic signature (contamination defect recorded from a graphic/epoxy test panel) is displayed. From the *Feature Tab*, the user can select feature routines from the SCDS Library and apply the feature routines to a particular user-defined gate or multigate. In Figure 1, five features are defined for the frequency domain (power spectrum), and one feature is defined for the time domain. By selecting the *Calculator Button* in the *Feature Tab,* SCDS will interactively perform the signal processing and gating, and display the calculated features in a table format.

The SCDS interactive environment temporarily records all signal definitions (data files, class assignments) and data processing routines (DSP, templates, gates, features) in a Microsoft® Access™ database. For permanent storage, this information is recorded in a project file. SCDS uses the project file while executing in a batch, non-interactive, mode to generate feature output files.

SCDS Signature Classification: SCDS Version 1.02 provides a user-friendly, graphical interface for a three-layer, feed-forward, back-propagation neural network. Other architectures and learning algorithms will be added in later versions of SCDS. In addition, SCDS links to the NeuralWorks Professional II/Plus, Version 4.2 for additional ANN learning algorithms [13].

SCDS Source Code Generation: A planned capability of SCDS is to provide the source code that implements the developed classification algorithm. The C source-code will include: 1) signal processing, templates, and gates, 2) generation of parameter features, and 3) the classification algorithm. All processing will be built into a single function call, requiring only the signature as the input and returning the classification of the signature as the output. The C source-code may then be integrated directly into automated NDE equipment in the field or laboratory.

APPLICATION TO COMPOSITE STRUCTURES: The U.S. Navy is increasing the use of fiber-reinforced composites in structures because of the potential for weight, cost, and signature reduction, as well as increased corrosion resistance [14]. Current and future applications of composites include: deckhouses, propulsion shafts, machinery foundations, air flasks, masts, and heat exchangers.

Unlike metallic structures, the effect of defects on the structural integrity of an advanced, composite material is difficult to assess because the inhomogeneity and anisotropy of the composite material does not produce a predominant failure mode. Each defect type results in a degradation of a specific mechanical property. For example, voids affect a wide variety of mechanical properties, but have the greatest effect on the interlaminar shear strength [15]. Olster [16] demonstrated that the interlaminar shear strength would decrease approximately ten percent for each one percent increase in void content. Rhodes [17] showed that a paper inclusion, such as the backing paper from prepreg material, decreased the compressive strength of the graphite/epoxy composite by 25

percent. Gerharz and Schutz [18] investigated the effect of delaminations on composites, and observed that delaminations also reduced the compressive strength of the material, resulting in a buckling failure mode. Lastly, matrix cracking may cause a reduction in shear, compressive and flexural strength [18] as well as initiate delaminations [19].

Because each defect type results in a degradation of a specific mechanical property, classification of the defect type will provide additional information to determine whether the composite material is suitable for the intended structural application. Furthermore, the ability to classify the defect type may reduce production and maintenance cost by permitting defect specific acceptance criteria, thus reducing or eliminating the rejection of benign defect types.

CASE STUDY: Seven, 4" x 4" x 1" graphite epoxy test panels were fabricated from 192 plies of Fiberite HYE2048A1A graphite epoxy prepreg tape in a [0/90]s lay up. Six of the panels were made with programmed defects, as shown in Table I. Two panels were produced containing delaminations, two with porosity, and two with contamination. Each panel was constructed by co-curing two fault-free sub-panels around a middle sub-panel containing the defect. The sub-panels embedded with porosity defects were twenty plies thick and the other defect sub-panels were eight plies thick. The middle sub-panels were placed at one-eighth and one-half of the depth of the test panels. Before insertion into the full-thickness test panel, each sub-panel was inspected using an ultrasonic C-scan system to verify the location and presence of the defects. A reference standard was constructed in a similar manner except that no defects were introduced into the middle sub-panel.

Table I. Graphite/epoxy test block construction

Id	Description	Sub-panel Thickness	Sub-panel Placement	Plies
A2	Reference	8-Ply	1/2 Thickness	92-99
B1	Delamination	8-Ply	1/8 Thickness	24-31
B2	Delamination	8-Ply	1/2 Thickness	92-99
D1	Porosity	20-Ply	1/8 Thickness	24-43
D2	Porosity	20-Ply	1/2 Thickness	87-106
F1	Contamination	8-Ply	1/8 Thickness	24-31
F2	Contamination	8-Ply	1/2 Thickness	92-99

Following lay-up and cure of the full-thickness test panels, a Sonix ultrasonic immersion system was used with a 5 MHz, 0.5-inch-diameter unfocused, immersion transducer to acquire the pulse-echo defect signature data from the seven panels. Before collection of the A-scans, each panel was C-scanned to guide the positioning of the transducer. A one-inch-square area was selected on each test panel as the site where a 10 x 10 grid was used to record A-scans at every 0.1 inch increment. A time gate was used to select the portion of the ultrasonic response containing the defect signature.

A total of 700 A-scans, 100 from each of the seven composite panels, were collected using the Sonix inspection system. The defect signatures were examined visually for signal strength and

validity. Twenty-three signatures collected from panel F1 and ten signatures from panel F2 did not appear to be representative of the defect class and were eliminated from the data set.

SCDS Feature Generation: Brown and DeNale [11] demonstrated that planar and volumetric type defects exhibit different spreads in the power spectrum of the ultrasonic signature. To quantitatively characterize the signature spectral-spread, seven power-spectrum features were selected from the SCDS Ultrasonic Feature Library for input to the ANNs. In addition, one feature from the time domain, the peak signal amplitude, was selected to measure the presence of a defect. The eight selected features are described as follows:

(1) Peak Signal Amplitude, mVolts- peak amplitude of the time-based, ultrasonic signature.
(2) Bandwidth @ -2.5 dB Down, MHz- difference in frequency at the minimum and maximum energy crossing at the -2.5 dB threshold (75% amplitude).
(3) Bandwidth @ -6 dB Down, MHz- difference in frequency at the minimum and maximum energy crossing at the -6.0 dB threshold (50% amplitude).
(4) Bandwidth @ -12 dB Down, MHz- difference in frequency at the minimum and maximum energy crossing at the -12.0 dB threshold (25% amplitude).
(5) Spectral Energy from 0 to 2 MHz- sum of normalized spectral energy from 0 Hz to 2.0 MHz.
(6) Power Spectrum Mean, MHz- mean frequency value of the normalized power spectrum.
(7) Power Spectrum Variance, MHz^2- frequency variance of the normalized power spectrum.
(8) Power Spectrum Skewness, MHz^3- frequency skewness of the normalized power spectrum.

Prior to calculating each power spectrum, a Hanning filter [20] was used to reduce leakage of the spectral energy. In addition, before calculating the features, the energy (amplitude) of the power spectrum was normalized to a peak value of one. SCDS writes the feature vectors to a text file in a format consistent with NeuralWorks Professional II/Plus. The signatures were randomly subdivided by defect class into a training and test set.

Classification Results: Several learning algorithms available in NeuralWorks Professional II/Plus were used in this study: the back-propagation, the radial basis function, and learning vector quantization (LVQ) ANNs.

The radial basis function network (RBFN) can provide better generalization capability than the standard back-propagation ANN when the training data is sparse [13]. In contrast to the back-propagation ANN, which has a sigmoid transfer function for the hidden layer nodes, the RBFN uses a radially symmetric transfer function.

An LVQ network uses a competitive-learning approach to classify the input vectors [13]. This algorithm includes a Kohonen hidden layer and an output layer. The Kohonen layer generally requires more nodes than the hidden layer in back-propagation networks.

Table II lists the ANN training information . Four training trials were performed for each network type. First, the network was trained on 25% of the signatures and tested on the remaining 75%. Second, the network was trained on 50% of the signatures and tested on the remaining 50%. In the third and fourth trials, the training and testing sets were reversed from the first and second trials.

Table II. Neural Network Training Information

Network Type	% Training	% Testing	Hidden Nodes	Training Iterations
Back Propagation	25	75	3	16,107
	50	50	3	28,772
	50	50	3	45,000
	75	25	3	30,000
Radial Basis Function	25	75	8	45,261
	50	50	8	50,748
	50	50	8	30,000
	75	25	8	92,000
Learning Vector Quantization	25	75	13	22,500
	50	50	52	15,030
	50	50	52	14,985
	75	25	40	22,500

All network architectures have three layers, an input layer, hidden layer, and output layer. The input layers have eight nodes, one node for each signature feature extracted by SCDS. Prior to training the network, the input features are linearly scaled between 0 and 1. The output layers have four nodes, one for each of the desired classifications: nominal, delamination, porosity, and contamination. The number of hidden nodes varies with the learning algorithm. Through trial and error, the number of hidden nodes was reduced until the network performance degraded. A hidden layer with fewer nodes is desirable for two main reasons: 1) training and classification speeds increase with fewer hidden nodes, and 2) over-training of the network is reduced, improving the classification accuracy on test and field data.

The networks were trained using the "Save Best" option in NeuralWorks Professional II/Plus. During training, NeuralWorks periodically measures the classification accuracy of the network on the test set, saving the network weights when the classification accuracy improves. Training stops when the network performance does not improve after a specified number of training iterations.

The performance of the networks on the training and test sets is summarized in Tables III and IV, respectively. The three different network types correctly classified the porosity defect case for both the training and test sets. This was true even when only 25% of the signatures were used for training. No cases of a nominal, delamination or contamination signature were misclassified as a porosity defect (individual cases are not shown in tables). Further research is required: 1) to

determine which features the ANNs clearly recognized as characteristic of the porosity defect population, and 2) to determine if the physical significance of these features is consistent with the differences in spectral spread observed in ultrasonic signatures from planar and volumetric weld defects [11].

Table III. Summary of Neural Network Performance on Training Set

Network Type	% Training	% Testing	% Correct Classified Training Set			
			Nominal	Delamination	Porosity	Contamination
Back Propagation	25	75	100.0%	100.0%	100.0%	**85.7%**
	50	50	100.0%	100.0%	100.0%	**85.7%**
	50	50	**98.0%**	100.0%	100.0%	**92.8%**
	75	25	100.0%	100.0%	100.0%	**81.6%**
Radial Basis Function	25	75	**92.0%**	100.0%	100.0%	**81.0%**
	50	50	100.0%	100.0%	100.0%	**77.4%**
	50	50	100.0%	100.0%	100.0%	**69.9%**
	75	25	100.0%	**98.0%**	100.0%	**75.2%**
Learning Vector Quantization	25	75	100.0%	100.0%	100.0%	100.0%
	50	50	100.0%	100.0%	100.0%	**97.6%**
	50	50	100.0%	100.0%	100.0%	**97.6%**
	75	25	100.0%	100.0%	100.0%	**96.0%**

Table IV. Summary of Neural Network Performance on Test Set

Network Type	% Training	% Testing	% Correct Classified Test Set			
			Nominal	Delamination	Porosity	Contamination
Back Propagation	25	75	**61.3%**	**90.7%**	100.0%	**92.0%**
	50	50	100.0%	100.0%	100.0%	**77.1%**
	50	50	**96.0%**	**97.0%**	100.0%	**83.3%**
	75	25	100.0%	100.0%	100.0%	**71.4%**
Radial Basis Function	25	75	**60.0%**	100.0%	100.0%	**80.0%**
	50	50	**98.0%**	100.0%	100.0%	**71.1%**
	50	50	100.0%	100.0%	100.0%	**77.4%**
	75	25	100.0%	100.0%	100.0%	**78.6%**
Learning Vector Quantization	25	75	**89.3%**	100.0%	100.0%	**91.2%**
	50	50	100.0%	100.0%	100.0%	**94.0%**
	50	50	100.0%	100.0%	100.0%	**90.5%**
	75	25	100.0%	**94.0%**	100.0%	**92.9%**

The networks have good success, although not perfect, distinguishing the two planar defects, delamination and contamination. In a few cases, the nominal case was misclassified as a planar defect. The training trials on only 25% of the population perform worse than the other trials, incorrectly classifying many nominal signatures. The variability in this training set does sufficiently represent the variability in the entire population.

The LVQ network performed better than either the back-propagation or radial basis function networks, with a higher classification accuracy achieved for the smaller training sets. The three LVQ networks trained on at least 50% of the data performed comparably, with the best

performance achieved on the first 50-50 trial. For this best case, 94% of the contamination signatures in the test set were correctly classified, 5% were misclassified as delaminations, and the remaining 1% were misclassified as nominal. All other signatures were correctly classified.

The two planar defects, delaminations and contaminations, result in reductions in the compressive strength of the composite. When these defects are considered equivalent, the network correctly classified over 99% of the planar defects.

The remaining 1% of the contamination signatures were misclassified as nominal, which may be indicative of the reference panel construction. Three defect-free sub-panels were co-cured to construct the reference panel. The interfaces of these sub-panels are not perfect, and may be representative of a minor planar defect.

CONCLUSION: SCDS provides a user-friendly environment to develop signature-classification algorithms. With SCDS, an engineer can start with raw data and perform digital signal processing and feature extraction to develop a sophisticated classification algorithm using a variety of neural-network technologies. SCDS performs, or facilitates, all of the data handling and management tasks. The interactive graphical interface allows rapid what-if analyses. SCDS successfully built a classifier that distinguishes ultrasonic signatures from several defect types in thick-section graphite/epoxy composite test panels. Future enhancements to SCDS include: expanded neural-network training and testing capability; user-defined signal processing, templates, and features; statistical classification routines; and automated source-code-generation of the developed classification routine.

REFERENCES:

[1] "Ultrasonic Inspection Procedure & Acceptance Standards for Hull Structure Production & Repair Welds," *NAVSEA 0900-LP-006-3010* (Jan 1966).

[2] Ditchburn, R.J., S.K. Burke, and C.M. Scala, "NDT of Welds: State of the Art," in *Welding Institute of Australia, 42nd National Welding Conference*, Melbourne Australia, Vol. 3, Paper 55 (October 1994).

[3] Lebowitz, C.A., "Evaluation of an Automated Ultrasonic Scanner," in *Review of Progress in Quantitative NDE*, edited by D.O. Thompson and D.E. Chementi, Plenum Press, New York, Vol. 10B, pp. 2045-2052 (1991).

[4] Crutzen, S., P. Jehenson, R.W. Nichols, and N. McDonald, "The Major Results of the PISC II RRT," *Nuclear Engineering and Design*, Vol. 115, No. 1., pp. 7-21 (1989).

[5] DeNale, R. And C. Lebowitz, "Weld Defect Sizing Using Time-of-Flight-Diffraction Ultrasonics," in *38th Defense Conference on Nondestructive Testing*, Kelly AFB, TX (1989).

[6] Brown, L.M. and R. DeNale, "Knowledge-Based NDE System," in *Review of Progress in Quantitative NDE*, edited by D.O. Thompson and D.E. Chementi, Plenum Press, New York, Vol. 14A, pp. 787-794 (1995).

[7] Singh, G.P., J.L. Schmalzel, S.S. Udpa, "The Application of Digital signal Processing and Pattern Reccognition to Ultrasonic and Electromagnetic Nondestuctive Testing and

Evaluation," Nondestructive Testing Information Analysis Center, Southwest Research Institute, San Antonio, TX (February 1991).

[8] Windsor, C.G., F. Anselme, L. Capineri, and J.P. Mason, "Classification of Weld Defects from Ultrasonic Images. A Neural Network Approach," *British Journal of NDT*, Vol. 35, pp. 15-22 (1993).

[9] Brown, L.M., R.W. Newman, R. DeNale, C.A. Lebowitz and F.G. Arcella, "Graphite Epoxy Defect Classification of Ultrasonic Signatures Using Statistical and Neural Network Techniques," in *Review of Progress in Quantitative NDE*, edited by D.O. Thompson and D.E. Chementi, Plenum Press, New York, Vol. 11A, pp. 677-684 (1992).

[10] Lorentz, M. and T.S. Wielina, "Ultrasonic Characterization of Defects in Steel using Multi-SAFT Imaging and Neural Networks ," *NDT&E International*, Vol. 26, No. 3, pp. 127-133 (1993).

[11] Brown, L.M., and R. DeNale, "Classification of Ultrasonic Defect Signatures Using an Artificial Neural Network," in *Review of Progress in Quantitative NDE*, edited by D.O. Thompson and D.E. Chementi, Plenum Press, New York, Vol. 10A, pp. 705-712 (1991).

[12] Microsoft Corporation, *Microsoft™ C/C++ Version, 7.0, Environment and Tools*, Microsoft Cooperation, Redmond, WA (1991).

[13] Neural Ware, Inc., *Neural Computing. A Technology Handbook for Professional II/Plus and Neural Works Explorer*, Neural Ware, Inc., Pittsburgh, PA (1993).

[14] Gagorik, J.E., J.A. Corrodo, and R.W. Kornbau, "An Overview of Composite Developments for Naval Surface Combatants," in *Proceedings of the 36th International SAMPE Symposium*, Corvina, CA, Vol. 36, pp. 1855-1861 (April 15-18, 1991).

[15] Judd, N.C.W., and W. Wright, "Voids and Their Effect on the Mechanical Properties of Composites - An Appraisal," in *Proceedings of the 14th International SAMPE Symposium*, Corvina, CA, Vol. 14, pp. 10-14 (Jan/Feb 1978).

[16] Olster, E.F., "Effects of Voids on Graphite Fiber Reinforced Composite, Technical Report" AVCO Corporation, Lowell, MA (August 1972).

[17] Rhodes, F.E., "The Influence of Manufacturing Defects on the Performance of Carbon Fibre Composites," Institute of Physics One Day Meeting on The Significance of Defects on the Failure of Fibre Composites, London (November 1979).

[18] Gerharz, J.J. and D. Schutz "Literature Research on the Mechanical Properties of Fibre Composite Materials - Analysis of the State of the Art," *Royal Aircraft Establishment*, Translation - 2045, Farnborough, Hants, England (August 1980).

[19] Reifsnider, K.L., E.G. Henneke, W.W. Stinchcomb, "Defect Property Relationships in Composite Materials," Technical Report AFML-TR-76-81, Virginia Polytechnic Institute and State University, Blacksburg, VA (April 1976).

[20] Shiavi, R., *Introduction to Applied Statistical Signal Analysis*, Irwin Publishers, Boston, MA (1991).

MICROELECTROMECHANICAL SYSTEMS (MEMS) OVERVIEW

LIQUID VISCOSITY AND DENSITY MEASUREMENT WITH FLEXURAL-PLATE-WAVE SENSORS

Stuart W. Wenzel[†], Ben J. Costello[†], and Richard M. White[†¥]

[†]Berkeley MicroInstruments, Inc.
1301 South 46th Street, Building 164
Richmond, CA 94804

[¥]Berkeley Sensor & Actuator Center
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley 94720

Abstract: Micromachined flexural-plate-wave (FPW) sensors can detect subtle variations in the density and viscosity of liquids. Liquid density causes a mass-loading effect that lowers the sensor operating frequency. In viscous liquids, the attenuation coefficient of the acoustic waves is proportional to the square-root of the viscosity.

We discuss recent results that explore the density-sensing precision and the viscosity-sensing range of FPW sensors. A micromachined FPW sensor (wavelength 100 μm, plate thickness 6.0 μm, operating frequency in water 5.7 MHz) detected the density of a variety of solvents and aqueous salt solutions with an error less than ±0.003 g/cm^3. Most of this error was likely due to the uncertainty with which we know the true density and sound velocity of each sample.

The same sensor was used to measure the viscosity of aqueous polymer solutions (poly(ethylene glycol) and hydroxyethyl cellulose). The sensor accurately measured the viscosity of solutions with molecular weights up to around 10,000. Polymer solutions with MW > 20,000 appeared water-like to the sensor (i.e., there was little or no attenuation of the acoustic waves).The decrease in apparent viscosity measured with the FPW sensor occurs for polymers with relaxation frequencies greater than or equal to the FPW frequency.

Key Words: Densitometer; density; flexural; oil; sensor; ultrasonic; viscometer; viscosity

Introduction: The density of liquids has traditionally been measured by either glass hydrometers or vibrating mechanical elements; viscosity has been measured with expensive laboratory instruments such as cone-and-plate viscometers, or simple glass apparatus such as capillary viscometers. All require large sample volumes, and non are amenable to continuous, on-line measurement.

Previously we have reported a silicon-based microsensor that can continuously measure the density and viscosity of tiny quantities of liquids [1][2][3]. The microsensor, shown in Fig. 1, utilizes ultrasonic flexural-plate waves (or Lamb waves), one of a family of acoustic modes that have been exploited for measuring liquid properties. Others in the family are shear-mode bulk acoustic waves [4][5][6], shear-horizontal acoustic plate waves [7], Love and Bleustein-Gulyaev surface waves [8], and torsional waves [9]. The flexural plate wave (FPW) differs from these other modes in that the wave

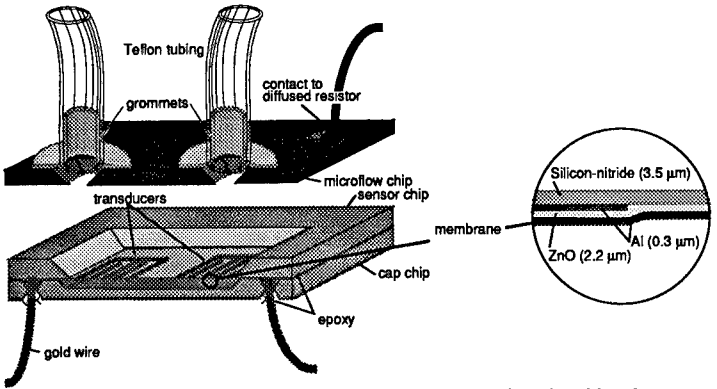


FIGURE 1. FPW density/viscosity sensor. The sensor chip, containing the thin plate or membrane, is attached to a cap chip which protects the metalized side of the sensor. A third chip forms an enclosure that contains liquids on the sensor, and has a diffused resistor for measuring liquid temperature.

motion has a component of motion that is normal to the sensor surface, as depicted in Fig. 2. The interaction between FPWs and an overlying liquid is primarily due to this normal motion rather than a shear interaction at the sensor surface. This makes the FPW sensor much more responsive to liquid density, even in inviscid liquids.

The FPW sensor has been used to measure the viscosity of dimethyl sulfoxide solutions at near-freezing temperatures [10] and to follow the density change associated with the fermentation of glucose by yeast [11] and diffusion in gels [12]. Here we describe the FPW sensor response to viscous fluids and quantify the accuracy with which prototype FPW sensors can measure the density of liquids. We also investigate the relation between the FPW-measured viscosity of polymer solutions and the molecular weight of the polymer.

Theory of operation: Figures 1 and 2 shows the key element of the FPW sensor: a thin plate that supports a traveling flexural wave. Interdigital transducers (IDTs) on either end of the membrane are used to excite and detect flexural waves, whose properties—such as velocity and attenuation—are affected by the measurand of interest. We use the IDTs to measure (1) center or synchronous frequency, which is proportional to wave velocity, and (2) ratio of output to input voltage, which is dependent on wave attenuation.

A simple and accurate approach for modeling sensor response is to solve a thin-plate differential equation of motion for the wave-propagation characteristics. This method yields simple, analytical expressions for the flexural-wave phase velocity in multilayered plates subjected to in-plane tension. It can be extended to include the effects of a viscoelastic medium contacting the plate, yielding simple expressions for the phase velocity and attenuation coefficient when the medium is a viscous or inviscid liquid. Here we summarize a thin-plate solution derived previously; for the full derivation see References [3] and [13].

1. **Free plate:** Using the coordinate system shown in Fig. 2, we derive an approximate one-dimensional equation of motion by balancing the forces acting on a plate segment. Within thin plates, the dominant forces arise from the flexural rigidity D and tension T in the plate, and from

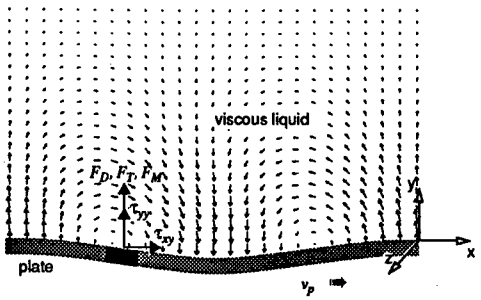


FIGURE 2. Forces acting on a plate segment are balanced to derive an equation of motion for the flexural wave, from which the phase velocity and attenuation coefficient can be calculated (see text). Stresses τ_{yy} and τ_{xy} induced in the viscous medium by the plate motion impede the transverse and rotational motion of the segment, respectively. The arrows represent the displacement or velocity field.

the resistance of the segment mass M to acceleration. We designate these forces (per unit plate area) F_D, F_T and F_M, respectively. Expressions for these terms are found with plate theory, assuming a harmonic traveling-wave with normal displacement

$$u_y = |U|e^{j(\omega t - kx)} = |U|e^{-\alpha x}e^{j(\omega t - \beta x)}. \tag{1}$$

Here, ω is the angular frequency, k is the complex propagation constant, and α and β are the attenuation coefficient and wave number, respectively, both of which are real quantities. Wave number is related to the wavelength λ of the flexural wave by

$$\beta = 2\pi/\lambda. \tag{2}$$

In typical operation, wavelength is fixed by the transducer period P ($\lambda \cong P$) and thus β is constant.

For a freely traveling wave, the forces acting on the plate must sum to zero, which yields:

$$v_p = \frac{\omega}{\beta} \cong \left(\frac{T + \beta^2 D}{M}\right)^{1/2}, \text{ and } \alpha = 0, \tag{3}$$

where v_p is the phase velocity. For a homogeneous plate of thickness d, Young's modulus E, Poisson's ratio ν, in-plane stress σ and mass density ρ, $D = Ed^3/12\left(1-\nu^2\right)$, $T = \sigma d$ and $M = \rho d$.

2. Inviscid liquids: When an inviscid liquid contacts the plate, it impedes the normal plate motion, adding an additional force term to that of the free plate. Summing the plate forces to zero in this case yields several possible modes with different attenuation coefficients. For density sensing, we are interested in the lossless mode, also called a "Scholte" mode, for which $\alpha = 0$:

$$v_p = \frac{\omega}{\beta} \cong \left(\frac{T + \beta^2 D}{M + \rho\delta}\right)^{1/2}, \text{ and } \alpha = 0. \tag{4}$$

Here,

$$\delta = \left[\beta\sqrt{1 - (v_p/c)^2}\right]^{-1} = \lambda/\left[2\pi\sqrt{1 - (v_p/c)^2}\right] \tag{5}$$

is a *real* number representing the decay depth of the *evanescent* displacement field in the fluid (plotted in Fig. 2). Thus, an inviscid liquid causes a mass loading $\rho\delta$ that lowers the phase velocity and frequency. For typical plate-wave structures (plate thickness 2-5 μm and transducer period P=100 μm) we find that $\delta \cong \lambda/2\pi \cong P/2\pi$, and that $\rho\delta$ is of the same order as the plate mass per unit area, M. This results in very large frequency and velocity shifts when liquids are introduced to the plate surface, and is the reason for the extraordinarily high density sensitivity of FPW sensors.

3. **Viscous liquids:** When the liquid has a finite viscosity, the retrograde elliptical particle motion of the plate induces both reactive normal stress (τ_{yy}) and shear stress (τ_{xy}) in the liquid that act on the plate surface. The shear stress τ_{xy} produces a moment that impedes the plate segment's rotary acceleration. Using linear elastodynamics and assuming that: (i) the particle displacement across the interface at $y = 0$ is continuous (the so-called "no-slip" condition); and (ii) the position of the neutral plane of the plate is unaffected by the presence of the viscoelastic medium, the force-balance equation yields:

$$v_p = \frac{\omega}{\beta} \cong \left(\frac{T + \beta^2 D}{M + \rho\delta + M_\eta} \right)^{1/2}, \text{ and } \alpha \cong \left(\frac{\omega}{2}\right)^{3/2} \frac{\delta^2 \beta \sqrt{\eta\rho}}{T + 2\beta^2 D + \omega^2 \rho\delta/\beta^2}, \quad (6)$$

where η is the shear viscosity and $M_\eta = (\beta\delta)^2\sqrt{\eta\rho/2\omega}$ is the effective mass contributed by the liquid viscosity. Assumption (ii) is valid if the magnitude of the viscous shear modulus $\omega\eta$ is small compared to the elastic moduli of the plate materials.

The fluid viscosity therefore adds additional mass loading, and causes wave attenuation. Note that, as we would expect, Eq. (6) reduces to the case of an inviscid fluid (Eq. (4)) when η =0, and to that of a free plate (Eq. (3)) when $\rho = \eta = 0$.

In practice, only ρ and η are unknowns in Eq. (6), and we have to take two measurements—of ω and α—to determine both. We operate the sensor at a constant wavelength $\lambda \cong P$, where P is the transducer period. This constrains $\beta = 2\pi/\lambda$ to a known constant. The plate parameters D, T and M are measured with a simple calibration procedure, discussed later. We also assume that $\sqrt{1 - (v_p/c)^2} \cong 1$, since typically $v_p \ll c$.

Sensor Fabrication: Each FPW delay line consisted of a composite thin-film membrane (Fig. 1, inset). Silicon wafers were coated with 3.5-μm-thick LPCVD low-stress silicon nitride and etched in KOH, leaving membranes and 3 mm by 8 mm in 500-μm-deep wells. Sputtered aluminum was patterned into interdigital transducers, having period P=100 μm, which were then coated with RF-magnetron-sputtered zinc oxide (2.2 μm thick) and a 0.3-μm-thick aluminum ground plane. This sensor had a synchronous frequency of 7.6 MHz when operated in air, and 5.7 MHz when in contact with deionized water on one side.

Since FPW sensors are equally sensitive to the environment on both sides of the membrane, a second encapsulation chip was fabricated to hermetically seal one side of the membrane and to protect the aluminum and zinc oxide from corrosion. The sensor and cap chips were bonded to each other with epoxy

(Epo-tek 302) that was spin-coated onto the cap wafer. Electrical contact was made to the sensor with 100 μm gold wires indium-soldered to the bond pads and then coated with epoxy.

A third chip encased the well-side of the sensor and had two etched holes to which grommets were epoxied. Liquids were introduced to the sensor via Teflon tubing fitted over the grommets. This chip also contained a diffused resistor whose resistance was used to monitor the temperature of the liquid in contact with the sensor. The composite chip was attached to an aluminum carrier that was bolted to a temperature controlled copper block (set point 25 °C).

Density Measurement: The change in the frequency of an FPW delay-line oscillator was recorded as the sensor was exposed to a number of organic solvents. The data were corrected for temperature using the readings from the temperature-measuring resistor, and then fitted to Eq. (4) to extract the parameters M and $B = T + \beta^2 D$. A comparison of the FPW measurements and the published values of the liquid densities appears in Fig. 3A. The discrepancies are plotted in Fig. 3B. Notice that the scatter in the measurements of density of each solvent is much less than the sample-to-sample error.

In order to have this level of agreement we need to know the speed of sound c in the fluid (which enters into the expression for δ). We can circumvent this problem in at least two ways: (1) design the sensor with lower phase velocity so that it is less sensitive to the speed of sound in the liquid, and then assume a constant c; (2) measure the sound velocity with two FPW sensors operating at different wavelengths. For many applications, assuming a constant speed of sound should be adequate. For example, if all the liquids being tested have similar sound speeds or if the sound speed is correlated with liquid densities, then the effect of sound speed can be ignored. This principle has been demonstrated by measuring the density of salt solutions. The densities of the solutions, 0 to 100 g/l NaCl in 25 g/l increments, were calculated assuming a constant speed of sound of 1497 m/s. The error (Fig. 3B) was again only about ± 0.2%.

There are two sources of error in both these examples. There is a small amount (less than 0.001 g/cm^3) of scatter between repeated measurements of the same sample. This seems to be related to the introduction and removal of liquids from the sensor surface. There is also sample-to-sample error that may be caused by the difference between the actual densities of the samples and the literature values that we used in our calculations. Both errors are absent when FPW sensors are used to measure the change in the density of a solution that is in continual contact with the sensor, as would be the case in continuous, in-situ density monitoring.

We have seen this improvement, effectively, when we monitored sodium chloride diffusion through a polyacrylamide gel. The gel, made with deionized water, was polymerized on the sensor surface. The deionized water covering the gel was replaced with a 0.1 M NaCl solution. The sensor frequency dropped as salt diffused through the gel toward the sensor surface. Experiments like this have been used to measure the diffusion coefficients of a number of gels and solutes [12]. The noise level in this experiment corresponds to density perturbations of only 0.0001 g/cm^3.

Viscosity Measurement: We measured the FPW sensor response to different viscous liquids. Frequency of maximum transmission (ω) and corresponding insertion loss were determined by driving one IDT pair with a 25-cycle RF tone burst while sweeping the RF frequency and measuring the voltage of the burst received on the second IDT with a digital oscilloscope. Insertion loss was determined

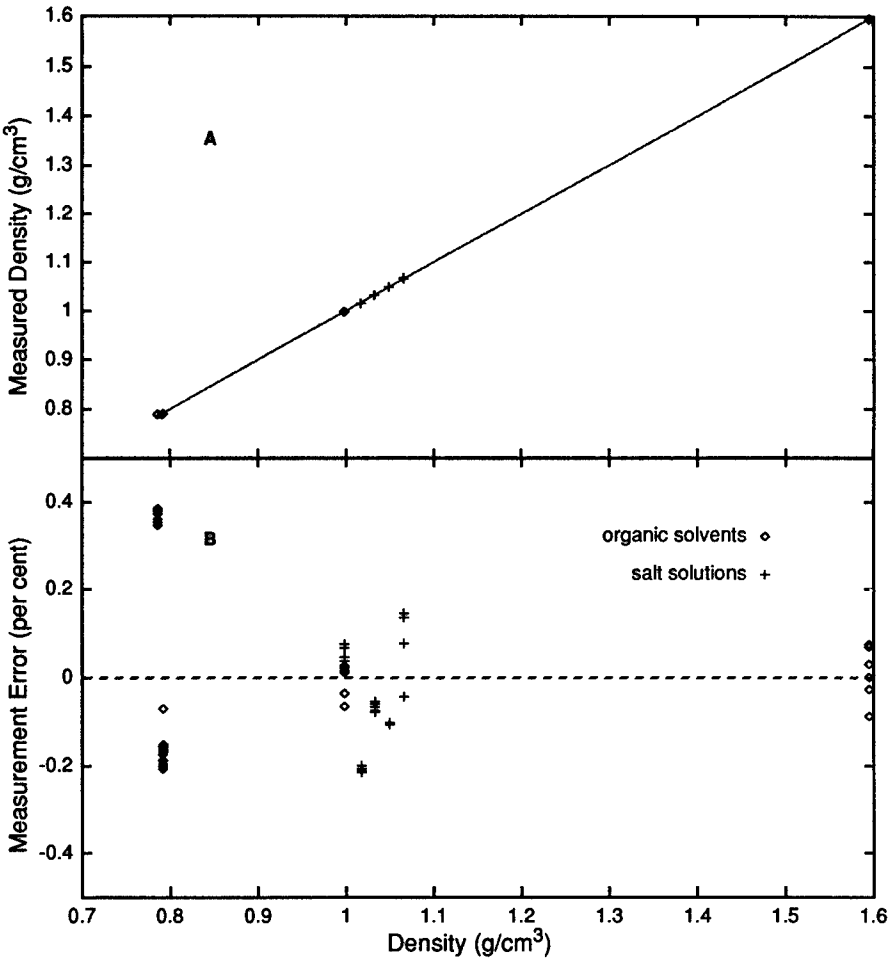


FIGURE 3. FPW-measured density versus literature values for methanol, acetone, deionized water, carbon tetrachloride, and aqueous salt solutions (A). The difference between the measured and literature density values are plotted in (B).

from the IDT RF voltage magnitudes by IL=20 log (v_{out}/v_{in}). Attenuation coefficient was calculated from the change of insertion loss $\Delta IL=20\log(e^{-\alpha L})=-8.686\alpha L$, caused by introduction of a viscous liquid on the sensor, where L=5 mm is the distance between IDTs.

The low-frequency, or low-shear-rate viscosities of all solutions were measured independently with a capillary viscometer. Glycerol-water solutions were used for reference, because of their frequency-independent (Newtonian) viscosity characteristics. The FPW-measured viscosity, determined from measured ω and α and Eq. (6), is plotted against the capillary-viscometer-measured viscosity in Fig. 4.

The data from solutions of poly(ethylene glycol), having average molecular weights 3350 and 15,000, and glycerol-water fall on the same line. The FPW-measured viscosity of higher-molecular-weight polymers is lower than the capillary-viscometer-measured viscosity. This effect is most pronounced with the 98,000 molecular-weight hydroxyethyl cellulose which shows near-water apparent viscosities. We have seen similar results for the FPW-measured viscosity of salmon-sperm DNA solutions.

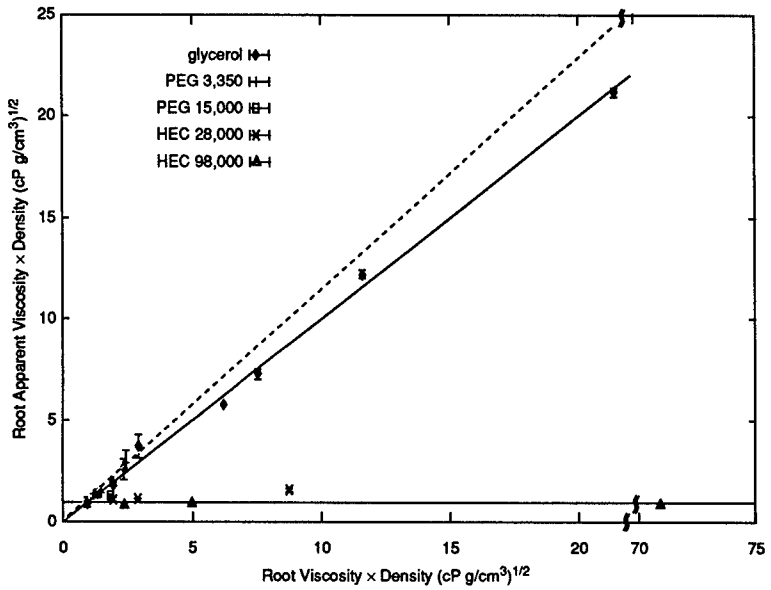


FIGURE 4. Plot of the viscosity-density product of a number of aqueous solutions of the polymers poly(ethylene glycol) (PEG) and hydroxyethyl cellulose (HEC). The response of the FPW sensor (vertical axis) is calibrated using glycerol solutions and is compared to measurements made with a capillary viscometer (horizontal scale). Error bars represent the maximum and minimum values observed. The dotted line represents the attenuation predicted by an approximation of Eq. (6) that assumes ω and δ are constant. This approximation is only valid at lower viscosities; at higher viscosities, ω decreases as viscosity increases, and the line over predicts the FPW-measured viscosity.

Discussion: Our results have demonstrated the sensitivity with which the FPW sensor responds to both density and viscosity of liquids. For low-viscosity liquids (like water) the effect of viscosity on the oscillation frequency is small. For the sensor used in this experiment, a viscosity increase of 0.3 cP would only lower the frequency by 1.3 kHz – the same as a density change of 0.001 g/cm^3. The same viscosity change would increase the transmission loss by 2.4 dB. Higher viscosity liquids increase the FPW attenuation and lower the phase velocity. By observing both quantities, it is possible to calculate both the viscosity and the density of the liquid.

The behavior of the FPW sensor when exposed to high molecular weight solutions is similar to that observed by Reed et al. with a shear-thickness mode quartz resonator [5]. They used a variation on the Maxwell model for the complex shear modulus of a viscoelastic medium to account for the lower than expected viscosity of high molecular weight polymers:

$$\eta = \eta_0 \frac{j\omega}{(1 + j\omega\tau)^{\gamma}}, \tag{7}$$

where τ is a viscoelastic relaxation time characteristic of the polymer and γ represents the distribution of relaxation times. We have shown previously [12] that the FPW sensor is relatively insensitive to the real shear modulus of an overlying solid (such as a gel) so long as the shear modulus is less than 10^6 Pascals. Assuming the attenuation we observed is due entirely to the imaginary part of the shear modulus and $\gamma = 1$ (a Maxwellian fluid), we have calculated a viscoelastic relaxation time of 830 ns for the 24,000 molecular-weight HEC. This is consistent with the relaxation time of solutions of polymers with similar molecular weights [14].

While this aspect of FPW behavior is similar to that of other acoustic sensors, the mechanism of the FPW interaction with liquids is unique. Shear stress at the interface of the sensor and the liquid (τ_{xy} in Fig. 2) has relatively little effect on the velocity or attenuation of the plate waves. This could have the benefit of making the FPW sensor less affected by surface properties (such as roughness or hydrophobicity) than shear-mode acoustic sensors.

Acknowledgments: The authors are indebted to Gregory Mulhern and the laboratory of David Soane for donation of the polymer samples, the use of the capillary viscometers, and insightful discussion on the high-frequency behavior of polymer solutions and melts. This research was supported by the Berkeley Sensor & Actuator Center (BSAC), an NSF/Industry/University Cooperative Research Center, funded by the National Science Foundation and a consortium of industrial corporations and government agencies. The sensors used in these experiments were made in the Berkeley Microfabrication Facility.

References:

[1] R. M. White and S. W. Wenzel, "Fluid loading of a Lamb-wave sensor" *Appl. Phys. Lett.*, vol. 52, No. 20, pp. 1653–1655, May 1988.

[2] B. A. Martin, S. W. Wenzel, and R. M. White, "Viscosity and density sensing with ultrasonic plate waves" *Sensors and Actuators*, vol. A21-A23, pp. 704–708, June, 1989.

[3] B. J. Costello, S. W. Wenzel, and R. M. White, "Density and viscosity sensing with ultrasonic flexural plate waves," presented at The 7th International Conference on Solid-State Sensors and Actuators, Yokohama, Japan, 1993.

[4] H. E. Hager, "Fluid properties evaluation by piezoelectric crystals operating in thickness shear mode" *Chem. Eng. Commun.*, vol. 43, pp. 25–38, 1986.

[5] C. E. Reed, K. K. Kanazawa, and J. H. Kaufman, "Physical description of a viscoelastically loaded AT-cut quartz crystal" *J. Appl. Phys.*, vol. 68, pp. 1993–2001, 1 Sept. 1990.

[6] H. Endo, K. Sode, I. Karube, and H. Muramatsu, "On-line monitoring of the viscosity in the dextran fermentation using piezoelectric quartz crystal" *Biotechnology and Bioengineering*, vol. 36, pp. 636–641, September 1990.

[7] S. J. Martin, A. J. Ricco, T. M. Niemczyk, and G. C. Frye, "Characterization of SH acoustic plate mode liquid sensors" *Sensors and Actuators*, vol. 20, pp. 253–268, 1989.

[8] P. Kielczynski and R. Plowiec, "Determination of the shear impedance of viscoelastic liquids using Love and Bleustein-Gulyaev surface waves" *J. Acoust. Soc. Am.*, vol. 86, pp. 818–827, Aug. 1989.

[9] J. O. Kim, Y. Wang, and H. H. Bau, "The effect of an adjacent viscous fluid on the

transmission of torsional stress waves in a submerged waveguide" *J. Acoust. Soc. Am.*, vol. 89, pp. 1414–1422, March 1991.

[10] T. K. Eto, B. Rubinsky, B. J. Costello, S. W. Wenzel, and R. M. White, "Lamb-wave microsensor measurement of viscosity as a function of temperature of dimethylsulfoxide solutions" in *Proc. of the ASME 28th National Heat Transfer Conference and Exhibition*, HTD-Vol 206-2, (San Diego, CA.), pp. 47–53, August 9-12, 1992.

[11] B. J. Costello, A. W. Wang, and R. M. White, "A flexural plate-wave microbial sensor" in *Technical Digest of the IEEE Solid-State Sensor and Actuator Workshop*, (Hilton Head Island, S.C.), pp. 69–72, June 21-25 1992.

[12] A. W. Wang, B. J. Costello, and R. M. White, "An ultrasonic flexural plate-wave sensor for measurement of diffusion in gels" *Analytical Chemistry*, May 15, 1993.

[13] S. W. Wenzel, *Applications of Ultrasonic Lamb Waves*. Ph.D. thesis, University of California, Berkeley, 1992.

[14] J. Ferguson and Z. Kemblowski, *Applied Fluid Rheology*. New York: Elsevier Science Publishing, 1991.

MEMS TECHNOLOGIES

PLASMA MICROMACHINED MEMS FOR INTELLIGENT DIAGNOSTIC SENSORS

Gregory J. Galvin and Timothy J. Davis
TMS Technologies, Inc.
22 Thornwood Drive
Ithaca, NY 14850-1263

Abstract: Microelectromechanical Systems (MEMS) is a new technology rapidly gaining in commercial acceptance. In MEMS, integrated circuit processing technology is utilized to fabricate silicon microstructures that are released from the underlying substrate such that they are free to move in one or more dimensions. This technology therefore enables the integration of mechanical and electronic components in a single integrated circuit style device.

MEMS is an ideal technology for intelligent diagnostic sensor applications. It provides the capability for high resolution sensors for a variety of physical phenomena (e.g.,, pressure and acceleration) in a small package with integrated electronics. In volume production, MEMS devices can reach low unit costs enabling large arrays of such devices to be used in practical industrial monitoring applications. On-chip electronics increases device intelligence as well as offering the possibility of remote sensing and local decision making.

TMS Technologies' proprietary plasma micromachining technology offers superior performance and reduced manufacturing costs when compared to other MEMS processes in use at the present time. It is hoped that cost effective machine diagnostic sensors can be developed by bringing the capabilities of plasma micromachining to the attention of the failure prevention community through forums such as the *1996 Technology Showcase.*

Key Words: Machine diagnostics; MEMS; micromachining; plasma etching; power generation; sensors; vibration analysis

Introduction: Maintenance of large industrial equipment presents both a technical challenge and a major expense in many industries. Of particular concern in most situations, such as power generation, military systems, or petrochemical plants, is the prevention of unexpected outages. Such equipment failures can be catastrophic in their impact on human life, the environment, and the financial position of the operating company. In the most demanding of applications, e.g., a nuclear power plant, very costly technologies can be employed to ensure continuous monitoring of the health of major equipment. In most other applications, fiscal reality constrains the tools that can be applied to this problem. One most often finds labor intensive preventative maintenance rather than real-time monitoring. Such routine maintenance is deliberately chosen to take place frequently enough to ensure emerging problems are detected and repaired prior to their resulting in equipment failure. Although this can be an effective solution, it is time consuming,

results in excessive equipment downtime, is not fool proof, and provides little information on real-time correlations between operating conditions and failure modes. A far superior solution would be to obtain a cost effective technology for real-time machine diagnostic monitoring and condition based maintenance [1].

This paper is concerned with the use of a new microfabrication technology to produce intelligent sensors for use in machine diagnostic applications. To provide a specific context we will consider the case of vibration analysis of heavy rotating equipment, such as found throughout a power generation plant. This type of equipment exemplifies the maintenance issues being discussed and is prevalent in many industries. However, the diagnostic technology described herein and the issues surrounding its use in industrial applications, are applicable to a much broader scope of machine maintenance situations.

Background: The following specific examples, and the general observations drawn from them, are based on visits by one of the authors to several electric power generation plants. In the case of rotating equipment found throughout such plants, vibration analysis [2] can provide critical information as to the health of the equipment and can be used to predict a variety of failure modes. Currently available systems utilize either case mounted accelerometers or shaft position (gap) measurements to determine relevant vibrational information. Data from these sensors are either provided in real-time or sampled at some regular interval. In most instances, if a certain threshold value of vibration is detected an alarm is indicated and maintenance personnel are brought in to ascertain the nature of the problem and to initiate corrective measures.

Consider the specific case illustrated in Figure 1. This data represents the annual maintenance expenditures, both preventive and corrective, on several medium size exhaust fans in an electric power generation plant. The actual dollar value of the expenditures has not been shown. However, the relative magnitudes do illustrate of the real situation that was observed. During the first seven years, no effort was made to measure vibration on the fans. In year eight, after experiencing severe increases in maintenance expenses, the vibration of the fans was periodically monitored with a portable instrument (sensors were not permanently affixed to the fans). This relatively simple condition based maintenance effort resulted in a five-fold reduction in maintenance expenditures in that year.

Largely as a result of the expense of presently available vibration monitoring equipment, and the cost of retrofitting such sensors to existing rotating equipment, real-time vibration analysis has not attained widespread acceptance. Rather, the approach has been to make periodic inspections and measurements with portable vibration monitors such as in the fan case above. To ensure that the power plant meets its reliability objectives, substantial redundancy was built into the plant's systems and a very large inventory of spare parts is maintained on site. Such an approach effectively meets the objectives but at a very high cost. In the case of power generation this cost is partially masked by the regulatory environment in which the industry is currently operating. For other industries, such capital expenditures would not be a viable approach to meeting equipment availability objectives.

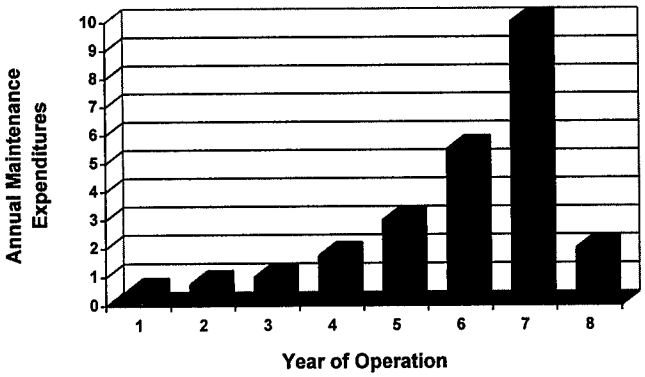


Figure 1. Maintenance Expenses for Exhaust Fans

A natural question therefore is why is vibration monitoring equipment so costly. The answer to which relates to the nature of the sensor industry. This is an industry that is characterized by low to medium product volumes and high levels of product and market specialization. Consequently, it is difficult to attain economies of scale in the sensor industry. A leading manufacturer of piezoelectric sensors with whom TMS has collaborated exemplifies the situation. This company has innovative accelerometer technology well suited to vibration analysis. However, the manufacturing of their products involves low levels of automation, a large amount of manual assembly, and a high degree of custom specialization—all factors that contribute to high prices.

Presently available sensors have very little, if any, local intelligence associated with them. Largely they simply measure a particular parameter and either relay that data in real-time to a control booth or store the data for later interrogation. A final illustration of the potential for improved technology came to the authors from an interaction with a leading supplier of valves and other pipeline equipment for such applications as oil refining and natural gas distribution. In this industry a major recent "innovation" was to incorporate a circuit card with a pressure sensor that allowed the sensor to determine whether or not the valve was operating within normal parameters. This improvement resulted in substantial operating cost savings because it eliminated the need to have personnel make a visual inspection of the pressure sensor on a weekly basis. Given the large distances, and numerous valves, involved in something like a natural gas pipeline, the savings can be substantial.

The above examples, although limited, portray the need for condition based maintenance approaches to reducing operating costs and preventing catastrophic failures. Furthermore, they suggest that industry is aware of the potential solutions and has implemented variations thereon in some cases. Widespread adoption of the technology, however, is being limited by the expense of presently available machine diagnostic tools.

Microelectromechanical Technology: A potential solution to the technical and financial dilemmas presented above can be found in the emerging field of microelectromechanical systems (MEMS) [3]. In MEMS integrated circuit processing technology is utilized to fabricate silicon microstructures that are released from the underlying substrate such that they are free to move in one or more dimensions. This technology therefore enables the integration of mechanical (sensors and actuators) and electronic components in a single integrated circuit style device.

Four techniques can be used to fabricate MEMS devices: bulk micromachining, surface micromachining, LIGA (a German acronym Lithographie, Galvanoformung, Abformung, which translates to lithography, electroforming (plating), and molding), and plasma micromachining. Each technique is briefly summarized in Table I. Although each has been successfully used in research programs and in a few initial commercial products, we believe TMS's plasma micromachining has the most promise for new MEMS products, such as those required in machine diagnostics applications, when two practical constraints are considered.

Table I. Summary of MEMS Fabrication Technologies

MEMS Fabrication Technology	Description	Possible Device Geometries	Mechanical Material	Mechanical Structure Aspect-Ratio	Process Integration with Micro-electronics
Bulk micro-machining	Anisotropic chemical etchant (e.g., KOH) is used to define structures bounded by certain crystal planes. Wafer bonding or ion implantation is often used to form etch stops.	very limited shapes; a function of the crystal planes	single crystal silicon	typically low; special cases can be very high	poor
Surface micro-machining	Polycrystalline silicon is deposited on a sacrificial layer of silicon dioxide. Mechanical structures are formed in the polysilicon and then the sacrificial layer is etched away to release them.	arbitrary shapes but limited in vertical height	polycrystal silicon	low	fair
LIGA	Synchrotron x-rays are used to expose very thick resists that are then electroplated or used as a mold for injection molding.	arbitrary	polymer or metal	high	poor
Plasma micro-machining	A sequence of anisotropic and isotropic reactive ion etches is used to fabricate released mechanical microstructures in single crystal silicon.	arbitrary	single crystal silicon	high	excellent

The first constraint is that of process integration. If MEMS devices of reasonable functionality are to be produced at low cost, then eventually the micromechanical devices and the associated electronic components must be integrated on a common chip with a very high yield in manufacturing (>80%). This constraint effectively precludes both the bulk micromachining and LIGA techniques. Bulk micromachining's wet chemical etchants are largely incompatible with pre-existing electronic devices on the wafer. Following etching of thin mechanical devices, the wafers can no longer reliably survive the physical handling inherent in standard automated electronic device fabrication processes. LIGA being a polymer injection molding process (or an electroplating process) has little potential for monolithic integration with silicon-based microelectronic devices.

Surface micromachining is in principle integrable with standard microelectronics fabrication techniques. In practice, however, this turns out to be very difficult to accomplish, particularly with high yields. The wet chemical etch used to remove the sacrificial layer in surface micromachining presents a number of process incompatibilities. Furthermore, the capillary forces generated during removal of the wet etchant often result in deformation of the thin mechanical structures (a failure mechanism known as stiction). Lastly, reproducible fabrication of the polysilicon layer has been found to be difficult in manufacturing. Evidence to date from those companies attempting volume manufacturing of MEMS devices by surface micromachining suggests that this technique cannot attain the required high yield.

The second constraint is that of device performance. A simple geometric factor, the aspect-ratio (height/width) of micromechanical structures plays a critical role in determining the resulting device's performance. Figure 2 defines the parameters of interest for the following calculations in which the role of the aspect-ratio will become clear.

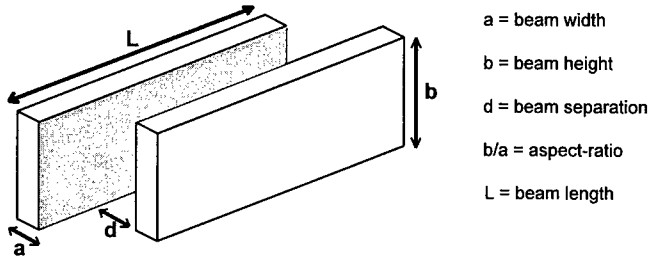


Figure 2. Definition of Key Geometrical Parameters

Many MEMS devices, be they actuators or sensors, involve some form of capacitive device involving pairs of parallel micromechanical beams as conceptually illustrated in Figure 2. If the device is to be used as a sensor, then the capacitance of the structure is the key parameter determining the device's sensitivity. The phenomena to be sensed (force, acceleration, heat,

acoustic energy, pressure, etc.) is coupled into the device in such a manner that one plate of the capacitor moves relative to the other. The change in capacitance is used to transduce the motion and hence provides a measure of the desired external stimulus. The capacitance, per unit area of silicon wafer, is given by:

$$C \approx n\varepsilon \frac{b \cdot L}{d} \cdot \frac{1}{a \cdot L} \approx n\varepsilon \left(\frac{b}{a}\right) \frac{1}{d} \quad (1),$$

where n is the number of plate pairs, and ε is the dielectric constant of the material between the plates, typically air.

If the device is to be an actuator, such as a comb-drive actuator, then the force generated is the key performance factor. The force per unit area of silicon for this type of actuator is given by:

$$F = n\varepsilon \frac{b \cdot V^2}{d} \cdot \frac{1}{a \cdot L} = n\varepsilon \left(\frac{b}{a}\right) \frac{V^2}{d \cdot L} \quad (2),$$

where V is the voltage applied between the beams.

A final important performance factor is the stiffness of the mechanical structures as this determines the resonant frequency and resistance to motion out of the plane. Higher resonant frequencies increase the operating bandwidth of the device, while high out of plane stiffness precludes unwanted mechanical modes. The stiffness of a micromechanical beam loaded at its midpoint is given by:

$$K = \frac{48 \cdot E \cdot I}{L^3} \cdot = \frac{4 \cdot E}{L^3} \cdot a^3 \cdot b \quad (3),$$

where E is the Young's modulus of the material and I is the moment of inertia.

The calculations presented above for sensing capacitance and actuating force were stated as per unit area of silicon. The relevance of this quantity is that the manufacturing yield of a silicon IC chip in conventional microelectronic device processing is essentially inversely proportional to the area of the chip. The larger the chip size, the lower the yield per wafer. Hence, capacitance or force per unit chip area provides a means for comparing surface and plasma micromachining for comparable manufacturing yields.

In surface micromachining, the height of the mechanical structures (b) is determined by the thickness of the polycrystalline silicon film deposited. Residual stress in polysilicon films establishes a practical limit of 2-5μm for their height. Whereas in plasma micromachining, the height of the structures is determined by the depth of the reactive ion etch. In practice these heights can be as much as 20μm. Recent research [4] advances in plasma micromachining suggest the possibility of structures in single crystal silicon as high as 100μm.

The beam width (a) and separation (d) are determined by lithography in either surface or plasma micromachining. Both techniques can attain minimum feature sizes of 2μm and below. Hence, the key distinction between surface and plasma micromachining is the greater beam heights and aspect-ratios attainable in the latter. A factor of ten higher beam height for plasma micromachining (e.g., 20μm versus 2μm) results in sensing capacitance and actuating force per unit area of silicon ten times that available with surface micromachining. Table II provides some representative values. The significant advantages of plasma micromachining are clearly illustrated in the right-hand column of the table.

Table II. Representative MEMS Geometries and Device Parameters for Surface and Plasma Micromachining

	Surface Micromachining (published Analog Devices process[5])	**Plasma Micromachining** (same parameters as previous column other than aspect ratio)	**Plasma Micromachining** (using typical process capabilities)
Beam width (a)	4 μm	4 μm	1.5 μm
Beam height (b)	2 μm	20 μm	20 μm
Beam separation (d)	1.3 μm	1.3 μm	1.0 μm
Beam length (L)	120 μm	120 μm	250 μm
Number of pairs (n)	46	46	100
Actuator voltage (V)	24 V	24 V	24 V
Device capacitance (C)	0.07 pF	0.74 pF	4.4 pF
Device actuating force (F)	0.37 μN	3.6 μN	10.2 μN
Single beam stiffness (K)	56.3 N/m	563 N/m	3.3 N/m

Plasma Micromachining: The foregoing analysis of MEMS device performance and process integration clearly suggests that plasma micromachining offers the best potential for commercially viable MEMS products. This proprietary process [6] largely developed by researchers at Cornell University [7] has been licensed exclusively to TMS Technologies.

In plasma micromachining all of the mechanical structures are fabricated in single-crystal silicon using dry etching (plasma etching) techniques. The process is fully compatible with conventional integrated circuit processing, enabling easier integration of mechanical devices with microelectronics for control and signal processing. Even more significant, the process can fabricate mechanical devices onto a pre-existing integrated circuit wafer without damaging the electronics. This is possible because no high temperature steps, which would ruin existing metal interconnects, or wet chemical etches, are necessary to produce the mechanical structures.

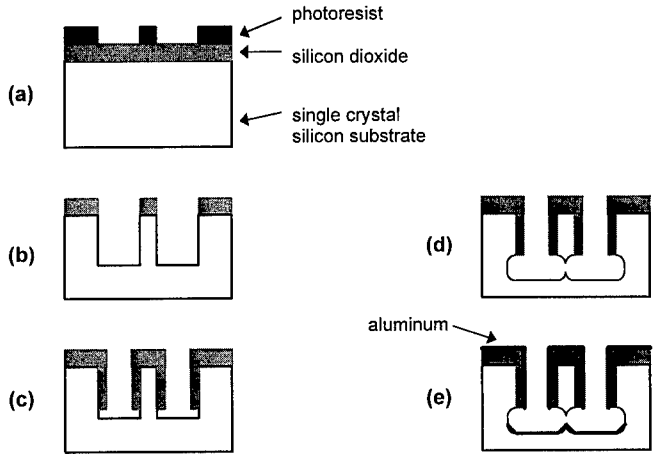


Figure 3. Illustration of Plasma Micromachined Silicon MEMS Process

The basic, single-mask, TMS micromachining process is illustrated in Figure 3. Photoresist is deposited and patterned on top of a layer of silicon dioxide (a). After the pattern is transferred using reactive ion etching, the silicon dioxide serves as a hard mask during a deep silicon trench etch (b). A layer of silicon dioxide is conformally deposited over the entire structure, and unwanted oxide on the floor of the trenches is removed using another reactive ion etch step (c). The silicon mesa, to become a released cantilevered beam, is undercut with a nearly isotropic reactive ion etch. The etch undercuts those areas not protected by silicon dioxide, leaving the beam in the center free to move laterally (d). Finally, a self-aligned metal layer is deposited over the entire structure to provide electrical connection (e).

Figure 4 is an electron micrograph showing some of the mechanical structures that can be fabricated with the plasma micromachining process. Illustrated in this figure are the high aspect ratio beams referred to earlier. The separation of the beams from the substrate is clearly visible.

Conclusions: Structures such as those shown in Figure 4 form the basis for micromechanical accelerometers which can be used in vibration analysis. These devices are capable of measuring acceleration over a wide range (fractions of a g to hundreds of g's) and can incorporate sophisticated electronics for signal conditioning and local information processing. In addition to vibration analysis, MEMS based sensors can be developed to measure a wide range of relevant operating parameters including pressure, temperature, rotation, and the like. From a practical perspective, integrated microelectromechanical sensors offer the opportunity to bring low cost, high volume integrated circuit manufacturing techniques to the sensor industry. This merger of fields can result in cost effective sensor technologies that will catalyze the use of new condition based maintenance techniques in industrial applications.

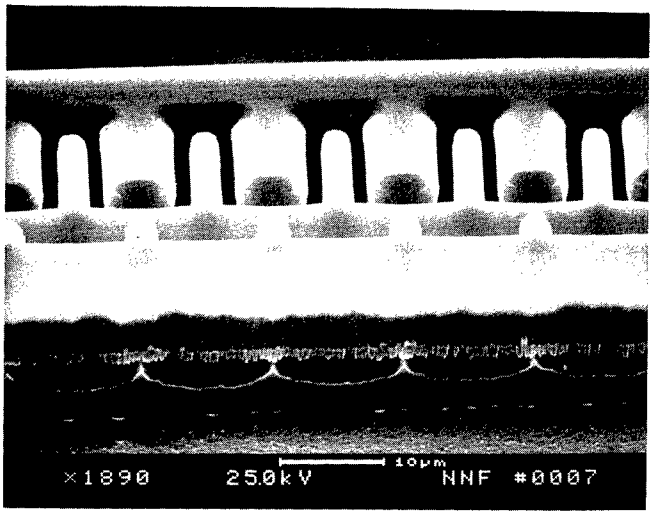


Figure 4. TMS Plasma Micromachined MEMS Structures

The authors would like to acknowledge the generous assistance of Don Wells of New York State Electric and Gas Company and the personnel at the NYSEG Kintigh and Milliken power plants. We would also like to express our appreciation to Ed Birdsall of SynEx, Inc. Fabrication of the MEMS devices shown in this paper was carried out in part at the Cornell Nanofabrication Facility at Cornell University.

References and Bibliography: 1. J.E. Birdsall, "EPRI/Utilities Develop Self-Evolving Failure Prediction System," *M&D Network*, Volume 7, No. 1, EPRI, 1994, pp. 4-6.

2. V. Wowk, *Machinery Vibration—Measurement and Analysis*, (McGraw-Hill, New York, 1991).

3. See for example: Kurt Petersen, "Silicon as a Mechanical Material," *Proc. IEEE*, **70** (1982), pp. 420-457; and Janusz Bryzek, Kurt Petersen, and Wendall McCulley, "Micromachines on the March," *IEEE Spectrum*, May 1994, pp. 20-31.

4. N.C. MacDonald and A. Jazairy, "Single crystal silicon: application to micro-opto-electromechanical devices," *Proc. SPIE*, **2383** (1995), pp. 125-135.

5. Wolfgang Kuehnel and Steven Sherman, "A surface micromachined silicon accelerometer with on-chip detection circuitry," *Sensors and Actuators A* **45** (1994), pp. 7-16.

6. U.S. Patents #5,179,499, #5,198,390, and #5,316,979.

7. Z. Lisa Zhang and Noel C. MacDonald, "A RIE process for submicron, silicon electromechanical structures," *J. Micromech. Microeng.* **2** (1992), pp. 31-38.

MEMS HYDRODYNAMIC BEARINGS: APPLICATIONS AND IMPLICATIONS TO MACHINE-FAILURE PREVENTION

Kristin L. Wood, Dean Neikirk, Ilene Busch-Vishniac, William Weldon, Chin-Seng Chu, Youndmin Kim, Vikas Gupta, William Maddox, and David Masser

Departments of Mechanical and Electrical Engineering
The University of Texas, Austin, TX 78712

Abstract: Microdynamical systems have been studied for a number of years. Only limited work, however, has been completed on integrating microdynamical components into systems that satisfy mechanical tasks on *macroscopic* scales. In this paper, we describe microdynamical components that are needed to produce a surface which is actively deformable on local scales. In particular, we consider the design and demonstration of smart journal and thrust bearings capable of using embedded sensors and actuators to dynamically change the surface geometry. The ability to actively deform bearing surfaces allows for the design of bearings which are less prone to failure, the design of bearings with greater load carrying abilities, and a fundamental study of the effect of surface geometries and fluid conditions on bearing performance, such as start-up and shut-down conditions. Results of our new bearing designs are presented, focusing on numerical bearing models, sensor and actuator design and fabrication, and physical experimentation.

Key Words: Hydrodynamic bearings; new technology; micro-electro-mechanical systems (MEMS); steady-state and dynamic analysis; turbomachinery

INTRODUCTION: Bearings are low-level devices that perform a wide variety of functions within higher-level systems, e.g., motors, engines, pumps, and turbines. These functions, such as support load, guide motion, and attenuate dynamic conditions, are usually critical to the higher-level system's performance and reliability. As a result, unsatisfactory bearing design or poor bearing selection can be devastating to the system's life cycle and maintenance. Despite current efforts to limit these circumstances, increasingly exotic high-level systems are encroaching on the technical limitations of conventional bearing designs (Bhat, 1982). A novel bearing design, implementing micro-electro-mechanical systems technology, is presented to address this need.

Role of Smart Bearing: Smart bearings offer a possible solution to the failure of conventional bearing styles. For the sake of definition, a smart bearing is a system that senses and adapts to instantaneous operating conditions by varying a design parameter (such as clearance in the case of a hydrodynamic foil bearing) or by exploiting one or more physical parameters (such as ferromagnetism in the case of an active magnetic bearing) (Ku, 1990; Siegwart, 1991). In essence, the bearing's design and/or control features possess the capability to sense instantaneous system stimuli and respond accordingly. These capabilities suggest that a smart bearing could control the performance of the overall rotor-bearing system. Since bearing design parameters and performance parameters are inherently coupled, the dynamic and static characteristics of a rotor-bearing system could be actively controlled. In turn, critical mode vibrations could be attenuated if the bearing could sense system conditions and actuate an appropriate bearing clearance. In the next section, we present a new smart bearing concept, designed specifically to address both dynamic and static performance issues.

MEMS SMART BEARING CONCEPT: Geometric design parameters have an effect on the performance of hydrodynamic bearing systems. These variations in geometric form

could be accommodated through macroscopic or microscopic methods of actuation and sensory control. This research effort explores the possibility of implementing microscopic devices in the role of sensors and actuators. The choice of micro over macro is based upon the recent advancements of VLSI manufacturing technology -- which have also lead to the sudden proliferation of integrable microelectromechanical devices. Such an integrated system could reduce or eliminate the need for external actuator and sensory devices. This advantage saves space and required power, while increasing the possibilities of local fluid control.

Studies suggest that hydrodynamic performance can be affected by microscopic variations in the bearing surface profile (Tzeng, 1991; Hargreaves, 1991, Srinivasin, 1995). This knowledge leads to the development of a hydrodynamic slider bearing with an actively deformable surface. An example of this smart bearing concept is illustrated in Figure 1. An actively deformable surface is defined as a surface that undergoes microscopic variations to achieve a desired profile or layout (Maddox, 1994). These surfaces are composed of an array of micron-scale sensors and actuators. A possible layout of such an array is illustrated in Figure 2.

MEMs surfaces can operate on global or local surface characteristics. In global surface control, an overall surface structure is actively prescribed by a central processor. Hypothetically, an array of pressure data could be collected from a slider bearing surface. Upon processing, the pressure profile could be modified by actively deforming the bearing from a flat to a crowned surface. In local surface control, sensing, control, and actuation are accomplished locally in order to produce either a purely local or a global effect. The surface is composed of *cellular automata*, where all necessary operations are carried out within each individual cell or a cell and its nearest neighbors, thus avoiding the difficulties of global interconnections. In such a system, local control could be used to minimize local pressure gradients across a slider bearing surface which, in turn, could globally affect the bearing's instantaneous load capacity.

Because of the potential advantages of local surface control, this research concentrates on creating microscale surface deformations on slider bearing surfaces to produce macroscale performance variations. The ability to actively deform bearing surfaces will allow for the design of bearings which may control one or more of the following performance characteristics: stiffness, damping, load capacity, power loss, cavitation, film thickness, clearance, etc.

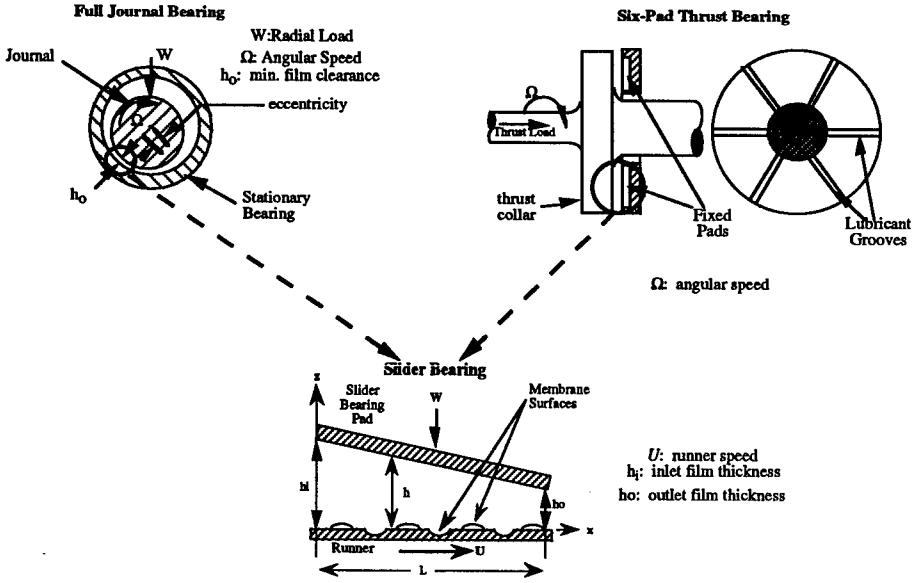
MICROFABRICATION AND MEMS: To create the deformable surfaces discussed above, micromachining techniques must be employed to produce the microdevices which comprise the surface. This section discusses the various surface features and sensors being developed for use in a smart bearing.

Surface Features: Based on the micromachining techniques that presently exist, two types of surface features are being investigated. Membrane surface features are diaphragm-type actuators which deform due to an applied pressure. The other type of surface features are fixed features, which are permanently etched into a silicon wafer.

To create the membrane features, alternating layers of silicon oxide and silicon nitride are deposited onto a silicon wafer using low pressure chemical vapor deposition (LPCVD). Typical individual layer thicknesses range from 0.5 μm to 2 μm. Next, the silicon wafer is back-etched to release rectangular sections of the dielectric films. The film sections released by the back etching process are the membranes which deform due to the applied pressure. To form the fixed features, a mask is deposited on the wafer surface to prevent undesired etching of the silicon wafer. An anisotropic etchant is then used to remove the unwanted silicon, controlling feature depth with careful timing.

Individual membrane and fixed features are combined to form membrane surfaces, fixed surfaces, or hybrid (combined membrane and fixed) surfaces. The overall surface and feature dimensions are constrained based upon the crystallographic orientation of the silicon wafers, the silicon wafer

thickness, the feature type, whether the feature is adjacent to the surface edge, and the type of adjacent features. These constraints are due to limitations in the micromachining processes. For example, if two membrane features are placed adjacent to one another on a (110) silicon wafer, the features must be placed at least 10 microns apart. This constraint is due to the minimum required bonding distance for the silicon wafer.



Note: the regions encircled on the journal and thrust bearings can be approximated by a slider bearing

Figure 1. Side view of a smart slider bearing approximation with a deformable surface.

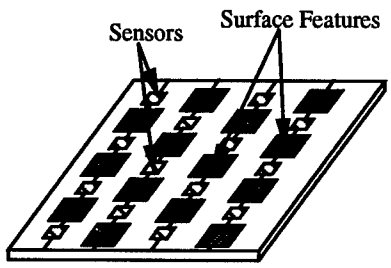


Figure 2. Planar view of an actively deformable surface with embedded sensors.

Sensors — Modeling and Fabrication: Besides the general surface design and fabrication, our work on sensors and actuators for active bearings has focused on a pressure sensor, which uses an optical transduction method, and a proximity sensor, which uses an inductive transduction approach. Each of these sensors is briefly described below.

Fabry-Perot Based Pressure Transducers: Microsensors embedded in a bearing surface should allow localized measurements, for example pressure and temperature, without disturbing the measurand. The local measurements will determine dynamic characteristics of the bearing fluid film and verify a model for bearing performance. The microsensor for pressure measurement is implemented using optical interferometry. Pressure applied to the membrane is measured by detecting the deflection of the membrane. Advantages of optical measurement are as follows: remote data acquisition is achieved without loss of signal to noise ratio, pressure averaging effect reducing sensitivity of piezoresistive pressure sensor is avoided, and dimensions of the device could be much smaller than capacitive pressure sensor.

The Fabry-Perot cavity and optical fiber are used as the sensing element and interconnect, respectively. The cavity is monolithically built by etching a sacrificial layer that lies between dielectric film stacks. The gap of the cavity can be precisely adjusted by controlling the thickness of a sacrificial layer grown using LPCVD. With LPCVD, multiple dielectric films (consisting of silicon dioxide and silicon nitride) can be deposited to form wavelength selective dielectric mirrors. The technique allows for batch fabrication of the pressure sensors with excellent alignment and parallelism of the two mirrors in the cavity.

Fabry-Perot cavity-based sensors have been widely used for their versatility; for example they have been used to sense both pressure and temperature (Halg, 1992; Dakin, 1987; Lee, 1991; Wolthuis, 1991). This kind of sensor detects changes in optical path length induced by either a change in the refractive index or a change in physical length of the cavity. Micromachining techniques make Fabry-Perot sensors more attractive by reducing the size and the cost of the sensing element. Another advantage of the miniature Fabry-Perot sensor is that low coherence light sources, such as light emitting diodes (LEDs), can be used to generate the interferometric signal, since the optical length of the miniature cavity is of the same order as the wavelength of the light.

In these devices, the cavity mirrors can be either dielectric layers or metal layers deposited or evaporated during the manufacturing process. The thickness of each layer must be tightly controlled to achieve the target performance of a sensor. However, there are unavoidable errors in thickness even though techniques of thickness control for thin films have rapidly improved (Macleod, 1986). For Fabry-Perot optical interference filters, it has long been recognized that the performance of the filter is greatly influenced by random thickness variations in the films used (Bousquet, 1972; Macleod, 1976). For instance, the resonant wavelength is very sensitive to thickness variations. We have considered the impact of manufacturing errors on the performance of such sensors. In particular, we have considered how random errors in thickness of the cavity mirrors influence the accuracy with which gap can be measured. We have found that an optimum combination of initial gap and mechanical travel of the cavity exists for a given mirror design which gives the least variation in response curve (Kim, 1994).

Proximity Sensors: One of the most difficult quantities to monitor in a bearing is the actual gap between the bearing pad and the runner. Previous work (Grover, 1929) suggests the use of inductive sensing techniques (eddy current sensors); our current bearing tester uses commercially available proximity sensors mounted externally to the silicon bearing pad. An initial calibration procedure with the runner fixed must be used to locate the reference plane of the sensors with respect to the actual pad location, leading to one of the most significant potential sources of error in the actual bearing gap determination.

To act as the transducer for an *in-situ* proximity sensor, a planar inductor is fabricated directly on the silicon bearing pad. Models developed by Grover (Grover 1929; Grover, 1962) and Greenhouse (1974) are used to calculate the inductance of different planar coil configurations. Experiments to validate these models of a variable gap between the coils and an electrically "floating" plate have been performed. The experiments and model show that there is a decrease in the inductance of the total system (coil and plate) as the gap between the coil and plate decreases.

This is due to the increase in the mutual inductance between the plate and the coil. Different coil configurations show a different amount of absolute inductance change for the same gap. This result indicates that the rectangular spiral may be the best design choice.

A configuration using two coils, forming a planar transformer, has been explored as an alternative to the conventional single-coil sensor. With this arrangement, the gain and the phase of the output changes as the plate gap decreases. Using Grover's (1962) and Greenhouse's (1974) model, values for the inductance of the coils have been calculated. Using these values, a SPICE model for the planar transformer has also been developed. The twin coil design avoids the problem of low Q in planar, small geometry spiral inductors. By measuring a phase difference between input and output coils, the operating frequency remains largely insensitive to the resistance of the transducer, which is critical in "miniaturized" IC inductors.

BEARING MODELS AND EXPERIMENTS: Using the MEMs surfaces and devices introduced above, the goal of this section is to demonstrate performance results of bearing modeling and experimentation. This section first presents the relevant steady state modeling results, followed by experimentation and dynamic modeling.

Steady State Modeling: To determine the performance parameters of a hydrodynamic bearing, fluid flow models based upon hydrodynamic lubrication theory are needed. The relevant hydrodynamic lubrication theory for a slider bearing is the normalized Reynolds' Equation:

$$\frac{\partial}{\partial \bar{x}}\left(\bar{h}^3 \frac{\partial \bar{p}}{\partial \bar{x}}\right)+\left(\frac{L}{B}\right)^2 \frac{\partial}{\partial \bar{y}}\left(\bar{h}^3 \frac{\partial \bar{p}}{\partial \bar{y}}\right)=6\frac{\partial \bar{h}}{\partial \bar{x}} \qquad (1)$$

where L and B is the bearing length and breadth, respectively, and the normalized film thickness equation is:

$$\bar{h}(\bar{x},\bar{y}) = H_x - (H_x - 1)\bar{x} - \bar{w}(\bar{x},\bar{y}) \qquad (2)$$

where H_X is the inlet to outlet height ratio. Reynolds' Equation is then solved numerically using successive over-relaxation (SOR). Due to the converging and diverging geometry of the surface features, the Reynolds' boundary condition is incorporated to form the rupture and reformation boundaries for any predicted cavitation regions (Hargreaves, 1991).

Based upon the lubricant pressure profile determined from Equation 1, performance parameters are calculated to determine the performance ranges of the smart bearing system. Examples of relevant performance parameters are the load capacity, friction force, flow rates.

Using these performance parameters, the operating ranges for the membrane and fixed bearing surfaces may be determined. To facilitate the design of membrane, fixed feature, and hybrid surfaces, surface parameters are developed (Maddox, 1994). The surface design parameters provide an efficient means of conveying the necessary design information of the individual features while describing the overall surface geometry. A major benefit of the surface design parameters is the ability to exchange information in an efficient manner between surface designers, modelers, and manufacturers. Another benefit is the ability to vary individual surface design parameters to determine the individual effects on bearing performance.

The surface design parameters are analyzed using a factorial design (Maddox, 1994). A factorial design is implemented due to its efficiency and ability to determine variable (surface parameter) interactions (Hicks, 1982).. While many design rules and inferences have been made with regard to the steady-state modeling, some of the significant results are given below (Maddox, 1994):

- With the membrane surfaces, the load capacity varies from -30 to 110% from the nominal smooth surface bearing.
- A low number (five) of membrane features are adequate for altering the performance characteristics. In fact, no significant variation occurs by adding additional features.
- Fixed features can alter the load capacity by 20 to 100% from the nominal smooth surface bearing.
- Hybrid features can alter the load capacity form -32 to 53% from the nominal smooth surface bearing.

The modeling results demonstrate that the MEMs surfaces can be used to alter bearing clearances, vary the bearing stability, and correct for wear within the bearing (Maddox, 1994). Thus, the MEMs surfaces should be useful in making a hydrodynamic bearing relatively insensitive to internal parameter variations and external disturbances.

Experimental Apparatus: The basic platform for the test apparatus consists of a Harig model 618 surface grinder. The grinder provides a rotating bearing surface, fluid delivery system and precise three axis translation, a novel approach compared to other *scaled-up* bearing experimental experiments (Hargreaves, 1991). A custom subassembly shown in Figures 3 and 4 is designed to provide rotational orientation and the measurement instrumentation required of the bearing.

The bearing subassembly manifold, instrumentation, and orientation systems have been built and assembled as a unit and are mounted to the grinder. These systems provide two rotational degrees of adjustment; mounting ports for up to four micrometer heads and four eddy-current proximity sensors; sixteen sealed airlines; and force measurement in the thrust and drag directions. The rotational degrees of freedom are used to achieve proper initial orientation of the stationary (non-rotating) bearing surface as well as to help compensate for any deflections due to loading.

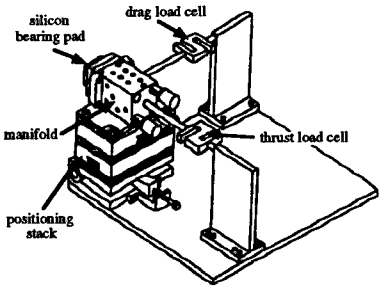


Figure 3 Detailed drawing of thrust bearing subassembly.

Wedge drive screws and wedges are designed to provide control of the apparatus as well as a useful range of motion. To restrict unwanted motion while allowing thrust and drag forces to be transmitted to the load cells, two single-axis crossed roller bearing tables are included in the system. In addition, two load cells anchor the manifold and prevent it from moving and altering the geometry of the bearing gap. Since these cells are the only structures preventing motion in the horizontal plane, they measure the thrust and drag forces acting on the bearing.

Measurement of the bearing gap is a particularly difficult task. The current manifold includes three commercially available eddy current proximity sensors. To calibrate these sensors, a bearing pad is first attached to the manifold, and the whole assembly placed in direct contact with the (non-

rotating) grinder wheel, i.e., the runner. Readings from the three proximity sensors establish the "zero gap" position of the bearing pad/manifold assembly relative to the runner surface. During actual operation, the noncontact eddy current sensors measure the position of the manifold assembly relative to this zero.

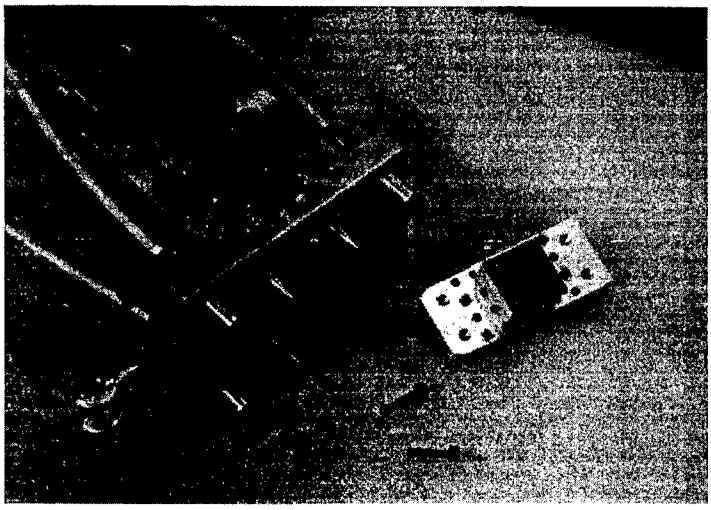


Figure 4: Photo of manifold assembly used to test micromachined bearing pads; bearing pad (fixed features) shown center-right.

Using the experimental apparatus described above, a number of bearing surfaces have been tested, with the objective to match, within experimental uncertainty, the results obtained computationally (Maddox, 1994). For example, Figure 5 shows the experimental tests obtained for a flat silicon surface and a (100) fixed featured surface at a constant angle of attack. The trends shown in Figure 5 closely follow the trends the computational results. These data indicate that loads can be varied significantly for featured bearing surfaces and for different feature depths.

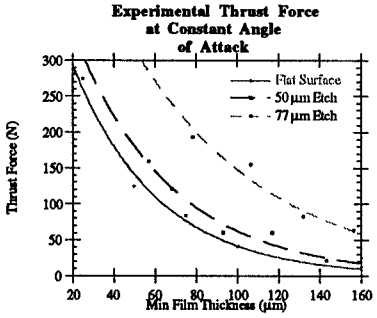


Figure 5: Load versus Film Thickness for Experimental Slider Bearing.

Dynamic Modeling: Steady state slider bearing results in the previous sections show that the presence of MEMs features in the bearing gap greatly affect bearing load, friction, and flow rates. These results also suggest possible dynamic characteristic variations due to the MEMs features (Maddox, 1994). To confirm the dynamic changes within the bearing, two dynamic issues are considered important to the practicality of smart MEMs bearings:

(1) effects of MEMs features on dynamic performance of fluid film bearings; and
(2) dynamic response of MEMs bearings under actual operating conditions.

These issues can be addressed with prior knowledge of the MEMs bearing dynamic characteristics. Consequently, the current modeling effort is focused on developing quasi-static dynamic models to calculate the stiffness and damping coefficients for the MEMs slider bearing. The remainder of this section details these dynamic modeling efforts.

Observations from an error analysis suggest the importance of higher order effects in bearing dynamic characteristics. As a result, a nonlinear dynamic model is developed to predict MEMs bearing dynamic characteristics (Chu, 1995). This model will allow bearing reaction expansion up to the desirable order of accuracy (Chu, 1995).

Analysis of dynamic model suggests that a fourth order model will produce accurate results up to a vertical displacement (Δz) of ± 0.3 times the minimum gap about a quasi-static position. Using this criterion, a fourth order dynamic model is applied to a bearing surface with four membrane features. Within the valid motion ranges ($\Delta z = \pm 0.3$ x minimum gap), comparisons between MEMs bearings and conventional bearings give the results shown in Table 1. Several conclusions may be drawn from these results:

- Within the valid motion ranges, MEMs bearings can increase the load capacity, stiffness, and damping (43 to 89%, 23 to 132%, and 0 to 66% respectively) of a conventional slider bearing with a similar configuration.
- Within the valid motion ranges, MEMs bearings can also lower the load capacity, stiffness, and damping (-18 to -58%, -21 to -60%, and -20 to -28% respectively) of a conventional slider bearing with a similar configuration.
- Actuation of membrane into the oil film especially near the location of maximum pressure will significantly increase bearing stiffness and damping.
- Actuation of membrane out of the gap will lower bearing load, bearing stiffness, and bearing damping.

Actuation Profile	Load Capacity	Stiffness	Damping
All Up	43 - 85% Higher	23 - 132% Higher	23 - 66% Higher
All Down	18 - 23% Lower	21 - 27% Lower	20 - 28% Lower
2 up, 1 flat, 2 down	54 - 58% Lower	50 - 60% Lower	Negligible Change
2 down, 1 flat, 2 up	71 - 89% Higher	55 - 112% Higher	0 - 18% Higher

Table 1. Load capacity, stiffness, and damping variations for four membrane profiles.

The dynamic model discussed in this section shows that an actively-deformable bearing has the potential for fine-level dynamic control. This characteristic of an active MEMs bearing has important implications in practice. In general, the dynamic characteristics of fluid film bearings govern the threshold for the onset of critical speeds and self excited whirl instabilities, rotor stability during operations, and rotor response to transients and external loading (Someya, 1989). Active MEMs bearings will have the ability to influence these dynamic conditions. Other applications might be maintaining clearance and correcting misalignment within the bearing pad.

CONCLUSIONS AND IMPLICATIONS: Mechanical machine components are the building blocks for automation. Perhaps the most prolific class of machine components is bearings. In this paper, a new concept is presented for hydrodynamic bearings based on MEMs technology. Fabrication approaches and results are shown for fixed and membrane surface features (actuators), for Fabry-Perot pressure sensors, and for inductive transducer sensors. These MEMs devices form the necessary components for an actively deformable surface. Modeling results of such surfaces show that bearing load carrying capacity, stiffness, and damping can all be changed by over 100%, compared to a nominal bearing configuration. Experimental results of fixed-feature surfaces support these results. Future work for this research will concentrate on further experimentation, on developing local small-scale power supplies for the sensors and actuators (Koeneman, 1995), and on system integration.

Acknowledgments: The research reported in this document was made possible in part by a grant from the Advanced Research Projects Agency (ARPA), Embedded Microsystems Program, Contract No. DABT63-92-C-0027. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES:

Bhat, B. N. and Dolan, F. J. "Past Performance Analysis of HPOTP Bearings". NASA TM-82470, 1982.

Bousquet, P., A. Fornier, R. Kowalczyk, E. Pelletier, and P. Roche, "Optical filters: monitoring process allowing the auto-correction of thickness errors," *Thin Solid Films*, vol. 13, pp. 285 - 290, 1972.

Chu, C. S., "Dynamic Modeling of a Smart Hydrodynamic Bearing," Master's Thesis, The University of Texas, Dept. of Mechanical Engineering, Austin, TX 78712, 1995.

Dakin, J. P., C. A. Wade, and P. B. Withers, "An optical fiber pressure sensor," *SPIE Fiber optics '87:Fifth International Conference on Fiber optics and Opto-electronics*, vol. 734, pp. 194 - 201, 1987.

Greenhouse, H. M. "Design of Planar Rectangular Microelectronic Inductors," *IEEE Transactions on Parts, Hybrids, and Packaging*, vol. Vol. PHP-10, 1974.

Grover, F. W. "The Calculation of the Inductance of Single-Layer Coils and Spirals Wound with Wire of Large Cross Section," *Proceedings of the Institute of Radio Engineers*, vol. 17, pp. 2053-2063, 1929.

Grover, F. W. *Inductance Calculations, Working Formulas and Tables.* New York: Dover, 1962.

Gu, A. "Process Fluid Foil Bearing Liquid Hydrogen Turbopump". AIAA Document 88-3130, 1988..

Halg, B. "A silicon pressure sensor with a low-cost contactless interferometric optical readout," *Sensors and Actuators A*, vol. 30, pp. 225- 229, 1992.

Hamrock, B. J. *Fundamental of Fluid Film Lubrication.* McGraw-Hill,New York, 1994. Pages 244-252.

Hargreaves, D. J. "Surface Waviness Effects on the Load-Carrying Capacity of Rectangular Slider Bearings". *Wear*, vol. 145, pp. 137-151, 1991.

Heshmat, H. "Smart Bearing Boosts Air Turbine Efficiency". *Design News.* pp. 101-102, 1993.

Hicks, C. R. *Fundamental Concepts in the Design of Experiments.* Holt, Rinehart, and Winston. New York, 3rd ed, 1982.

Kim, Y. and D. P. Neikirk, "Design for Manufacture of Micro Fabry-Perot Cavity-based Sensors," *submitted to Sensors and Actuators A*, 1994.

Koeneman, P., "Conceptual Design of a Micro Power Supply," Master's Thesis, The University of Texas, Dept. of Mechanical Engineering, Austin, TX 78712, 1995.

Ku, C.-P. R. and Heshmat, H. "Compliant Foil Bearing Structural Stiffness Analysis: Part I -- Theoretical Model Including Strip and Variable Bump Foil Geometry". *Journal of Tribology, Transactions of the ASME*, vol. 114, pp. 394-400, 1990.

Lee, C. E. and H. F. Talyor, "Fiber-optic Fabry-Perot Temperature Sensor Using a Low-Coherence Light Source," *Journal of Lightwave Technology*, vol. 9, pp. 129 - 134, 1991.

Lund, J. W. " Review of the Concept of Dynamic Coefficients for Fluid Film Journal Bearings". Journal of Tribology, 109:37-41, 1987.

Macleod, H.A. "Thin film narrow band optical filters," *Thin Solid Films*, vol. 34, pp. 335 - 342, 1976.

Macleod, H. A. *Thin-film optical filters.* New York: McGraw-Hill, 1986.

Maddox, W. E. "Modeling and Design of a Smart Hydrodynamic Bearing with an Actively Deformable Surface". Master's Thesis, The University of Texas at Austin. Austin, Texas, 1994.

Nagaya, K. and S. Takeda. "Active Control Method for passing Through Critical Speeds of Rotating Shafts by Changing Stiffness of the Supports with Use of Memory Metals". Journal of Sound and Vibration, 113(2):307-315, 1987.

Siegwart, R. and Truninger, R. "Electromagnetic Bearings, a Resonable Element in Turbomachinery?". *Rotating Machinery Dynamics, Proceedings of the Third International Symposium on Transport Phenomena and Dynamics of Rotating Machinery.* pp. 253-267, 1991.

Someya, T. *Journal-Bearing Databook.* Springer-Verlag, Germany, 1989. Pages 288-294.

Srinivasan, R. S. and Wood, K. L. "Geometric Tolerancing for Mechanical Design Using Fractal-Based Parameters". *ASME Journal of Mechanical Design*, vol. 117, no. 1, pp. 203-205, 1995.

Tabata, O., K. Shimaoka, and S. Sugiyama, "In Situ Observation and Analysis of wet etching process for Microelectromechanical Systems," *Proc. IEEE Workshop on Microelectromechanical Systems*, 1991, pp. 99-102.

Tzeng, S. T. and Saibel, E. "Surface Roughness Effect on Slider Bearing Lubrication". *ASLE Transactions*, vol. 10, pp. 334-338.

Wolthuis, R.A., G. L. Mitchell, E. Saaski, J. C. Hartl, and M. A. Afromowitz, "Development of medical pressure and temperature sensors employing optical spectrum modulation," *IEEE Trans. on Biomedical Engin.*, vol. 38, pp. 974 - 980, 1991.

*A Distributed, Wireless MEMS Technology for Condition Based Maintenance**

UCLA/Rockwell - Low Power Wireless Integrated Microsensor (LWIM) Team
K. Bult, A. Burstein, D. Chang, M. Dong, M. Fielding, J. Ho, W. J. Kaiser, E. Kruglick, F. Lin, T. H. Lin, H. Marcy*, R. Mukai, P. Nelson, K. S. J. Pister, G. Pottie, H. Sanchez, O. M Stafsudd, K. B. Tan, C. M. Ward, S. Xue, J. Yao*

UCLA Electrical Engineering Los Angeles, CA 90095-1594	*Rockwell Science Center Thousand Oaks California

Abstract: Distributed MEMS networks can revolutionize critical military, industrial, and civil surveillance, transportation, manufacturing, and environmental management systems. Low cost sensor network development, coupled with the high performance of compact computing systems can provide new monitoring and control capability. Embedded microsensors and microactuators may provide control for improved dynamical response of large structures, for reduced requirements on dimensional precision, and for health monitoring and failure prediction of airframes, powerplants, buildings, and other structures. The installation of wireline networks for sensors raises important questions for condition based maintenance (CBM) system cost. For example, the installation of sensor network cables may require major modification to capitol equipment and vehicle systems, particularly for rotating component diagnostics. Thus, the development of autonomous, low power, wireless microsensors offers an opportunity to provide CBM at low cost to a wide range of applications. This presentation will describe fully integrated, wireless MEMS devices implemented with new RF communication and MEMS integration methods.

Key Words: Condition Based Maintenance, MEMS, wireless microsensors, low power electronics, infrared, vibration, and acoustic sensors.

Introduction: New product opportunities and new system capabilities are enabled by the development of a low cost distributed Microelectromechanical Systems (MEMS) technology. The development and deployment of distributed monitoring and controls has been hindered in the past by the requirements of complex installation and communication network requirements. Conventional distributed sensors have required cable interface, and therefore, extensive modification to structures and capitol equipment for their installation. The applications of distributed MEMS are being expanded by a new technology, Low Power Wireless Integrated Microsensors (LWIM). A wireless microsensor network may be distributed rapidly and without modification to large systems. In particular, LWIM nodes may be applied to rotating machinery without the complex slip-ring systems that would normally be required for a sensor electrical interface. In this paper the unique requirements and solutions emerging for wireless microsensor based CBM will be described.

A set of unique requirements exist for distributed wireless microsensor networks. The individual low cost sensor nodes must be 1) reconfigurable by their base station, 2) autonomous to permit local control of operation and power management, 3) self-monitoring for reliability, 3) power

efficient for long term operation. In addition, sensor nodes must incorporate diverse sensor capability with highly capable microelectronics.

LWIM intelligent node technology, based on commercial, low cost CMOS fabrication and bulk micromachining, has demonstrated capability for multiple sensors, electronic interfaces, control, and communication on a single device. LWIM nodes are fabricated by the new CMOS Integrated MicroSystems (CIMS) process. CIMS provides high sensitivity devices for vibration, acoustic signals, infrared radiation and other diverse signal sources. The central challenges for low cost, manufacturable, LWIM devices are the requirements for micropower operation and the complete integration of a CMOS RF transceiver. The CIMS process, micropower measurement, and micropower RF communication systems are described below.

Low Power Wireless Integrated Microsensors: A wireless microsensor technology has been developed that combines new micromachining methods with commercial CMOS to provide high performance, diverse, sensor capability combined with an integrated wireless interface. Low power electronics enables self-powered, autonomous, nodes for mobile applications. Bidirectional communication permits remote programmability. Integrated digital control provides adaptability and scalability. Finally, digital control of each LWIM node enables power and health management. The network architecture described here is a simple star network with a single, powerful base station, supplied by utility power, and numerous distributed wireless microsensors. Network architecture and communication protocols are developed to exploit the asymmetry of distributed sensor communication. Specifically, most information flow is from the sensor nodes to the base station with drastically less flow in the form of commands to the sensor nodes from the base station.

Typical CBM applications may be optimally serviced by sensor networks having local signal processing by sensor nodes. Thus, individual nodes may propagate condition measurements periodically to the base station. Only upon an alarm condition will continuous data transmission be required. This method permits a base station to service a much larger network than would be possible for simple continuous communication with sensor node. Instead, periodic updates of the network base station, by distributed network sensor nodes permits detection of changes in system (machinery) operation. For example, individual sensor nodes may provide continuous *measurement* of a vibration spectrum, while only *transmitting* the observation of a *change* in this spectrum. By exploiting the low duty cycle requirements for sensor communication, large efficiencies may be obtained in sensor node and base station operation. The remote programmability (enabled by including transmit and receive capability at each node) permits high data rate operation at any time that a request is made by the base station.

Completely independent LWIM nodes must operate at micro-ampere current levels and low voltage. This allows long operating life from compact battery systems. Alternatively, for some CBM applications, with nodes mounted directly on a motor or drivetrain shaft, LWIM nodes may receive power by continuous or periodic reception of RF energy from a nearby power source via an inductive coupling. Typical low duty cycle, low data rate (10kbps) and short range (10 - 30m) communication permit 30 μA average current for an LWIM node operating at 3V. A conventional Li coin cell provides this current level for greater than a three-year unattended operating life.

LWIM Node Fabrication Technology: The capability and flexibility required for LWIM intelli-

gent nodes demands that LWIM microsensor systems employ conventional commercial CMOS technology. CMOS technology now provides the embedded control and micropower digital systems needed for LWIM nodes. Challenges remain for low noise, micropower analog measurement and RF communication systems. Demonstrated micropower CMOS measurement systems[1] for MEMS devices will be described below. The first integrated wireless microsensors have been implemented with the CMOS Integrated MicroSensors (CIMS) method. CIMS provides devices ranging from inertial to infrared,[2] and acoustic sensors in the same process (Figures 1, 2 and 3). CIMS combines commercial CMOS (post-processed after foundry-fabrication by XeF_2 micromachining)[3] with high performance bulk micromachined sensor and actuator structures (Figure 2 and 3) by flip chip bonding.

The CIMS process offer several advances over previous techniques. First, by separating the CMOS and bulk micromachining processes, conventional low cost CMOS technology may be directly applied. This offers the system developer great flexibility and the capability to update the circuit technology rapidly as progress continues in this rapidly changing area. In addition, the separation of CMOS and sensor element fabrication permits the introduction of novel materials, for example pyroelectric systems,[2] without disturbing critical CMOS processing.

CIMS also offers performance advantages for CBM sensors. Specifically, the use of bulk micromachined structures enables large proof mass values for inertial and vibration sensors. Single crystal flexures provide stability. Finally, the long standing problem of parasitic capacitance is directly addressed by a unique suspended electrode structure (Figures 1 and 2). The suspended electrode eliminates the need for Si-on-glass for many applications. Also, the proper design of the suspended electrode permits control of squeeze-film damping.

The acceleration and infrared sensors, implemented by the CIMS process, are shown in Figures 1, 2, and 3. Figure 1 displays the CMOS interface die component of the CIMS process. The interface die contains low parasitic capacitance measurement, data conversion and RF communication capability. The CIMS accelerometer relies on a single crystal flexure supporting a proof mass with a resonance frequency that may be adjusted (by design) from 10kHz to 100Hz. The CIMS infrared sensor relies on a new silicon-compatible $PbTiO_3$ sol-gel material system that is applied to the CIMS sensor die in a modular process separated from CMOS fabrication.[2] Direct measurement (Figure 3) yields a responsivity of 4V/Watt for this pyroelectric sensor. Integrated measurement systems, with micropower operational amplifiers and data conversion systems (Figure 4) provide switched-capacitor measurement capability for CIMS devices.

LWIM Systems and Wireless Communication: The wireless sensor technology reported here differs from prior work in that the sensor, control, and communication system are *integrated together in a single unit.* In addition, we have addressed the challenge of combining low noise, high performance measurement and data conversion sensor interfaces with micropower operation in CMOS systems. An important technology challenge is the development of micropower, integrated CMOS RF systems. This requires advances in both level of integration, sharp reduction in operating power, and a conversion from conventional bipolar technology to CMOS.

A new micropower communication system has been demonstrated based on the first CMOS, surface acoustic wave (SAW) stabilized receivers and transmitters (Figure 4). These high selectivity micropower receiver systems have been operated at 320MHz with detector currents of less than 30 μA at 3V supply bias. This device relies on the regenerative receiver principle. However, it

differs from previous systems in that analog sampling of the regenerative wave form (controlled by digital systems) is employed to provide demodulation. Thus, stable operation through automatic gain control is now possible. Current prototypes of this new receiver operate with sampling clock rates as high as 1MHz. Thus, receiver bit rate capability greater than 100kbps is permitted. Prototype LWIM nodes operating with 1mV transmitter power, and 10m range, require single loop antenna of less than $1cm^2$ area.

Summary: The wide range of Condition Based Maintenance (CBM) applications requires sensors carrying diverse measurement capability and network distribution capability. The prohibitive cost of connectors, cables, cable installation, and powerplant, vehicle, and structure modification creates the requirements for wireless sensor networks. A low power integrated wireless microsensor (LWIM) technology has been developed using new processes, circuits, and communication systems. Both RF induction and battery powered nodes may be implemented with the micropower system demonstrated here. LWIM nodes are under development for military and commercial CBM applications.

*** Research supported by the Advanced Research Projects Agency**

References:

1) A. Burstein and W. J. Kaiser, "Mixed Analog-Digital highly sensitive sensor interface circuit for low cost microsensors", Sensors and Actuators (in press)

2) J. Ho, P. R. Nelson, F. Lin, D. Chang, W.J. Kaiser, and O. M. Stafsudd, "Calcium lead titanate pyroelectric detectors on Si MEMS structures". Proc. SPIE (in press).

3) F. Chang, R. Yeh, P. Chu, E. Hoffman, E. Kruglick, K. Pister, M. Hecht, ``Gas-phase silicon micromachining with xenon difluoride", Proc. SPIE Microelectronic Structures and Microelectromechanical Devices for Optical Processing and Multimedia Applications, 117-128 (Oct. 1995).

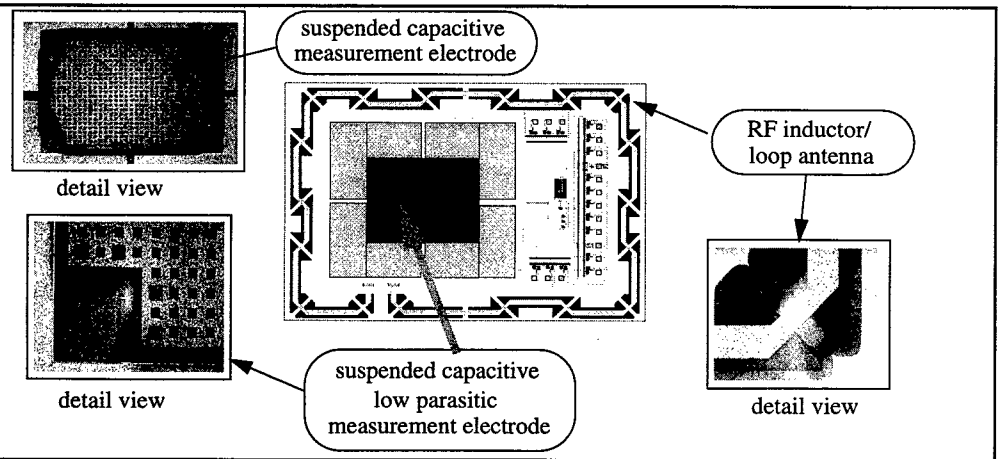


Figure 1. **Wireless Microsensor Interface Die.** The Interface Die provides sensor measurement, sensor interface, control, and RF communication functions in a single integrated component. The Interface Die, fabricated in commercial 2.0μ CMOS, is integrated with high performance sensor elements by flip chip bonding. These micrographs display the low parasitic capacitance, suspended measurement electrodes, and suspended inductor / loop antenna. The CMOS die was processed after fabrication by a mask-less etch in XeF_2.

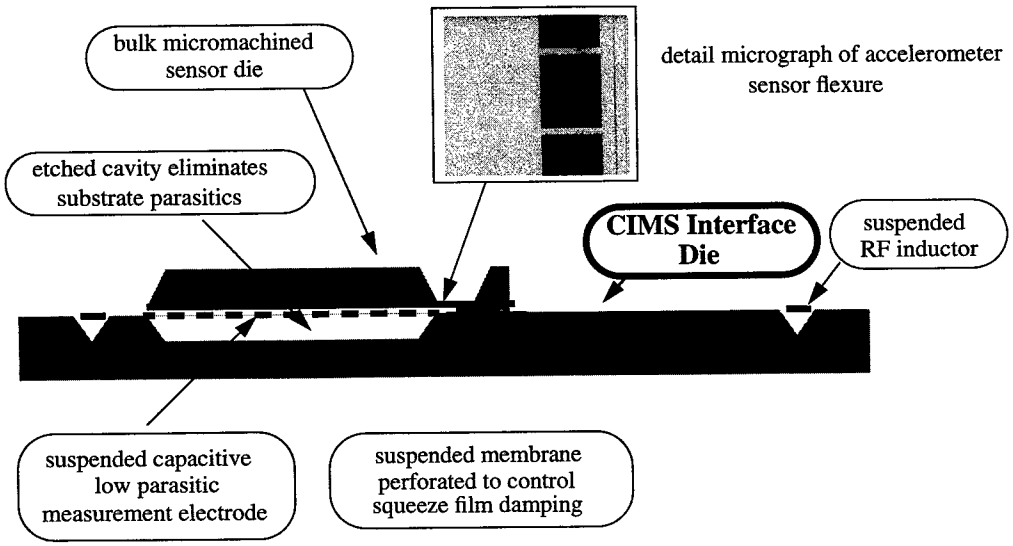


Figure 2. **Wireless Accelerometer Cross-Section.** The accelerometer employs capacitive sensing with a suspended measurement electrode. A micromachined cavity drastically reduces substrate parasitics that limit conventional devices. Membrane perforation provide control of squeeze film damping without requirements to perforate the proof mass.

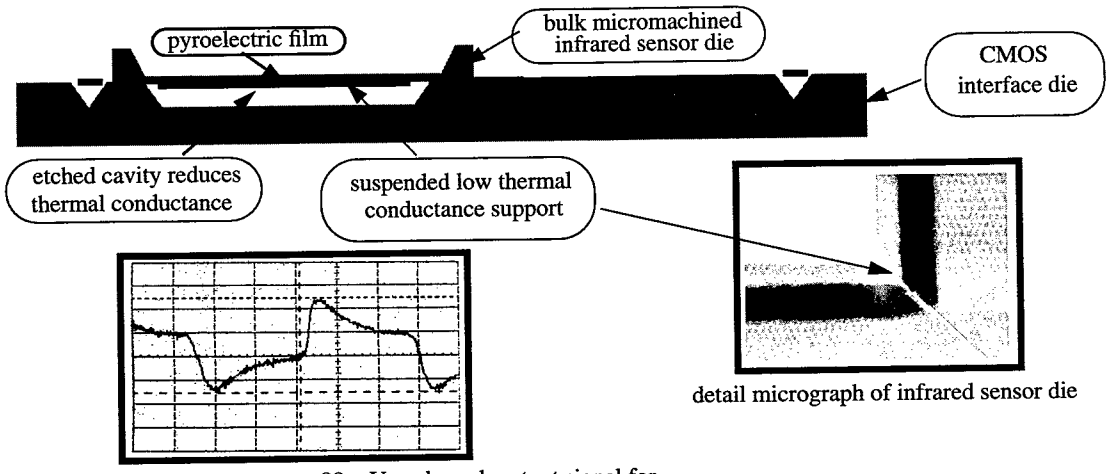


Figure 3. **Wireless Infrared Sensor Schematic Cross-Section.** The wireless infrared sensor employs a low thermal conductance structure supporting a sol-gel thin film pyroelectric infrared sensor. Flip chip bonding of sensor to interface die permits separation of the high temperature, ideal infrared material processing, from commercial CMOS fabrication.

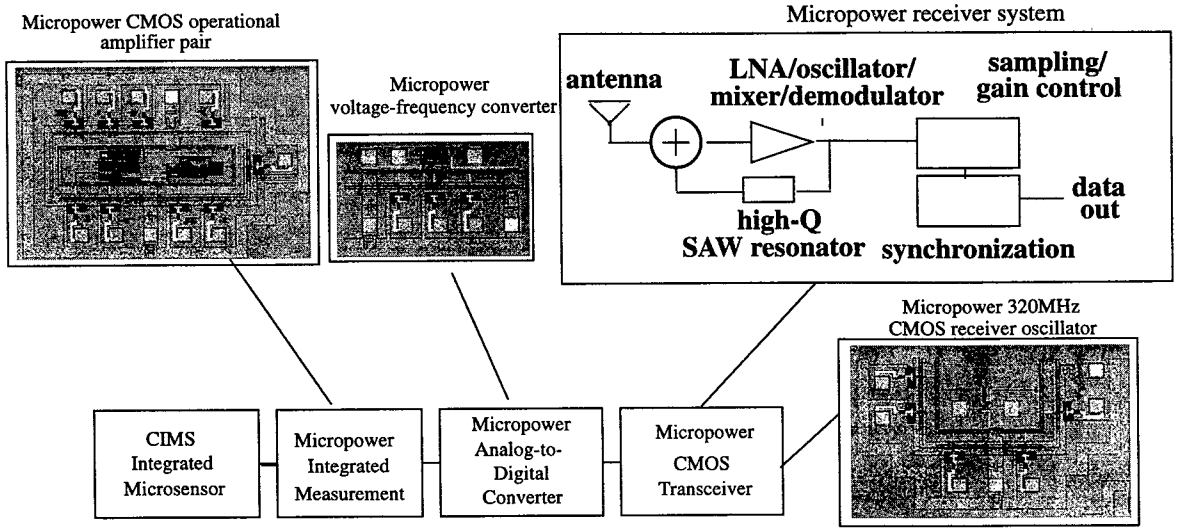


Figure 4. **Wireless Integrated Sensor System.** Measurement, data conversion, and communication systems have been demonstrated in commercial 2μ CMOS. A micropower operational amplifier (supply current 1μA) is combined with micropower voltage-frequency data converters (supply current 2μA). The transceiver is based on a low power CMOS, SAW stabilized and digitally-sampled regenerative architecture (peak supply current 30μA).

MEMS-BASED RESEARCH IN INTEGRATED MONITORING AND ACTUATION AT CASE WESTERN RESERVE UNIVERSITY

Michael A. Huff and Mehran Mehregany
Micro-Fabrication Laboratory
Department of Electrical Engineering and Applied Physics
Case Western Reserve University
Cleveland, OH 44106

Abstract: The paper discusses the development of MEMS at CWRU for the active monitoring and control of systems or structures. The ultimate objective is the realization of high density array architectures of microfabricated sensors and actuators for distributed monitoring and control of a system or its components. Our program addresses the research needs for the advancement of the technology toward realization of low-cost, reliable, integrated sensor/actuator devices and arrays capable of real-time signal processing and control. Our program is also addressing the need for high-temperature MEMS required for use in many aerospace, automotive, and industrial process control applications. Toward this goal, we have research and development activities in high-temperature sensors and actuators made from silicon carbide (SiC), a wide-band-gap semiconductor material capable of functioning in environments where the ambient temperature is well over 250 C, which is the limit of silicon devices. We are also investigating high-phase transformation temperature shape-memory alloys as integrated actuation mechanisms suitable for various high-temperature applications.

Key Words: Microelectromechanical Systems; MEMS; microsensors; microactuators; microsensor arrays; high-temperature microsensors; high-temperature microactuators.

Introduction: As a elemental material, single-crystal silicon has probably been better characterized than any other material known to man, mainly due to the commercial interest of the integrated circuit (IC) industry. The electrical properties of silicon have been exploited to create several large and successful industries. However, silicon is also an excellent mechanical material with a yield strength twice that of stainless steel and a strength-to-weight ratio exceeding that of aluminum. Silicon has a higher modulus of elasticity than steel, yet one-third its density [1]. The excellent mechanical properties of silicon combined with the high sensitivity to external and environmental factors allows it to be used successfully in various microsensor and microactuator applications.

The ability to fabricate micromechanical elements such as microsensors and microactuators from silicon has inherent benefits which make the technology extremely attractive from a manufacturing perspective. Silicon micromachined sensors and actuators can be made using the same techniques of the integrated circuit industry and several important consequences result from this technological leveraging: the ability to batch fabricate; and the existence of the very sizable integrated circuit infrastructure.

Another important benefit of making sensors and actuators from silicon is that the sensor and actuation elements can be readily merged with integrated circuits to form microelectromechanical systems or MEMS. In the most general form, MEMS is the merging of sensors, actuators, and electronics onto the same silicon substrate. The sensors provide information about the environment based on electrical, physical, chemical, or

biological measurements; the electronics process the information derived from the sensors and provide a decision making capability for the system based on this processed information; and the actuators respond to control signals from the electronics and manipulate the system or environment for a desired outcome or purpose. While integrated circuit technology has brought unprecedented computational power ever closer to the point of use, revolutionizing the design of electronics products and enabling the creation of entirely new product categories, MEMS promises to do the same for electromechanical systems through miniaturization, batch fabrication, and integration with electronics, thereby enabling the development of smart systems by providing the required interface between the available computational power and the physical world through the perception and control capabilities of microsensors and microactuators.

MEMS technology is expected to have enormous opportunities in the commercial markets due to the low-cost, high functionality, and small size/weight of the devices. MEMS technology allows much more functionality to be placed within a given space than conventional technologies. Alternatively, sophisticated functionality can be placed where it never could be placed before. For the last two decades, researchers at various university, government, and industrial labs having been working on silicon micromachines and silicon micromachining technology mostly in the form of discrete sensors or actuators. The fruits of this research are evident by a rapidly growing and vibrant industry which currently has annual components sales estimated to be in excess of one billion dollars and several tens of billions of dollars in value-added products made possible by the technology. MEMS devices are emerging as product differentiators in markets such as automotive, aerospace, industrial process control, electronics instrumentation, office equipment, appliances, and telecommunications. Research is now focused on merging the sensors, actuators and electronics onto silicon substrates to realize fully integrated smart systems. As research in this area progresses, the world market for MEMS is expected to reach fourteen billion dollars by the year 2000 with an additional one-hundred billion dollars in high-value added product sales [2].

At CWRU, we are addressing the development of MEMS, embedded into structures or mounted onto structural surfaces, for the active monitoring and control of static and transient phenomenon. Our program addresses the need for the advancement of technology toward realization of low-cost, reliable, high-density, integrated sensor/actuator devices and arrays capable of real-time signal processing and control, while also addressing the need for high-temperature MEMS required for use in many aerospace, automotive, and industrial process control applications. This paper will review current research program at CWRU in these topic areas.

Embedded Tire Sensors: Devices enabling improved monitoring of tire air pressure are of interest to trucking and insurance companies since the maintenance of correct air pressures results in increased fuel economy, improved tire wear, and a higher level of highway safety. The preferred method of employment is to mold the pressure sensing device and associated readout electronics into the truck tire during manufacture and then remotely interrogate the device using a mobile or fixed telemetry unit. A particularly important requirement of the sensor is its stability over extended periods of time, since recalibration of the device after insertion into the tire is difficult. Another difficult requirement is the ability of the sensor and electronics to withstand the harsh pressures and temperatures of the tire molding process.

The embedded MEMS tire air pressure sensor program originated out of research at CWRU on devices for highly-stable, long-term continuous measurement of intra-cranial pressure in hydrocephalus patients by Ko [3]. Leveraging this technology, CWRU researchers in joint

cooperation with Goodyear Tire and Rubber, have developed a smart tire pressure sensor [4]. The embedded MEMS tire air pressure sensor is a low-cost, easy-to-use, and fast method for the continuous measurement of truck tire air pressures. The sensor is embedded in the tire sidewall during manufacture and remotely interrogated using either a handheld or fixed telemetry unit. A capacitive-based sensor is employed for the pressure measurements in order to take advantage of the higher sensitivity, higher stability, and larger dynamic range capacitive pressure sensors exhibit over competing technologies. The sensor consists of a bulk micromachined deformable thin silicon diaphragm which acts as one electrode of the capacitor (Figure 1). The silicon is anodically bonded to a glass substrate having a metal electrode on the surface, thereby forming the other capacitor plate. As the deformable thin silicon membrane deflects under pressure loading, the membrane deflection results in a measurable change in the capacitance. The change in capacitance is converted to an output voltage over a dynamic range of 0 to 80 psig. The extremely small size of these devices (e.g., 2mm x 2mm x 1mm) allows them to be embedded into a tire during manufacturing to provide life-time tire pressure monitoring capabilities.

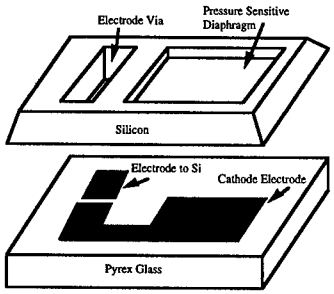


Figure 1. Illustration of the silicon micromachined CWRU embedded truck tire pressure sensor [4].

Microfabricated Shear Stress Sensor: The measurement of shear stresses is of primary importance in the design of aerodynamic and hydrodynamic components since it determines a major part of the drag and has significant influence on many of the performance characteristics. Further, as progress advances on aircraft employing non-fixed airfoils for active wing positioning, continuous measurement of shear stress may become desirable. Several indirect methods for the measurement of wall shear stress have been developed including, the Preston tube, hot wire/films, and liquid surface techniques. However, no universal method for the accurate and direct measurement of shear stress has yet been realized.

A microfabricated sensor enabling the direct measurement of wall shear stress has been developed at CWRU [5]. Both active and passive designs have been demonstrated, and integration of the sensors with on-chip electronics is pursued. Furthermore, in a collaboration between CWRU and Analog devices, Inc., Waltham, MA, integrated shear stress sensors have been produced which combine surface micromachining with a 3 μm BiCMOS process. These sensors have the transducer element and electronics for signal amplification and conditioning integrated on the same chip. The CWRU device employs a surface micromachined polysilicon floating element, of dimensions 100μm x 100μm, suspended by 100μm-long polysilicon tethers separated from the substrate surface by a micron-sized air gap (Figure 2). The floating element displaces laterally when loaded by the force of a flowing fluid, and this displacement is measured by the change in capacitance between the element and interdigitated sensing electrodes. In the passive device, the

floating element displaces in response to the fluid flow, and this displacement is detected by a differential capacitive transducer. In the active device, an electrostatic drive is used to restore the floating element to its null displacement position, thereby improving the sensor performance. Microfabrication of the device in silicon enables an fully integrated active sensor to be realized having capability for self-testing and self-calibration, and thereby improving long-term reliability and stability.

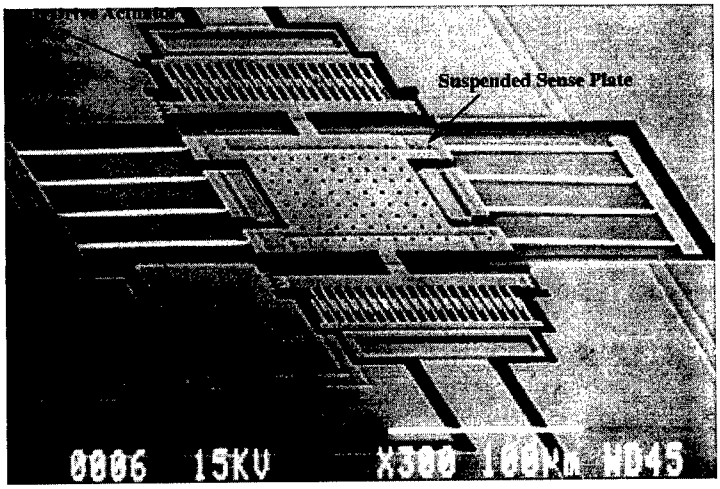


Figure 2. SEM photo of the transducer element of the CWRU shear stress microsensor.

The graph in Figure 3 shows test results for a sample of shear stress sensors. Sensors with 100µm and 120µm folded beams were tested, illustrating that the sensitivity of the sensor can be tailored by varying the stiffness of the suspended support structure. The suspended plate with the longer support beams is more compliant, exhibiting a higher sensitivity to shear stress. Separate data points indicate test results of different sensors, showing that a similar response is achieved from one sensor to the next. Testing of the integrated CWRU/Analog Device sensors demonstrated an electrical sensitivity of 6 Pa/Volt. The first resonant frequencies of the transducer elements was near 10 kHz.

Smart Microvalves: The ultimate goal of the miniature valve project is to develop "intelligent" or "smart" microvalves and microvalve arrays. Our work on smart microvalves and microvalve arrays is focused on two application areas based on the interest of our commercial partners, combustion-engine fuel delivery systems, particularly gas-turbine engines, and closed-loop control drug delivery systems. However, it is expected that the results of this work will positively impact many other application areas as well.

The smart microvalve consists of a micromechanical fluid valve (for liquids and gases), along with associated flow sensing and integrated electronics for closed-loop feedback control of the fluid delivery rate. The microvalve employs an electrically-activated, fully-integrated, shape-memory alloy actuation mechanism, thereby enabling the valve to generate sufficient actuation force to control fluid at pressures typical for fuel delivery systems, while simultaneously achieving a sufficiently large stroke thereby reducing the open flow resistance to liquids. An important aspect of our development is fully-integrated,

electrically-activated, arrays of smart valves for redundancy, spatial flow control, and macro flow control with microactuators, all of which are required for intelligent fluid management and distribution. Further, we are exploring technologies enabling the smart microvalve to function in extremely harsh conditions, such as high temperatures and corrosive environments which are common to many fuel distribution applications. For low-temperature (below 100 C) operation, we are using TiNi shape-memory alloy for the valve actuation. For higher temperature applications, we are investigating a high phase transition temperature shape-memory alloy technology, such as TiPdNi. We are also exploring the use of SiC to coat valve seats and internal chambers. These improvements are expected to reduce sticking and wear of the valve seat during cycling, as well as allow the microvalve to operate in corrosive environments.

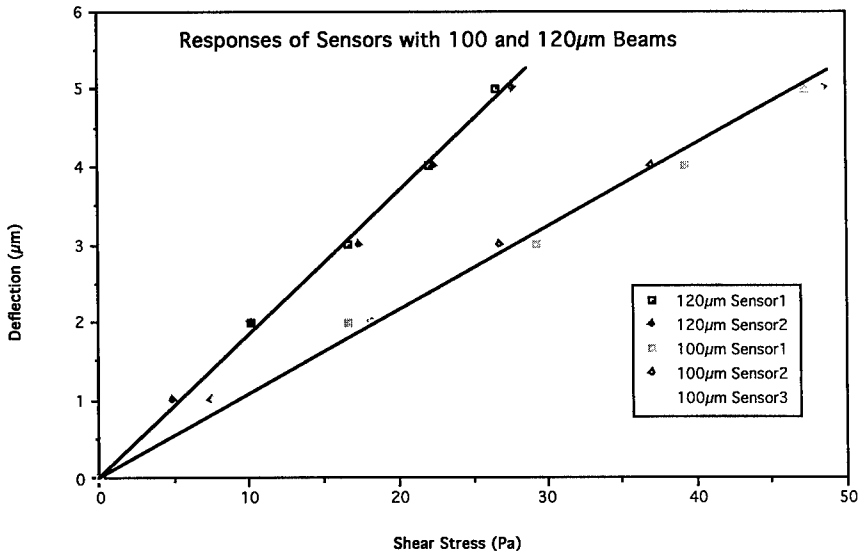


Figure 3. The element deflection as a function of applied shear stress is indicated for three sensors with 100μm-long and two sensors with 120μm-long support beams. The sensors with longer beams are more compliant, resulting in a greater sensitivity.

Currently, our TiNi films are rf sputter deposited in an ultra-high vacuum chamber, using a 2-inch TiNi target [6]. The films are amorphous as-deposited at room temperature. Therefore, an anneal at 520 C for 30 minutes, is performed directly after deposition, without breaking vacuum, in order to crystallize the films. The crystallized TiNi films display good shape memory behavior, with phase transformation temperatures around 60 C and temperature hystereses as low as 10 degrees. The transformation is between the austenite and martensite phases, which will achieve the maximum recovery force. We have recently purchased a pulsed-laser deposition (PLD) system. This system will employ two lasers, whose powers can be determined independently, thereby enabling very precise control over the relative deposition rates of the two elements and the resulting shape-memory behavior. The resultant thin films of TiNi are expected to have uniform thickness and very well controlled composition (e.g., to within 0.1 at %). In addition, this new

system will allow deposition over a four inch diameter area, and since no ambient gases are used, the resulting films will contain less impurities.

A conceptual drawing of one of the microvalve designs we are developing is shown in Figure 4. To fully exploit the large strain levels possible with the use of shape-memory alloys, the deformable part of the actuator mechanism must consist entirely of TiNi, since a silicon diaphragm would fracture at strains larger than 0.5%. The design uses a TiNi diaphragm completely released from the silicon substrate in the area of the valve chamber. Consequently, the stroke of the valve is limited only by the maximum strain level of the shape-memory alloy. Resistive heating of the TiNi will be used to initiate the martensite to austenite phase transformation causing the plunger to lift off of the outlet orifice as the material recovers its initial shape and actuating the valve. Closure of the valve is achieved by turning off the heating power, allowing the TiNi to cool. As the TiNi cools, it once again returns to the more compliant martensite phase. The silicon spring applies a reset force on the TiNi layer causing the plunger to reseal over the outlet orifice.

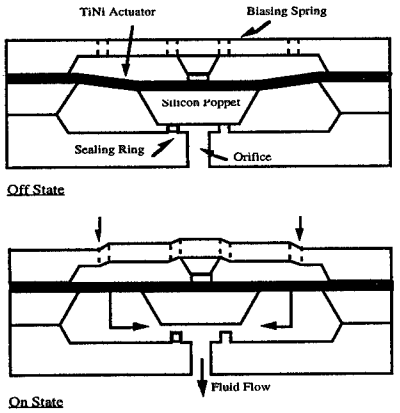


Figure 4. Shape-memory alloy actuated microvalve employing integrated silicon reset spring.

Silicon Carbide (SiC) MEMS: Our program at CWRU is developing silicon carbide (SiC) as a basic mechanical material for high-temperature (e.g., up to 550°C) MEMS. We recently built a cold-walled SiC APCVD reactor for the deposition of single-crystal 3C-SiC onto 4 inch substrates which has been fully operational for about one year (Figure 5). High quality single-crystal SiC films are routinely deposited in the reactor with good uniformity onto 4 inch substrates [7]. Augmented with these in-house capabilities for SiC deposition, our approach is to move many of the concepts from the integrated silicon-based MEMS technology onto SiC. The high-temperature sensor devices we are currently exploring include; pressure sensors, temperature sensors, and flow sensors. We are also investigating the use of SiC coating to improve the wear properties of silicon micromechanical elements and SiC-based microactuators.

The cornerstone of our effort is the development of a MEMS-based, in-cylinder pressure sensor for engine monitoring using SiC. The design for the first prototype of the micromachined SiC pressure sensor utilizes ion-implanted SiC piezoresistors fabricated on a SiC bulk-micromachined diaphragm. The fabrication for this device is nearly complete.

We are also extending our current SiC surface micromachining capability toward an integrated high-temperature MEMS process based on SiC surface micromachining on SiC substrates. To do this, we are using many of the already established technologies in SiC electronics to develop an integrated MEMS technology on SiC. The SiC surface micromachining process enables the fabrication of sealed-cavity, capacitive SiC pressure sensors, as well as lateral-resonant-type devices suitable for high temperature applications.

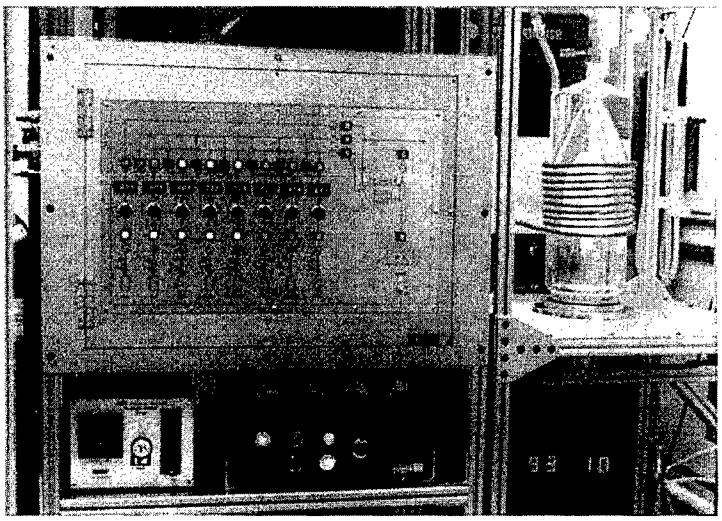


Figure 5. Picture of the APCVD SiC Reactor.

Using a SiC surface micromachining process, we recently fabricated a poly-SiC lateral-resonant-type electrostatic microactuator [8]. After growth of the poly-SiC on polysilicon, a 5000Å-thick aluminum masking layer was deposited, patterned, and wet-etched. Next, with the patterned aluminum layer as the etch mask, poly-SiC microstructures were defined by reactive ion etching in a CHF_3/O_2 mixture. After etching, the aluminum mask was removed using HF. The poly-SiC microstructures were then released by etching the polysilicon sacrificial layer in 40 wt% KOH at 40°C. The etching time is controlled so that the polysilicon was etched from everywhere but under the anchor regions of the poly-SiC devices. Figure 6 shows an SEM image of a released lateral resonant microactuator fabricated using the above-mentioned process. Devices with lateral dimensions as large as 500 μm were fabricated, with no noticeable residual stress induced deformations. The resonant frequencies of the nominally 2 μm-thick poly-SiC microstructures ranged from 20 kHz to 60 kHz, depending on the geometry.

Integrated Optical Switches: A novel variation of micromotor technology was recently developed at CWRU, the diffraction grating microscanner [9]. The diffraction grating microscanner is potentially a low-cost, highly reliable and small sized integrated optical scanner which may have many advantages over conventional electromechanically rotated optical prisms or grating commonly used in optical switches, bar-code label laser scanners, and computer interconnects. The diffraction grating microscanner is shown in Figure 7. A diffraction grating is etched in the top surface of a solid micromotor rotor using dry plasma etching. The microscanner uses a relatively thick polysilicon layer to form the diffraction grating rotor. This was necessary because the diffraction grating required a

relatively deep etch in the top surface of the rotor and the drive voltages must be relatively low. The increased thickness of the motor poles and the rotor allows for increased electrostatic coupling and therefore reduced drive voltages. The microscanners have a minimum operating voltage of 18V and operate at scan speeds over 5000 rpm. The devices have over one meter of working distance, and the diffracted beam profiles are of high quality.

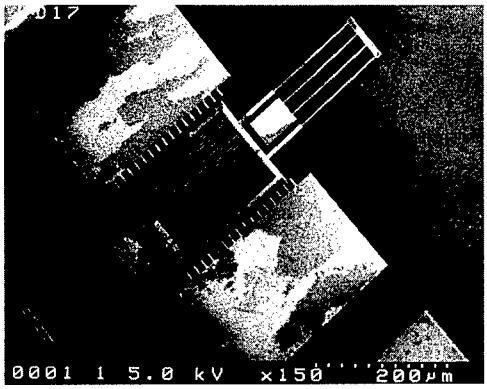


Figure 6. SiC surface micromachined lateral resonant actuator.

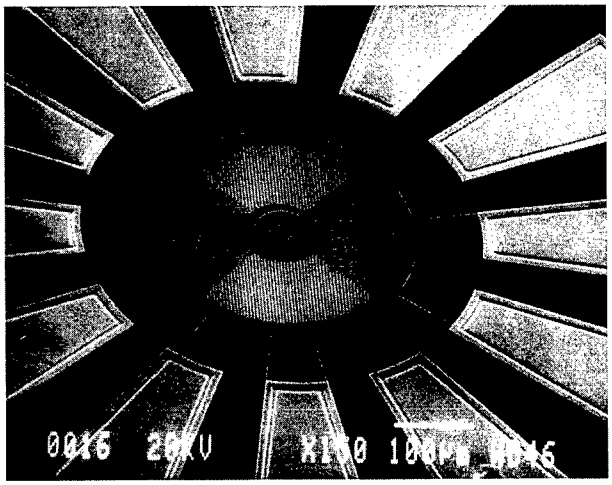


Figure 7. SEM of the polysilicon surface micromachined diffraction grating microscanner.

Micromechanical Relays: Micromechanical relays have many applications in communications, power distribution, and industrial control. Micromechanical relays have several advantages over conventional mechanical relays including: smaller size; faster

response time; low-cost; and lower power. Furthermore, micromechanical relays, unlike solid-state relays, have no leakage, are more rugged, and can handle larger signals. A micromechanical relay developed at CWRU is shown in Figure 8 [10]. The device was fabricated by electroless plating of nickel in a micromold pattern formed in a relatively thick positive photoresist layer. A standard UV exposure combined with a highly transparent and viscous resist (Hoechst AZ 4620) allows a 3.5 μm linewidths with 2.5 μm spaces to be formed in resists layers 25 μm thick with a high aspect ratio. The CWRU micromechanical relays have a measured contact resistance of 10 ohms, can handle current loads over 150 mA, and have a switching bandwidth over 1 kHz.

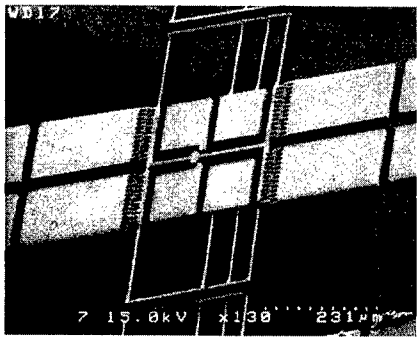


Figure 8. SEM of the CWRU micromechanical relay.

Conclusions: MEMS technology promises to radically change our capabilities in the design and implementation of systems and products requiring sensors and actuators. The low-cost, high performance and high functionality of MEMS compares extremely well with conventional technologies, and consequently the market for MEMS components and systems is expected to grow at a rapid pace in the future.

Acknowledgments: We wish to thank ARO, ARPA, NSF, NASA Lewis Research Center, and industrial sponsors for supporting our research efforts at CWRU and Professor Wen Ko for providing material on the embedded tire air pressure sensor.

References:

[1] K.E. Petersen, "Silicon as A Mechanical Material," Proceedings of the IEEE, 70, 1982, p. 420.

[2] Market research funded and provided by the United States Advanced Research Projects Agency (ARPA), Department Electronic Systems Technology Office.

[3] F. Nulsen, A. Leung, D. Fleming, R. Lorig, J. Bettice, K. Dolin, and W. Ko, "Chronic Intracranial Pressure Monitoring by Telemetry: Clinical Experience," Annals of Biomedical Engineering, vol. 8, 1980, p. 505.

[4] M. Mehregany and C. Bang, "MEMS for Smart Structures," in Proc. of the SPIE Conference on Smart Structures and Materials, San Diego, CA, Feb. 26-March 3, 1995.

[5] T. Pan, D. Hyman, M. Mehregany, E. Reshotko, and B. Willis, "Calibration of Microfabricated Shear Stress Sensors," The 8th International Conference on Solid-State Sensors and Actuators, Technical Digest, Stockholm, Sweden, June 1995, p. 345.

[6] R.H. Wolf and A.H. Heuer, "TiNi (Shape Memory) Films on Silicon for MEMS Applications," Journ. of MicroElectroMechanical Systems, Vol. 4, No. 4, December 1995, p. 206.

[7] C.A. Zorman, A.J. Fleischman, A.S. Dewa, M. Mehregany, C. Jacob S. Nishino, and P. Pirouz, "Epitaxial Growth of 3C-SiC on 4-inch Diameter (100) Silicon Wafers by APCVD," J. Appl. Phys., 78 (8), October 15, 1995, p. 5136.

[8] A.J. Fleischman, S. Roy, C.A. Zorman, M. Mehregany, and L.G. Matus, "Polysilicon Silicon Carbide for Surface Micromachining," accepted for poster presentation, IEEE MEMS Workshop, San Diego, CA. Feb. 1996.

[9] A.A. Yasseen, S.W. Smith, M. Mehregany, and F.L. Merat, "Diffraction Grating Scanners Using Polysilicon Micromotors," Proceedings, IEEE Micro Electro Mechanical Systems, Amsterdam, Netherlands, January 29-February 2, 1995, p. 175.

[10] S. Roy and M. Mehregany, "Design, Fabrication, and Characterization of Electrostatic Microrelays," accepted for presentation at the Micromachining and Microfabrication Conference, SPIE, Austin, Texas, to be held October 1995.

SIGNAL ANALYSIS

WAVELET DOMAIN CHARACTERIZATION & LOCALIZATION OF MODAL ACOUSTIC EMISSIONS IN AIRCRAFT ALUMINUM

Grant A. Gordon, PhD and Randy K. Young, PhD
The Pennsylvania State University
Applied Research Laboratory
PO Box 30
State College, PA 16804 USA
gag100@psu.edu

Abstract: As machines age, broadband acoustic emission (AE) signals are emitted from unknown locations, often due to microscopic material degradation. Not only is the AE source location unknown, but the source mechanism and hence signal characteristics are also unknown. These signals then propagate along multiple, complex geometric paths, in various distorting modes experience mode conversion and wave guide selection, before finally reaching the sensors. In addition to this wave distortion, the signals are corrupted by the noisy and interference-rich environment. If we wish to improve our understanding about the "health" or state of a mechanical system, we should employ AE analysis methods which can more effectively account for these physical effects. This paper presents a technique for detecting, classifying and localizing these acoustic emissions. Due to the significant uncertainty surrounding these sensed signals, high fidelity, well verified models are uncommon. Starting from wide-band, high-fidelity AE data we have shown that it is possible to determine source location with greater accuracy than is possible by traditional source location techniques. This improvement was achieved by considering the multimode wave nature of the AE events, and by applying wide band signal processing techniques that specifically accounted for the dispersive nature of the signals.

Key Words: Acoustic emissions; condition based maintenance; Integrated Predictive Diagnostics; multi-modal propagation; wavelet transforms

INTRODUCTION: Heuristically, or perhaps superficially, it is easy to understand the phenomena of acoustic emission and to acknowledge the importance of this phenomena to a condition based maintenance strategy for machinery. After all, acoustic emission is the internal generation of transient elastic waves by the rapid release of energy from a localized source or sources from within a material. The source mechanisms can vary considerably, but are generally associated with a discontinuous, irreversible, microscopic event. Given this description, one would expect that cracks produced in a material from loading or environmental attack during

service, could be ideally studied via acoustic emission. It is not difficult to imagine that a network of permanent sensors could permit continuous monitoring of defect growth. Used in concert with fracture mechanics principles, a crack growth rate could be measured and the remaining life before the crack reached a critical size determined. Indeed, there has been considerable research in this area, with typical early work [1-3] beginning in the mid 1970's. However, this problem has proven to be more difficult to solve than was initially anticipated. Relationship which link relevant damage measures to an acoustic emission parameters rarely work except in very special situations [3, 4]. Furthermore, the use of AE parameters to identify imminent failure has been even less successful.

Historically AE technology has been primarily concerned with assigning and collecting AE signal 'parameters'. Signal characteristics, such as rise time, peak amplitude, and energy, are evaluated from the filtered response of a resonant transducer used to sense the transient signals. The parameters are evaluated by mathematically describing the sensed waveform as a damped sinusoid. This procedure, which originally appeared warranted when AE was in its infancy due to hardware limitations, is predicated on the assumption that the essential acoustic emission features can be well represented with the single model structure for the data. Unfortunately there is no theoretical basis to support this contention [5], and although researchers have tried to find meaningful correlations between these simple parameters and physically relevant values, the results have not been compelling. For example, the literature is full of conflicting results over the significance of AE amplitudes and source mechanisms [6]. The purpose of the waveform parameterization is to reduce the AE events to their essential characteristics, however many claim that just the opposite occurs. Instead of preserving information, the resonant transducer sensing, nonlinear filtering and amplification, followed by simplistic waveform modeling, actually destroys much of the true information that can be extracted from an AE signal.

Recently a new approach to practical field-based acoustic emission has been proposed [5], and is being pursued [6-9], since it holds considerable promise in being able to extend the limitations of traditional AE. In contrast with the traditional technique, the new approach places emphasis on recording the entire acoustic emission event with as much fidelity and bandwidth as possible. This distinction is significant. By accurately recording the entire waveform, wave propagation theory can be employed for superior source identification and determination of source location. This greatly enhances the ability to distinguish noise from relevant AE signals; a particularly vexing issue in practical monitoring situations. The purpose of this paper is to demonstrate how wide band signal processing techniques, along with wave propagation theory can be used to exploit this rich data set and improve the acoustic emission source identification and location.

ACOUSTIC EMISSION IN PLATES: In acoustic emission the structure being studied is often geometrically platelike; the structure's thickness is much smaller than its length or width. An ideal approximation to these structure is the infinite plate waveguide with traction free boundary conditions. Through a process of multiple reflection, the boundaries of a waveguide actually act to guide the wave energy along the direction of the constraining structure. In this manner the wave modes supported within the guide can follow curvatures, provided the radius of curvature is much larger than the wavelength. This situation is encountered in many aircraft fatigue crack problems, where AE systems are being developed to monitor in-flight crack growth. Cracks are

initiated at fatigue critical regions such as the point where fasteners are used to bond the aircraft skin (platelike) to underlying ribs, spars and/or other plates. When the crack progresses, AE energy is released and guided wave modes are generated which travel through the skin material.

The solution to the problem of elastic wave propagation in an isotropic solid plate, with traction free boundary conditions, has been know for many years. There are three distinct types of guided waves, horizontally polarized shear waves, longitudinal (also known as extensional) plate waves with displacements symmetric about the plate center line and flexural plate waves having antisymmetric displacements. For each of these wave types many possible modes can exist, the solution of which is dictated by a transverse resonance principle. Only those combinations of frequency and phase velocities which generate standing waves in the thickness direction will exist. By mapping out this relationship between frequency and phase velocity a set of dispersion curves can be obtained which prescribes the speed at which a particular frequency component of a transient wave packet will travel if its mode is known.

Often simplifications can be made. In has been observed experimentally [5, 6, 9] that acoustic emission signals in thin plates propagate primarily in the lowest order symmetric (S_0) and antisymmetric (A_0) Lamb wave modes. The dispersion relations for these extensional and flexural plate wave modes are given by the Rayleigh-Lamb equation [10]. For the symmetric modes,

$$\frac{\tan\left(\beta\frac{d}{2}\right)}{\tan\left(\alpha\frac{d}{2}\right)} = -\frac{4\alpha\beta k^2}{\left(k^2-\beta^2\right)^2} \tag{1}$$

and for the antisymmetric modes,

$$\frac{\tan\left(\beta\frac{d}{2}\right)}{\tan\left(\alpha\frac{d}{2}\right)} = -\frac{\left(k^2-\beta^2\right)^2}{4\alpha\beta k^2} \tag{2}$$

where

$$\alpha^2 = \frac{\omega^2}{C_L^2} - k^2 \tag{3}$$

$$\beta^2 = \frac{\omega^2}{C_T^2} - k^2 \tag{4}$$

d is the plate thickness, k is the wavenumber, w = 2pf is the angular frequency, C_L is the bulk longitudinal wave velocity, and C_T is the bulk shear wave velocity. Given a particular isotropic plate material of known thickness, C_T, C_L and d are defined. Using these values equations (1) to (4) can be numerically solved for the two unknowns w and k. This then can be used to determine the phase velocity

$$c = \frac{\omega}{k} \tag{5}$$

which is usually plotted against the dimensionless abscissa variable fd. All modes, other than the two fundamental modes, S_0 and A_0, exhibit cut off frequencies above zero frequency, f = 0.

These cut off frequencies can be determined by forcing k to approach 0 in (1) - (4), which allows us to determine that the next closest mode is the second antisymmetric mode (A_1) with a cut off frequency of $f=C_T/2d$.

Although these dispersion equation appear straight forward, they are quite ill-behaved and present a number of difficulties even for numerical analysis. As an alternative to this exact description for the dispersion curves, a number of authors [5-7, 9] have been studying high fidelity wide-band AE data using approximate plate theories to describe the wave propagation behavior of the fundamental modes. These solutions

$$C_{ext} = \left[\frac{E}{(1-\upsilon^2)\rho}\right]^{\frac{1}{2}} \tag{6}$$

$$C_{flex} = \left[\frac{D}{d\rho}\right]^{\frac{1}{4}} \omega^{\frac{1}{2}} \tag{7}$$

where

$$D = \frac{Ed^3}{12(1-\upsilon^2)} \tag{8}$$

E is Young's modulus, r is density, and n is Poisson's ratio, yield much simpler dispersion relationships, but they only apply to the fundamental modes, and are only accurate when the wavelength is much larger than the plate thickness.

Using the value given in Table 1 for 7075 - T6 aluminum, the dispersion curves for the lowest order symmetric (S_0) and antisymmetric (A_0) Lamb wave modes are plotted in Figure 1. Classical plate theory predicts that the S_0 mode is dispersionless while the A0 mode is highly dispersive. This dispersion makes the A_0 mode data more difficult to handle and it is often argued in traditional AE analysis that the flexural wave should be filtered and removed in order to facilitate accurate source location. Unfortunately the existence of a strong extensional mode is not guaranteed since the source itself dictates the nature of the AE emission generation. Figure 1 also shows the Rayleigh-Lamb dispersion calculations which are exact for homogeneous isotropic material. It can be seen that the lowest order modes approach the Rayleigh wave velocity asymptotically as the fd product grows large. Notice also, that the classsical estimates of the wave mode velocities are only accurate when the frequency thickness, fd, product is small. Calculating the cut off frequency of the next "nearest" mode (A1 mode) we find that no higher order modes can exist below a fd product of 1.6 MHz*mm. Thus we can conclude that the classical theory estimates can lose accuracy at fd values which are smaller than those required for admission of higher order modes. In other words, although classical plate theories are not expected to accurately predict the fundamental Lamb wave modes phase velocity response at high fd values, they can also be inaccurate for values of fd which preclude the existence of other higher order modes.

ACOUSTIC EMISSIONS FROM AIRCRAFT WING SPAR: Acoustic emission data was collected on an experimental test bed designed to simulate a section of an aircraft lower wing. The single layer skin material, 3.175mm 7075-T6 aluminum, was attached to frame members through bolt fasteners at the locations shown in Figure 2. A starter notch to seed crack growth was generated at the midpoint between the rows of fasteners via an access hole. When loading was applied, any AE signals that exceeded a preset threshold level, triggered the simultaneously storage of the AE signals at each of the four sensor locations surrounding the starter notch. The coordinates, in inches, of these transducers and the notch are also shown in Figure 2. A new measurement system called the Fracture Wave Detector manufactured by Digital Wave Corporation, Englewood, Colorado was used to monitor, amplify and store the data recorded by the transducers. This multi-channel transient capture system digitized each transducer responses at a rate of 12.5 MHz for a total record length of 1024 bytes per waveform.

ACOUSTIC EMISSION DATA AND COMPARATIVE ANALYSES: Typical examples of data collected during the experiment, labeled event #1 through #4 for convenience, are shown in Figure 3. Each event shows the simultaneously collected response of the four transducer located at the positions indicated in Figure 2. The data shown in events 1 and 2 are indicative of cracking while the data from events 3 and 4 are generated from noise. This determination is made by considering such factors as frequency content, temporal relationship between the four responses, the duration, amplitude and shape of the signals. For our purposes noise will be considered as any AE signal that arises from a source other than the crack growth mechanism. This can be due to testing machine noise, grip induced damage, electro-magnetic interference, rubbing and mechanical sources. Since cracking is associated with the rapid release of energy we expect the time duration to be short. In carefully controlled ideal experiments [e.g. 9], the two lowest order plate wave modes were identified and their frequency content observed to be different. The flexural waves exhibit more low frequency content than the extensional waves, and because of this, suffer less attenuation as they travel along the plate since attenuation increases as the square of frequency. These waves also shows considerably greater spreading in the time domain due their highly dispersive nature; the lower frequencies lag the faster high frequency components. These shape and frequency features can be used to discriminate between noise and crack signals. In addition, from the sensor location, we can determine the path length difference between the anticipated AE source and the various receivers. This can be used to quantify the expected offset time between the four records, and to further screen out spurious events.

Table 1. Properties of a 7075-T6 aluminum plate.

Young's Modulus	Poisson's Ration	Density	Longitudinal wave velocity	Shear wave velocity	Thickness
71 GPa	0.30	2700 kg/m^3	5950 mm/us	3180 mm/us	3.175 mm

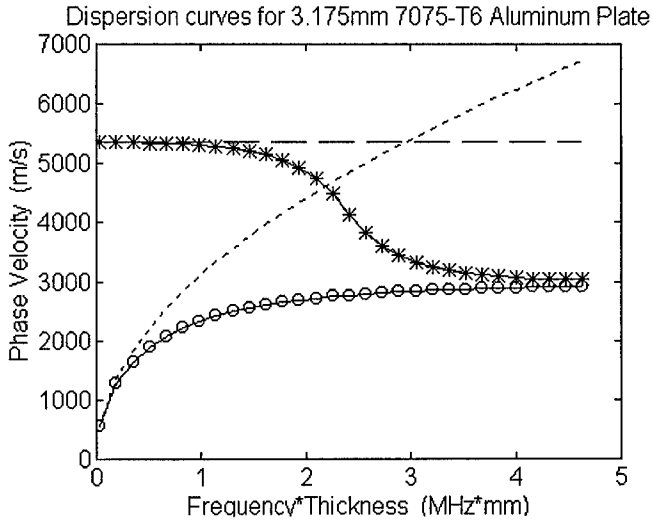


Figure 1. Dispersion curves for the S_0 and A_0 modes in 3.175mm thick 7075 - T6 aluminum plate. The dashed and dotted lines are for the S_0 and A_0 modes respectively, calculated using classical plate theory. The solid line with stars and the solid line with circles are for the S_0 and A_0 modes respectively, calculated using the transcendental Rayleigh-Lamb equations.

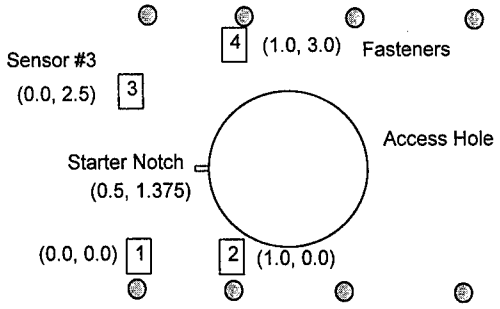


Figure 2. Experimental setup showing sensor location and notch geometry.

Many of these wave mode features can be observed in the power spectra and spectrogram plots of our data. In Figure 4 the power spectra of two typical AE signal are compared. The top half of the figure shows the power spectrum of an AE crack signal. This plot shows a significant amount of energy both above and below 1 MHz. The energy above 1 MHz is probably due to the S_0 mode. However, we can not rule out the possibility that this contribution is due to higher order modes, since its fd product exceeds the A1 mode cut off limit. Notice that the amplitude of the high frequency content is smaller than the lower frequency content due to the flexural wave. The noise signal shown in the bottom half of the figure has much lower frequency content. Figure 5 shows how the frequency content of a typical crack signal varies with time. Lighter areas indicate higher energy concentration. Again we see two main area of frequency content; a strong concentration below 1 MHz and a weaker concentration above 1 MHz. Initially these 'packets' arrive at the sensor at roughly the same time, as shown by the two lightest areas of the spectrogram. However as time progresses, the low frequency content dominates. This is consistent with our model of a multimode signal comprised of a weak extensional mode and a strong flexural mode. The fact that the low frequency components of the A_0 mode lag the faster high frequency components is also seen in the spectrogram.

To further analyze the data we will consider the problem of source location. Since the source location is assumed to be known, the accuracy of the location estimate will be used to evaluate the effectiveness of various processing approaches. Let (x_0 , y_0) be the position of the source, which is to be estimated. For each sensor located at a unique (x_i , y_i) the following relation will be satisfied

$$(x_0 - x_i)^2 + (y_0 - y_i)^2 = (ct_i)^2 \tag{9}$$

when a signal is sensed; ti is the time for the wave to reach the sensor and c is the wave velocity. Since only the time differences between the sensors can be measured from the data, we substitute $t_i = t_0 + \Delta t_i$, where t_0 is the time required for the wave to reach the closest sensors. By substituting each of the four sensors into equation (9) and subtracting any ith sensor expression from a jth sensor expression, we produce a set of six linear equation. From this overdetermined set of equations, the three unknowns (x_0, y_0, t_0) can be found given the wave speed, time differences and sensor location, by the method of least squares.

Traditionally the time of flight differences between sensors is determined using threshold triggering. The time difference is measured from the first occurrence at which the sensed signal rises above a threshold level on each sensor. Since the extensional wave travels faster than the flexural wave it is customary to assume that this wave mode is responsible for the threshold triggering and to assign an extensional wave velocity (equation (6)) for use in the location algorithm. As already pointed out, the type of wave mode generated is dependent of the AE source. In-plane cracking yields higher extensional mode amplitudes, while source mechanisms that produce displacement perpendicular to the plane give rise to flexural modes. Clearly it is not always accurate to assume that the extensional mode contributions are significant enough to rise above the threshold level. In our data we have determined that contributions from the antisymmetric mode dominate, especially at low frequencies. To make use of this data we used a

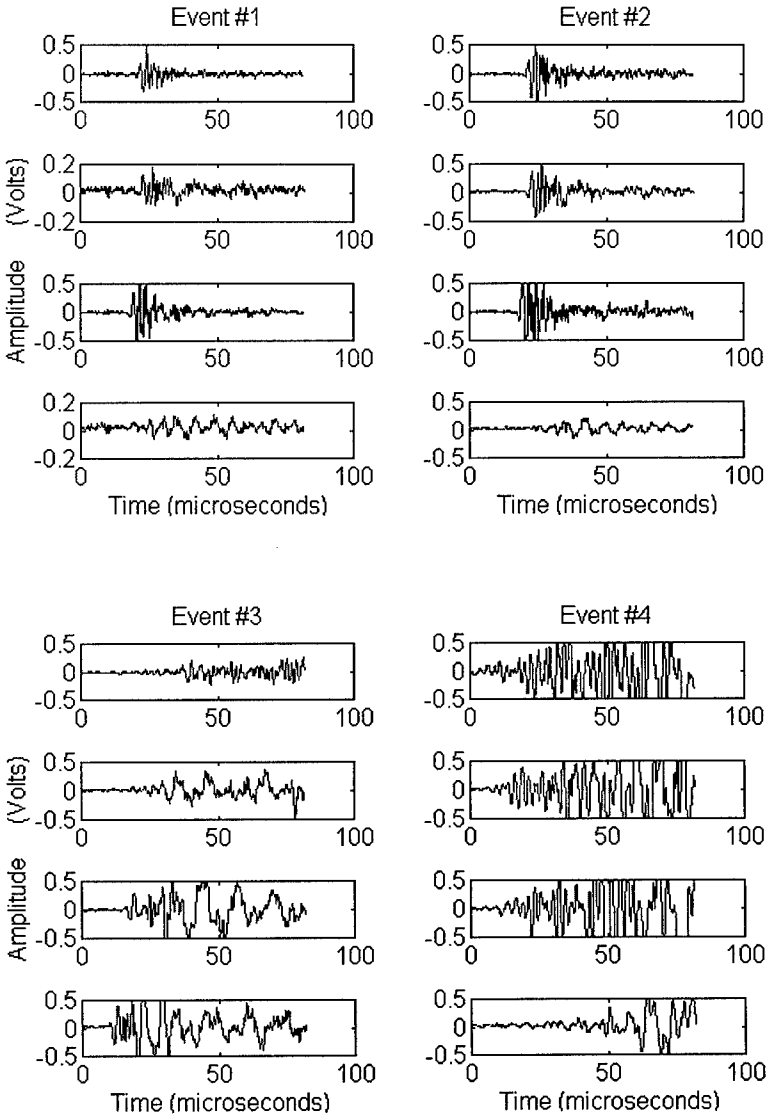


Figure 3. Sample acoustic emission data. All four transducer responses for an AE event are given in a single column, ordered from top to bottom to correspond with the responses of the first through fourth transducer.

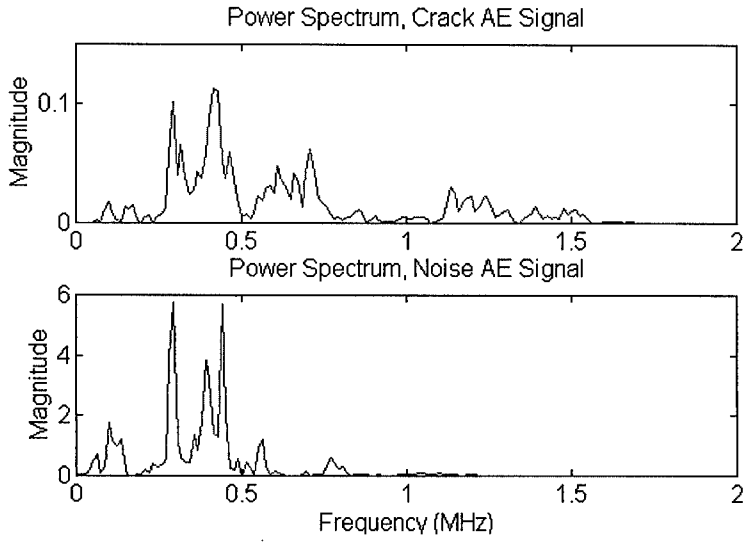


Figure 4. Typical power spectral density curves for the crack and noise data.

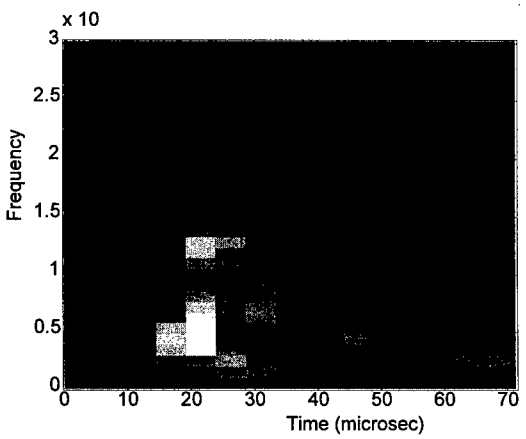


Figure 5. Typical spectrogram for the crack AE signals showing the distribution of frequency with time.

novel approach based on wide band processing. The method can be considered a time domain spectroscopy (TDS) technique where the sampled signals are processed by a bank of the narrowband filters defined by the Morlet wavelet

$$\Psi(t) = \exp(j\omega t)\exp\left(-\frac{t^2}{2}\right) \tag{10}$$

This is nothing more than a complex sinusoid windowed modulated by a Gaussian envelope where the frequency to be extracted from the data is used to define the center frequency of the wavelet. Subsequent filtering of the wide band signals is performed by multiplication of the Fourier transformed filter response by the spectral response of each sensor record. By performing this operation repetitively, at various prescribed frequencies, we built up a filter bank decomposition of the data. The time difference between sensors was found, for each filtered component of the data, by performing the inverse Fourier transform and cross correlating the resulting waveforms. Cross correlation will show a peak at the corresponding time lag position which is equal to the delay between the two signals. Since the wide band data has been filtered into narrowband responses the cross correlated time differences can be taken as phase delays. To determine the source location from these new narrowband time differences, equations (2)-(4) were used to define the appropriate phase velocity. This mechanism provide a rich set of location estimates that could be averaged to yield a robust estimate of the crack source location.

Table 2. Comparison of AE source location estimates

Data	Threshold Value	Location Estimate	Error	Data	Frequency (KHz)	Location Estimate	Error
Event #1	.1	(.666, 1.00)	.406	Event #1	600	(.317, 1.32)	.190
	.05	(-.071, -5.62)	1.24		500	(.437, 1.37)	.063
					400	(.570, 1.26)	.130
					300	(.534, 1.26)	.110
				average	n/a	(.465, 1.30)	.082
Event #2	.1	(.671, 1.01)	.403	Event #2	600	(.419, 1.32)	.100
	.05	(1.95, 1.57)	1.46		500	(.470, 1.27)	.109
					400	(.377, 1.28)	.151
					300	(.441, 1.31)	.110
				average	n/a	(.427, 1.29)	.112
Event #3	.1	(-3.47, -2.73)	5.72	Event #3	600	(.613, .735)	.650
	.05	(-3.09, -.795)	4.19		500	(.119, -.336)	1.75
					400	(.985, 1.89)	.710
					300	(.226, 2.55)	1.20
Event #4	.1	(1.70, 1.71)	1.24	Event #4	600	(.603,1.27)	.144
	.05	(6.16, -4.79)	8.37		500	(.195,1.19)	.356
					400	(.517, 1.24)	.129
					300	(.301, 1.27)	.223

The left hand side of Table 2 shows the results of using the traditional thresholding approach to AE source location. An S_0 wave velocity of 5375 mm/ms was used in the calculation and the error was determined by finding the distance of the estimate from the assumed AE crack source location at the starter notch. Table 2 also shows the results of the new time domain spectroscopy approach for determining AE source location. Four frequency values, centered around the low frequency spectral content of the AE signals (Figure 4) are shown. It can be seen that this approach gives a much better estimate of the source location, especially when the various estimates are combined via averaging. From both techniques we can see that the location of AE events #3 was placed far from the crack site. This further confirms our assessment that this signal was not due to a cracking event. Event #4 is also placed far from the crack site by the thresholding technique, but not by the spectroscopy technique. We believe that this event is also due to noise but that the spectroscopy technique experienced difficulty because of the strongly clipped nature of the signal, Figure 3.

CONCLUSIONS: Starting from wide-band high-fidelity AE data we have shown that it is possible to determine source location with greater accuracy than is possible by traditional source location techniques. This improvement was achieved by considering the multimode wave nature of the AE events and by applying wide band signal processing techniques that specifically accounted for the dispersive nature of the signals.

ACKNOWLEDGMENTS: The authors would like to thank Paul Lacjak and Michael Gorman of Digital Wave Corporation for providing the data and for illuminating discussions.

REFERENCES:

[1] Harris, P.H. and Dunegan, H.L. "Continuous monitoring of fatigue-crack growth by AE techniques," Expl Mechs, **14**, Feb. 1974, pg 71.
[2] Tetelman, A.S. and Evans, A.G. "Failure prediction in brittle materials using fracture emchanics ans AE," UCLA-ENG-7365, Aug. 1973.
[3] Williams, J.H, DeLonga, D.M., and Lee, S.S. " Correlations of AE with fracture mechanics parameters instructural bridge steels during fatigue," Materials Evaluation, **40**, p1184, 1982.
[4] Scott, I.G. Basic Acoustic Emission, Gordon and Breach Science Publishers, NY, NY, 1991.
[5] Gorman, M.R., "New Technology for Wave Based Acoustic Emission and Acousto-Ultrasonics," pp. 47-59 in AMD-vol **188**, Wave Propagation and Emerging Technologies, ASME, NY, NY, 1994
[6] Prosser, W.H., Jackson, K.E., Kellas, S., Smith, B.T., McKeon, J., and Friedman, A. "Adavnced waveform-based acoustic emission detection of matrix cracking in composites," Material Evaluation, **53**, Sept 1995, p1082.
[7] Searle, I., Ziola, S., and Rutherford, P. "Crack detection in lap-joints using acoustic emission," SPIE, **2444**, p212.
[8] Martin, C.A., Van Way, C.B., Lockyer, A.J., Kudva, J.N., and Ziola, S. "Acoustic emission testing on an F/A - 18 E/F Titanium bulkhead," SPIE, **2444**, p204.
[9] Gorman, M.R. "Plate Wave Acoustic Emission," pp. 358-364, JASA, **90**, No. 1, 1991.
[10] Graff, K. Wave motion in elastic solids, Dover Pub., Inc., NY NY, 1975.

OVERVIEW OF WAVELET / NEURAL NETWORK FAULT DIAGNOSTIC METHODS APPLIED TO ROTATING MACHINERY

Jose E. Lopez
Inna A. Farber Yeldham
Kevin Oliver
ALPHATECH, Inc.
50 Mall Road
Burlington, MA 01803-4562

Abstract: New technology in the form of wavelet-based methods coupled with intelligent classification schemes built around neural networks, can drive the development of substantially improved fault detection and identification (FDI) methods. Such systems represent important next generation FDI kernels for integration into advanced condition based maintenance systems for rotating machinery. This paper presents an overview of the results obtained by ALPHATECH in a program aimed at developing wavelet/neural network based FDI systems for vibrating machinery. The paper presents the performance results of these methods applied to a range of platforms including helicopter transmissions, turbopumps, and gas turbines. In addition, enhancements to the basic fault detection and identification system are presented and include overviews of multi-sensor wavelet-based differential features and improved FDI performance through classification fusing using hierarchical neural networks.

Key Words: monitoring; rotating machinery; fault detection; helicopter transmissions; prognosis; condition based maintenance; wavelets; neural networks

INTRODUCTION: As the 21st century approaches, global competition will continue to produce tremendous economic pressures on all industrial powers. The results of this competition are shorter time to market and thinning profit margins with a simultaneous demand for increased product quality and improved overall efficiency. Unfortunately, the coupling of these phenomena to military domains will be quite direct. Decreases in real GNP for individual industrial powers directly affects their ability/determination for increasing resources for maintaining viable high-tech military forces in an ever increasing complex world. The military domain under such circumstances will continue to experience increasingly tighter budget appropriations, cancellation of important new systems while simultaneously struggling with the problems of sustaining a high state of readiness in the face of rapidly aging platforms.

In the commercial domain, knowledge of enterprise wide machinery health and usage are crucial to avoid expensive down time. Timely fault detection and identification at a local level can prevent more serious damage to other parts of the enterprise that are coupled. In the military domain, the mathematics of decreased budgets translates to being able to maintain the status quo with fewer resources. For personnel intensive operations, methods for streamlining processes, increasing automation, and computerizing monitoring/reporting will be at the top of the priority list. More sophisticated/automated machine monitoring systems that require reduced human intervention will be a very desired commodity.

Presently, the dramatic decrease in the cost of powerful computational capability is fueling the investigation and uses of significantly more advanced system and signal processing methods for application to the domain of machine monitoring and prognostic systems [1]-[3]. A particularly promising set of methods relies on the application of wavelets and neural networks to the development of next generation fault detection and identification (FDI) systems [4]-[12]. The current form of the technology involves wavelet-based methods for decomposing system vibrations and coupling intelligent classification schemes that rely on neural networks for identifying machine condition and state of deterioration.

This paper presents an overview of these wavelet/neural network based FDI methods. The paper presents the performance results of these methods applied to a range of platforms, such as helicopter transmissions, turbopumps, and gas turbines. Additionally, enhancements to the basic fault detection and identification system are presented and include overviews of multi-sensor wavelet-based differential features and improved FDI performance through classification fusing using hierarchical neural networks.

WAVELET/NEURAL NETWORK BASED METHODS: This section presents the application of wavelet-based techniques coupled with neural networks to develop a fault detection and identification system. Continuous wavelet transforms and the selection of wavelet basis functions appropriate for real-time feature extraction are discussed. Examples are given for complex platforms providing formidable FDI challenges. The successful development of advanced fault monitoring processes for these platforms provides substantial benchmarks for the viability of the wavelet-based tools being developed.

Continuous Wavelet Transforms: To develop viable FDI schemes, means of extracting significant discriminate features from the vibration signal plays a critical role. Harmonic analyses in the form of a Fourier transform proves problematic for several reasons. First, the transform is global in that localized events in time can affect the entire frequency spectrum. Additionally, the Fourier transform is fundamentally not applicable to real-time monitoring applications due to the mathematical formulation of the transform that operates on the entire time axis. Windowing schemes are thus required to address the real-time feature extraction requirements for capturing important events localized in time. Unfortunately, fixed windowing schemes imply fixed time-frequency resolution in the time-frequency plane. The problem this poses is the selection of a single window that provides sufficient fidelity discriminating important events in the vibration signal that are separated by large orders of magnitude along the frequency axis. This scenario is exemplified by main helicopter transmissions where important information concerning bearings can be on the order of tens to hundreds of Hertz, whereas mesh frequencies and important fundamentals associated with gearing of the engine inputs can be on the order of tens of thousands of Hertz (i.e. order of ~10^4).

The continuous wavelet transform (CWT) resolves the window selection problem with a "zoom-in" and "zoom-out" capability that generates a flexible time-frequency window that automatically narrows (along the time axis) at high center-frequencies and expands (along the time axis) at low center frequencies [13]. The continuous wavelet transform provides this flexible time-frequency analysis by decomposing the vibration signal over dilated and translated wavelet basis functions. A wavelet is a function with finite energy, or a member of the function space $L^2(R)$, i.e., a wavelet function satisfies:

$$\int_{-\infty}^{\infty} |\psi(x)|^2 dx < \infty \tag{1}$$

The wavelet function has a zero average or essentially no DC component. A set of basis functions is obtained through dilation's and translations of a base wavelet and takes the form:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \tag{2}$$

where u is the translation parameter and s is the dilation parameter. The wavelet transform is then achieved via the inner product of the respective vibration signal, $f(t)$, with the wavelet basis function of eq. (2):

$$W_\psi f(u,s) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-u}{s}\right) dt \tag{3}$$

There are an infinite number of wavelet basis functions that satisfy eq. (1) and contain no DC component. The particular analyzing wavelet basis functions used in this work were mathematically inspired from biological systems that are effective in their decomposition and detection of vibration signals. The wavelet basis functions mimic the auditory nerve neuron's impulse response. This particular wavelet family has semi-infinite support in the time domain and can be modeled using causal real-rational transfer functions. The immediate implication is the ability to develop, on the individual wavelet basis function level, real-time feature extractors that can be efficiently implemented using auto-regressive moving average techniques (ARMA).

Wavelet-Based FDI: Figure 1 provides a simplified block diagram of a wavelet-based FDI system. The goal of the system is two-fold. First, the wavelet-based feature extraction provides the important role of extracting the essential projections of the system dynamics in an efficient manner. Second, the wavelet feature set essentially reduces the dimension of the information from the input space (real-time, continuous, analog vibration signal) to a robust lower dimensional representation that simplifies the design of the adaptive neural network classification scheme. In the design phase (as indicated in Figure 1), CWT analysis is performed on the vibration data sets to identify a set of robust wavelet features (in time and frequency) that provide discrimination between normal operation and failure conditions. In addition, these wavelet features provide sufficient separation of the failure conditions in feature space for reliable identification of the fault condition using capable classification technologies such as neural networks. Once selected, these features form the components of a feature vector extracted from the vibration signal on a regular basis. The time interval between wavelet-based feature vectors extracted from the vibration signal is user and application dependent [6].

As part of the extraction process, the wavelet basis function is adapted relative to slowly varying mesh fundamentals to compensate for input engine RPM fluctuations. Feature post-processing in the form of nonlinear transformations of the feature components is employed to provide enhanced separation of failure conditions in feature space, and hence improved classification. Finally, extracted feature vectors train an adaptive, neural network classifier. During real-time monitoring applications, the classifier output provides basic/raw classification results indicating the state of the system under test. Depending on the actual system requirements and needs, this basic classification information may be subject to further processing to enhance higher level diagnostic decision systems, or may be conjoined with other auxiliary information or systems in a fully integrated diagnostic/intelligent monitoring system.

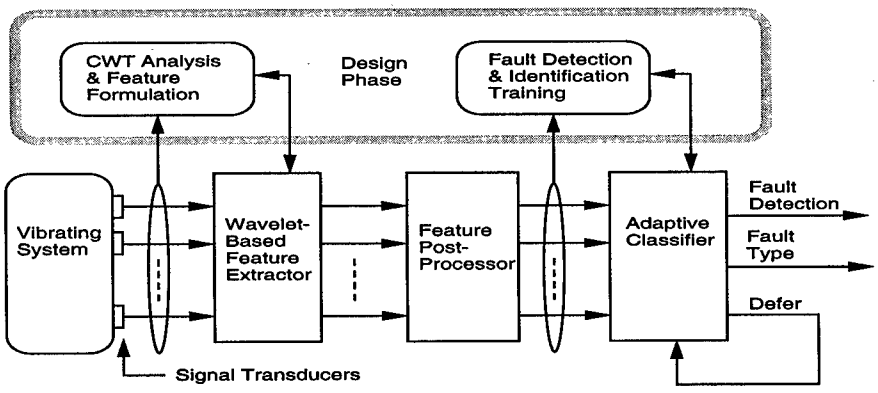


Figure 1. Wavelet-based FDI System

APPLICATION TO ROTATING MACHINERY:

Helicopter Main Transmission: Data sets from the main transmission of a helicopter with seeded faults were acquired from a major helicopter manufacturer. The transmission involves multiple gears, multiple shafts, and multiple meshes, and thus produces complex vibration data. The challenge is determining the physical phenomena that can be inferred from the wavelet visualizations, and whether a robust feature set can be extracted for FDI purposes. It is necessary to relate the CWT visualization structure to physical phenomena when selecting wavelet based features. Despite the complexity of the vibration structure in this helicopter main transmission, the CWT provides a vivid portrayal of the physical phenomenon occurring during the transmission operation.

CWT Analysis Main Transmission: Figure 2 is a CWT visualization of the first second of normal operation of the helicopter main transmission as recorded by a sensor on the transmission housing [9] at a shaft input power of 1000 HP. The Figure 2 visualization contains 512 wavelet filters distributed in an octave (log base 2) fashion along the scale axis (the vertical axis with units of Hz). The magnitude of the wavelet outputs were processed by a smoothing filter configured to have 5 millisecond smoothing time constant. The outputs were down-sampled to 512 Hz and the magnitudes color-coded using a hue-saturation scheme that maps red to large magnitudes and black to small magnitudes.

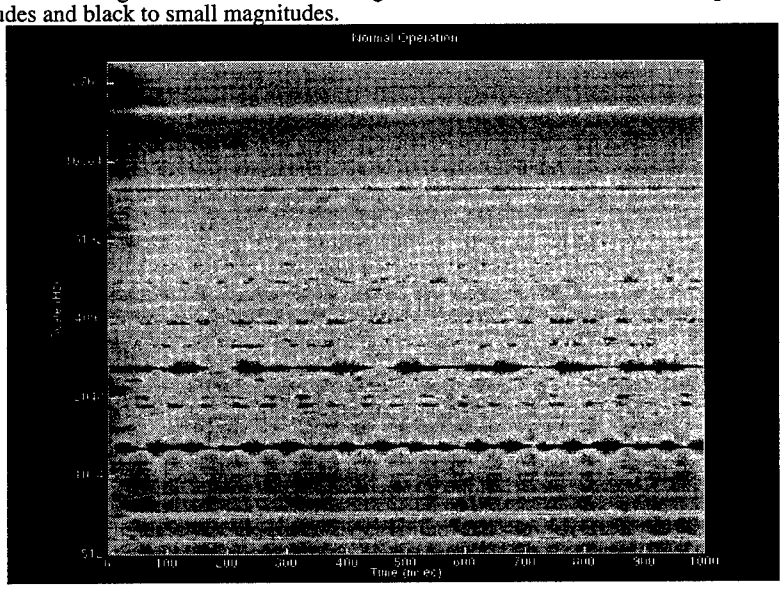


Figure 2. CWT of Normal Operation for Helicopter Transmission

The most notable structure in Figure 2 is the frequency-modulated and amplitude-modulated structure at around 1323 Hz. This corresponds to the mesh between the main input bevel gear and the top of the main bevel gear. This modulation is a very common effect in the vibrations of drive systems [14], and is usually (but not always) related to the shaft frequency. For the CWT of Figure 2, the average modulation period is of the order of approximately 52 milliseconds. This corresponds to a frequency of 19.2 Hz, which is nearly the frequency of rotation of the Main Bevel Gear. The second harmonic of the mesh at 1323 Hz is once again very visible at approximately 2646 Hz and exhibits a smearing effect of the modulation observed in the

fundamental. The third fundamental at 3969 Hz is still very visible. Another important vibration at 12920 Hz is prominent in Figure 2. This vibration corresponds to the mesh of the input engines and spur gears. The second harmonic of this mesh is also visible at 25840 Hz. A modulated vibration at approximately 1949 Hz corresponds to the mesh of the bottom of the main bevel gear and main bevel output gear to the tail rotor. The lines associated with the second harmonic are readily visible at 3898 Hz. There is a relatively strong line at approximately 5708 Hz, and it's second and third harmonics appear visible at 11416 Hz and 17124 Hz, respectively.

To determine a viable set of robust wavelet features, the normal operation CWT must be compared against the CWT visualizations for all fault cases. Additional processing tools based on morphological operations of the CWT images assist the feature identification process. The selection is primarily data driven (i.e., driven by differences detected in the visualizations), but the analysis of the previous section provides important links to the physics of the vibration structure observed. This ensures the capturing of important system dynamics that bear a causal relation to the underlying physics of the fault vibration mechanisms.

Extracted features were normalized by the magnitude of the wavelet output associated with the linking feature to account for recording variation. The wavelet magnitude outputs were smoothed using a 5 millisecond smoothing time constant and features were extracted every 10 milliseconds. Features were extracted from approximately 2.9 seconds of data for each case, and each case generated 286 feature vectors. More details concerning the features extracted and feature post processing are found in [9].

Neural Network Classification and Results: A total of 858 feature vectors (286 from each of three classes: normal, pre-overhaul, and post-overhaul) resulted from the feature extraction process. Three hundred (300) feature vectors, or 100 from each class, were used to train a global neural network with backpropagation. The neural network was trained until a sum-squared error of .975 was attained. Once the neural network was trained, all 858 feature vectors were presented to the neural network for classification. The performance results appear in Table 1. The cost of obtaining this extremely accurate classification is modest. The neural networks consisted of a small number neurons. The relative training time and convergence to small sum-squared errors was very fast given the small number of neurons used.

TABLE 1. SUMMARY OF NEURAL NETWORK PERFORMANCE RESULTS

Performance Results	
Prob. of False Alarm	.00699
Prob. of Missed Detection	.00350
Prob. of Misclassification	.00466
Prob. of Deferral	0
Number of Feature Vectors Used in Training	300

Performance Results	
Number of Feature Vectors Used in Classification	858
Number of layers	1
Number of total Neurons	3
Training Error (SSE)	.975838
Number of Training Epochs	8550

Complete Non-Parametric Approach: To illustrate that one can build wavelet-based FDI systems with essentially no detailed mechanical information concerning the underlying platforms, a second set of vibrational data from a different helicopter transmission was obtained. Unlike the previous helicopter example, this helicopter data set was unaccompanied by any mechanical information. The goal was thus to apply our wavelet methodology to develop a high performing FDI system in the absence of any detailed mechanical information.

The data was supplied with the following brief information: vibration data for normal operation ("nor") and multiple fault conditions were recorded at different torque levels by eight accelerometers positioned at various unspecified locations. The multiple fault operations included bearing corrosion at the planetary pinion ("fault 2") and at the spiral bevel input pinion ("fault 3"), tooth spalling at the spiral bevel input pinion ("fault 4"), tooth chipping at the helical input pinion gear ("fault 5"), and crack propagation at the helical idler gear ("fault 6"), at the collector gear ("fault 7") and at the quill shaft ("fault 8"). The data from the eight accelerometer sites was multiplexed with a reference and tachometer signal. Each signal contained 2.261 million samples with a 116.5 kHz sampling rate. No mechanical information was provided.

An initial approach consisted of selecting an intermediate torque setting. Not every fault was recorded at every torque level. Therefore, an intermediate torque value with a representative number of faults was selected. CWT visualizations were then generated at all accelerometer sites to identify the most observable sites. Accelerometer 7 was selected as the most desirable accelerometer site and CWT analysis was performed using the data for 100% torque values that included normal condition and six (out of seven) fault conditions.

Discriminating features were obtained by comparing the normal and fault operation CWTs. These features were then used to train a global neural network. The neural network architecture used a backpropagation learning rule with an adaptive learning rate and momentum. The neural network consisted of two layers and twenty-two neurons (fourteen neurons in the first layer, eight neurons in the output layer). The network was deemed trained when a sum squared error of 0.02 was achieved. The data for the fault scenarios and the normal condition at accelerometer site seven at 100% torque was then applied to the neural network for classification.

A confusion matrix is presented in Table 2. The network trained on 100 feature vectors and classified 920 feature vectors. Wavelet-based feature vectors were extracted every 10 ms. To understand the tabulation method (i.e., confusion matrix) used to display the results of this FDI simulation, individual table elements indicate the following: an element in the confusion matrix indicates the number of feature vectors from a given class (row label) that were classified as the class with the corresponding column label. Perfect fault detection and identification results in a confusion matrix tabulation were all numbers aligned on the main diagonal.

Results as tabulated in Table 2 indicate that very high performing wavelet/neural network based FDI was obtained. The probability of false alarm was zero and the probability that a fault condition will be classified as normal operation is 0.078%. The probability of fault misclassification is 0.068%.

FDI analysis was therefore successfully performed non-parametrically. Mechanical information about the helicopter transmission system was unknown and the only information supplied was the number and types of faults present in the transmission. A feature set was derived merely by comparing CWT visualizations from normal and fault operation. This generic methodology was then used to generate feature vectors that trained a global neural network. The classification performance of the neural network was highly accurate as shown by the confusion matrix and the time history of the neural network performance.

Turbopump Applications: This section focuses on applications of the wavelet based FDI system to detect bearing failures within High Pressure Oxygen Turbopumps (HPOTP) of the Space Shuttle Main Engine (SSME). The SSME turbopump rotates at up to 30,000 rpm on two pairs of bearings, one pair at the pump end and one pair at the turbine end.

Vibration data was collected during several test firings at two different rated power levels (104% and 109%) of one new turbopump and six faulty turbopumps (i.e., these pumps were rejected by the flight center due to the evidence of bearing degradation in their dynamic signatures). The

TABLE 2. WAVELET / NEURAL NETWORK BASED FDI SIMULATION RESULTS

	Estimated Classes								
Correct Classes	nor	fault 2	fault 3	fault 4	fault 5	fault 6	fault 7	fault 8	def
nor	920	0	0	0	0	0	0	0	0
fault 2	0	920	0	0	0	0	0	0	0
fault 3	0	0	920	0	0	0	0	0	0
fault 4	0	0	0	920	0	0	0	0	0
fault 5	0	0	0	0	920	0	0	0	0
fault 6	5	0	0	0	0	915	0	0	0
fault 7	0	0	0	0	0	0	920	0	0
fault 8	0	0	0	0	0	0	0	920	0

vibration signals were measured from two accelerometers located at 135^0 from the pump inlet and 45^0 from the pump inlet. The vibration data was recorded and used in the FDI analysis. Each turbopump was inspected after test firings and it was determined that each pump had different bearing failures and ball wear; thus, the goal of the wavelet-based FDI system was to correctly classify the different pumps to differentiate between the various faulty pumps and a normal pump.

The analysis was performed on 13.9 seconds of data collected from the accelerometer located at 135^0 from the pump inlet during the firing test at 104% rated power level. The distinct features were determined by comparing the CWTs of the normal operating turbopump against the faulty pumps. These features were then used to train a global neural network. The neural network consisted of two hidden layers and twenty-one neurons (fourteen neurons in the first layer, seven in the second layer). The neural network was trained on 130 feature vectors and classification was performed on 1391 feature vectors. The training was determined to be complete when the sum-square error goal of 0.02 was reached.

Table 3 shows a confusion matrix presenting the classification results. The normal operating pump is designated as "nor" and the faulty pumps are represented by their flight unit number. Table 3 indicates good classification results through wavelet/neural network based FDI that includes a probability of false alarm of 0.07% and a probability of missed detection of .12%, and a probability of fault misclassification of 2.19%. Important points to note here is that these results are the basic output classifications occurring every 10 ms without any deferral mechanisms or post classification algorithms performed. Simple deferral, averaging and trending techniques will provide nearly flawless detection and identification at the sacrifice of aggregating classification decisions that are occurring every 10 ms. Another, significant point is that the data was collected from physically different pumps. Hence these results illustrate the excellent performance that can be achieved using these methods when applied to different physical units of the same type (i.e., the interoperability of the FDI systems across different units is not a problem).

Gas Turbine Example: The application here focused on investigating trending and prognostics using wavelet-based methods for detecting gas turbine engine blade tip shroud failures. Endurance tests are performed on aircraft engines to evaluate and detect component failures by operating the engine at a variety of speeds over extended time intervals. During endurance testing of the F110-GE-129 at the General Electric (GE) Aircraft Engine Division, data was collected from six accelerometers. Accelerometers measuring horizontal vibrations and vertical vibrations were positioned at each of three different locations. These three locations included the front end of the fan frame, the rear end of the fan frame, and the rear end of the low pressure turbine frame (Figure 3).

TABLE 3. WAVELET / NEURAL NETWORK FDI RESULTS FOR HPOTP SSME

	Estimated Classes							
Correct Classes	nor	un2325	un2321	un2224	un2322	un4402	un4009	def
nor	1390	1	0	0	0	0	0	0
un2325	1	1334	3	18	0	20	15	0
un2321	0	2	1376	3	0	4	6	0
un2224	0	17	0	1371	0	0	3	0
un2322	3	0	0	0	1385	3	0	0
un4402	5	74	4	0	4	1293	11	0
un4009	1	10	2	0	0	4	1374	0

During the endurance tests, a shroud failure (i.e., partial separation of the shroud) occurred on a blade tip in the second stage of the low pressure turbine (LPT). This failure caused a rotor imbalance, which consequently generated engine vibrations. At the time of failure, the turbine was operating at approximately 8600 rotations per minute (rpm). GE Aircraft Division provided the raw data from all six accelerometer sites to ALPHATECH, Inc., for analysis. The purpose of the analysis was to determine whether the failure could be detected, and whether that failure could be predicted.

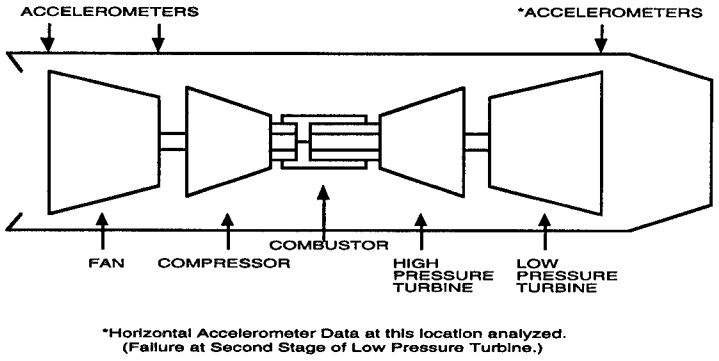


Figure 3: Instrumented Turbine Engine

Analysis of the GE turbine engine data began by examining the accelerometer data from all six accelerometer locations using time-scale analysis (i.e., CWT visualizations), while considering turbine engine mechanics and general turbine engine operation. Although the time-scale analysis indicated all six sensors detected the fault, the horizontal accelerometer at the turbine fan frame location generated the most prominent reaction to the blade tip shroud failure. The attenuated response at the remote accelerometers (relative to the LPT stage 2 turbine blade) is explained by the mechanical structure of the turbine engine.

Using data from the horizontal accelerometer at the turbine frame, time-scale analysis was performed on a four second region, eighty seconds prior to the shroud failure, and a four second period, one-hundred forty-one seconds following the shroud failure. By observing the Normal (i.e., pre-failure) operation CWT and comparing with the Fault operation (i.e., post-failure) CWT, several distinguishing characteristics appear. It was thus clear from the CWT visualizations that fault detection was possible.

As mentioned previously, the turbine speed at the fault instant was 8600 rpm, or 143.3 Hz. This frequency component is noticeably affected by the shroud failure, increasing in magnitude and bandwidth in the post-failure CWT. Additionally, the second harmonic of this frequency, approximately 280 Hz, appears with increased strength in the post-failure condition. Other frequencies of interest are noted by the post-failure decreases in magnitude at approximately 250 Hz, 500 Hz, 4550 Hz, and 6800 Hz.

Although CWT visualizations indicate that wavelet technology produces information to recognize the fault scenario, being able to reliably predict an impending failure before it actually occurs is a highly desired commodity in the turbine engine diagnostic community. The information obtained by comparing pre and post-failure CWT time-scale images was therefore used to train a global neural network. Once the neural network was satisfactorily trained (i.e., sum-squared error of .02), GE data from the failure period was applied to the neural network for processing and classification. A fifty-eight second interval was selected. The blade tip shroud failure occurred at approximately fifty-one seconds into this record.

Output from the neural network occurred every 10 ms during the course of the fifty-eight seconds monitored. The raw outputs of the neural network were processed by a set of causal algorithms (i.e., algorithms that could be used in an on-line, real-time monitoring system). Figure 4 presents the results of this processing.

As the top plot (Normal Indicator) of Figure 4 illustrates, at the beginning of the fifty-eight second interval, or nearly 40 seconds before the actual failure, the system is generating a strong indication of normal activity, with no indication (bottom plot, Fault Indicator) of any fault; however, almost immediately thereafter, the Normal Indicator begins declining at approximately the same rate as the rise in the Fault Indicator, thereby illustrating a system with a progressively increasing fault. By approximately second 26 on the relative time plots, both Normal and Fault Indicators are at a value of 0.5, strongly implying a faulted system, particularly when compared to the Normal operation as evidenced at relative time 0-5 seconds. By second forty, or eleven seconds before actual failure, the Fault Indicator, at value 0.7, heavily outweighs the Normal indicator at less than 0.3. By second forty, the system outputs have moved closer spatially to the fault condition.

The trending results are extremely promising for detecting faults before actual failure. By implementing thresholding logic upon globally trained neural network data as shown in Figure 4, system operation may be accurately characterized for reliable and robust prognostic systems.

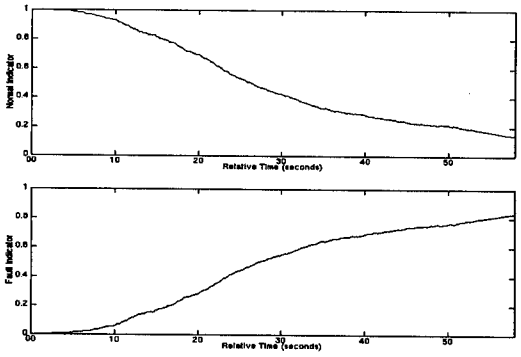


Figure 4: Fault Indicators for F110-GE-129 Turbine

SYSTEM ENHANCEMENTS: Wavelet visualizations in the form of CWTs from different sensor sites provide varying perspectives and insight into the mechanical operation of the transmission [10]. In fact, it is highly desirable to distribute the sensors around the transmission to guarantee observability of all important mechanical phenomena available from the vibration signals that could provide useful in health monitoring operations. Previous work has indicated that it is possible to achieve comparable, well performing, wavelet-based FDI systems from more than just a single sensor site [11]. An issue to consider would be how this information from multiple sensors might be combined to produce higher performing wavelet-based FDI systems.

One method for combining information from multiple sensor sites to achieve higher performing wavelet-based FDI systems would be to combine information extracted from the sensor sites at the feature vector level before performing classification. The next section on using multi-sensor wavelet-based differential features discusses such an approach and presents results achieved.

Another method for improved FDI consists of combining information from multiple sensors after the classification stage. A number of sensor sites can lead to reasonably well performing, single sensor, wavelet-based FDI systems. All such FDI systems are continuously making assessments about the operational condition of the platform (i.e., whether or not it is in a normal regime). When a failure is detected these systems, localized to an individual sensor site, are trained to decipher which particular fault the transmission is currently suffering from. However, different sites allow viewing the underlying mechanical phenomenon in different ways or from different aspects. Combining these multiple wavelet-based FDI results in some meaningful way may prove useful in developing health monitoring redundancy and increased robustness for an overall, higher performing, wavelet-based FDI system. The section dealing with hierarchical neural networks discusses one such method for combining individual sensor site wavelet / neural network FDI information for improved system performance.

Using Multi-Sensor Wavelet-Based Differential Features: To improve upon fault identification, multiple sensor wavelet extraction was investigated. The technique involves identifying a primary sensor site from which wavelet features are extracted. Using a secondary sensor site (or possibly multiple sensor sites), wavelet features at various scale settings are differenced against the primary sensor values. These differential features are conjoined to the original primary feature vector, thereby increasing its dimension. This enhanced set of feature vectors is then used as the basis for classification.

This method was applied to two channels of helicopter intermediate gearbox accelerometer data and the results were compared with results obtained using single channel data with no differential wavelet-feature augmentation. The single channel case resulted in 14 curvature/power features being extracted; however, when 15 additional differential features were added, the multiple channel case expanded to 29 curvature/power features.

To compare the separation power of the two feature sets, Fisher Linear Discriminants, which provide a one dimensional metric (linear functional) indicating the maximum separation between classes, were computed. The maximum separations computed using the differential feature element set were improved 15.2% with a resulting improvement in the FDI performance of 17.3% without deferral processing and 42.3% with deferral processing. For more details see [7].

Improved FDI Through Hierarchical Neural Networks: A natural question to be considered, in light of the fact that a select set of sensor sites have the ability to perform reasonable fault detection and identification, is the following: is there a reasonable way to combine or aggregate this FDI information from the individual sensor sites to produce a more reliable, robust, higher performing, overall FDI system? One method might be to collect the resulting FDI information from these wavelet / neural network systems and devise some ad-hoc methodology for integrating, aggregating or effectively combining the results to produce better fault detection and

identification decisions. Another method might be to defer making any decision on what this ad-hoc aggregating and/or combining algorithm should be and design a hierarchical neural network to figure out an appropriate strategy.

A hierarchical neural network would be designed to take information from a number of these complete, individual, wavelet / neural network FDI systems localized to a particular sensor. This hierarchical neural network would then be effectively trained to automatically devise an appropriate aggregation and/or combining algorithm (which could effectively be highly non-linear) to process multiple FDI data from multiple sensor sites with the goal of improving the overall robustness and performance of the eventual FDI system. Figure 5 provides a block diagram of the proposed hierarchical neural network system.

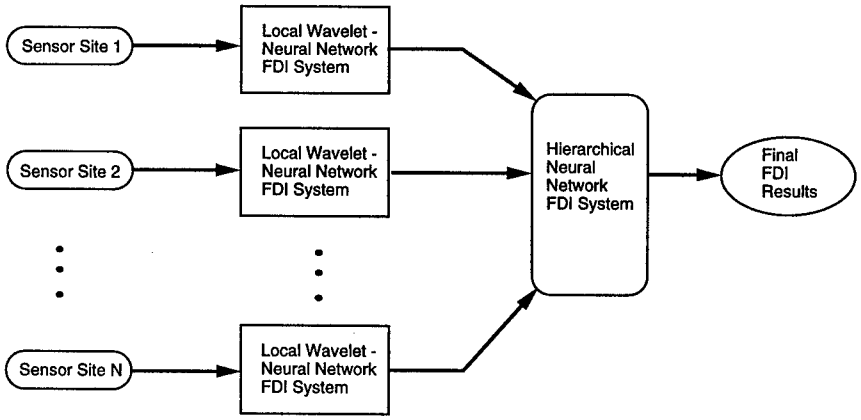


Figure 5. Block Diagram of Hierarchical Neural Network FDI System

Table 4 summarizes the neural network performance for helicopter main transmission vibration data that was recorded for normal operation and multiple fault conditions at three sensor sites. Results are given for the individual sensor wavelet/neural network FDI system (i.e., sensor 8, sensor 9, sensor 10) and the hierarchical system designed using all three sensor sites.

TABLE 4. PERFORMANCE METRICS FOR ALL FDI CONFIGURATIONS

Metric	Sensor 8	Sensor 9	Sensor 10	Sensor 8,9, & 10
PFA	0.0	0.002594	0.009079	0.0
PMD	0.003886	0.002594	0.0	0.0
PMC	0.003886	0.007782	0.01038	0.0

The networks were trained on thirty feature vectors from each fault class. Classification was conducted for each fault case across 2.52 seconds of vibration data. Table 4 shows the probability of false alarm (PFA), the probability of missed detection (PMD) and probability of misclassification (PMC). The Table 4 results indicate that the overall performance of the neural network improves as the sensor sites are combined. For more details on this methodology and simulations performed see [12].

CONCLUSIONS: Wavelet technology, when coupled with intelligent classification schemes constructed from neural networks, provides a basis for designing powerful fault detection and identification systems. This technology when applied to challenging vibrational systems, such as helicopter transmissions, resulted in highly accurate classification results. Similar results were obtained when applying these methods to other platforms such as turbopumps. These techniques have been extended to provide a promising prognostic / trending technique for failure prediction and fault severity indicators as demonstrated by the gas turbine application overviewed in this paper. Further robustness and improved system performance are achievable through aggregation and fusing methods across multiple sensors at both the feature extraction level and classification level. Finally, depending on the actual system requirements and needs, these basic wavelet / neural network FDI kernels, designed for real-time monitoring operations, can be integrated to support higher level diagnostic decision systems, or may be conjoined with other auxiliary information or systems in a fully integrated diagnostic/intelligent monitoring system.

REFERENCES:

1. Yan, Tinghu, Zhong, "Artificial Neural Network Technique and Its Applications to Rotating Machinery Fault Diagnosis", *J. of Vib. Engrg.*, Vol. 6, pp. 205-212, 1993.
2. Tansel, I.N., Mekdeci, C., and McLaughlin, C., "Detection of Tool Failure in End Milling With Wavelet Transformations and Neural Networks", *Manufacturing Science and Engineering ASME, Production Engineering Division*, Vol. 64, pp. 369-374, 1993.
3. Rohrbaugh, R.A., "Application of Time-Frequency Analysis to Machinery Condition Assessment", *Proc. 27th Asilomar Conf. on Sigs., Syst.* , Vol. 2, pp. 1455-1458, 1993.
4. Lopez, J.E., R.R. Tenney, and J.C. Deckert, "Wavelet Feature Extraction For Real-Time Neural Network Condition Based Maintenance," *Proc. US Navy Conference on Neural Network Applications*, Arlington, VA, 16-17 June 1994, pp. 109-121.
5. Lopez, J.E., Deckert, J.C., Tenney, R.R., "Condition-Based Machinery Maintenance: Interim Report", Alphatech, TR-663, August 1994.
6. Lopez, J.E., R.R. Tenney, and J.C. Deckert, "FDI Using Real-Time Wavelet Feature Extraction," *Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Philadelphia, PA, 25-28 October 1994, pp. 217-220.
7. Lopez, J.E., R.R. Tenney, and J.C. Deckert, "Improved Fault Identification Using Multi-sensor Wavelet-Based Differential Features," *Annual Symposium on Machinery Failure Prevention Technology*, 18-20 April 1995, Virginia Beach, VA.
8. Lopez, J.E., Mc Kenney, A., "Next Generation Testing and Machine Monitoring Systems Based on Application of Wavelet and Neural Network Technologies," *Test Technology Symposium XXI: Testing in lhe 21st Century*, Laurel, MD, pp. 11-13 April, 1995.
9. Lopez, J.E., Polyak, N., "Wavelet-Based Diagnostics for Helicopter Main Transmission", *American Helicopter Society 51st Annual Forum*, Fort Worth, TX, 9-11 May 1995.
10. Lopez, J.E.,"Performance of Wavelet / Neural Network Fault Detection Under Varying Operating Points", *Proceedings of the 66th Shock and Vibration Symposium,* Vol. 1, pp. 209-217, Oct.30 - Nov. 3, Biloxi, MS 1995.
11. Lopez, J.E., Oliver, K., "Improved Analysis Tools for Wavelet-Based Fault Detection", *IASTED International Conference, Signal and Image Processing - SIP- 95*, Las Vegas, NV, November 20-23, 1995.
12. Lopez, J.E., Yeldham, I.F., Oliver, K., Protz, M., "Hierarchical Neural Networks for Improved Fault Detection Using Multiple Sensors", *to be presented at the American Helicopter Society 52nd Annual Forum*, Washington, D.C., June 1996.
13. Grossman, A., R. Kronland-Martinet, and J. Morlet, "Reading and Understanding Continuous Wavelet Transforms," in *Wavelets, Time-Frequency Methods and Phase Space*, J.Combes, et. al. (Eds.), Springer-Verlag, 1989.
14. Smith, J.D., *Gears and Their Vibration*, MacMillan Press Ltd.,1983.

ACKNOWLEDGMENTS: This work was supported by the Office of Naval Research under contract N00014-93-C-0077.

HELICOPTER TRANSMISSION DIAGNOSTICS USING VIBRATION SIGNATURE ANALYSIS

B.E. Parker, Jr., M.P. Carley, R.J. Ryan, and M.J. Szabo
BARRON ASSOCIATES, INC.
3046A Berkmar Drive
Charlottesville, Virginia 22901-1444

and

H.V. Poor
PRINCETON UNIVERSITY
Department of Electrical Engineering
Princeton, New Jersey 08544-5263

Abstract: Fault diagnostics represent a vital task in the monitoring of mission critical systems, as well as for condition-based maintenance of machinery in general. The focus of this report is on the early detection, and subsequent classification, of small changes in the behavior of mechanical systems. Such changes, known as *incipient* faults, portend the development of more serious failures.

Physical models of machinery processes, which are useful for model-based fault detection and isolation, are not generally available in most applications. Instead, the approach to fault detection considered in this study involves the application of *statistical change detection.* Statistical change detection is essentially the problem of homogeneity testing within a time series. In particular, statistical change detection algorithms seek to detect situations in which a given model that describes the initial behavior of a time series, eventually fails to describe that time series accurately. The performance of non-likelihood-ratio techniques are evaluated on a CH-47D helicopter combiner transmission (non-seeded) fault; results indicate that the fault is detected in its incipient stage.

The approach to fault *isolation* (i.e., classification) discussed herein is based on the use of minimum-logistic-loss polynomial neural networks (PNNs). The fault isolation capabilities of PNN classification networks are investigated using seeded-fault data taken from CH-46E helicopter combiner transmissions. Perfect fault classification results are achieved.

Key Words: Condition-based maintenance, Fault diagnostics, Health monitoring, Incipient faults, Nonparametric statistics, Pattern recognition, Polynomial neural networks, Statistical change detection

Introduction: Mechanical system maintenance is inherently a safety issue for critical systems such as helicopter transmissions, since even a single in-service failure is intolerable. The maintenance strategy in such systems is made more difficult by the need to minimize unnecessary inspections, since this requires expensive teardown and rebuild procedures. Alternatively, it is also expensive to have unexpected unavailability of such systems. What is needed to address this dilemma in commercial, industrial, and military systems, such as aviation and manufacturing process control, is transition from a "time-based" to "condition-based" maintenance (CBM) strategy. Such a transition is critical because of the costs associated with maintenance. Enterprises employing such diagnostic capabilities would realize enhanced responsiveness, competitiveness, profitability, and consumer image. The need for diagnostics capabilities is, if anything, even greater for military systems, since maintenance requirements during military conflict impart a tactical disadvantage. At such times, spare aircraft, for example, may simply not be available.

In the case of helicopters, on-board health and usage monitoring systems (HUMS) could be integrated with groundstations, analyzing data that are collected and stored for post-flight analysis. Affordable computer workstations could readily provide the needed processing power (special-purpose *hardware* is *not* needed). For most mechanical system faults, *need* for an on-line (i.e., airborne) system would be determined by the possible risk of the off-line (i.e., groundstation) system not detecting a fault sufficiently early for the operator to take maintenance action, or not having significant ground support available (such as the situation of forward deployment). Rare faults can initiate and progress to complete failure within the time of a single mission (e.g., < 4 hours), so a trade-off must be made between the costs and benefits of on-board (on-line) and post-flight (off-line) analyses.

The goal of this work is to develop capabilities that will lead to affordable systems that can diagnose faults sufficiently early to significantly enhance safety, improve the likelihood of accomplishing mission objectives, prevent the loss of assets, and reduce maintenance costs.

Diagnostics: Model-based fault detection and isolation (FDI) is the method of choice when physical models of machinery processes are available (see, e.g., [6]). With the present state of the art, however, such models are available only for very simple and idealized mechanisms, which do not capture the necessary complexities of real-world processes. Even relatively simple gearboxes, such as those of helicopter tail rotors, can have a dozen or more shafts when auxiliary equipment is included, along with dozens of gears and bearings.

Although pattern-recognition techniques are often used to implement fault detectors via classifiers that process feature-set data, such an approach is not entirely satisfactory. First, it is not possible to ensure with *ad hoc* feature selection that one has chosen a sufficient feature set with which to distinguish normal behavior from all possible abnormal behaviors. Even if the features selected perform well on the available training and test data (a necessary condition), this provides no information regarding the sufficiency of such features for detecting previously unseen types of faults. Validation of the feature set using available data does not ensure that abnormal data will not be encountered that the system is unable to classify correctly.

Second, pattern-recognition techniques used for detection that rely on extracted features often require access to a significant amount of training data to achieve acceptable performance. In many applications such data are expensive to acquire (e.g., flight vehicles) and may therefore be available only in very limited quantities. Additionally, feature-based approaches require an inordinate amount of labor for designing each detector/classifier system, and it is unlikely that features engineered for one system can be used reliably in other systems of different design. Even though the same features may, in part, be useful, the entire process of validating the features on test data has to be repeated for each such system, once again with no real assurances regarding their sufficiency.

A better approach, both from the viewpoint of increasing the reliability of detection and in reducing development costs (through reduced engineering manpower needed to synthesize such detectors), is to use a general methodology that can be applied more readily to different systems. *Statistical change detection* represents such an approach that can be used for the early detection of small changes in systems. Statistical change detection does *not* require a database of fault examples, allowing *novel* situations and faults to be detected. Following fault detection, fault classification may be performed by utilizing pattern recognition techniques (e.g., neural networks). With fault classification, "feature engineering" also may be avoided by using parameters of the models employed in the statistical change detection algorithms as "features;" these parameters would need to be adapted on line so as to continually "fit" the data. (Note that the detector would have to be isolated from such adaptation, since the detector relies on a fixed whitening model — even if all of its coefficients are set to zero.) The motivation for such use of the parameters is that it obviates the need to seek *ad hoc* feature sets for fault classification in machinery, which is the approach currently taken in nearly all diagnostic systems that are not based on parametric models. The justification for such use of these parameter values is that many modern spectral analysis techniques are based on nonparametric (e.g., autoregressive, moving average, autoregressive-moving average, etc.) models; these techniques utilize model parameter estimates to obtain spectral densities.

Statistical Change Detection: Generally speaking, the detection algorithms consist of two parts: signal processing to reduce the observed times series data to a single real number, i.e., the computation of a *detection statistic*; and the comparison of this real number to a threshold to make a decision. The signal processing can, in principle, be implemented with no *a priori* data, by synthesizing whitening filters (both their structure and parameter values) on-line. The use of neural network algorithms can be particularly beneficial here. The second part of detection requires determination of the detection threshold, for which access to some *a priori* data is generally needed, but in certain cases this threshold can be set using theoretical values that do not depend on data (e.g., use of χ^2 statistics); additionally, useful thresholds can also be learned adaptively on-line by monitoring false-alarm statistics. In monitoring mission critical systems for potential faults, it is vital that false alarms be minimized, since mission abort decisions have their own price (e.g., when aborting a mission over water or hostile territory).

Statistical change detection thus provides a framework within which vibration monitoring of mechanical systems can be approached. Discrete-time change detection problems can be formulated as parametric hypothesis testing problems based on a series $y_1, y_2, \ldots, y_n$ of random measurements. We assume that these measurements are generated by a statistical model M_0 up until an unknown time $t_0 - 1$, and then that they are generated by another statistical model M_1 thereafter.

There are two types of change-detection paradigms of interest: off- and on-line detection. In off-line change detection, the problem of interest is to examine a fixed-length set of measurements $y_1, y_2, \ldots, y_n$, and to decide whether or not $t_0 \leq n$. In this framework, the detection criterion is to maximize the probability of detecting a change, within a constraint on the false-alarm probability. In the on-line detection framework, the problem of interest is to continuously monitor the observations to detect the change point t_0 as quickly as possible after it occurs, again within a constraint on the allowable rate of false alarms.

Because of the similarity between off- and on-line change detection methodologies, the signal-processing algorithms used in the two approaches are similar structurally, and in fact can be identical if computational issues are not of concern. Where the algorithms differ is in the decision-making after the signal processing has been performed. This difference in decision-making procedures stems from the difference in their performance criteria: maximum detection sensitivity for off-line algorithms (maximization of the power of the test subject to the constraint of a fixed probability of false alarm), and quickest detection for on-line algorithms (minimization of the delay for detection for a given mean time between false alarms).

For general change detection, the detection statistic is based on a fixed-length sliding window of observations $y_{n-n_0}, y_{n-n_0+1}, \ldots, y_n$, and is a function of the form:

$$\Delta^n = \max_{n-n_1 \leq j \leq n-n_0} \Delta_j^n, \tag{1}$$

where, $n_1 - n_0 + 1$ is the length of the sliding window and n_0 is selected arbitrarily. The detection statistic Δ^n is suitable, in the general case, for deciding whether or not $t_0 \leq n$. For each $j = 1, 2, \ldots, n$, Δ_j^n is a suitable detection statistic for deciding between the hypothesis that the model changes from M_0 to M_1 at exactly time $t_0 = j$ versus the hypothesis that the model does not switch at all during the n observations (i.e., that $t_0 > n$).

The explicit structure of the detection statistic depends on the two models M_0 and M_1, on the nature of the difference between them, and on the complexity that can be tolerated by the detection system. In this context, there are four general types of detection statistics: the *log-likelihood ratio*, the *generalized likelihood ratio*, the *locally optimum* statistic, and *non-likelihood ratio (NLR)*-based statistics. These statistics were reviewed recently in [11, 12]; extensive comparisons of their performance were also provided via simulations. Emphasis herein will be placed on NLR-based detection statistics since these require the fewest assumptions; in particular, investigation is made into the Zhang, Basseville, and Benveniste (ZBB) algorithm [19, 12], and of the Basseville-Nikiforov (BN) algorithm [2, pp. 415 - 417].

After computation of the detection statistic from Eq. (1) through any of the above methods, the decision algorithms for change detection are quite simple. For off-line change detection, the presence of a change during the observation time window is announced only if $\Delta^n > \tau_{\text{off}}$, where τ_{off}

is a threshold set as small as possible while satisfying a false-alarm constraint, $P(\Delta^n > \tau_{\text{off}}|M_0) \leq \alpha$, where $P(\cdot|M_0)$ denotes probability computed under model M_0, and α is the desired constraint on the false-alarm probability. Alternatively, for on-line detection, a change from model M_0 to model M_1 is announced at the *alarm time*, t_a, given by $t_a = \min\{n|\Delta^n \geq \tau_{\text{on}}\}$, where τ_{on} is a decision threshold chosen to control the rate of false alarms. In general, the threshold in this process must be chosen to balance the desire for detection efficiency — i.e., a low threshold to achieve quick detection in on-line cases, high probability of detection in off-line cases — (which would indicate a low threshold) with the high threshold needed to minimize false alarms. The choice of threshold requires knowledge of the probability distribution of the statistic Δ^n under the model M_0. It is usually not possible to determine this distribution in most problems of vibration monitoring in mechanical systems; however, estimates based on Brownian motion approximations to the statistics Δ_j^n are available in most cases (see, e.g., [2, 19]). Useful thresholds can also be learned adaptively on-line by monitoring false-alarm statistics.

Nonparametric Models: In selecting a modeling approach, we are concerned with models whose function is to whiten (i.e., remove correlation from) the observed process (e.g., sensed vibration data), to produce an innovation or residual sequence that is used by the detection algorithm. For this purpose, *any modeling approach that is practical can be used to fit the data*; estimation neural networks are particularly useful here. The model should not, of course, overfit the data; moreover, parsimonious models are generally desired to allow fast on-line parameter estimation.

The consideration of external inputs as *nuisances* is paramount in the modeling of time-series data of systems where these inputs are varying, but for which no observability is provided (e.g., changing helicopter flight regime). *Instrumental variables (IV)* techniques exist that provide for parameter estimation under such circumstances (see, e.g., [7]). In particular, Basseville and Nikiforov [2] have shown that "the vibration monitoring problem is nothing but the problem of detecting and diagnosing changes in the eigenstructure of a nonstationary multivariable system in state-space form, or equivalently, in the AR part of a multivariable ARMA model with a nonstationary MA part." Because the AR parameters of systems are associated with its dynamics and the MA parameters with the (generally time-varying) input excitation, the latter are often treated as nuisance parameters and the AR process estimated using IV techniques. Instrumental variables techniques are especially important when reduced-order process models are used, as under-parametrization of models can result in situations where changes in the estimated system dynamics reflect only variations in conditions that are not being monitored [19].

It is also important to note that high-accuracy models are generally *not* needed for detecting incipient faults and small changes in machinery conditions; as a result, even non-harmonic processes, which are difficult to model well with a reasonably small number of parameters, can be monitored for purposes of detecting changes. Model consistency, at whatever level of accuracy, is the important attribute.

Second-order statistics are, in general, adequate when data being fitted are Gaussian. When data are non-Gaussian, or represent the output of a nonlinear process, higher-order statistics (HOS) may be more useful in characterizing a process. Part of the elegance of the statistical change detection approach is its ready extensibility to include higher-order statistical moments where necessary to exploit additional information in the data. In order to extend these methods to HOS, it is necessary only to produce a parametrized model based on HOS. One can then apply a generalization of the methodology used for second-order statistical estimation for the parameters of this model.

An essential element of the statistical change detection algorithms is that, although they too may exploit only second-order statistical information, they do *not* require reducing the "feature set" further. If higher-than-second-order statistics are to be used, feature-based methods grow even more problematic, for example, as bi-spectra now have on the order of N^2 spectral lines, and tri-spectra on the order of N^3 spectral lines, which must be reduced severely for inputting to a classifier.

Fault Isolation: To isolate (i.e., classify) faults, *classification* neural networks that employ a constrained minimum-logistic-loss criterion are most appropriate. The Barron Associates, Inc. *Algorithm for Synthesis of Polynomial Networks for Classification (CLASS)* [4] is used herein to

synthesize these neural network classifiers. With *CLASS*, network outputs represent true estimates of the *a posteriori* probabilities of class membership.

Using an *estimation* neural network to perform classification is not optimal since it imposes unnecessary constraints on the solution [16]; for example, for binary classification, the network may be trained arbitrarily to output a "one" for a fault, and a "zero" for normal data. In essence, use of the squared-error loss function corresponds to the maximum likelihood rule only in the case of a Gaussian probability model for the distribution of the errors [7]. However, for multiclass classification problems with categorical variables, a multinomial probability model in regular exponential form is more suitable than the Gaussian model [1]. This approach is based on the estimation of nonparametric probability density functions using minimum-logistic-loss polynomial neural networks. The advantage of this approach is that the decision surfaces are more general and reflect distributions found in the data. In this case, the *CLASS* subnetwork (polynomial) functions are used to model the log-odds associated with the conditional probability of each class given the observed inputs. In this setting, the maximum likelihood rule corresponds to the choice of the logistic-loss function. Additionally, logistic discrimination has been shown to perform well on both Gaussian and non-Gaussian data. Another advantage of the general multiclass logistic model in regular exponential form is that, for classification problems, it forces satisfaction of the probability constraints $0 < p_i < 1$ and $\sum_i p_i = 1$.

In summary, the constrained minimum-logistic-loss criterion, which is explicitly designed for classification problems, provides performance superior to classifiers fitted using estimation criteria. Estimation networks place emphasis on estimation accuracy; minimum-logistic-loss networks instead place emphasis on maximizing the likelihood of correct class discrimination. Whereas true probabilities are always between 0 and 1, estimation networks are unbounded; in contrast, the logistic-loss criterion correctly maps the network outputs onto [0,1]. A significant advantage of the minimum-logistic-loss classifier is it gives the system a complete view of the problem at hand, with the coefficients in all nodes fitted simultaneously to the entire synthesis data set, instead of using separate fitting of partitioned data sets. This property also forces the trained nodes to be consistent with each other. The logistic-loss network is a completely general way of reflecting the natural distribution of the data, without imposing any assumed structure on the data.

Ensemble Processing for Cyclostationary Signals: The models discussed in the preceding sections were assumed to be stochastically stationary (within pre-change or post-change regimes); that is, their underlying statistical behavior was assumed to be invariant to arbitrary translations in time. However, some of the data sets considered in this study are not stochastically stationary, but rather they are *cyclostationary,* by which we mean that their underlying statistical behavior is invariant to time translations that are integral multiples of a basic time period, C. (See, e.g., [5] or [9] for a discussion of cyclostationarity.) For example, in the monitoring of a helicopter gearbox, cyclostationarity results from the cyclic motion of the gear shaft. In this section, we discuss modifications of the statistical change detection techniques for use on such cyclostationary signals.

It should be noted that a change in the statistics of a signal may not be manifested in the same way in every phase of the cyclic structure; evidence of a flaw in a gear, for example, may appear only in phases of the measurement signal during which the flawed part of the gear is engaged. Thus, processing the different phases of the signal together may reduce the detectability of such flaws. For this reason, it is of interest either to consider separate processing of the different phases of the cyclostationary signal, or to consider joint processing techniques that view the different phase signals as components of a C-dimensional vector time series.

In the first case, we can essentially treat each phase as an independent channel that can be processed by any of the scalar means described in the preceding sections. In this approach, the channels are combined only after per-channel processing has been performed. This combining can be either pre-decision or post-decision. For example, a detection statistic $\Delta^n(c)$ can be computed for each channel $c = 1, 2, \ldots, C$, and then threshold comparison can be performed with the combined statistic

$$\max_{1 \le c \le C} \Delta^n(c). \tag{2}$$

Alternatively, each per-channel statistic $\Delta^n(c)$ can be compared with a per-channel threshold λ_c, to produce C per-channel decisions, which can be combined to produce an overall decision. Such

approaches will be designated *ensemble* approaches. The goal of ensemble processing is to improve fault detectability through signal-to-noise ratio enhancement resulting from coherent averaging. One such method in which a ZBB/NLR algorithm is used to perform per-channel processing is demonstrated below.

Boeing Helicopter Gearbox Data: The Boeing helicopter gearbox data analyzed in this report are based on vibration measurements performed on a CH-47D helicopter combiner transmission spiral bevel input pinion gear that exhibits a classic bending fatigue failure as a result of test-cell overloading (156 percent rated torque). Therefore, this data does *not* represent a seeded-fault test. Data collection techniques and other analyses performed previously on this data have been reported in [13].

The data available for this study were comprised of two accelerometer channels digitized from high-speed analog tape recordings at a sample rate of 121,212 Hz; one accelerometer was mounted on the combiner collector gear along an axis parallel to the output shaft center line. The other accelerometer was mounted on the lefthand combiner input pinion on an axis perpendicular to the input shaft center line. Both accelerometers had bandwidths of five to 20,000 Hz. A tachometer signal was also available that provided a pulse for each revolution of the gear experiencing the bending fatigue failure. The 23 Boeing data records available represent consecutive (but not contiguous) 30-second segments extracted from each minute of time-series data; these data segments initially represent normal operating conditions, and eventually include the initiation of a gear microcrack at the root of a tooth, and its progression to a complete tooth failure, at which time testing was terminated due to a sudden increase in the test cell external noise level.

Analysis Results Using Stewart Hughes Ltd. MSDA Analyzer: Based on metallurgical analyses of the gear tooth that fractured in this experiment, Boeing has determined that a microcrack initiated approximately 13.25 minutes prior to complete gear tooth failure and termination of the test data recording [13]. Alignment of the data available in the present study with that of the Boeing analyses indicates that the microcrack initiated midway in file #12.

Analysis of the same data by Boeing using the Stewart Hughes Ltd. Mechanical Systems Diagnostic Analyzer (MSDA) [13] led to post-fault initiation detection delays ranging from 2.85 minutes to 9.45 minutes (detection in files #15 and #22, respectively), depending on the failure indicator (i.e., figure-of-merit (FOM)) monitored. These time delays are based on the ten MSDA FOM, out of the 32 investigated, that showed positive indications (i.e., significant and continuous change from the long-term value developed during a period of normal operation) of the microcrack failure progressing with time.

Statistical Change Detection Using Time-Average Statistics: Because the data files are not contiguous segments of the original recording, an off-line change detection approach was used. Each of the available data files #1 - #23, with the exception of file #2, was subdivided into 139 segments of 25,000 observations each; data file #2 was not used in these analyses because one of the two digitized accelerometer channels was not readable by computer. Close examination of the raw data also revealed several sections of the second accelerometer (i.e., channel 7) time-series data that were saturated. In particular, files #6 and #20 contained regions in which the data appeared to be magnified in bursts relative to the majority of the recorded data. These sections were, therefore, eliminated. The first 25,000-sample segment of file #1, which represents normal operational data, was reserved for estimating parameters and training (i.e., learning the mean and covariance values used in computing the test statistics during evaluation).

Using the BN/NLR estimation algorithm [2], the AR parameters of a two-dimensional ARMA (2,1) process were found; the BN/NLR algorithm requires that the order of the MA process, q, be one less than the order of the AR process, p, thus $q = p - 1$. The MA parameter(s) are then not actually estimated, but suppressed using instrumental variables (IV) estimation techniques. Subsequently, all of the other segments of the two channels of data were evaluated individually based on the information provided by the training data sequence.

Based on the ratios of the mean values of the detection statistics, $\overline{\Delta^n}_{ev}/\overline{\Delta^n}_{tr}$ (where the subscripts ev and tr imply evaluation and training data respectively), which are provided graphically in Fig. 1, it is seen that the fault appears to be detectable first in file #20; it is detectable consistently

thereafter (i.e., in files #21 - #23). Based on this result, the time duration between the estimated time of fault initiation (i.e., midpoint in file #12) and the time of detection (i.e., detection delay) was determined from the test time interval data to be approximately eight minutes.

Off-line change detection was also performed on the univariate (i.e., scalar) time-series data; these results are illustrated in Fig. 2. The results for the univariate algorithm are slightly worse than those for the multivariate algorithm in that the fault appears to be detectable consistently by file #21. The BN/NLR algorithm scalar detection results therefore indicate a detection delay of approximately nine minutes.

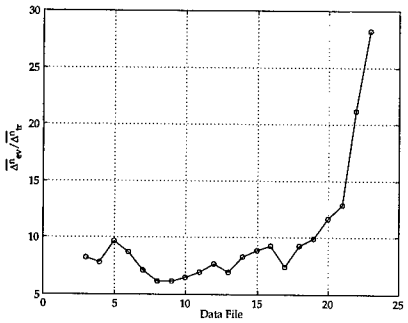


Figure 1: Off-Line Boeing Helicopter Gearbox Results (BN/NLR Multidimensional Algorithm)

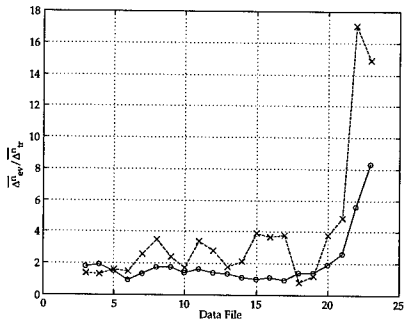


Figure 2: Off-Line Boeing Helicopter Gearbox Results (BN/NLR Algorithm); solid line represents accelerometer channel 3, dashed line accelerometer channel 7.

Statistical Change Detection Using Ensemble-Average Statistics: Analysis was performed also using the ZBB/NLR Ensemble algorithm. This algorithm differs from the time-average algorithm in that ensemble averages are used to calculate the detection statistics. The tachometer signal was used to divide the data into distinct records; l additional observation samples from the previous record were added to the beginning of each new record so that detection statistics could be calculated for the first samples of each record. These records were grouped into ensembles containing M records each. Breaking the data up in this way allows the basic statistics to be phase dependent, providing C different values, one for each phase of a record.

The use of ensemble averages is intended to help remove background noise, making the detection statistic more sensitive to change detection. Also, using a nonstationary bias and covariance matrix helps to improve modeling accuracy. A distinct detection threshold, λ_c, may be chosen for each phase within a record, further increasing the sensitivity of the detector.

Unfortunately, the time interval between tachometer pulses was not always constant, providing record lengths that varied slightly due to test stand speed fluctuations (+0.34% to −1.01%). To perform the off-line analysis, the two most common record lengths, 591 and 592 samples respectively, were selected and then decimated by a factor of two, to achieve equal record lengths of 296 samples (plus the l extra samples added to the beginning of each record). Use of this procedure allowed about 70 percent of all data records to be included in the analysis. More sophisticated techniques for synchronously averaging a signal are discussed in [14]. The decimated records from the first two files were grouped into ensembles containing $M = 50$ records. These ensembles were used, along with an AR(10) model, whose parameters were set to zero, to estimate the training parameters. The remaining data were then evaluated using ensembles containing $M = 100$ records.

The resulting statistics are graphed in Fig. 3. The ratio of the means of the test statistics for accelerometer channel 3 show a potentially significant change from the earlier statistics for data

files #14 - #23. The fault was not as manifest on accelerometer channel 7, as the statistics do not show substantial change from a stable baseline until file #20. The results using the ensemble-average ZBB/NLR algorithm, therefore, provide for a best-case detection delay of two minutes, better than the other algorithms, including the MSDA analysis.

These results are based on use of the mean of the detection statistics computed over the entire record, which does not fully exploit the algorithm. A set of C thresholds, λ_c, one for each sample time (i.e., phase) within a record, were therefore established; each was arbitrarily set equal to the maximum value of the detection statistic seen in files #1 - #8. To avoid having to plot C different detection statistics, fault detection was based instead on the maximum number of consecutive times *any one* of the C phase-dependent detection statistics exceeded its corresponding λ_c. With this method, a significant change in the detection statistics may be seen as early as file #11, as shown in Fig. 4. This suggests that the microcrack may have initiated even earlier than the Boeing metallurgical analyses may have indicated, a finding that is not inconsistent with the approximate nature of such analyses [15].

For purposes of comparison, results were obtained next for the time-average ZBB/NLR algorithm using the same training and evaluation data and window lengths that were used to obtain the results for the ensemble-average ZBB/NLR algorithm. For training and evaluation the decimated data was used and M was set equal to 100. The ensemble-average ZBB/NLR algorithm (see Fig. 4) was found to detect the fault up to eleven minutes before the time-average ZBB/NLR algorithm. Thus, ensemble averaging clearly improves detection performance.

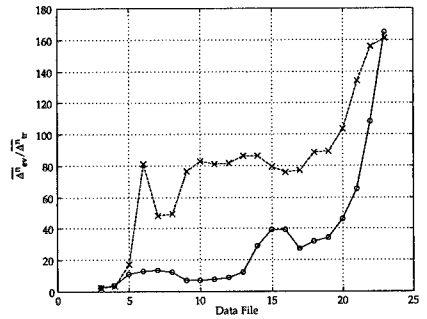


Figure 3: Off-Line Boeing Helicopter Gearbox Results (ZBB/NLR Ensemble Algorithm; solid line represents accelerometer channel 3, dashed line accelerometer channel 7.)

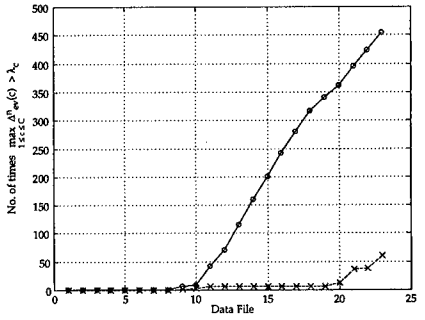


Figure 4: Off-Line Boeing Helicopter Gearbox Results (ZBB/NLR Ensemble Algorithm with Multiple Thresholds; solid line represents accelerometer channel 3, dashed line accelerometer channel 7.)

Fault Classification: Another important aspect of helicopter gearbox diagnostics is fault isolation (i.e., classification); in other words, determining the type of fault present once a fault has been detected. Contemporary fault classification, using neural networks, involves three basic steps. The first is feature extraction, in which pertinent features that can be used to distinguish one fault category from another are drawn from the data. A common method for extracting features is the short-term Fourier transform (STFT). Feature selection is the second step in classification. To simplify the neural network classifier and to avoid overfitting the data, the number of features input to the network must be kept to a minimum. Therefore, principal component analysis (PCA) [8] was used to reduce the dimension of the feature vector. The final step is synthesizing a classification neural network; here, the Barron Associates, Inc. *Algorithm for Synthesis of Polynomial Networks for Classification (CLASS)* [4] was used.

Classification Results for Westland Helicopter Gearbox Data: The Westland helicopter gearbox data analyzed in this report are part of a data set collected on the Westland Helicopter Universal Test Rig (see [17, 18]). The test rig was used to measure the vibration of a CH-46E helicopter combiner transmission under several different test conditions. These conditions were partitioned into nine different categories, as listed in Table I. Since "Fault Types" one and nine both represent "No Defect" cases, these categories were combined and are referenced as a single class (*viz.* Class 1).

Table I: Westland Helicopter Gearbox Data Description

Fault Type	Description
1	No Defect
2	Planetary Bearing Corrosion
3	Input Pinion Bearing Corrosion
4	S.B. Input Pinion Spalling
5	Helical Input Pinion Chipping
6	Helical Idler Gear Crack Prop.
7	Collector Gear Crack Prop.
8	Quill Shaft Crack Prop.
9	No Defect

Table II: Westland Helicopter Data Files, Fault Type vs. Torque Level

	Torque Level (%)									
Fault Class	27	40	45	50	60	70	75	80	100	Total
1	0	0	0	0	0	0	0	2	0	2
2	0	0	0	0	0	0	0	0	3	3
3	4	4	4	3	3	5	5	3	4	35
4	3	3	4	3	2	4	2	4	5	30
5	0	0	0	0	0	2	2	2	1	7
6	0	0	0	0	0	1	4	4	6	15
7	1	1	0	2	4	2	4	4	3	21
8	5	5	5	5	5	7	5	5	3	45
9	2	3	3	2	3	2	3	3	3	24
Total	15	16	16	15	17	23	25	27	28	182

Data were collected at different torque levels between 27 percent and 100 percent of the full torque level. In total, 182 different experiments were run. Table II shows the number of files obtained for each fault category as a function of the different torque levels. Each data file consists of approximately 22.6 seconds of data sampled at 100 kHz. Ten channels of data were recorded, which included eight accelerometers, one tach pulse, and one test signal. The tachometer was placed on the aft transmission in place of the rotor position motor. The tach signal is a 256 pulse-per-revolution signal with a once-per-revolution signal superimposed on it. Based on its position in the gearbox, one revolution describes a complete rotation of the rotor position output, not that of the main shaft.

All of the available data files, except for file #24 on tape 1, were divided into one-revolution periods using the tachometer signal. File #24 was not utilized because it contained sections that were unreadable. With the 100 kHz sampling rate, there were between 897 and 904 samples within the period defined by the tachometer signal; all of these samples were used and then zero-padded to obtain 1,024 samples per period. Subsequently, 1,024-point power spectral density (PSD) estimates were computed for each period, after smoothing the data using a Kaiser-Bessel window. The PSDs were then clustered into groups of 20 and ensemble averaged. The resulting data vector of dimension 512, however, was too large for practical computation of the necessary PCA transformation matrix. Therefore, the PSDs were decimated by a factor of three to ease the computational burden; this resulted in reducing each PSD from 512 values to 179. The data were not decimated in the time domain to avoid the effects that this would have on the STFTs. Decimation by three in the time domain would cause loss of all information in the upper two-thirds of the STFT,[1] whereas decimation in the frequency domain maintained information at

[1]For example, with a 100 kHz sampling rate, spectral values are obtained from zero to 50 kHz; at a sampling

all frequencies, albeit at a courser resolution. The decimation operation is similar to averaging adjacent spectral lines to form a reduced set that still spans the original bandwidth. The resulting features were next normalized over the interval [0,1]. Lastly, the data were split into training and evaluation databases, each containing roughly half of the exemplars. To test the robustness of the classification algorithms, three different methods were used to effect this division; in the first approach, different data *files* were assigned randomly to the training and evaluation databases; in the second approach, data *exemplars* were apportioned randomly; in the third approach, data *exemplars* from the first half of individual time series were used for training, and exemplars from the second half were used for evaluation.

Since data were available at multiple torque levels within each fault category, a method was needed to reduce the within-class variance caused by changing torque levels, while simultaneously maintaining the between-class variance. Methods to accomplish this have been used before in sonar target recognition. In one study [3], a method was needed to reduce variation associated with the aspect angle of a target, while still being able to distinguish between different targets. In the present case, PCA was used to extract features within a class that tended to vary least with respect to torque level; these features are associated with the *smallest* eigenvalues. This process was used to find the 50 principal components within each class; these principal components account for the least amount of variance. Use of eight different one-vs.-all, binary-output neural networks, each specialized to recognize a single fault (or no-fault) class, were synthesized using *CLASS* [4] to provide a capability to distinguish between classes.

As discussed earlier, *CLASS* neural networks output estimates of the true *a posteriori* classification probabilities. To use the eight neural network outputs to reach a classification decision, the highest output probability was deemed to indicate the correct class, and the given exemplar was said to be of that class.

Use of any single accelerometer channel was found to lead to at least a few misclassifications. Therefore, all eight accelerometer channels were employed. To enable utilization of all accelerometer channels, eight binary-output neural networks were synthesized for each fault category (one for each accelerometer channel); the corresponding network output probabilities within each class were then averaged. This therefore required training of 64 different neural networks. A final fault category decision was reached based on the resulting probabilities in the same fashion as with the single channel method — i.e., that network whose output was largest was selected as winner.

Use of all accelerometer channels led to perfect classificatory results on both the training and evaluation data (where the latter were not used in training the classifier). Indeed, for *all* three partitionings of the data considered herein, interrogation of the neural networks with both training data and evaluation data exemplars always produced perfect classification results (i.e., diagonal confusion matrices). A typical confusion matrix, shown here for the random *file* partitioning, is provided in Table III.

Table III: Westland Helicopter Data Evaluation Confusion Matrix (Multiple Channels, Random File Partitioning)

True Class	Model Output 1	2	3	4	5	6	7	8	Total
1	1500	0	0	0	0	0	0	0	1500
2	0	250	0	0	0	0	0	0	250
3	0	0	2125	0	0	0	0	0	2125
4	0	0	0	1875	0	0	0	0	1875
5	0	0	0	0	375	0	0	0	375
6	0	0	0	0	0	875	0	0	875
7	0	0	0	0	0	0	1250	0	1250
8	0	0	0	0	0	0	0	2625	2625
Total	1500	250	2125	1875	375	875	1250	2625	10875

rate of 33.3 kHz, spectral values range from zero to 16.7 kHz.

It is noted that when there are only a few example data files for a given fault class, such as in the case of Class 2 (see Table II), classificatory results can be dependent somewhat on the assignment of particular files to the training or evaluation databases. For example, in the case of Fault Class 2 where there were only three fault examples, there are six possible ways in which the three files can be assigned to the training and evaluation databases (assuming that each database contains at least one file). In the case of Fault Class 2, it was found that two of the six possible partitionings led to perfect classification results on the evaluation data, whereas the other four partitionings resulted in a small number of misclassifications, which never exceeded 0.229% (14 out of 6,125 exemplar evaluations). Such results might be expected when so few data are available for training purposes.

Conclusions: The results documented in this report demonstrate success in two areas of fault diagnostics for helicopter transmissions: (1) rapid and accurate fault detection through the application of statistical change detection algorithms; and (2) perfect fault isolation (i.e., classification) using polynomial neural network classifiers synthesized with a constrained minimum logistic-loss criterion. Several examples of helicopter gearbox data, both seeded and non-seeded, were evaluated. The algorithms developed (i.e., statistical change detectors and trained neural network classifiers) are practical computationally, are data dependent rather than system dependent, and use nonparametric time-series models for change detection and polynomial neural networks for classification.

A vital characteristic of the non-likelihood ratio (NLR) fault detection algorithm is its robustness when utilizing reduced-order models. This is important in practice, as models used to monitor systems will necessarily be approximate. Although under-parametrization leads to some loss of information, if the model reduction is done well, detection of changes of interest can be achieved reliably, and all other changes can be treated as nuisances [19]. This provides a capability that is important in many practical applications (e.g., helicopter rotor transmission health monitoring). Inadequacy in the whitening model then becomes a secondary issue, affecting mainly detection *efficiency*. The use of neural networks is expected to play a significant role in the practical and flexible application of statistical change detection techniques. Here neural networks may be used as estimators to enhance detection when processing nonlinear and/or non-Gaussian data, and to automate the syntheses of both linear and nonlinear detectors.

Other desirable attributes of the NLR fault detection algorithm are its capabilities for quickest possible change detection (on-line algorithm) and maximum detection sensitivity (off-line algorithm) when the true models are *unknown*. The methodology is sufficiently general to be applicable across mechanical system designs without significant re-engineering effort. The algorithm does *not* rely on *ad hoc* feature characteristics, acting instead as a novelty detector, thereby obviating the need to collect examples of possible fault signature patterns for training purposes. The algorithm should enable more automatic syntheses of change detectors, placing the design burden on the adaptive algorithm, rather than the human. Without such automation, it is unlikely that the huge cost-savings potential of condition-based maintenance can be captured.

Best results were obtained with use of the ZBB/NLR multivariate algorithm and with ensemble-averag-ed, rather than time-averaged, detection statistics; the former are applicable whenever an adequate shaft-rate signal, such as a once-per-revolution tachometer pulse, is available (or can be derived [14]). Fault detection results on the CH-47D helicopter data are especially promising, because little attempt was made to use an optimal model structure to capture appropriate feature characteristics of the data or to whiten the data. In terms of the former, a simple AR(10) model was used, causing the algorithm to rely on the first ten autocorrelation coefficients for detection. In terms of the latter, the AR(10) model parameters were all set to zero, eliminating data whitening entirely.

Based on analyses of other data not reported on herein (see [10, 12]), it is likely that the difference statistically between different "normal" machines will be as large as those between normal and faulted machines, at least where faults in the latter case are incipient. This suggests that the detector may need to be "trained" on data taken from the particular machine in which changes are to be detected.

The excellent results obtained for fault classification also attest to the robustness of logistic-loss classifiers, since no special effort was required to achieve the demonstrated results. (Indeed, only

one approach to data pre-processing and to training of a classifier was attempted for the CH-46E helicopter gearbox data; still, perfect *single-look* fault classification results were obtained!) This suggests that the polynomial neural network classifiers are also likely to perform well with other input features, including parameters obtained on-line through adaptation of the fault detection model.

References:

[1] A.R. Barron, "Statistical properties of artificial neural networks," *Proc. IEEE Conf. on Decision and Control*, Tampa, FL, December 1989.

[2] M. Basseville and I.V. Nikiforov, *Detection of Abrupt Changes*, 1993. (Prentice-Hall: Englewood Cliffs, NJ)

[3] P.C. Chestnut and R.W. Floyd, "An aspect-independent sonar target recognition method," *J. Acoust. Soc. Am.*, Vol. 70, No. 3, September 1981.

[4] *CLASS: Algorithm for Synthesis of Polynomial Networks for Classification Users' Manual*, Version 1.61, June 1995, Barron Associates, Inc.

[5] W.A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Magazine*, Vol. 8, No. 2, pp. 14 - 37, April 1991.

[6] C.J. Li and T. Kim, "Linear model-based fault detection and isolation for a screw compressor," *Mechanical Systems and Signal Processing*, Vol. 8, No. 3, 1993, pp. 259 - 273.

[7] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, 1983, p. 4. (MIT Press: Cambridge, MA)

[8] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*, 1980. (Academic Press: London)

[9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 1984. (McGraw-Hill: New York)

[10] B.E. Parker, Jr., T.M. Nigro, and M.P. Carley, *Machinery Diagnostics Using Vibration Signature Processing*, Barron Associates, Inc. Final Technical Rept. for Office of Naval Research, Contract N00014-92-C-0060, November 1992.

[11] B.E. Parker, Jr., H.V. Poor, M.P. Carley, N.A. Nigro, and R.J. Ryan, "Detection of small changes in mechanical systems," *Proc. of the U.S. Navy Small Business Innovation Research Conf.*, Arlington, VA, June 16 - 17, 1994, pp. 93 - 107.

[12] B.E. Parker, Jr., H.V. Poor, M.P. Carley, R.J. Ryan, M.J. Szabo, J.W. Stayman, N.A. Nigro, and P.D. Mooney, *Helicopter Transmission Diagnostics Using Vibration Signature Analysis*, Barron Associates, Inc. Final Technical Rept. for Office of Naval Research, Contract N00014-93-C-0074, July 1995.

[13] H.J. Rose, "Vibration signature and fatigue crack growth analysis of a gear tooth bending fatigue failure," *Current Practices and Trends in Mechanical Failure Prevention — Proc. 44th Meeting of the MFPG*, Virginia Beach, VA, April 3 - 5, 1990. (Vibration Institute: Willowbrook, IL)

[14] G.P. Succi, "Synchronous averaging of multiple gear shafts using an interpolating filter," *Life Extension of Aging Machinery and Structures — Proc. 49th Meeting of the MFPT*, Virginia Beach, April 18 - 20, 1995, pp. 113 - 125. (Vibration Institute: Willowbrook, IL)

[15] R.S. Teal, Boeing Defense & Space Group, personal communication.

[16] D.G. Ward, "Generalized networks for complex function modeling," *Proc. IEEE Systems, Man, and Cybernetics Conf.*, October 2 - 5, 1994, San Antonio, TX.

[17] Westland Helicopter, Ltd. Research Paper RP907, *Final Report on CH-46 Aft Transmission Seeded Fault Testing*, Vols. 1 and 2, September 1, 1993.

[18] Westland Helicopter, Ltd. Mechanical Research Report MRR20168, *CH-46 Aft Transmission Seeded-Fault Testing Analysis of Vibration Recordings*, January 1, 1994 (DTIC Document Number AD-B180 276).

[19] Q. Zhang, M. Basseville, and A. Benveniste, "Early warning of slight changes in systems," *Automatica*, Vol. 30, No. 1, 1994, pp. 95 - 113.

Acknowledgments: This investigation was funded by the Office of Naval Research under Contract N00014-93-C-0074. The authors thank Dr. Thomas M. McKenna for his kind support of this work. Gratitude is also expressed to Natalie A. Nigro of Barron Associates, Inc. for producing the figures contained herein, and to Susan L. Woodson, also of Barron Associates, for her technical typing assistance.

MULTISCALE STATISTICAL MODELING APPROACH TO MONITORING MECHANICAL SYSTEMS

Kenneth C. Chou

Applied Control and Signal Processing Group
SRI International
Menlo Park, CA 94025

Abstract

Signal processing for condition based maintenance and equipment monitoring has focused in recent years on non-stationary signal analysis using time-frequency representations of the signal. These representations are used to identify non-stationary events in the signal that indicate some change in the state of a structure or a machine. It is important to be able to reliably detect such changes in real-time to do necessary preventive maintenance and also to minimize unnecessary maintenance. While transformations such as the Wigner-Ville, Gabor, and wavelet transforms are useful in highlighting time-frequency features of the signal, the application of such transforms to the monitoring problem requires additional tools for making decisions concerning the condition of the object being monitored. In particular, the interpretation of the transform coefficients in terms of physical events is essential to making such decisions. We develop a methodology for identifying the physical state of the object based on statistical models of the signals, which could comprise, for example, multiple outputs from devices such as accelerometers, strain sensors, and acoustic emission sensors. Classification of machine states based on monitoring signals is performed by comparing likelihood scores for each machine state. We present examples of applying our system to various data, including damped sinusoids and noisy chirps, as a way of illustrating system performance for the case of transient monitoring signals. We compare our system to one which is trained using a DFT-based (non-time-frequency-based) representation (in particular, LPC coefficients) and show that our system exhibits both superior performance as well as greater robustness to noise in the signals. We also compare results using multiscale parameters versus LPC coefficients for the case of synthesized autoregressive signals and for the case of actual, measured signals from a weld depth monitoring system.

Keywords: multiscale, nonstationary signals, statistical models, time-frequency, wavelets

1 Motivation

In machinery monitoring applications, such as tool wear monitoring for example, the signal representation component of the monitoring system is fundamental to the classification performance of the monitoring system. In these applications, characterization of transient signals is key to classifying machine wear states [9]. Furthermore, the ability to train a pattern recognition system using limited amounts of data is important, especially those encountered in flexible manufacturing. In standard frame-based systems, spectral-based coefficients (e.g. LPC, cepstral, melcepstral parameters) are used to model the data within each frame, the assumption being that this data is stationary. To address the need for characterization of transients and other types of nonstationary signals for machine state classification, work has been done toward using time-frequency representations as a basis for features [10]. In order to address the need for controlling the degrees of freedom of our representation to match the available data (e.g. limited training time, low SNR), we propose the use of a parameterized model for representing the time-frequency characteristics of the data in each frame. By varying the structure and the number of parameters, the model is rich enough to capture a wide range of signals, both stationary and non-stationary. The structure of the parameters, however, can be constrained, e.g. in terms of number of parameters, to accommodate limited training data.

2 Multiscale Models

For nonstationary signals, it has been shown that the statistical correlation between the wavelet coefficients of a process has a great deal of structure [12]. For example, for many processes of interest the strongest

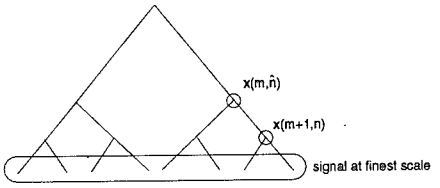


Figure 1. Binary tree representation of a multiscale stochastic process

correlations exist between different scale coefficients located at the same point in time. This type of structure is well-modeled by the class of multiscale stochastic models (MSMs) developed in [1, 2, 3]. These models are represented on a binary tree in which the levels of the tree may be taken to represent projections of the signal onto multiple scales (Figure 1). Our MSM is characterized by the following state equation in which the state, $x(m,n)$, is a vector (this allows for the representation of multichannel data) at the node located at scale m and location n in the binary tree, while $y(m,n)$ is the observed value corresponding to the signal value at that node.

$$x(m+1,n) = F_{m+1,n}x(m,\hat{n}) + w(m+1,n) \quad (1)$$
$$y(m+1,n) = H_{m+1,n}x(m+1,n) + v(m+1,n) \quad (2)$$

where $F_{m+1,n}$ is the matrix that maps the state at scale m and location $\hat{n}$ to the state at scale $m+1$ and location n, and $w(m+1,n)$ is the random variable uncorrelated with $x(m,\hat{n})$, representing the detail that is added. The matrix $F_{m+1,n}$ represents, for example, the weighting of one of the branches in Figure 1 in going from a coarse node to one of the finer nodes below it. The matrix $H_{m+1,n}$ and the random variable $v(m+1,n)$ allow for the observed signal value at node m,n to be a linear function of the state plus white noise.

For the purpose of modeling features in a frame of data, the observations at the bottom or finest level of the tree would represent the data in a particular frame. In this case, the observations $y(m,n)$ are non-zero only at the finest level. The state vector $x(m,n)$, however, exists at every node on the tree. By varying $H_{m,n}, C_{m,n}$, the variances of $w(m,n)$, and the variances of $v(m,n)$, we can model a rich class of random phenomena including fractal Brownian motion and Markov random processes [8]. This leads us to investigate the use of the model parameters of our MSM as features for classification of nonstationary signals.

As a final note on multiscale tree models, we summarize the second-order statistical properties of the models and make a comment about parameterization of these models. Note that the following discussion will be in terms of the notation in Figure 2. To begin with, we describe the local covariance $C(t)$ (i.e. the covariance of the state-vector at a particular node t) as it propagates from one node to the next in the following way.

$$C(t) = F(t)C(\gamma t)F^T(t) + P(t)$$

where γ is a backward shift operator on the tree and $P(t)$ is the variance of the white noise term $w(t)$. Note that the covariance of the state at a particular node is a function of the covariance at the parent node dynamically driven by the covariance of the white noise term added at that node.

We now describe the cross-covariance between states at two different nodes, i.e. the correlation between the state at two different nodes. For two distinct nodes on the tree, $s \neq t$,

$$E[x(s)x(t)^T] = (\prod_{i\in\Gamma_{us}} F_i)C(u)(\prod_{j\in\Gamma_{ut}} F_j^T)$$
$$u = s \wedge t$$

where $\wedge$ is a function that maps two nodes s and t into their unique common (least) ancestor node, and Γ_{us} is the ordered set of nodes $\{s, \gamma s, \gamma^2 s, ...\tilde{u}\}$, where $\gamma\tilde{u} = u$. Referring to Figure 2, note that the cross-covariance between states at nodes s and t is simply a function of the state covariance at the node corresponding to

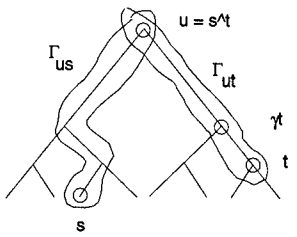


Figure 2. Illustration of dependencies of correlation function on trees

their first common ancestor and the dynamic matrix, F, along the two paths joining s and t, respectively, to their common ancestor node (denoted u in Figure 2).

The correlation characteristics of the MSM's point to several observations. First, if we consider F to be a constant, independent of node location, then the characteristics of the MSM become quite specific. In particular, for example in Figure 2, node s is correlated with each of the four nodes in the right half of the bottom level of the tree in *precisely the same way.* Note that regardless of how the variance at each node, $C(t)$, evolves, if F is constant, the correlation between states at two nodes is simply a function of $C(t)$ at the ancestor node and the number of branches connecting the two nodes to their common ancestor. In [2], it is shown that this correlation structure has eigenfunctions corresponding to the Haar wavelet transform. Clearly, if we want to expand our MSM's to include a wider class of processes, we must consider a richer class of parameterizations.

In this paper, we explore the idea of varying the dynamic matrix F in our MSM's to allow for richer descriptions of signals in terms of their correlation structure. In particular we will show examples of allowing F to vary arbitrarily over nodes starting from the top node down to and including the nodes at some specified scale. Thus, by varying F we can vary the order of our MSM, much in the same way as one would vary the order of a parameterized statistical model such as the class of autoregressive (AR) models. By allowing F to vary across both scale and time we would expect non-stationary behavior, and in particular behavior such as changing spectral characteristics within one's analysis window, to be better captured by our models than they would using stationary models.

3 Multiscale Parameter Identification

We use an efficient Maximum-Likelihood procedure for identifying the parameters of our MSM based on a frame of data. It is based on the Expectation Maximization (EM) algorithm [4] as applied in [5]. The basic idea behind the algorithm is to first perform the expectation step of the algorithm, i.e. computing the expectation

$$E[\mathcal{L}(X,Y,\theta)|Y] \tag{3}$$

using the current MSM parameters θ, where $\mathcal{L}$ is the likelihood of the data in the frame Y and the state variables X given θ. Then, in the next step, $\mathcal{L}$ is maximized with respect to the parameters θ. In iterating this procedure, one is guaranteed to converge to a local maximum of $\mathcal{L}$ and furthermore, at each iteration $\mathcal{L}$ is guaranteed to increase. The expectation step can be performed rather efficiently for our models using efficient smoothing algorithms developed in [2]. These algorithms are based on a Kalman filtering scheme that exploits the parallel and local structure of the scale dynamics of eq.'s(1,2). This smoothing operation can be performed in $O(m^3 log N)$ operations using parallel processing, where N is the data frame length and m is the state dimension. The maximization step is straightforward, resulting in the computation of sums of outer products involving state vectors at each node and observations at the finest scale.

The basic points about the EM algorithm for Maximum-likelihood estimation of MSM parameters can be summarized as follows.

- Expectation step: Compute

$$E[\mathcal{L}(X,Y,\theta)|Y] \tag{4}$$

- Involves computing expectations of MSM state x and local correlations conditioned on data Y
- Fast algorithm exploits Markov structure of model: $O(N)$ complexity, $O(logN)$ with parallel processing

- Maximization step: Compute

$$\max_{\theta} E[\mathcal{L}(X, Y, \theta)|Y] \tag{5}$$

- Direct solution exists
- Sufficient statistics easily computable, $O(N)$ complexity

3.1 Classifying Features

Once the signal is modeled at the frame level using MSM parameters as features, we need a way of modeling statistically the ensemble of frames for a given signal. We model the sequence of frames of features as samples of a multivariate distribution. Moreover, by doing this we provide a way of mitigating the effects of sensitivity of our multiscale models to time shifts. The Gaussian Mixture Model [11] provides a convenient multimodal model for our ensemble of frames. In particular, the GMM leads to an efficient recognition procedure for classifying signals by processing successive frames of observation vectors and evaluating likelihood functions [6]. Also, we use an efficient procedure for training our GMM based on applying the EM algorithm to the problem of identifying the GMM parameters. This iterative procedure, as was the case for the MSM parameter identification procedure, yields locally optimal, maximum likelihood parameter estimates by ensuring a nondecreasing likelihood function after each iteration.

4 Numerical Examples

We present numerical examples to give an indication of how our system performs as compared to a system using features based on a spectral representation. We will show examples from a variety of data sets, but the basic procedure for analyzing the capabilities of our model will be the following. For a particular type of data (e.g. damped sinusoids), we either synthesize or measure signals from two different classes (e.g. representing two different machine states). We then analyze and compare the classification performance for two different representations: 1) our MSM's and 2) LPC models. To characterize classification performance, we will use bounds on classification error, visualization of classification surfaces, and simulations of classification performance based on actual training using data. Note that unless otherwise stated, our frame size is taken to be 64 samples.

For our first example, we generate data from two different classes of transient signals, representing noisy damped sinusoids. The transients will be modeled as follows for classes $j = 1, 2$:

$$y_j(t) = \sum_{i=1}^{N_j} K_{i_j} e^{-\alpha_{i_j} t} \sin(\beta_{i_j} t) u(t) + \eta_j \tag{6}$$

where N_j controls the density of transients within a given data segment and η is Gaussian white noise. Figure 3 is a plot of typical signals from each of the two classes.

4.1 Bounds on Classification Error

We now examine the performance of a classifier using our MSM parameters as features versus using features based on a spectral representation. For the examples in this section our MSM's are fixed at order 4, i.e. the parameters consist of a fixed dynamic parameter, F, the observation parameter, H, and the two noise variance corresponding to the white noise driving term w, and the white noise observation term v, respectively. In later examples we will allow F to vary in order to demonstrate the capabilities of higher-order MSM's. As an example of spectral features we use Linear Predictive Coding (LPC) coefficients, set at order 4 in order to normalize the number of parameters to be equal to the number of MSM parameters. To give an indication

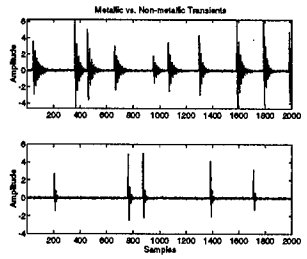


Figure 3. The top plot shows an example of a typical signal in Class 1 and the bottom plot shows a typical signal in Class 2. Note the number of transients in the Class 1 signal is higher, and the damping factor is smaller.

Data scenario	*MSM features*	*LPC features*
noiseless	6.8099e-4	.3227
SNR 12dB	.0072	.4161
SNR 6dB	.0083	.4519

Table 1. Comparison of Bhattacharyya bounds using MSM versus LPC features for transient data. 4000 sample sequences used.

of the performance of a classifier based on a particular set of features, we using Bhattacharyya bounds [7] to estimate the classification error. The Bhattacharyya bound gives an upper bound on the probability of error in a 2-class classification problem. We can compute the bound using sample means and sample covariances of the set of features derived from the two classes of signals. Table 1 contains bounds on the classification error probability for the 2-class data set in Figure 3 for the case of MSM features versus LPC features. Results are shown for the noiseless case as well as for SNR's of 12dB and 6dB. From these results we see not only that MSM features leads to better discrimination between the two types of transients but also that they seem to be less sensitive to noise than the LPC features.

We now show performance comparisons based on other data sets. Table 2 shows results for two different data sets. One data set consists of signals from a weld penetration monitoring system. Figure 4 shows optical signals for the case of a full penetration weld and a partial penetration weld. Note that the signals differ quite subtly in character and that the MSM parameters seem to discriminate better than the LPC parameters do. Also in Table 2 are results on synthesized autoregressive data. Two classes of AR(1) processes were created with lag coefficients of .97 and .9, respectively. Perhaps surprisingly, the MSM features outperform the LPC features, even though the signals are synthesized to be stationary and are of the LPC model class. However, it is important to note that we are measuring discrimination performance rather than representational

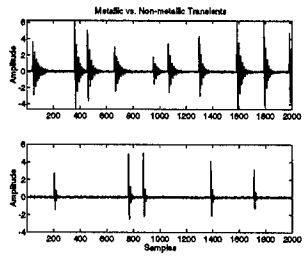


Figure 4. The top plot shows an optical signal corresponding to a full penetration weld while the bottom plot shows an optical signal corresponding to a partial penetration weld.

Data scenario	*MSM features*	*LPC features*
weld depth data	.2890	.4371
AR process	.2230	.3835

Table 2. Comparison of Bhattacharyya bounds using MSM versus LPC features for optical weld depth signals, AR(1) processes. 4000 sample sequences used.

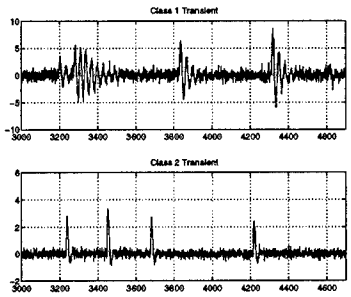


Figure 5. Transient signals

performance.

4.2 Visualization of Classification Surfaces

In this section we provide further examples of differences in classification performance between using MSM's and using LPC models, and do a comparison of performance for different order MSM's and LPC models. In addition to bounds on classification error, we provide in the following a visualization of the actual estimated densities of the MSM and LPC parameters for each of the two classes corresponding to a particular data set. In particular, we compute the sample means and covariances of the features (both MSM and LPC) for each class and present a 3D plot of the modes corresponding to these means and covariances for these two classes. Since we can only visualize modes of at most 2 features, we project the multivariate modes onto 2 features representing: a) the 2 features giving the best mode separation, corresponding to best classification and b) the 2 features giving the worst mode separation, corresponding to worst classification.

Figures 5-7 show sample paths for two classes of signals corresponding to three different data types: 1) transient signals generated as before using noisy damped sinusoids 2) 6th order AR signals and 3) chirp signals. The plots of the modes based on sample means and covariances are depicted in Figures 8-10. Note that in each case the mode separation is greater, indicating better classification performance, in the case of MSM parameters versus the case of LPC parameters. This is consistent with the bounds computed for these examples as given in Table 3.

Table 3 also gives bounds for the case of 6th order and 10th order MSM and LPC models applied to the AR signals as well as the chirp signals. It is interesting to note that even in the case of the AR signals, one in which we might expect the LPC model to do better, the MSM features do a much better job of classification based on the bounds. This may be due to a much greater sensitivity on the part of LPC models to edge effects due to small window sizes. Note that in the case of the chirp signals we compare MSM and LPC performance for orders 4-6 for the case of a larger window (256 samples). For this particular window size, LPC does better than MSM for 4th order but worse for orders 6 and 10. Of course, the bounds can be deceiving and thus, we plot the modes for the order 6 case in Figure 11 (best two features). The 6th order MSM features show considerably more mode separation than the 6th order LPC features.

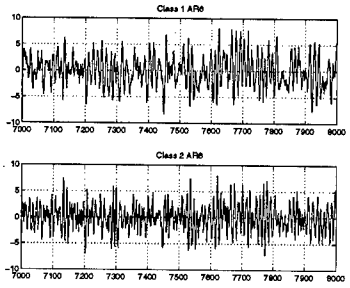


Figure 6. Autoregressive signals

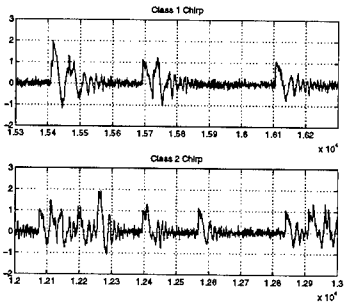


Figure 7. Chirp signals

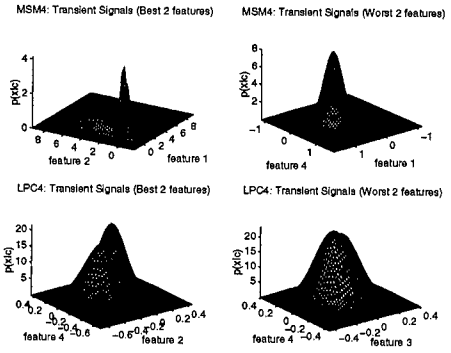


Figure 8. Single mode densities for transient signals

SIGNAL TYPE	MSM 4	LPC 4	MSM 6	LPC 6	MSM 10	LPC 10
Transients	.1387	.4709				
AR 6 Processes	.1670	.1231	7.1639e-10	.1552	1.7498e-19	.2047
Chirps	.3636	.4723	.0057	.4620	.3711	.4602
Chirps (256 pt. frame)	.4040	.3628	.3346	.3570	.0515	.3345

Table 3. Comparison of Bhattacharyya bounds using MSM versus LPC features for three signal types, orders 4,6 and 10.

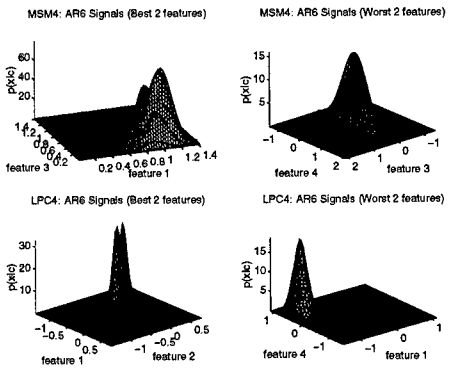


Figure 9. Single mode densities for AR6 signals

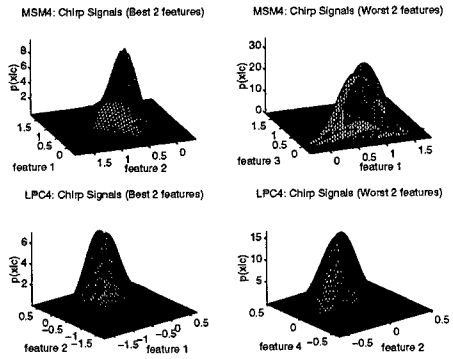


Figure 10. Single mode densities for chirp signals

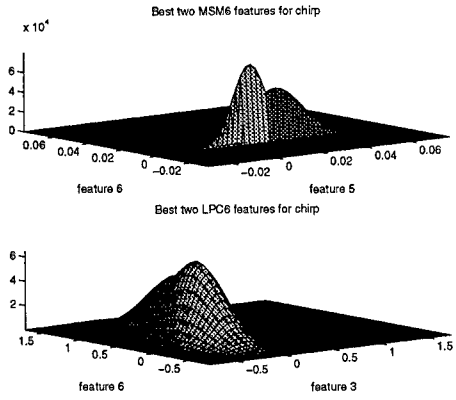


Figure 11. Single mode densities for chirp signals: 6th order models

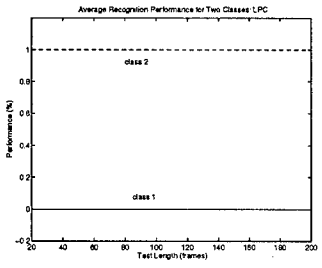


Figure 12. LPC performance curve

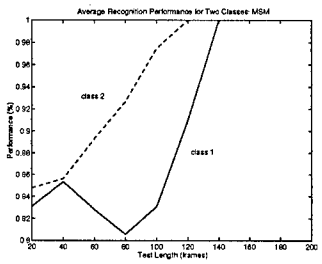


Figure 13. MSM performance curve

4.3 Classification performance based on simulation of classifiers

Finally, we show the results of training a classifier based on using MSM parameters as features versus using LPC parameters as features. We train both a 4th order MSM model and a 4th order LPC model on 24,000 samples of each class of chirp signals. To do the training, a 64 sample frame is used in each case, each frame giving rise to 4 features corresponding to the MSM (LPC) parameters for that frame. Furthermore, we use a 3-mode Gaussian mixture model to model the features over the entire collection of frames. We then use an independent set of 24,000 samples of each class of chirp signal to perform a classification test using MSM features versus using LPC features. The classification is performed by doing a likelihood ratio test based on the Gaussian mixture model. Figures 12,13 give performance plots for tests given each chirp signal type, using both MSM and LPC features. Average performance is given for different amounts of data used to perform the likelihood ratio, where each point on the curve reflects average detection performance using a fixed number of frames. For example, for test length equal to 100, the performance value is equal to the percentage of correct classifications for a particular signal class, computed over the entire 24,000 sample data set, using 100 frame chunks to make each classification decision. In the case of LPC features, for this particular set of data the classifier chooses class 2 every time, yielding .5 probability of error, independent of the number of frames used to make the decision. This of course indicates the worst possible performance (note that the bound indicated .4723 as an upper bound on the probability of error; violation of this bound is due to finite-sample effects). For the MSM features, performance starts at around .82 probability of detection and eventually converges on zero percent error. Again finite-sample effects come into play, especially when the ratio of the number of frames used to make the decision to the total number of test samples becomes large.

5 Acknowledgment

We would like to thank Ken Pietrzak at UTRC for providing the weld penetration monitoring data.

References

[1] K.C. Chou, S.A. Golden, and A.S. Willsky, "Multiresolution Stochastic Models, Data Fusion, and Wavelet Transforms," *Signal Processing*, Vol. 34, Issue 3, 1993.

[2] K.C. Chou, A.S. Willsky, and A. Benveniste, "Multiscale Recursive Estimation, Data Fusion, and Regularization," *IEEE Trans. on Automatic Control*, Vol. 39, No. 3, March 1994.

[3] K.C. Chou, A.S. Willsky, and R. Nikoukhah, "Multiscale Systems, Kalman Filters, and Ricatti Equations," *IEEE Trans. on Automatic Control*, Vol. 39, No. 3, March 1994.

[4] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM algorithm," J. R. Stat. Soc. Lond., Vol. 39, pp. 1-38, 1977.

[5] V. Digalakis and K.C. Chou, "Maximum Likelihood Identification of Multiscale Stochastic Models Using the Wavelet Transform," *Proc. of ICASSP*, Minneapolis, MN, 1993.

[6] L.P. Heck and K.C. Chou, "Gaussian Mixture Model Classifiers for Machine Monitoring," *Proc. of ICASSP*, Adelaide, Australia, 1994.

[7] T. Kailath, "The Divergence and Bhattacharrya Distance Measures in Signal Selection," *IEEE Trans. of on Communication Tech.*, vol. Com-15, no. 1, Feb. 1967.

[8] M. Luettgen, W. Karl, A. Willsky, and R. Tenney, "Multiscale Representations of Markov Random Fields," *IEEE Trans. Signal Processing*, Vol. 41, No. 12, December 1993.

[9] S. Rangwala and D. Dornfeld, "Sensor Integration Using Neural Networks for Intelligent Tool Condition Monitoring," *Trans. of ASME, Journ. of Eng. for Industry*, Vol. 112, No. 3, pp. 219-228

[10] R. Priebe and G. Wilson, "Application of 'Matched' Wavelets to Identification of Metallic Transients," *IEEE-SP Symp. on Time-Freq. and Time-Scale Anal.*, Victoria, B.C., October 1992.

[11] D.A. Reynolds, *A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification*, Ph.D. Thesis, Georgia Institute of Technology, September 1992.

[12] A. Tewfik and M. Kim, "Correlation structure of the wavelet coefficients of fractional brownian motions," *IEEE Trans. Informat. Theory*, vol. 38, Mar. 1992.

SYNCHRONOUS PHASE AVERAGING METHOD FOR MACHINERY DIAGNOSTICS

Jen Jong
AI Signal Research, Inc.
3322 South Memorial Parkway, Suite-67
Huntsville, AL 35801

James McBride, Jess Jones, Tony Fiorucci, Thomas Zoladz
NASA, Marshall Space Flight Center, ED-23/32
MSFC, AL 35812

Abstract: This paper discusses a new diagnostic signature analysis algorithm called Synchronous Phase Averaging (SPA) which performs signal enhancement on quasi-periodic waveforms common in machinery and/or gearbox diagnosis. Time domain averaging (TDA) is an effective technique for extracting periodic signals from noisy or complex waveforms associated with machinery vibration. Due to the quasi-periodic nature of machinery shaft rotational speed, TDA is usually initiated with a trigger-pulse synchronous signal in order to synchronize the averaging process with the shaft rotation. This process is called Synchronous Time Averaging (STA), and is a common diagnostic tool in mechanical signal analysis. The Synchronous Phase Averaging method discussed in this paper can provide such signal enhancement performance but does not require a tachometer, or once-per-revolution input, as the reference signal. This is achieved through a novel synchronization process called the Phase Synchronized Enhancement Method (PSEM) which transforms a quasi-periodic synchronous (Sync or RPM) component within a vibration signal into a pure-tone discrete component. Within this newly transformed PSEM signal, since all the Sync-related components become discrete, standard Time Domain Averaging can be directly applied to the new signal to achieve signal enhancement. The elimination of tachometer monitoring reduces both instrumentation and data acquisition requirements, but most importantly, allows the effective implementation of non-intrusive sensors (accelerometer, acoustic) in applications such as gear box diagnostics, etc. Application examples using the SPA method for gearbox diagnostics will be demonstrated in this paper.

Key Words: Diagnostics, Gear Box; Signature Analysis, Synchronous Phase Average, Turbomachinery, Non-intrusive Sensing.

NOMENCLATURE

PSD Power Spectral Density Function
PSEM Phase Synchronized Enhancement Method
STA Synchronous Time Average
SPA Synchronous Phase Average
N Synchronous (Rotating) Frequency
Sync Synchronous Frequency Component or RPM Frequency
FM Frequency Modulation
IF Instantaneous Frequency

INTRODUCTION: Time domain averaging (TDA) is a well-known and powerful technique for extracting periodic signals from noisy or complex waveform. For this process to be coherent, it requires that the period of the signal to be extracted be known or assumed. It is based on averaging points one period apart, where the period is that of the signal to be extracted. However, in machinery diagnosis, the quasi-periodic nature of shaft rotational speed does not allow direct application of TDA. Therefore, TDA is usually performed by utilizing a trigger-pulse synchronous signal, such as a once per revolution tachometer signal, in order to synchronize the averaging process with shaft rotational motion. Such Synchronous Time Averaging is generally considered a key method for analyzing gear behavior [1, 2, 3]. Though an important point to be made is that when performing synchronous time averaging only the first sample of each period of the ensemble signal is synchronized to the rotating shaft.

In theory, the most perfect form of time averaging utilizes external sampling where the sampling process speeds up or slows down with the shaft rotational speed [4]. When engine rotational speed(s) varies, the sampling does also yielding non-uniform sample intervals in acquired data. However, the rotational phase angle between consecutive samples maintains absolutely constant. When time averaging is applied, the complete cycle of data points within each ensemble is synchronized to the rotating shaft rather than just the first sample of the collection as is the case with synchronous time averaging. Such an ideal averaging process provides superior diagnostic signal enhancement capability. Unfortunately, the luxury of an optical encoder is usually not available. In addition, the requirement to have a speed measurement as the reference signal for STA is somewhat impractical within some operational environment where intrusive instrumentation is not possible, such as gear box diagnostics using remote acoustic measurements. Therefore, it is highly desirable to develop a waveform enhancement technique which does not require tachometer or key-phaser input as the STA would require.

This paper discusses a new algorithm called Synchronous Phase Averaging (SPA) which provides the same degree of signal enhancement as external sampling without sophisticated and intrusive external sampling instrumentation and data acquisition systems. Moreover, SPA does not even require a tachometer based reference signal for triggering data acquisition. This is achieved through a novel signal transformation technique called the Phase Synchronized Enhancement Method (PSEM) [5]. The PSEM

utilizes the instantaneous phase information associated with the quasi-periodic shaft rotational motion to transform the synchronous (Sync or RPM) component into a pure-tone discrete component. This Sync discretization process generates a highly desirable effect on the entire diagnostic signal where all other Sync-related components automatically become discrete. Within this newly transformed PSEM signal, all spectral components which are nonlinear correlated with Sync, such as Sync harmonics, are automatically discretized. In a rotor system, different rotational mechanisms could interact with one another due to some nonlinear process [6, 7, 8, 9]. Such nonlinear correlation plays a significant role in accurate mechanical signature diagnosis, and can be identified from its dynamic response signal using nonlinear spectral analysis techniques such as bi-coherence, tri-coherence, and hyper-coherence, etc. [10, 11, 12]. Once all the sync-related components within the PSEM signal become discrete, standard Time Domain Averaging can be directly applied to the PSEM signal to achieve signal enhancement. Thus, Synchronous Phase Averaging provides the same signal enhancement performance which external sampling achieves without the related expense and tachometer requirement.

PHASE SYNCHRONIZED ENHANCEMENT METHOD (PSEM): A vibration signal in a rotor system can be modeled as an FM signal with multiple spectral components at different center (carrier) frequencies. The instantaneous frequency (IF) and/or Instantaneous Phase (IP) information of each component can be recovered using several digital frequency demodulation methods, such as the complex demodulation (heterodyne) techniques. [13] and Phase Lock Loop (PLL) [14] method. During steady state operation of most machinery systems, the instantaneous frequency of Sync tends to fluctuate about some center frequency and is not a constant frequency. Since the amplitude of such frequency variation could be much smaller than the bandwidth within PSD analysis, the Sync spectral component still appears as a very discrete peak in its PSD. Based on this phenomenon, the PSEM Method was developed to take advantage of such IF variation phenomenon. The basic principle of the PSEM is to force the narrow-band spectral component of Sync whose frequency is fluctuating slightly around some center frequency into a pure-tone discrete component of constant frequency. This discretization process thus generates a highly desirable effect on the entire signal, i.e., all other spectral components which are correlated with Sync (such as Sync Harmonics, Gearmesg, etc.,) automatically become discrete.

The signal model can be formulated as:

$$x(t') = A\cos[\underset{\Uparrow}{\Psi}(t')] + B\cos[\gamma * \underset{\Uparrow}{\Psi}(t')] \qquad (1)$$

SYNC SYNC-RELATED

Where $\Psi(t)$ is the instantaneous phase of sync
γ is a constant

The first term A*cos[$\Psi(t')$] represents the Sync frequency component, while the second term B*cos[$\gamma*\Psi(t')$] represents a Sync-related component whose phase variation is synchronized with Sync. The sampling process of the original signal can be viewed as an observer traveling at a constant speed V(t) along the time-axis observing the waveform x(t'). If the observer is traveling at a time-dependent non-constant speed V(t), he will then observe a slightly different waveform y(t) as:

$$y(t) = x[t' = V(t)t] = A\cos\{\Psi[V(t)t]\} + B\cos\{\gamma * \Psi[V(t)t]\} \quad (2)$$

This new waveform y(t) will generate new phase information for both the sync and the sync-related components. However, the new phase of the Sync-related component is still synchronized with the phase of Sync as shown in equation (2). This relationship indicates that phase correlation between two spectral components is independent of the sampling process.

The waveform of the Sync frequency component under a constant sampling rate is typically quasi-periodic. This sync component then can be discretized by resampling with a unique time-dependent sampling rate so that the newly-generated waveform will become periodic instead of quasi-periodic. This unique time-dependent sampling rate can be found from the instantaneous phase information of Sync. The signal model of Sync can be rewritten as:

$$x(t) = A(t)\cos[\omega_C t + \Phi(t)] \quad (3)$$

Where: ω_C is the center frequency of Sync
$\Phi(t)$ is instantaneous phase variation

Since time and phase are directly related to each other, the resampling rate can obtained through the following relationship:

$$\Phi(t) = \omega_C \Delta(t) \quad (4)$$

By replacing the instantaneous phase variation $\Phi(t)$ with $\omega_C\,\Delta(t)$, equation (3) becomes:

$$x(t) = A(t)\cos[\omega_C t + \omega_C \Delta(t)]$$

$$= A(t)\cos\{\omega_C[t + \Delta(t)]\} \quad (5)$$

Equation (5) states that, by treating the original quasi-periodic signal as a resampled signal represented by equation (5), the resampled signal will become periodic since the original phase variation is now transformed into its corresponding time variation $\Delta(t)$.

Once the Sync frequency component becomes discrete in the resampled signal, all the other Sync-related components will automatically become discrete.

Figure 1 shows the PSEM algorithm. The input signal is composed of a reference Sync component along with several other Sync-related and non-Sync-related components. The band-pass filter first removes all the other spectral components in order to generated the time history of a quasi-periodic Sync component. The instantaneous phase variation signal $\Phi(t)$ of Sync is then estimated with the Hilbert Transform method. With the relationship between phase and time, this phase variation signal can now be converted into a realignment time signal $\Delta(t)$. The uniform sampling interval of the original signal is then remapped onto an array of non-uniform sampling intervals based on the realignment time. This realignment process will transform the original quasi-periodic signal into a periodic signal. An interpolator is then used to recover the uniformly sampled periodic signal from the non-uniformly sampled signal. Notice that, once the realignment time signal $\Delta(t)$ is identified, the remaining realignment process will be performed on the original signal rather than the band-pass filtered signal. As a result, all the other Sync-related components in the signal will automatically become discrete.

With its unique capability of transforming a quasi-periodic Sync into a pure-tone discrete component, PSEM generates a highly desirable effect on an entire diagnostic signal with all Sync-related components becoming discrete. With this enhanced frequency resolution, the PSEM signal can recover detailed spectral information useful in diagnostic evaluation. This method is especially useful for machinery diagnostics, in which case a quasi-periodic driving process (Sync) generates many other rotational components. These related components such as Sync harmonics, bearing element spin and passing frequencies, modulation and sideband components can provide useful information about a machine's operational condition.

In this section, test data from an application example is used to demonstrate this PSEM technique. The PSD shown in figure 3-a is taken from a spindle motor during a vibration test. The FFT block size is 16K, which gives a very high bandwidth resolution of 1.6 Hz. In this PSD, the sync frequency component N, and its harmonics such as 2N, 3N, 8N and 9N are readily observable. In addition, many other Sync-related components such as the cage frequency component C, the Inner ball pass component I, the ball spin frequency component B, and many other related components such as their harmonics and modulations are evident. During steady state operation, the fundamental Sync frequency is very constant. Therefore, a PSD of spindle motor vibration data shows a very discrete peak at synchronous frequency. A joint time-frequency plot (isoplot) of the test data also shows a very stationary peak at sync whose frequency never moves outside an analysis frequency bin (unit bandwidth).

However, instantaneous frequency analysis indicates that the fundamental Sync frequency component is by no means a perfect discrete. Figure 2 show the instantaneous frequency of Sync, Inner Ball Pass (IBP) and the 8-th harmonic of Sync. The IF of these components are all moving periodically. In addition, the frequency variation of Sync is

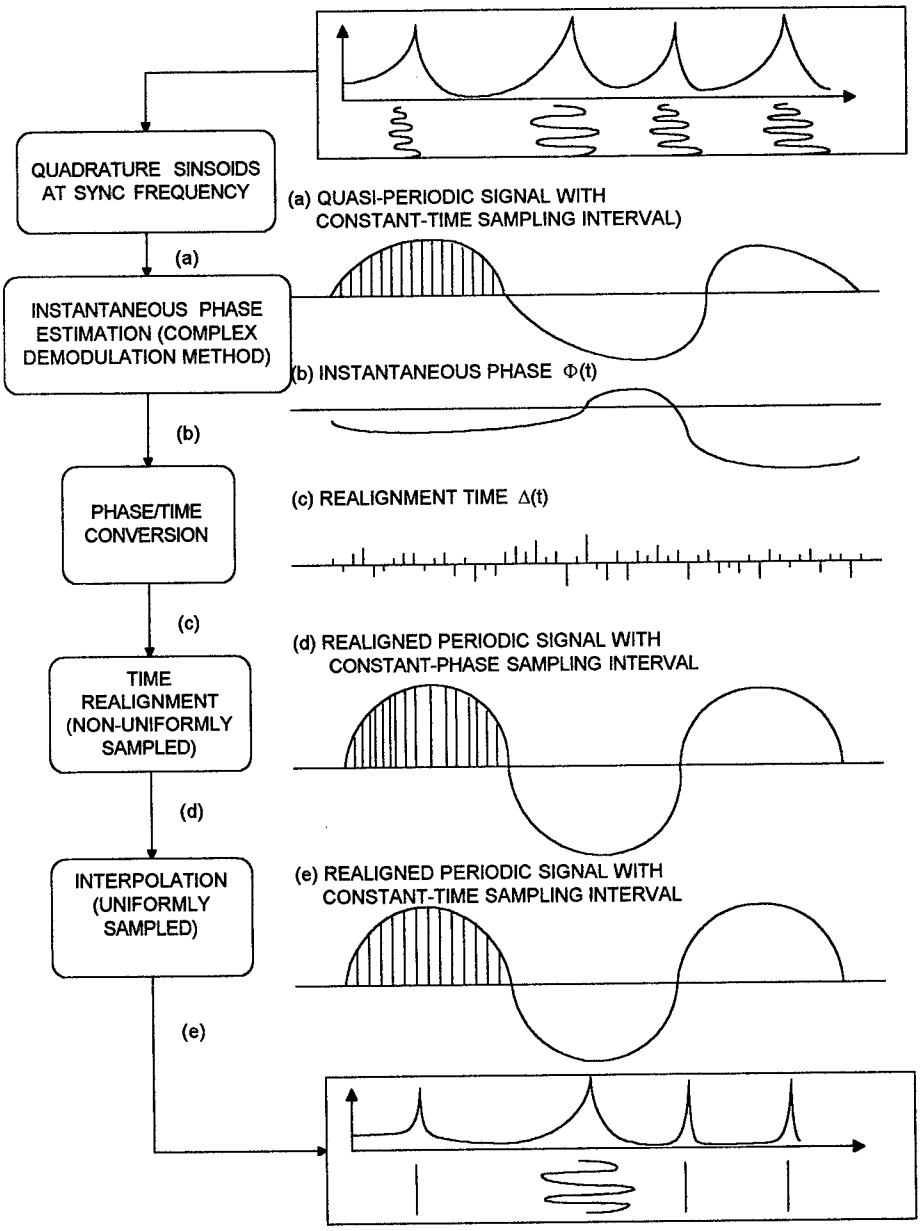


FIGURE 1 PHASE SYNCHRONIZED ENHANCEMENT (PSEM) ALGORITHM

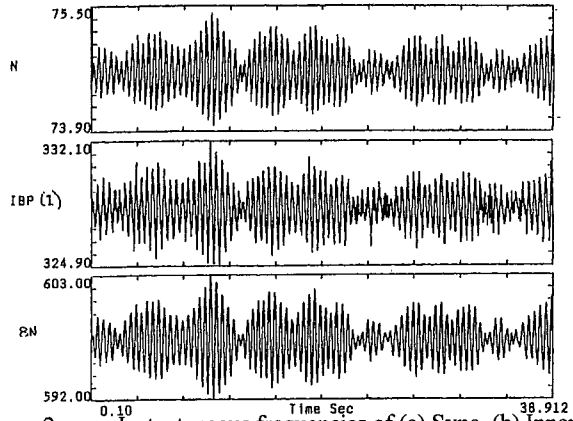


Figure 2: Instantaneous frequencies of (a) Sync, (b) Inner Ball Pass, (c) 8N, of Spindle Motor Data.

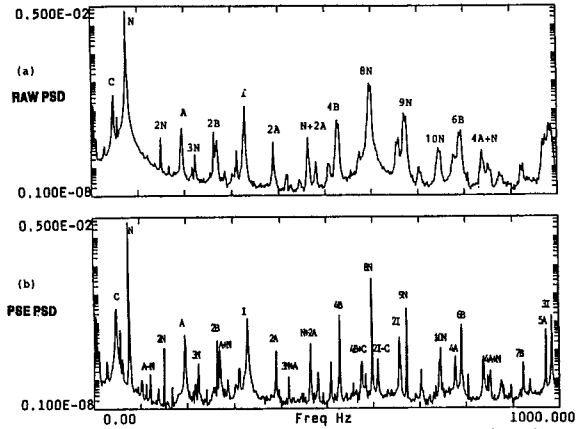


Figure 3: PSDs of (a) Raw Signal (b) PSEM Signal from 0 to 1000 Hz of Spindle Motor Data

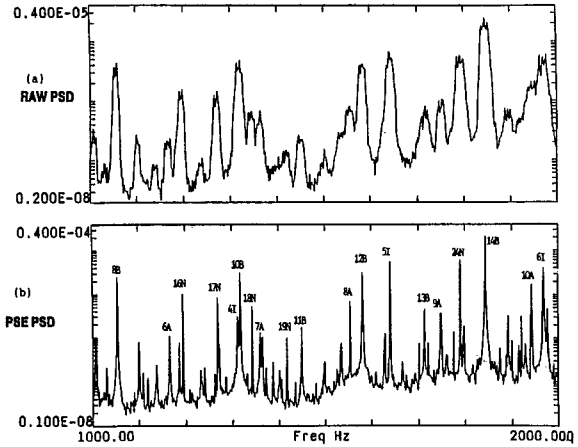


Figure 4: PSDs of (a) Raw Signal (b) PSEM Signal from 1000 to 2000 Hz of Spindle Motor Data

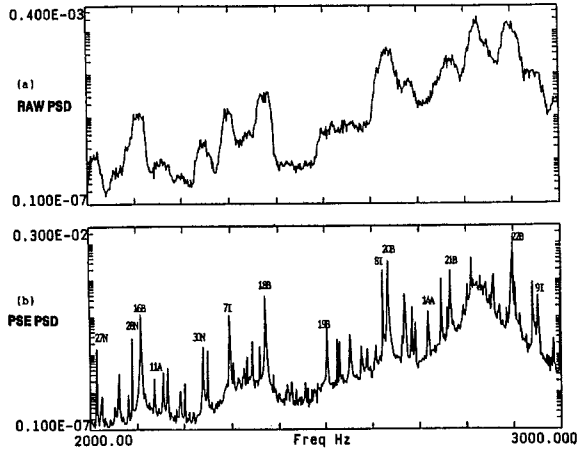


Figure 5: PSDs of (a) Raw Signal (b) PSEM Signal from 2000 to 3000 Hz of Spindle Motor Data

within 1.5 Hz, which is less than the bandwidth of the PSD. Based on this IF information, the PSEM algorithm forces the frequency of Sync from this periodic variation into a constant frequency. As a result, a highly desirable effect is generated on the entire signal.

Figure 3-a and 3-b show the ordinary RAW PSD and the enhanced PSEM PSD of the test data. The frequency range is from 0 to 1000 Hz. Since all the Sync-related components become discrete, the signal's frequency resolution becomes much sharper. The original raw PSD shows a broad peak near 1000 Hz, but the enhanced PSD resolves it into two spectral peaks with the first one being the 5-th harmonic of the Outer Ball Pass (OBP) component while the second one is the third harmonic of the IBP component.

The signal enhancement effect becomes much more noticeable in the higher frequency region. Figures 4-a and 4-b are the raw and enhanced PSEM PSDs of the same data set from 1000 to 2000 Hz. In the raw PSD, all the peaks become so broad that all the spectral information is smeared. However, the enhanced PSD recovers all the sync-related components such as 8 times, 10 time, 12 times and 14 times of Ball Spin (BS) components, 6, 7, 8, 9, and 10 times of OBP frequency, and the harmonics of Sync at 16N, 17N, 18N, up to 24N.

At higher frequencies, PSEM signal enhancement capabilities become even more evident. Figure 5 shows respectively the Raw and enhanced PSEM PSDs from 2000 to 3000 Hz. Again, the enhanced PSEM PSD recovers all the spindle motor Sync-related components such as 16, 18, 20 and 22 times of BS frequency component. These components are completely hidden in the original spindle motor mechanical signature analysis PSD.

SYNCHRONOUS PHASE AVERAGING (SPA) METHOD: As discussed in the last section, the PSEM can transform a quasi-periodic sync into a pure-tone discrete component. Within this newly transformed PSEM signal, since the sync and sync harmonics become discrete, standard Time Domain Averaging can be directly applied to the PSEM signal to achieve signal enhancement. Thus, Synchronous Phase Averaging can provide signal enhancement performance without the related expense and tachometer requirement. The elimination of tachometer monitoring reduces instrumentation, data acquisition, and most importantly would allow direct application to remote sensing such as gear box diagnostics using a single acoustic or accelerometer measurement.

Successful application examples are utilized to demonstrate the performance of the SPA techniques for gearbox diagnostics. The test data were captured from externally mounted accelerometers of a gearbox with both input pinion and output gears having 27 teeth. The shaft rotational frequency is running at approximately 71.7 Hz during the tests, therefore, the predicted gear mesh frequency is approximately 1935 Hz (27N). Six sets of test data from the gearbox are used to demonstrate this SPA application.

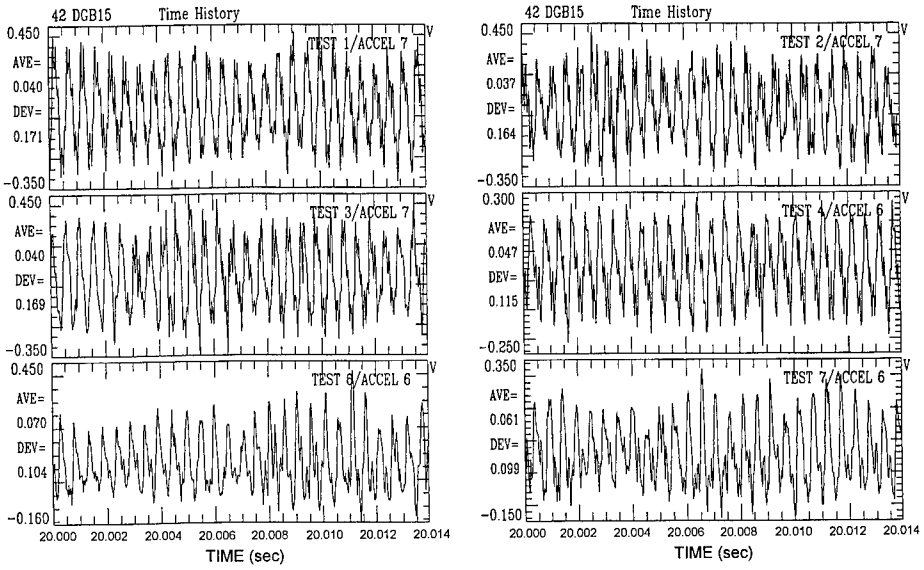


Figure 6 **RAW Signal of 42 Degree Gearbox Data with 500 Averages (Fundamental Frequency=Sync)**

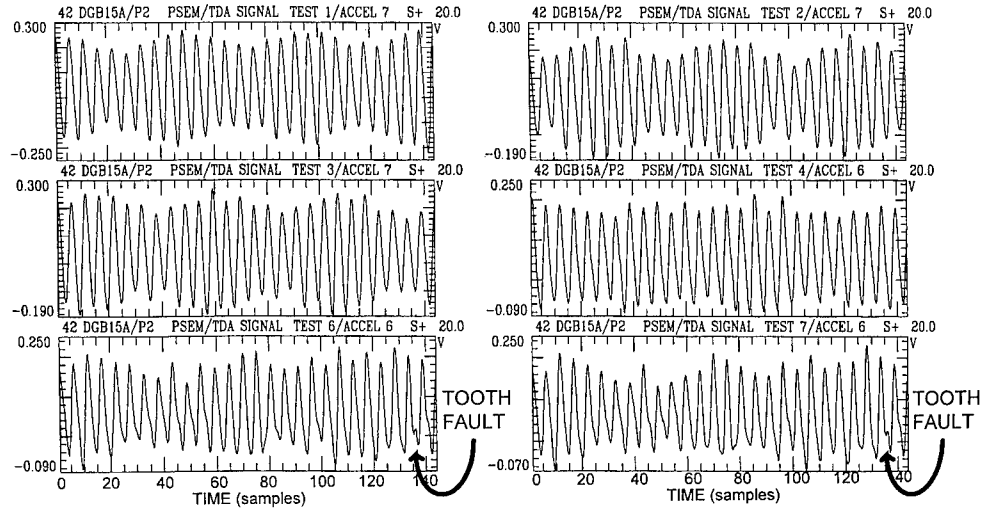


Figure 7 **SPA Signal of 42 Degree Gearbox Data with 500 Averages. (Fundamental Frequency=Sync)**

Figure 6 shows the raw time history of an accelerometer measurement for these 6 sets of gearbox data during one revolution of shaft rotation. The waveforms of these raw signals are all corrupted by background noise. Now, in order to perform SPA, the raw accelerometer data is first synchronized to the gearmesh frequency using the PSEM method. Since the PSEM data has now been synchronized to Sync, Time Domain Averaging is then directly applicable to the PSEM data using a TDA fundamental frequency of 71.7 Hz (Sync). Figure 7 shows the six resulting enhanced SPA signals over a full cycle of shaft rotation. The number of ensemble averages performed in the analysis is 500. Within this full cycle of shaft revolution, an enhanced signal with 27 cycles of oscillation is observed which corresponds to the tooth mesh waveform. In examining the smoothness of this waveform, a noticeable discontinuity or irregularity is readily observed in tests 6 and 7 at the 26th oscillation cycle. This discontinuity repeatedly shows up in the SPA enhanced signal at other time periods within the test. This indicates a gear tooth fault at the 26th gear tooth within the hardware used in tests 6 and 7. This results of SPA analysis correctly identify the conditions of the test gearbox of tests 6 and 7.

CONCLUSIONS: By converting the instantaneous phase signal into an equivalent resampled time signal, the PSEM technique transforms a quasi-periodic spectral component of Sync into a pure-tone discrete component. This discretization process generates a highly desirable effect on the remaining signal with all the other Sync-related components automatically becoming discrete. With this enhanced frequency resolution, the PSEM signal can recover detailed spectral information useful in diagnostic evaluation. Within this newly transformed PSEM signal, since the sync and sync harmonics become discrete, standard Time Domain Averaging can be directly applied to the PSEM signal to achieve additional signal enhancement. The SPA analysis discussed above does not require tachometer or key-phaser input as the STA would require. Such synchronization can be directly performed on any gearmesh (GM) frequency of a chosen rotor within a measured vibration signal. Since the amplitude of the GM component measured by a vibration sensor is typically strong, synchronization to various RPM within a single measurement can be achieved by tuning the PSEM reference frequency to the desired GM frequency. Therefore, the complexity of instrumentation and data acquisition requirement is reduced. Furthermore, non-intrusive sensing with applications such as gear box diagnostics using a single acoustic or accelerometer measurement becomes feasible.

REFERENCES

[1] McFadden, P.D. "Interpolation Techniques for the Time Domain Averaging of Vibration Data With Application to Helicopter Gearbox Monitoring," Aero Propulsion Technical Memorandum 437, AR-004-488, DoD-Science and Technology Org. Aeronautical Research Lab., Commonwealth of Australia, 1986.

[2] Braun, S. and B. Datner. " Analysis of roller/ball Bearing Vibrations," ASME paper No. 77-WA/DE-5, 1977.

[3] Leon, Robert L. and Martin J. Dowling. On-line Monitoring of Turbine Blades for Incipient Failure Detection. EPRI Final Report RP1854-1,2, 1986.

[4] Martin Dowling and Keith Trainor. "High Frequency Spectral Analysis Study for Helicopter Bearing and Gear Failure Detection.", Liberty Technology Center Final Report for NAVAL Air Development Center. Report No. NADC-90033-60. January 1990.

[5] Jong J., Jones J. "Phase Synchronized Enhancement Method for Space Shuttle Main Engine Diagnostics", NASA Conference on Advanced Earth-to-Orbit Propulsion Technology, 1994

[6] Jong J., Jones J., Jones P., Nesman T., Zoladz T., Coffin T., "Nonlinear Correlation Analysis for Rocket Engine Turbomachinery Vibration Diagnostics", 48th Meeting of the Mechanical Failure Prevention Group (MFPG). April, 1994.

[7] Jong J. Y. and Coffin T., "Diagnostic Assessment of Turbomachinery by the Hyper-Coherence Method", NASA Conference Publication 2436, Advanced Earth-to-Orbit Propulsion Technology, May 1986.

[8] Jong J., Jones J., McBride J., Coffin T., "Correlation Identification Between Spectral Components in Turbomachinery Measurements by Generalized Hyper-Coherence", 3rd International Machinery Monitoring And Diagnostic Conference, December, 1991.

[9] Bendat J.S. and Piersol A.G., Engineering Application of correlation and spectral analysis. John Wiley & Son, 1980

[10] Bendat J.S. Nonlinear system analysis & Identification John Wiley & Son, New York, 1990

[11] Collacott R.A., Mechanical Fault Diagnosis and Condition Monitoring, Chapman and Hall, London, 1977.

[12] Eshleman, R.L., "The role of sum and difference frequencies in rotating machinery fault diagnosis," Second intl. Conf. on vibration in rotating machinery, Churchill College, 1980

[13] Dowling, Martin. AM and FM Digital Demodulation Technique for Electric Power. Final Report to Naval Ship Systems Engineering Station, FRC Project 5847-05,-13, February, 1984.

[14] Taub, H. and Schilling D.L. Principles of Communication Systems. McGraw-Hill, New York, 1971

RUSSIAN TECHNOLOGY INTERCHANGE #1

PRINCIPLES AND EXPERIENCE OF THE ON-CONDITION APPLICATION OF OILS AND WORKING FLUIDS TO AERONAUTICAL ENGINEERING

Boris Bedrik
Doctor, Petroleum Center under the Russian State
Standard Committe, 103006, Moscow, Dolgorucovskaia,21

Michael Yampolsky
Doctor, Scientific & Technical Centre "EXPERT"
140003, Liubertsy-3, Moscow region, a/b 8, Russia

Jean Szydywar
Doctor, Nico Company
75008, Paris 66, Avenue des Champs-Elysees /51, Ponthiev, France

Youri Ereomine
Nico Company (Representative office in Moscow)
75008, Paris 66, Avenue des Champs-Elysees /51, Ponthiev, France

Abstract: With dvances in technology, increasingly important become the operational properties, consumption, cost and shortage of oils, working fluids (subsequently referred to as oils) as well as the role of oil parameters as diagnosis indications for an "oil-machinery" system as a whole. Hence it follows that the man-hours required for maintenance and improvement of ecological situation near machinery operation sites should be reduced.Especially important is the problem regarding aircraft (AC), which is associated with elevated requirements to their reliability and operational safety. The analysis of the problem has shown that the major way to its solving is the creation and implementation of a methodological and information-analysis base providing the on-condition oil application in place of scheduling oil changes and, as a consequence, the more complete realization of oil quality potential in service. The paper is prepared based on the analysis of the results of multiyear investigations of oil properties in the course of their accumulation for machinery of various types, mainly aeronautical engineering, and their diagnosing. It is the continuation of works presented previously at analogous conferences held in Swansea (England), Buenos Aires (Argentina),Pensacola (USA) [1-3].

Key Words: Oil; fluid; application; condition; operation; destruction; oxidization; viscosity,

Investigations and implementation of the on-condition working-fluid application to AC's hydraulic systems: A purposeful investigation of working-fluid properties under operational conditions in AC's hydraulic systems has shown the possibility of implementing the

principle of their on-condition application instead of scheduling their changes (every 200-250 hours).In so doing, it was found that a decrease in kinematic viscosity of АМГ-10-type fluids, containing a stiffening components, takes place. The duration and the allowable level to which the viscosity decreases under the conditions existing in a hydraulic system, and consequently, the service life of a working fluid are dependent on the intensity with which various factors affect the fluid causing the mechanical destruction of a stiffening polymer and influencing destruction kinetics. Despite the fact that oxidization processes occur in the fluid accompanied by the accumulation of oxidization products increasing the acid number, the condition of the АМГ-10 hydraulic fluid is dominated by destructing the stiffening component and decreasing its viscosity.

This conclusion is supported by the experiment conducted on the simulation facility (pump rig) and ultrasonic instrument with variable actions (pressure 21 MPa, test duration up to 50 hours, fluid volume 18dm^3, temperatures 125 and 150°C - the maximum allowable for the АМГ-10 fluid; the ultrasonic action was exerted during 1 hour). As for synthetic hydraulic fluid 7- 50c-3 containing no stiffening agent, its condition was dominated by the accumulation of oxidization products. In so during its acid number increased and viscosity increased a little. An increase in the intensity of mechanical action upon the АМГ-10 fluid due to increasing pump speed from 2900 to 4000 r.p.m. resulted in decreasing fluid viscosity by 10-15% relative to its initial level accompanied by a minor variation in acid number. Replacing nitrogen with air in the experiment led to augmenting the oxidization process in the fluid and increasing its acid number by a factor of three.

With ultrasinic action an increase in oscillation amplitude from 40 to 65 mk resulted in sharply amplifying mechanical destruction of the polymer stiffener (fluid viscosity decreased by 30-40%). From the aforesaid it follows that parameters of hydraulic fluids in AC's systems differ from their initial levels and the main task is to establish allowable levels of governing paramrters for concrete "fluid-AC's hydraulic unit" system and introduce their operational monitoring. Thus we have worked out the following list of governing parameters and their allowable levels (Table I).

Table I. Governing parameters of working fluids for AC's hydraulic systems and their allowable values

Parameter	Working fluid, allowable/standard values			
	АМГ-10	МГЕ-10А	7-50с-3	Non-flammable hydraulic fluid
Kinematic viscosity at 50°C, cSt	7 / 10	7 / 10	17-26 / 22 at 20°C	6 / 8.7
Acid number, mg KOH/g, no more than	0.015 / 0.05	0.9 / 0.4-0.7	0.8 / 0.1	0.2-0.3/ 0.08
Fraction of mechanical additives and water, %, no more than	0.005 / 0.003	0.005 / 0.003	0.005 / 0.002	According to documents
Time to stabilize governing parameters	30-40	40-50	60-80	60-80

As current parameters reach the values indicated in Table 1, the fluid in a hydraulic system should be replaced with the fresh one. This rule of using hydraulic fluids is now introduced throughout the aircraft fleet in Russia, which has provided a 2-3 times reduction in their operational consumption.

A methodology was craeted ensuring the on-condition use, regeneration and reuse of fluids in AC's hydraulic systems including their cleaning, recovering the parameters deviated beyond the allowable limits. The methodology uses a complex system incorporating the computer-aided processing of information and reference data.

Investigation and implementation of the on-condition oil application to aircraft engines: Based on the analysis of normally operating "oil-AC's engine" systems during scheduled period before oil change, the following main indications for implementing the on-condition oil application were established: continuous circulation and contact of a limited oil volume with oil system parts; accumulation in the oil of the information and diagnosis indicators of the condition of the system as a whole; refreshment of oil quality with fresh oil during machinery operation.

In destincton to the joint USA program JOAP, in Russia the activities in this field are carried out according to purposeful programs, the most successful of them is realized in aviation. In pursuing these aims, put in operation is the network of laboratory and rig centers destined for testing petroleum products, assessing and monitoring their parameters, including Petroleum Product Test and Sertification Centers under Russian State Standard Committee.

The implementation of the methodology for the on-condition oil application in technical equipment is a complex of long-term, time-consuming and expensive measures, among them there are: analysis of operating conditions, substantiation of the list and levels of acting factors for concrete "oil-engine" systems as well their functional and governing parameters; forecasting the regularity of variations in oil properties, establishing allowable levels of parameters, methods and means for monitoring them during operation; choice of aeronautical equipment, development of programs and directions aimed at conducting test work in various stages; esrablishing the rules for assessing the condition of technical equipment, oil, oil systems in operation; taking decisions regarding the ahead-of-schdule oil changes and stoppage of equipment operation (if necessary); removing the monitored engine from service, supervision of defects in the engine's oil system and transmission parts at repair works; drawing up a conclusion based on the results of the investigation.

Taken as governing parameters can be those relating to both the oil and the equipment but provided that the parameters define the reliability of a system as a whole. The analysis of the results of investigation has shown that the products of wearing can not alwais be considered as dominating factor in comparison to oil quality parameters associated for example with oxidization and destruction processes in oil components and oil foaming.

Such parameters as viscosity, acid number, flash point, and filterability can turn out to be more sensitive and making it possible to obtain the information about the predicted condition of the "oil-engine" system elements earlier than the information about the concentration of wearing products in the oil.

For the oils usable in aircraft engines a list of governing parameters and their allowable levels was established (Table II) on exceeding of which the oil should be changed with the fresh one.

Table II. Governing parameters of oils for their on-condition application to aircraft engines and their allowable values

Oil specification	Governing parameters, allowable/standard values				Time to stabilize parameters, hr.
	Kinematic viscosity, cSt, at temperature: $50^0 C$	$100^0 C$	Acid number, mg KOH/ g	Wearing products, g/t	
МС-8п	10.0 / **8.0**	-	1.0 / 0.05	5.0(Cu) 6.0(Fe)	50-60
МС-8рк	11.0 / **8..0**	-	1.0 / 0.15	5.0(Cu) 6.0(Fe)	50-60
ИПМ-10	-	5.5 / **3.0**	4.0 / 0.05	4.0(Cu) 6.0(Fe)	80-90
ВНИИ НП-50-1-4ф (ВНИИ НП-50-1-4у)	-	5.5 / **3.2**	3.5/0.22 (0.25)	-	80-90
МН-7, 5у		**6.6**; 11.0/**7.5**	- / 0.1	-	100-120
Б-3в	-	6.2 / **3.0**	2.0 / 4.4	-	-
Л3-240	-	6.2 / **4.8**	2.0 / 5.0	-	-

Note: bold type - " not less", other - " not more"

Given in Table III are the data on implementing the on-condition application of oils to aircraft engines.

Table III. Data on implementing the on-condition application of oils to AC's engines.

Engine	Oil (base)	Number of oil changes in engine	Time between oil changes, hr		
			Primary	Intermediate	Established
For 1- and 2-engined AC	Petroleum	4-6	50-100	200	On-condition
	Synthetic	4-6	100	300	- " -
For 3- and 4-engined AC	Petroleum	5-6	300	600	- " -
	Synthetic	5-6	600	1500	- " -
For cargo ang assenger AC	Petroleum	3-4	200	600	- " -
	Synthetic	3-4	300	600	- " -
For trainer AC	Petroleum	3-4	300	500	- " -
	Synthetic	-	-	-	
For helicopters	Petroleum	3-4	one year	two year	- " -
	Synthetic	3-4	one year	two year	- " -

The implementation of such an approach has enabled a 2-3 times reduction in oil consumption and provisions for other advantages.

Assessment of the possibility for applying the oil Turbonicoil-210A of the French company Nico to Russia-produced aircraft engines: Currently accomplished in Russia are the measures and investigations on a comparative assessment of the qualities of domestic and foreign oils, lubricants and special fluids, on estimation of the possibilities of the application of these products to Russian aircraft engines, on the adaptation of the procedures and means for testing and monitoring the quality of petroleum products during their direct use

and on establishing business contacts with corresponding companies.

This is justified by searching alternative foreign oils in connection with exporting Russian aeronautical equipment of civil and military purpose and by the need for covering the deficit of main oils at Russian internal market, as well as by other factors.

It turned out that the oil analogues produced by the French company Nico are the most compositionally and qualitatively close to the main Russian oils. According to our evaluations, the oils of this company are also the most acceptable from technical, economic and cost considerations.

Thus, investigations have shown that the polyisoolefine-acid-based Turbonicoil-210A oil (TH-210A) is practically a full analogue for the isoparaffinic oil ИПМ-10 basically used in highly heat-loaded engines, including the complex of additives and agents.

According to the results of qualification tests, the oil TN-210A features the higher levels of thermal stability (up to 200°C) and foaming resistance with practically equivalent physical and chemical properties, which facilitates the adaptation of monitoring techniques and means used in laboratory, rig and operation testing of this oil under Russia's conditions.

At present together with the French party we are treating a question regarding the full-scale operational tests and introduction of the oil in highly heat-loaded engines of maneuverable AC interchangeably with Russian oils. It is planned to implement the complex of works and investigations according to Russia-accepted methodology for the on-condition application of this oil to AC's engines.

Some theoretical aspects of the on-condition application of oils and fluids: The methodology of the on-condition oil application is based on the substantiation of a functional interconnection between oil and equipment parameters, which enables an "oil-equipment" system to be considered as an interconnected object. This interconnection within an "oil-equipment" system is represented in the columns of Table IV.

Note to the table IV: F_{op} and F_s are respectively the operational and structural factors; t is the temperature, p is the pressure; P_v is the vibration parameter; Q_s is the supply amount (volume); R_s is the supply rate; I_f is the filterability index; t_s is the solidification temperature; ν_t is the current kinematic viscosity; ψ is the ambient humidity; q_d is the deposition amount; K is the acid number; R_w is the wearing rate; C_{wp} is the wearing product concentration; φ is the friction coefficient; $I_c(R_c)$ is the corrosion index (rate); t_f is the temperature foam formation.

Table IV. Interconnection within an "oil-equipment" system.

Process in the equipment	Oil operational property	Factors acting in the system	Functional parameters		Governing parameters
			Equipment	Oil	
Supply	Destruction of oil components	$F_{op}(t;p;P_v)$	$Q_s; R_s; I_f$	$t_s; I_f; \nu_t$	$Q_s; R_s; \nu_t$
Circulation ability	Structural and mechanical stability	$F_{op}(p;t;\psi_0)$ F_s	$I_f; Q_s$	$\nu_t; t_f$	$Q_s; \nu_t$
Oxidization of oil components, deposition and sedimentation in engine	Physical and chemical stability	$F_{op}(t;\psi;\varphi)$; F_s	$I_f; q_0$	$K; I_f$	$I_f; q_d; K$
Wearing	Wear-resistant properties	$F_{op}(p;t)$; F_s	$R_w(C_{wp})$; φ_t	$R_w(C_{wp})$; φ_t	$R_w(C_{wp})$; φ_t
Corrosion	Corrosion-resistant properties	$F_{op}(p;\psi)$ F_s	$I_c(R_c)$	$I_c(R_c); K$	$I_c(R_c)$

The "oil-engine" system is categorized as the complex and multifactor one. The solution of system optimization problems is aimed at the substantiation of the system governing perameter (P_{op}) and determination of its allowable level in operation according to relationship (1):

$$P_{op} = f(F_{op}; F_s; F_e; P_f), \qquad (1)$$

where F_e is the oil operational properties factor; P_f is the system functional parameter.

The model of the "oil-engine" system (Table IV) is the basis of the methodolody of the on-condition oil application which is implemented in accordance with the following principles: establishing interconnections inside concrete systems (factors and functional parameters); substantiation of a governing system parameter and its allowable value for operational conditions; technical and economic evaluation of the effectiveness of oil application to the engine.

This approach is illustrated in Fig. 1 where the operational region is shown for kinematic viscosity of the oil MH-7.5y ranging from 6.6 to 11.0 cSt with an initial level of 7.5 cSt. The oil is applied to two engines of different types. In this case, dominating in the engine of the first type are the loads causing the destruction of oil components, whereas operational conditions of the engine of the second type are dominated by the temperatures at rear engine bearings, which is accompanied by intensifying oxidization processes and leads to accumulating oxidization products and increasing the oil viscosity. The problem is solved on establishing allowable values of the MH-7.5y oil viscosity: it should be no less than 6.6 cSt for the engine of the first type and no more than 11.0 cSt for the engine of the second type. In cases of oil viscosity being beyond the limits indicated, the oil is to be changed with the fresh one.

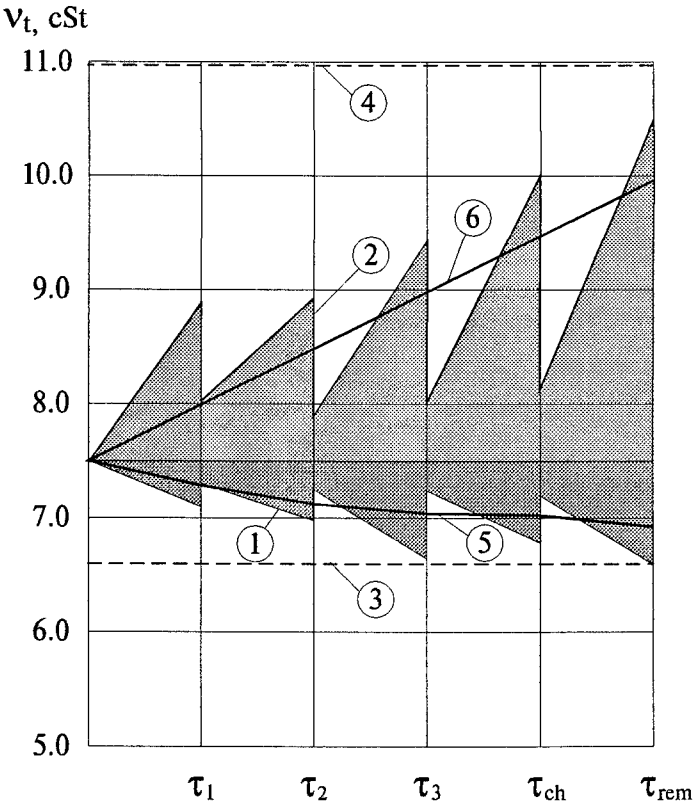


Fig. 1 Dynamics of varying the MH-7.5y oil kinematic viscosity taken as the governing parameter in the engines of the first (1) and second (2) types

Note to the figure 1: In-service time of an engine expressed in periods of maintenance monitoring and reoiling is $\tau_1, \tau_2, \tau_3 \cdots \tau_{ch}$; τ_{rem} is the time for engine removal).

1 , 2 - bounds of the working range of viscosity variation providing reliable operation of the engines of the 1st and 2nd types; 3 , 4 - allowable viscosity levels for the oil applied to the engines of the 1st and 2nd types, respectively; 5 , 6 - design viscosity variations of the oil when applied to the engines of the 1st and 2nd types, respectively.

The following design formulas for determining current viscosity values of the oil MH-7.5y in its application to above indicated engines were obtained:

for the 1st engine $$\nu_t = 7.5 - 0\ 001\tau + 2.337 \cdot 10^{-7}\ \tau^2 \qquad (2)$$

for the 2nd engine $$\nu_t = 7.5 + 0.01\tau + 6.3 \cdot 10^{-6}\ \tau^2 \qquad (3)$$

Transition to definite steady levels of oil properties in the course of its use in techincal equipment is most likely associated with structural and energetic mutual adaptation of the elements of tribomechanical systems.

With a definite combination of phases, operational conditions and structural features of a system and with accounting for repetitive fresh-oil replenishment in operation, current steady-state mode, featuring system self-adjustment, takes place. It allows a conclusion to be made that with a partial oil changes the tribosystem is a less long-run one because it operates in a nonlinear mode characterized by fluctuating oil properties from where follows the feasibility of the current self-adjusting system stabilization mode.

The realization of self-adjustment mode of operation of various "oil-engine" systems is supported by the results of numerous investigations and conclusions on a two-stage variation of the parameters of systems elements: short-duration evolution of their properties (operating time of 40-120 hours) and long-duration transition to a steady mode (well over 60-120 hours). This is also confirmed by the results of testing on simulation facilities with imitation of the operation of actual aircraft units and subassemblies. For every "oil-engine" system the stages of the evolution of oil properties and their adaptation should be established. It is of importance therewith to exclude from changing the oil in the stage of the evolution of its properties.

At present we are implementing this approach in aviation.Involved in this process are aircraft industry works, research and servicing organizations, and created under the Russian State Standard Committee are the special test and certification centres each oriented to a certain type of machinery and associated oils and hydraulic fluids.

Conclusions: 1. Established and implemented on the basis of experimental and theoretical investigations is the possibility of the on-condition application of oils and working fluids to aircraft systems in place of schediling their changes, which has enabled a 2-3 times reduction in oil consumption and other operating costs, diagnosing oils' and systems' condition in operation. Such an approach makes it possible to correct perameters of oils during their production.

2. It was found that the implementation of the on-condition oil application to aircraft engines

is based on the adaptation of oil properties to engine operational conditions and the associated necessity for establishing the interconnection between factors and functional parameters within concrete "oil-engine" systems and the stages of evolution and stabilization of oil properties, which permits excluding draining the oil from technical equipment when its quality potential is not exhausted yet.

3. The possibility is demonstrated of the application of the oil Turbonicoil-210A of the French company Nico to Russia-produced aeronautical engineering.

References:

1. B.Bedrik "Methodical principles of increasing of Fuel-Oil Materials (FOM) use efficiency in systems in service". International conference on condition monitoring'94. Pineridge Press Swansea, UK,1994. pp. 81-88, 1994.

2. B.Bedrik "Methodical principles of increasing of efficiency in the use of Fuel and Lubricating Materials (FLM) in systems during service". International conference'94. Buenos Aires, Argentina, p. 4, 1994.

3. B.Bedrik, M.Yampolsky "Study and procedure of oils application in aeroengines according to the condition. Joint oil analysis profram". International Condition Monitoring Conference. Pensacola, USA, pp. 209-215.

4. Allison M.Toms "A Preliminary Report on the Evaluation of FTIR for Lubricant Condition and Contamination Determination in Support of Machinery Condition Monitoring. Synthetic Lubricants. Joint Oil Program Technical Support Center".Pensacola,Florids, USA, p. 120,1994.

5. Gerd Koolhaas, Jurgen Frank. "Experiences in the Condition Monitoring Program of the Tornado Aircraft". International Condition Monitoring Conference, JOAP, Pensacola, Florida, USA, pp. 24-25, 1994.

6. Ron Greig "US Air Force Program Overview". International Condition Monitoring Conference. Pensacola, Florida, USA, p. 22, 1994.

DETERMINATION OF TECHNICAL STATUS, CAUSES OF WEAR AND FAILURE OF MACHINE PARTS USING THE METHODS OF METALLOPHYSICAL ANALYSIS

V. Lozovsky
Professor, Scientific & Technical Centre "EXPERT"
RUSSIA, 140003, Liubertsy-3, Moscow region, a/b 8

Abstract: The use of the scanning electronic microscopy and other methods of metallophysical analysis as well as the methods of nondestructive testing allows objectively identifying failure causes connected with the wear and breakdowns of machine parts and implementing the monitoring of their status in operating tectnical equipment.

Key words: Technical status; metallophysical analysis; wear; failure; machine part.

The on-condition operation of technical equipment implies the availability of reliable methods and means allowing in field conditions the life-limiting defects of machine parts to be detected. The analysis of results obtained over the 15-year examination of the causes of failures of aeronautical engineering associated with breaking down or wearing parts makes it possible to conclude that these cases are as a rule of law-governed in character and attributable to one of the following causes: imperfection of a design or manufacturing techniques, low quality of part manufacture, or operating errors.

A system of techniques, criteria and potrable flaw-detector devices was developed and introduced in the practice of operational engineering service which allow both promptly detecting the pre-failure status and causes of breaking down and wearing machine parts based on the metallophysical analysis. The effectiveness of this system is well illustrated by an example presented of estimating the serviceability of standard-type machine units like rolling-contact bearings, gear transmissions, spool-and-sleeve pairs in early stages of their operation or tests.

As applied to precision spool-and-sleeve friction pairs of control devices of fuel and hydraulic units of aeronautical and other complicated equipment operating under conditions of oscillatory movements of their parts, considered as hezardous are the seizure taking place even at microscopic segments of their actual contact.

In spool-sleeve control devices even a temporary augmentation of friction violates the principle of monitoring the variations in pressure or working fluid consumption and leads to defective operation of the entire system.

The character of damages appearing in the course of

operating fuel-control and hydraulic units could not be explained by actions of solid particles becoming trapped between parts. The cause of part damages which result in disrupting friction stability of little-moving spool-sleeve pairs is the seizure of mating surfaces over the segmant of their contacting.

Considered in terms of modern ideas regarding the conditions of the origination of inter-part seizure in friction, this phenomenon as applied to above-mentioned spool-sleeve pairs did not find explanation, which hampered woking out the measures aimed at its prevention.

Finding out the seizure causes and possible failures of spool-sleeve pairs created a demand for studing both the peculiarities of part damages and the loading conditions including the absorbed contact loads and relative movements.

The damages caused by the seizure of mated moving parts operating under the conditions of contact-vibratory loading is characterized by the following indications.

Such damages have the form of separate breakaways, accumulations of metal particles and scratches directed along the generatrixes of cylindrical surfaces of the parts. Sometimes damage dimensions are so small that they can only be revealed with the aid of a metallographic microscope with a magnifying power of 30-100.

The damage of above-mentioned character on one of the part of a precision pair have alwais the reciprocal ones on the mated surface of the other part. This being so, by the shape, dimensions and relative positions of the damaged segments on mated surfaces of the parts look like the mirror images of each other.

The dimensions of individual local damages on the spool and reciprocal ones on the sleeve are incommensurably less than the working stroke of the spool relative to the sleeve, which is indicative of appearing the seizure as a result of micromovements vibratory in character. In controlling spool-sleeve pairs this corresponds to a steady-state mode of controling device operation.

During seizure on the furfaces of the parts in the contact zone changes in metal state occur accompanied by its significant hardening. When etching the surfaces of the precision pair, the regions not etched are revealed called white layers, which are formed in the damag ezone. This suggests that phase transformations take place in part's matarial cccompanied by not only hardening but also changing magnetic properties in microvolumes of the material subjected to seizure with the mated part.

All the aforesaid enables the methods of nondestructive testing to be used with the aim of revealing seizure indications considered as the causes of failures of precision control devices. In so doing, the identification of objective indications of the changes in physical and mechanical

properties of parts' material makes it possible to predict potential seizure-induced failures of movable joints in further operation.

Concrete methods for identifying microindications of structural and phase changes in parts' material with the aid of the methods of nondestructive testing are the know-how and a constituent of the database of the expert-examination computer system intended for assessing the technical status of machine parts and structures.

The metallographical analysis of fractures and materials is the necessary means for objectively ascertaining the causes of breaking down the parts of various techinical equipment. Scanning electronic microscopy, X-ray diffraction analysis, metallography and other methods including methods of nondestructive testing are successfully used for these purposes.

Taken as a pictorial example can be the employment of metallophysical analysis for identifying the cause of breaking down the titanium engine compressor disk. As a result of the analysis, the cause of the disk breakdown associated with a metallurgical material defect was found out and the measures for preventing its manifestation in other aircraft engines were developed.

The presence of extraneous impurities of the same density as that of the main material with no interface discontinuity is reliably detected with the aid of the ultrasonic inspection method. In so doing, the diagnostic chech of engine compressor disks was proposed and successfully accomplished immediately in the course of aircraft operation.

A SYSTEM OF METHODS FOR IDENTIFYING THE CAUSES OF FAILURES AND ACCIDENTS OF AERONAUTICAL ENGINEERING

Yu. Korovkin,
Doctor, Scientific & Technical Centre "EXPERT"
RUSSIA, 140003, Moscow region, Liubertsy-3, a/b 8

Abstract: The scientific and methodological foundation as well as the substance are presented of the system of methods for identifying the causes of failures and damages of aviation materiel developed on the basis of theoretical and experimental investigations and generalization of multiyear experience in practical work in this field.

Key words: Accident; aviation; methods; failure; cause; engineering.

The identification of the true causes of hazardous failures and accidents of technical equipment is the necessary condition for developing and accomplishing effective measures aimed at precluding their recurrence.

The identification of these causes is a highly complex task especially when the accidents are accompanied by the secondary destruction of equipment units taking place as a result of explosion, fire, collision of damaged parts, etc.

Aviation belongs to the area of mankind's activity where the solution of such a problem is of particular urgency. ICAO's documents (Supplement 13) only establish a general order of organizing the investigations of aviation accidents (AA) and serious incidents.

Created by Russian scientists is the system of mathods for examinating faulty pieces of aeronautical engineering (FPAE). These examinations feature a series of principal peculiarities which differ them from other investigations of technical equipment.

The major peculiarities formulated on the basis of generalizing a multiyear experience in post-accident examinations of FPAE are as follows.

1. The picture of the origination and development of the phenomenon under investigation, which is as a rule resulted from a variety of causes, should be reproduced.
2. The sources of information used in failure cause identification (parts and units of FPAE, flight recorders, etc.) have considerable damages and destructions.
3. Functional connections between separate elements of aircraft (AC) systems are disrupted and some of these elements are often not found at the accident site (due to being completely destructed, burned, fallen into water, etc.).
4. The piece under examingation where the gailure occured

resulting in the accident is a single one.
5. The information used in failure cause identification esists in many cases in the form of microscopic tracks.
6. The AC's systems under examingation have often different principles of functioning (from mechanical to electronic).
7. The questions relating to defferent sciences (from mathematics to criminalistics) have often to be solved in the course of a post-accident investigation.
8. Significant part of the investigation is commonly carried out in field conditions, directly at the accident site.
9. The investigation has to be conducted in the shortest possible time.

The complexity, comprehensive, multiaspect and multidisciplinar character, elevated labour content, involvment of large number of various specialists, joint application of formalized, empirical and heuristic methods are responsible for the nees to use a system approach to these investigations and to create a system of methods ensuring the solution of entire complex of problems arising in all stages of investigating AA.

The system of methods created is based on the methodology developed of examinating RPAE (Fig. 1). In developing this methodology the experience in examinating thousands of AA was generalized, factors acting on pieces of aeronautical engineering (AE) during AA were revealed, laws of the phenomena occuring in accidents were determined, available data on variations in AE condition during operation and repair were used, structural and functional features of AE pieces were taken into account, methods of natural sciences were employed for examinating objects and phenomena.

Taken as an important element of the methodology is the clear statement of the problem on examinating FPAE on the basis of considering an AA model combined with defining possible kinds and sousces of information (Fig.2).

In developing scientific and methodological basis of examinating FPAE, the principles of conducting sich examinations were formulated. They include both known general principles of scientific investigation and specific principles accounting for the peculiarities of examining RPAE. The most important of them is the assurance for the credibility of results to be obtained. This is a major requirement imposed on the methods which, in particular, is satisfied with the following ways:

using a variety of indications and methods based on different physical principles in solving the problem;

maximum possible recovery of the information distorted as a result of acting AA factors;

carefully checking the information used for its validity by means of excluding its inherent contradictions;

documenting all the results obtained during an investigation;

strictly determining the sequence of operations which excludes the errors capable of resulting in loss of failure information;

combining analysis and synthesis of phenomena under

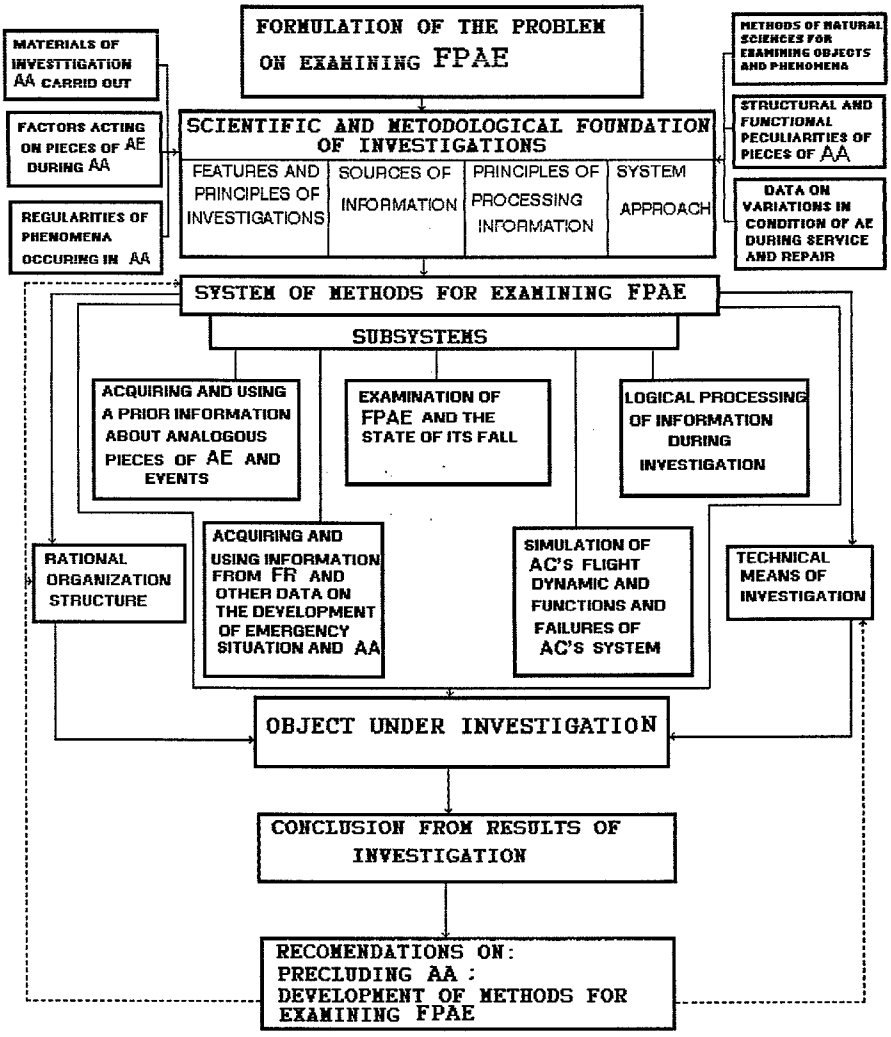


Fig. 1 Methodology of examining faulty pieces of aeronautical engineering (FPAE).

Note: *AA - aviation accident; AE - aeronautical engineering; FR - flight recorder; FPAE - faulty piece of aeronautical engineering.*

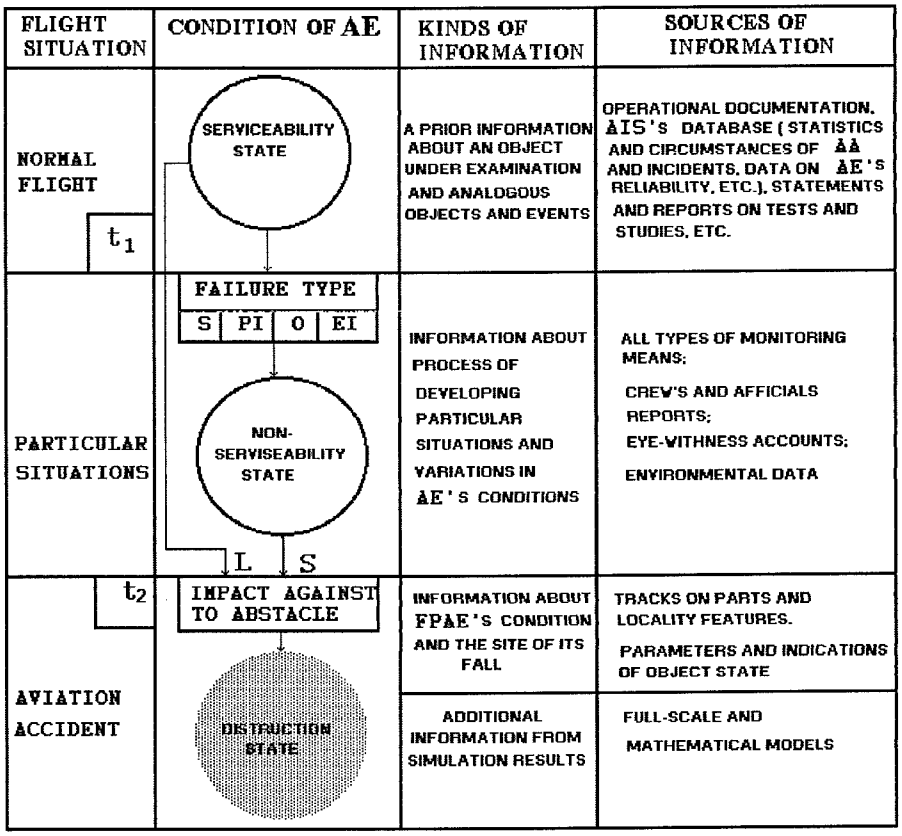


Fig. 2 Formulation of the problem on examining FPAE

Known: *1. At the instant* t_1 *the piece of AE was in serviceable state (SS).*
2. At the instant t_2 *the piece of AE was in destruction state (DS) as a result of impact against an obstacle.*
To be determined: *1. Way of transition (S or L) from SS to DS (that is, whether the failure was in flight).*
2. What is the failure cause:
a) device failed (unit, block, assembly, element);
b) failure cause (structural (S), production-induced; (PI), operational (O), externally-induced (EI))?

examination.

The system of methods for examining FPAE constitutes the major content of the methodology. According to the types of information used in examinating FPAE (see Fig. 2) and the peculiarities of its acquisition, application and processing, the system of methods of investigations includes the subsystems depicted in Fig. 1. Consider each of them.

1. The first one is based on the automated information system (AIS) "Safety" that include civil and military subsystems. The databese of this AIS contains the information about known AAs, incidents, results of investigations carried out.

In addition, widely used are the albums of illustrations of failures made up for basic types of AC and containing the copies of monitoring records, photos, schemes showing the circumstances under which the malfunctions arised and manifested themselves as well as the information reports issued for servicing organizations concerning AAs and incidents.

2. At present a main place in the subsystem of the mathods for obtaining and applying the information about developing the emargency situation and the AA is occupied by the methods ensuring the effective use of monitoring records, especially the data of flight recorders (FR). For the FR information to be used with the purpose of determining AC's flight parameters and assessing the serviceability of an AE and identifying failure causes, a series of user guides were worked out covering practically all basic types of airplanes and helicopters now in service in Russia. On modern AC equipped with magnetic recorders an automated input of flight information in a personal computer was assured and associated algorithms and programs were developed, which provide the maximally full use of recorded data for both evaluation of AE's serviceability and solution of flight -dynamics problems. By way of example, displayed can be the AC's flight trejectory and its spatial attitude. It should be noted that the PC-based system created for processing flight information with the aim of assessing the AC's condition and preventing failures is applied not only in Russia. Thus, for instance, successfully employes in Germany since April 1995 is the system APM OK with the aid of which the data on several thousands of flights of Luftwaffe's MiG-29 aircraft have already been processed.

But the effective use of flight information in investigating AA is often hampered by the fact that the protected airborne recorders in many cases do not assure the complete retention of recorded information. In this connection, developed and widely used are the methods for recovering and obtaining information from both optical and, especially, magnetic recorders. Corresponding user giudes give the recomendations for those people participating in investigation of AA, which provides the retention of recorded information at an accident site.

3. Flight recorders (even in cases of their undamaged state)

carry the information about variations of a limited number of flight perameters and AE's operation. Because of this for evaluating the serviceability of AC's systems, locating a failure and identifying its cause, there is a need to examine the condition of the FPAE and the site of its fall. With this purpose a complex of numerous methods and procedures based on different physical and chemical principles are used which constitute the third subsystem. Used as its basis is the regularity established of reflecting events and phenomena occuring in AE failures and AAs in various tangible tracks on parts and locality features. As a result of generalizing the materials of a rich variety of AAs and experimental studies, the regularities of the formation of tracks on FPAE were revealed based on which are the methods and procedures allowing one to find informative tracks on damaged AC's fragments and "make them tell".

For example, it was found that in the case of impact against an obstacle one can find the tracks reflecting the following characteristics of the AE condition:

kinetic energy of rotating structure members (of engines, generators, pumps, etc.);

pressure of working media (fluids, gases);

amount (volume) of working media and expendable fluids, gases in AC's systems;

position of actuator devices;

position of the elements of automatic systems;

position of the elements of instruments and indicators;

presence (or absence) of residual magnetizationof a definite level;

degree of parts' heating.

Also, structure parts carry the tracks of the conditions of their destruction, which enables making a distincion between damages, fires, explosions that can take place in flight and the same phenomena occuring in the case of impact against an obstacle.

For the information about the serviceability, mode and conditions of operation of a unit under examination to be obtained on the basis of analyzing the tracks, special procedures were developed giving the relationship between track features (colour of oxide film, arrangement of metallized coatings, relative positions of the elements of spool-and-sleeve pairs, etc.) and some parameters characterizing the condition of an object under examination. Such relationships were found on the basis of special studies (flight and groun-based tests, simulations). These relationships in the form of nomographs, plots, tables, sets of standard specimens are given in user guides for concrete AC.

4. In modern conditions simulations are widely used in investigating AA. Primarily it is the mathematical simulation of flight dynamics and AC's trajectories using the data from FR and other information about the AA and the condition of the FPAE.

Physical simulation (both pilot-in-the-loop and hardware-in

-the-loop) is employed in cases that the information from FR and data obtained on the FPAE condition are insufficient for identifying causes of failures in AE. This type of simulation is intended for establishing:

possibilities and durations of altering the revealed part's defect to the failure of the unit or subassembly;

the nomenclature of malfunctions leading to the failure registered in flight;

actual operational conditions of the part and boundary conditions under which the malfunction appears;

correspondence between part's damages and a definite unit state (serviceable or non-serviceable);

possibilities of practical realization of stated hypotheses regarding the causes of failures.

The totality of the mathematical and physical simulation methods constitutes the fourth subsystem of the methods for examining FPAE.

5. Finally, the fifth subsystem includes the methods ensuring logical processing both the primary information in all stages of investigating AA and the analysis of all data obtained in the course of the investigation for formulating a conclusion on the basis of its results. Among them we would like to set off the method for constructing the schemes of the connection betweem causes and effects which is used in every examination of FPAE and ensures a purposeful acquisition and processing of information.The construction of a cause-effect linkage scheme is accomplished on the stage-by-stage basis beginning with a final effect derived from the analysis of the circumstances of the AA and data from the monitoring system and ending with the prime cause. Currently available is a special software for PCs and methodological guides containing standardized schemes of cause-effect schemes that are to be refined and added with fresh data gained during a concrete investigation.

For attaining a required logical succession in acquiring and processing information, algorithms were developed relating both to separate stages on an investigation and to entire process of examining AE during the investigation of AA.

For conducting examinations of FPAE, developed on the basis of the above-presented methods and practically used is the system of technical means which includes both specialized tools and equipment usable in other areas of science and engineering. In military aviation special-purpose flying laboratories are brought into use for in-site investigations which are equipped with devices necessary for work in field conditions.

Conclusion : The system outlined of the methods for examining faulty pieces of aeronautical engineering provides a sufficiently comprehensive and plausible solution to the complex of the problems associated with determining the serviceability of aeronautical engineering and identifying the causes of its failures in all stages of investigating aviation accidents.

EXPERT - EXAMINATION COMPUTER SYSTEM FOR ESTIMATING OPERATIONAL STATUS OF TECHNICAL EQUIPMENT WITH ACCOUNTING FOR THE RELIABILITY OF ITS LIFE-CRITICAL PARTS

Vladislav Lozovsky
Professor, Scientific & Technical Centre "EXPERT"
140003, Liubertsy-3, Moscow region, a/b 8, Russia
Michael Yampolsky
Doctor, Scientific & Technical Centre "EXPERT"
140003, Liubertsy-3, Moscow region, a/b 8, Russia

Abstract: The paper describes the principles of creating the expert-examination computer system for estimating the operational status of technical equipment which, on the basis of generalizing a multi-year experience in operating aeronautical and other engineering with the use of the metallophysical analysis methods, enables making conclusions on the causes of the wear and breakdowns of machine parts and structural members as well as predicting potential failures and choosing the best suited flaw detection method for similar technical equipment.

Key words: Expert system; computer; technical equipment; reliability; operational status.

The on-condition operation and repair of technical equipment call for developing a package of techniques, criteria and associated means which would provide the distinct and credible in-operation detection of defects in life-limiting parts and units of concrete machines. In many cases these techniques and means are immediately connected with the peculiarities of operating concrete technical equipment and must be continuously upgrading as an operational experience is gained and the equipment itself is renewed. All this demands the availability and regular replenishment of a corresponding database as well as the development of algorithms for forming standardized expert decisions on the status of technical equipment and the causes of its wear, failure or breakdown based on the impartial information acquired during its monitoring. With this in mind an expert-examination system was created for examining operational status of technical equipment which was developed on the basis of the analysis and generalization of the results of a large body of investigations of faulty and failed aviation and other materiel.

For the initial information required for running the expert-examination computer system to be obtained, one uses methodological procedures, criteria, portable flaw-detection and other material-testing instruments which make possible to reveal in proper time the pre-failure status of machine parts and to identify the causes of their wear and breakdown.

The expert-examination computer system is multifunctional. Through an interactive dialog of a user with the computer, it

enables:
estimating the status of life-critical machine parts;
receiving a conclusion on the causes of the wear and breakdown of parts and structural members;
offering the best acceptable flaw-detection method for similar technical equipment;
estimating, in a number of cases, a residual life.

The estimation of the technical status of a part or structural member with the aid of the expert system is accomplished through an operator-computer dialog. In response to computer's query, the operator enters available inspectorial data on the objective indications recognized of changing physical and mechinical properties of the material which allow predicting possible failures of the parts and joints under examination.

In the case of the part breakdown, the inspectorial information is to be entered according to the following divisions:
features of the macro- and microrelief of the failure surface;
features of the fracture;
peculiarities of developing plastic deformation;
peculiarities of part's loading;
quality of welding and assembling the part or structure;
operational conditions.

Each of the divisions indicated consists of smaller and more specialized groups, each including in turn a set of indications characterizing the part status. For example, included in the division for features of the macro- and microrelief of the failure surface are the following six indications:
microrelief of the failure surface (fracture);
microrelief of the failure nucleation site;
microrelief of the zone of gradual crack propagation;
microrelief of the fracture of homogeneous structure or the zone of gradual crack propagation (failure);
macrorelief of the zone of accelerated crack propagation;
macrorelief of the rupture zone.
Each of these groups contains 10-20 indications.

It is just the formulation of the set of indications of various types characterizing the status of a part or structural member that was one of the most important and complex problems in developing the expert-examination computer system. Another problem of great importance was to form such aggregates from these set of indications which would allow predicting a failure and, if a breakdown has occured, uniquely characterizing the kind of the fracture and identifying its cause.

The stationary database of the expert-examination computer system contains a great body of data, gained in many years of operating aeronautical and other equipment, on appearance status, matarial quality, assembly quality, operational conditions of parts, units and structures.

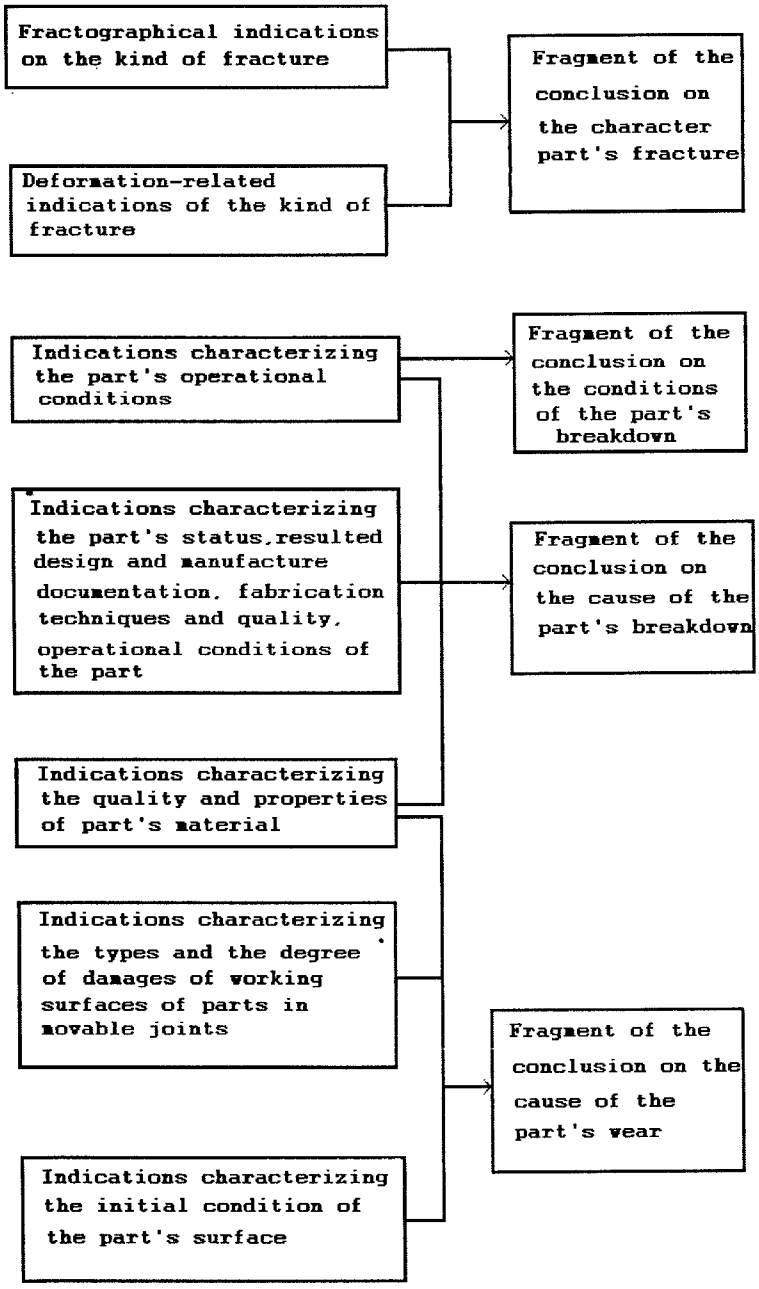


Fig.1 Sheme of forming a conclusion on the character, conditions and cause of the pert breakdown

All these data are stored in definite combinations which are referred to as "standards". Each standard characterizes a concrete cause of failure, wear or breakdown of a piece of machinery.

The comparison of inspectorial information, that is, the data obtained during the monitoring or examination of a failed (broken down) part, with the standards stored in the stationary database is executed by the computer according to a special program.

The variants of standardized conclusions produced by the expert system, can have different numbers in the priority list depending on the number of absent (not obtained during inspection) data and the number of disagreements between inspectorial data and the information (combination of indications and actual data) of the standard chosen by the system. The variant of expert decision having the highest rank corresponds to the minimum number of disagreements with a standard is always the most preferable. But the analysis of other variants is also quite useful for the investigation because with any priority rank the expert conclusion contains the information about which of the groups of indications and actual data these disagreements relate to (see Fig 1).

In entering inspectorial data into the computer the invertigator is provided with an opportunity for making help requests to get a reference information stored in the computer memory (comments, schemes, photos) which assists the operator to correctly form a set of indications characterizing the status of a piece of machinery.

The use of the expert-examination computer system permits developing the measures aimed at improving the reliability of technical equipment and prolonging its life with the use of the methods of metallographical analysis of failed parts.

The scientific and methodological base, experience in practical investigations and multifaceted reference information gathered in the system enable it to be used as a learning system for training specialists in the field of diagnosing machine parts and structures.

RUSSIAN TECHNOLOGY INTERCHANGE #2

METHODS FOR DIAGNOSING THE FATIGUE DAMAGE RATE OF MATERIALS OF MACHINE ELEMENTS AND ESTIMATING THEIR FATIGUE LIFE

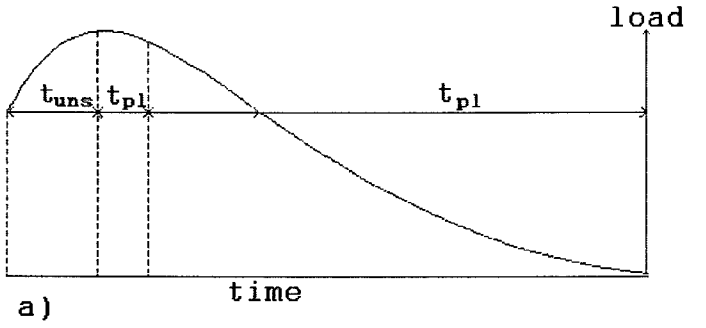
Alexander Koltunov
Doctor, Scientific & Technical Centre "EXPERT"
Russia, 140003, Liubertsy-3, Moscow region, a/b 8

Abstract: A pre-failure process of expiring the resistance of metal materials to elastic, plastic and unstable deformations in machine parts are similar to those in short-duration break tests of material specimens. Since the only similarity criterion for damage accumulation in a material element at different failure types is the ratio of characteristic times, the interconnected processes of expiring the resistance to elastic, plastic and unstable deformations in a material element and associated processes of accumulating damages can be estimated using the values of accumulated material damages in the short-duration break, namely, in the form of ratios of the times for elastic, plastic and unstable deformations to the total time of specimen failure. The methods, developed on the basis of this regularity, for estimating materials' fatigue damage rate in short-duration break tests enable not only quantifying the conditions of fatigue failures of machine parts but also determining their operational fatigue life and thereby obviating the need for time-consuming, laborious and expensive tests of machine members.

Key words: Mechanical tests; resistance to deformation; damage rate; fatigue strength .

When determining mechinical properties of materials of aeronautical engineering parts it was noticed that in separate cases a reduced impact resistance was observed despite the fact that main material tensile-strength mechanical characteristics met the techinical specifications. In analysing tension test diagrams of specimens, substantial distinctions in the time to failure formation were revealed which corresponds to the time interval in the diagram between reaching yield and tensile strengths (all tests were conducted on the same tast machine with the same tape advance rate). Based on theoretical ideas concerning deformation and failure processes, one can conclude that this time interval in tension test diagrams caracterizes the resistance of the specimens' material to plastic deformation in the course of expiration of which crack nucleation and propagation take place under static loading. The minimum time for failure formation was observed in tension test diagrams of the specimens cut from the part with reduced impact toughness and the maximum one was observed in the diagrams of the specimens made of the same parts with the impact toughness corresponding to the requirements of technical specifications (Fig. 1).

Thus it was established that taken as the criteria characterizing the resistance of materials to elastic, plastic and unstable deforamations can be the durations of



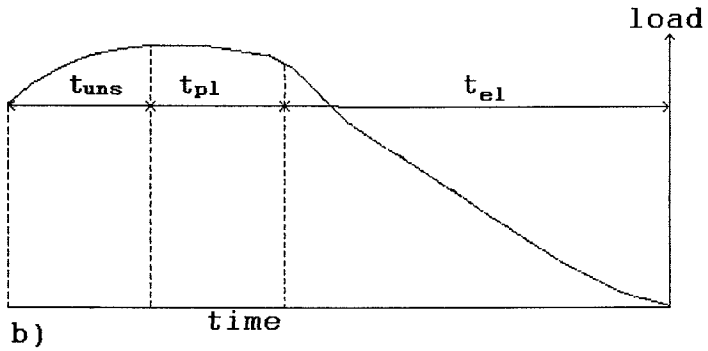


Fig.1 Tension tast diagram of the speciments cut from the material of the hydraulic accumulator cover which have an impact tonghness of no less than 50 j/cm^2 a) and no more than 50 j/cm^2 b).

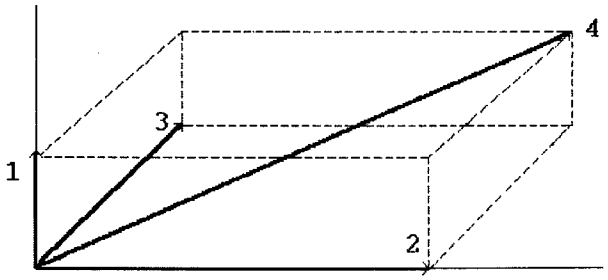


Fig.2 Element illustrating the limit fatique damage rate.

deformations determined from tension tast diagrams. The potentials inherent in any material of the resistance to elastic, plastic and unstable deformations are in sequence expirated in the course of any loading to failure. Hence, it is advisable to estimate them in the form of the ratios of the times for elastic t_{el}, plastic t_{pl}, and unstable t_{uns} deformations to the total time to failure t_f, while the accumulated damages a_d produced throughout the loading spectrum can be determined as a sum of these ratios squared:

$$\frac{t_{el}}{t_f}+\frac{t_{pl}}{t_f}+\frac{t_{uns}}{t_f}=1 \quad ; \tag{1}$$

$$a_d=\sqrt{(\frac{t_{el}}{t_f})^2+(\frac{t_{pl}}{t_f})^2+(\frac{t_{uns}}{t_f})^2}. \tag{2}$$

The regularities revealed became a prerequisite for developing a number of principally novel methods for diagnosing the fatigue damage rate of machine parts' material under loading.

Method for determining the times for the nucleation and propagation of fatigue cracks in machine parts: In identifying the cause of the fatigue failure of a part for which the operating time T expressed in any units is known, the following times can be estimated based on the linear damage integration model represented in the form of (1) and (2): for the phase of the fatigue crack nucleation (T_{nuc}) during which the expiration of the resistance to elastic deformation goes on in the course of its loading:

$$T_{nuc}=\frac{t_{el}}{t_f}\cdot T; \tag{3}$$

for the phase of the crack propagation T_{pr} during which the expiration of the resistance to plastic deformation of the material goes on at an apex of a crack as it rows:

$$T_{pr}=\frac{t_{pl}}{t_f}\cdot T; \tag{4}$$

for the phase of final failure (rupture) T_{fr}:

$$T_{fr}=\frac{t_{uns}}{t_f}\cdot T; \tag{5}$$

The technical essence of the method for estimating the times to nucleation ant time to propagation of a fatigue crack in machine parts is as follows. In the vicinity of the fatigue fracture one cuts the pieces of the material of the part under examination and makes of them the specimens of standard shape and size for short-duration break

tests. Cutting is accomplished so that the specimen's longitudinal axis be normal to the part's fracture surface. The specimens thus made are to be tested for break with a constant deformation rate. During the loading up to a final failure one records time variations in the load in the form of a load-time tension diagram. In addition one measures the times for elastic t_{el}, plastic t_{pl}, and unstable t_{uns} deformations as well as the total time to specimen failure t_f. The time to crack nucleation T_{nuc}, time to fatigue crack propagation T_{pr} and time to final failure T_{fr} are calculated with formulas (3), (4), and (5), respectively. The persent [1] method was practically approved in a series of investigations and was shown to be rather effective as compared to the well-known fractograghy method for estimating the time for fatigue crack propagation.. In particular, this method was employed in identifying the cause of fatigue failure of the wing panel of an aircraft occured after 1927 flights.

In this case the nucleation sites of fatigue cracks were the rivet holes located on the lower panel surface near the wing trailing edge. Crack propagation went on in the cross direction, that is, the failure-producing stress-strain state was formed by longitudinal forces. Because of this the mechanical material properties of the lower wing panel were determined using the short-duration break tests of the specimens cut from the panel in longitudinal direction near the beginning of the panel failure. In the course of testing the specimens with a constant loading rate one recorded on the recorder strip the time changes of the load and measured the times for elastic, plastic and unstable deformations as well as the time to specimen failure. From the results obtained and measurements performed it was found that the material under examination had a tensile strength of σ_B = 600 MPa, a yield strength of σ_t = 550 MPa, a percentage elongation of δ = 7.9%, with the parameters of the resistance to elastic, plastic and unstable deformations being respectively:

$$\frac{t_{el}}{t_f}=0.39; \qquad \frac{t_{pl}}{t_f}=0.58; \qquad \frac{t_{uns}}{t_f}=0.03.$$

From the fact that the expiration of the panel's material resistance to elastic, plastic and unstable deformations had been in progress during 1927 flights it followed that for 751 flights the crack nucleation was taking place

($\frac{t_{el}}{t_f} \cdot T$=0.39 x 1927=751), for 1118 flights the fatigue crack was propagating due to expiring the resistance to plastic deformation, and,finally, for the rest of the flying time (58 flights) the accelerated propagation of the fatigue crack took place due to expiring the resistance to unstable deformation resulting ultimately in the wing panel rupture.

During the investigation the assessments were performed of the times to crack nucleation and crack propagation in the wing panel using the fractography method by taking into account the width of fatigue furrows on the microrelief of the panel

fracture surface.

As a result of measurements and calculations performed using this method it was found that the fatigue crack was propagating over 1111 flights and the time to crack nucleation, defined as the difference between the total flying time and the time to crack propagation, was equal to 817 flights. The result obtained by these two methods under consideration are in good agreement, but the method based on the linear damage integration model yields more plausible and comprehensive data on the time to crack nucleation and time of its accelerated propagation. As for the fractography method, it only allows the time of fatigue crack growth to be determined.

The method for determining the total number of cycles to failure : In operation, machine parts are as a rule subjected to complicated modes of loading where the sequence of amplitude and mean stress values vary randomly. Such a loading is referred to as the random loading.

In random loading the total number of cycles to failure depends on loading character, that is, maximum stresses, fraction of maximum stresses in the total number of cycles, mean stress, frequency of cycles, sequence of high and low stresses and other factors. This generated a need for developing methods for fatigue tests of materials in laboratory conditions with simulating main features of a random loading and for creating computational methods enabling the random-loading fatigue strength to be assessed using the results of the regular-loading fatigue tests.

For the effect of the loading mode on the metal fatigue strength to be taken into account, a variety of hypotheses are used. The most widespread is the linear cumulative damage hypothesis [2] :

The linear hypothesis is not always in good agreement with experimental data, but its simplicity and the absence of additional parameters favoured to its use in practice. Its main shortcomings lie in an insufficient accuracy associated with the impossibility to account for the loading time history and the fact that the stresses less than the fatigue strength are excluded from calculations as well as the laboriousness of accomplishing the test procedure.

Obtaining the totality of the damages accumulated throughout the loading spectrum have made it possible to simplify and improve the accuracy of the determination of the total number of cycles to failure using the results of short-duration break tests of a given material, measurements of times for elastic, plastic and unstable deformations and the total time to specimen failures [3].In so doing, the totality of damages accumulated in the material in loading to failure are calculeted from equation (2).

We take as an example the results of estimating the total numbers of cycles to failure of the material of the aircraft landing wheel hub. When investigating, made from the wheel hub, fabricated of the aluminium alloy AK6, were standard round specimens which were tested for short-duration break and fatigue using the method of combined rotation and bending. The tests were carried out at constant stresses of σ_i = 150 MPa, 200 MPa, and 300 MPa. The number of cycles to failure for the specimens were obtained to be respectively $108 \cdot 10^4$; $15 \cdot 10^4$ and $3 \cdot 10^4$.

The material specimens undergone to short-duration break were tested on the machine ИМ-4Р with a loading rate of 1.7 mm/min. In the course of loading to the point of failure, one recorded time changes of the load and measured the times for elastic, plastic and unstable deformations as well as the total time to specimen failures. As as result of tests and measurements performed the following mechanical properties of the wheel hub's material were obtained (Table I).

σ_B MPa	$\sigma_{0.2}$ MPa	δ_5 %	$\frac{t_{el}}{t_f}$	$\frac{t_{pl}}{t_f}$	$\frac{t_{uns}}{t_f}$	a_d
42.8	32.2	16.1	0.15	0.68	0.17	0.51
43.3	32.1	15.1	0.15	0.71	0.14	0.55
41.9	32.8	14.0	0.17	0.73	0.16	0.57

Table I. Results of testing and measuring mechanical properties of the wheel hub's material (alloy AK6)

The total number of cycles to failure was obtained from the formula:

$$T_s = a_d[\sum_1^n \frac{K_i}{T_i}]^{-1}, \qquad \text{where} \quad K_i = t_i \Big/ \sum_1^n t_i; \quad K_1 = 0.024; \quad K_2 = 0.12; \quad K_3 = 0.85;$$

then $T_s = (1.22...1.37) \cdot 10^6$.

Thus, as a result of the tasts and measurements, it was found that the total number of cycles to failure of the wheel hub's material in random loading was $(1.22...1.37) \cdot 10^6$.

When using the prototype-related method the total number of cycles to failure was $(1.0...1.17) \cdot 10^6$ with the calculated total accumulated damage being 0.44...0.49. The use of the method proposed for determining the total number of cycles to failure simplifies the estimation procedure and increases the accuracy of the results obtained. It can be employed in diagnosing machine parts with the aim of determining their fatigue life and failure causes in operation.

Methods for determining the fatigue strength of metals : The methods presented are based on the following theoretical and experimental results. When a gradually increasing load is applied to a part or specimen, the deformation of the material goes on simultaneously in a set of glide planes and is accompanied by the bending and rotation of the glide planes. The vector of the stresses produced in reciprocal planes add up and reach limit values for the time less than the time to the failure of the part or specimen. In so doing, the process of reaching the limit stresses caused by the bending and rotation of glide surfaces is shortened in time due to

accumulating the damages produced by dislocation and disclination movements and by expiring the resistance to elastic, plastic and unstable deformations.Consider the behavior of the specimen's material in short-duration break. Let us denote by

$\frac{t_{el}}{t_f}, \frac{t_{pl}}{t_f}$ and $\frac{t_{uns}}{t_f}$ respectively the magnitudes of the vectors of accumulated elestic, plastic and unstable deformations (damages) of the broken specimen and the part (1,2,3 on Fig.2). Then the magnitude of the resulting vector was obtained from the formula (2) (4 on Fig.2.)

Depending on loading conditions (Fig. 2) the material element can lose stability as a result of either a critical change in volune (rotation instability) or a critical change in shape (translation instability). The translation instability is realized on reaching a critical dislocation density of the material in plane accumulations.Under the conditions of macroscopic deformations of a specimen or a part, this type of instability manifests itself at the boundary between elastic and plastic material behavior prior to the beginning of plastic yielding in the form of longitudinal translation in the tensile force field. The value of the maximum stress τ_{st} , required for longitudinally translating the material is equal to the scalar product of the vectors of tensile strength and limit elastic damage:

$$\tau_{st} = \frac{t_{el}}{t_f} \cdot \sigma_B \qquad (6)$$

The rotation instability is realized on reaching a critical disclination density. Under the conditions of macroscopic deformation of a specimen or apart, this instability manifests itself prior to the formation of the neck due to the cross translation of the material which by that instant was in the state of longitudinal plastic yielding in the tensile force field. The value of the maximum stress τ_{cros} , required for the cross translation of the material is equal to the scalar product of the vector of maximum longitudinal stress into the unesulting vector of elastic, plastic and unstable deformations a_d:

$$\tau_{cros} = a_d \frac{t_{el}}{t_f} \cdot \sigma_B \qquad (7)$$

When comparing the results of fatigue and short-duration break tests of the material it was found that the value of the maximum stress required for longitudinally translating the material in short-duration break is equal to the tensile and bending fatigue strengths of the material: $\tau_{st} = \sigma_{-1}$, and the value of the stress required for cross translation in break tests is equal to the torsion fatigue strength: $\tau_{cros} = \tau_{-1}$.

The results of tests and measurements are given in Table II.

Specimen's material	$\frac{t_{el}}{t_f}$	$\frac{t_{pl}}{t_f}$	$\frac{t_{uns}}{t_f}$	a_d	Calculated	Experimental
					σ_{-1}/τ_{-1}	σ_{-1}/τ_{-1}
40XH2CMA	0.43	0.33	0.24	0.59	786.9 / 464.3	750-790 / 450-470
30XГCH2A	0.41	0.25	0.34	0.59	676.5 / 399.1	630-690 / 380-410
30XГCA	0.31	0.30	0.39	0.58	384.4 / 222.9	-
12X2H4A	0.35	0.23	0.42	0.59	427.0 / 251.9	-
O3XIIHIOM2T	0.49	0.14	0.37	0.63	739.9 / 466.1	720-785 / 450-480
B-95	0.39	0.58	0.03	0.70	234.0 / 163.8	190-240 / 130-170
AK4	0.38	0.62	0	0.73	163.4 /119.3	120-178 / 80-120
AK6	0.16	0.71	0.13	0.74	68.8 / 50.9	-

Table II. Fatigue strengths of materials calculated from the results of short-duration break and fatigue tests

Thus the above-presented results are indicative of the similarity of the processes of damaging the materials in short-duration break and cyclic loading. The distinctions lie in the fact that in the short-duration break the processes of expiring the resistance to elastic, plastic and unstable deformations are realized in the macrovolumes of the material, whereas in cyclic loading they go on in local volumes of the material.

The technical essence of the methods for determining tensile (bending) and torsion fatigue strengths is as follows[4,5]. One makes the specimens of standard shape and dimensions from the material to be studied in short-duration break tests. Then one carries out tensile tests with a constant deformation rate. In the course of loading to failure of the specimens one records time changes in the load P in the form of a load-time diagram. In so doing, one measures the times for elastic t_{el}, plastic t_{pl} , and unstable t_{uns} deformations as well as the time to failure t_f and the maximum load which the specimen withstanded during the loading. Based on the parameters obtained one calculates the tensile (bending) σ_{-1} and torsion τ_{-1} fatigue strengths:

$$\sigma_{-1} = \frac{t_{el}}{t_f} \cdot \frac{P_{\max}}{F}; \quad (8) \qquad \tau_{-1} = a_d \frac{t_{el}}{t_f} \cdot \frac{P_{\max}}{F}; \quad (9)$$

Where F is the cross-section area of the specimen's working portion; P is the maximum load.

The use of these methods substantially simplifies the procedure of determining material fatigue strengths as compared to conventional methods based on fatigue tests and provides the possibility of estimating these parameters using broken machine parts when identifying the causes of their breakdowns.

Bibliography:

1. Inventors' certificate No. 1696956, 1991, USSR. Method for estimating the fatigue damage rate of material. P.F.Tyugashov, V.A.Belokopytov, A.YE.Koltunov.

2. V.T.Troshchenko, L.A.Sosnovsky. Fatigue strength of metals and alloys. Handbook, Pt.I.- Kiev, Naukova Dumka Publishing House, 1987, pp.190-201.

3. Inventors' certificate No. 1744583, 1992, USSR. Method for diagnosing the fatigue failure of a machine part. P.F.Tyugashov, V.A.Belokopytov, A.Ye.Koltunov.

4. Inventors' certificate No.1733964, 1992, USSR. Method for determining the fatigue strength of materials. P.F.Tyugashov, V.A.Belokopytov, A.Ye.Koltunov.

5. Inventors' certificate No. 1665278, 1991, USSR. Method for determining the fatigue strength of material. P.F.Tyugashov, V.A.Belokopytov, A.Ye.Koltunov.

HAZARDOUS WEAR PREVENTION SYSTEM FOR FRICTION UNITS OF GAS TURBINE ENGINES

V.A. Stepanov*, I.F.Touloupov**

V&RS, Company Ltd, 13, Marshala Golovanova Str., ap.409, 109144 Moscow, Russia

*General Director (Head of Dept., CIAM), **Leading Engineer

Abstract: The hazardous wear prevention system is comprised the automated continuous monitoring system of gas turbine lubricated friction units, the operation of which is based on the change in the number of wear particles, as well as the everage wear particle size in oil. The system includes the cyclonic separator, self-cleaning sensors from ferromagnetic particles, the control unit and the measured data accumulation unit.

The system's sensor is installed in the case of cyclonic separator of the wear particles that is incoporated in oil scavenge line.

The sensor's operation is based on the well-known principle of the ferromagnetic wear particles accumulation between the two totally isolated from each other magnetic (plus and minus) poles. They are at the same time used for ensuring the electrical contacts. In case sufficient electroconducting ferromagnetic particles are accumulated the contacts are shortened.

A completely new sensor's design, new algorythmic equations and the measured data processing allow to icrease decisively the sensor's sensibility and to apply it for measuring the ferromagnetic wear particles velocity multiplied by the average size under the real time condition.

The paper presents the first tests results of the hazardous wear prevention system as a part of PS-90GP gas turbine engine.

Keywords: Monitoring; particles; pole; sensor; separator; system; wear.

The field experience has demonstrated that the automated continuous quantative monitoring of debris (ACQMOD) as well as of the wear particles sizes washed out from the surface of the friction units is one of the most promising and perspective modern diagnostic method used for condition monitoring of the friction units in gas turbine engines (GTE).

For quite a long period of time the method of monitoring wear particles in GTE oil was being developed based on the improvement of devices and methods applying electrical ferromagnetic chips detectors (ECD). But however less interest has been revealed for the development of the systems based on ECD and to the instruments applying ECD of the latest generation (demagnetizing chip detector (DCD) which were never used for practical purposes after the invention of firm Tedeco. It has developed a new quantative monitoring method of particles (QDM) applying the magnet inductive sensor of the accumulative character.

But however DCD has certain advantages and a new approach to the design and the method of applying it can allow the creation of effective AQMOD on their basis.

The current paper presents some properties of DCD sensors, peculiarities of it's design and the preliminary test results of AQMOD with contact sensor magnet controled, which has been specially designed for PS-90 GP engine.

PREREQUISITES FOR DEVELOPING THE SYSTEM

The first AQMOD with the applied DCD was designed in 1978 [1]. Two iron electromagnetic pole pieces with the gap between them served as the sensitive elements of the sensor's system. The gap had the shape of the slot and

was δ wide. The magnetic poles pieces were isolated from each other and performed the role of sensitive electrical contacts. Following the appropriate command of an auxilary electronic unit the magnetic poles pieces could be magnetized or demagnetized.

In case the magnetic poles pieces were in the magnetized status ferromagnetic particles being moved by the oil flow could be captured and arranged in the chain between magnetic poles by the magnetic field. As long as the length of the line reached the size of the gap δ the poles-contacts of the sensitive elements were shortened. In case the magnetic poles pieces were demagnetized the particles were washed out and carried away by the oil flow.

The demagnetisation command was sent after equal periods of time t_o from the timer and was not related anyhow with the detector of shortening the contacts of the sensitive elements of the sensor. The magnitization of the magnetic poles pieces was performed automatically following the demagnetization after a short period of time which was necessary for removing the accumulated particles. In case time t_δ (needed for the formation of the shortening chain) was higher than the set time t_o the shortening of the contacts did not occur during the sensor's operation, as the growing chain colapsed long before they could reached the length δ. While the particles in oil grew in number time t_δ decreased. When t_δ becomes less t_o the sensor activated sending the signal about shortening to the memory of the auxilary electronic unit.

In case the time between two sensor's activations was less than t_o, which is an evidence of the non-random character of the activations, the auxilary electronic unit was sending signals of alarm irrespective of the further behaviour of the sensor.
The device could be adjusted for any set level of wear particles quantity in oil, by just adjusting the value t_o.

The described device is of a threshold character and does not allow to monitor wear processes trend, which deprives the user of the possibility to estimate adequatly the importance level of the sent warning signal and to prognose the residual life time of the monitored unit. Most probably that was exactly the disadvantage which did not allow it to compete with QDM [2] which was designed two years later and was well recognized.

But however DCD can be applied with a success in AQMOD systems if the control of the sensor's unit will be changed and effective algorythms for processing the measured data will be used.

Let us consider the physical phenomena to which DCD responses.
The magnetic field that is located across the sensitive sensor's gap, determines the location of the particles in the area: there is an attempt of the particles to arrange themselves along the magnetic force lines forming chains of particles making them as long as possible by being connected along their shortest sides. Assuming this is true the number of particles in the chain is big enough irrespective of the spectrum distribution mode of the biggest particles sizes the everage velocity of shortening the chain will be equal:

$$v = k \cdot v(t) \cdot l(t)$$

where $v(t)$ and $l(t)$ - the everage frequency of the particle's arrival and the everage length of the particle, approaching the area of the sensor's location from the monitorred unit under real time t;
k - the coefficient of particles flow fullness. It is equal to the ratio between the number of particles which formed the shortened chain and the total number of particles coming to the area of the sensor's location within the period of time when the chain was formed.

As the transformation of the normal wear mode into the accelerated wear mode by all means is accompanied by the increase in $v(t)$ or $l(t)$; and it is a more frequent case

when the growth of both values is observed, the increase of V is a good criterion for revealing the intensive wear of the monitorred unit.

Dividing value V by the width of the operated gap δ one arrives at:

$$V/\delta = 1/t_{\delta}$$

based on the above equation it becomes clear that if the commanded by the timer cleaning of the sensor is omitted and the cleaning of the sensor is done only following the sensor's command, then in case of the stationary character of the monitorred process the sensor's activation will be done with the constant average frequency $f=1/t_{\delta}$, which is proportional to the established level of the wear intensiveness. Due to this it may be stated that the system with such type of control of the sensor and the measurement of the average frequency of it's activation is capable of monitoring linier slowly changing trends that can be presented as $v(t)\cdot l(t)$. The system AQMOD can be created using the principle similar to the above, but with an exception that sliding non-linier algorythms should be applied for processing current values of $1/t_{\delta}$. This will allow to reveal the non-linier relatively quickly changing trends for the conditions under which the oil is contaminated by foreign lengthy particles which cause the false activation of the sensor.

DCD sensor has the integrating character which explains it's certain properties and capabilities. The proper choice of the gap's working width will provide the flexibility in adjusting it for the wear processes that are typical for different engine types. For example, in case the oil is finely filtered DCD could be adjusted for operating aimed at small size particles within the sizes range that can not be detected by QDM systems. But however in comparison with the latter, DCD system will ignore big-size single particles which can be found in oil as a result of a sudden hazardous failure in the unit being monitorred.

The possitive feature of DCD is explained by it's capability to be operated within the wide range of dynamic flow of ferromagnetic particles. Due to it's capability of self cleaning it will operate normally under the condition where the sensor of QDM system will be overaccumulated by particles too quickly. At the same time if the tendency to the critical trend growth is observed DCD can be deliberately switched off (under the status of magnetisation) for accumulating wear particles for the laboratory analysis. And for all that the particles accumulated by the sensor will be meaningful in the contrast with the particles accumulated under the same conditions by QDM sensors.

The disadvantage of DCD in comparison with the QDM system sensors is explained by it's lower sensitivity due to the smaller coefficient of particles flow fullness.

THE SYSTEM'S DESIGN

The hazardous wear prevention system for the friction units (HWPS) is designed for the operation as a part of the gas pumping station drive on the basis of PS-90 engine. The system's sensor is installed in the case of the cyclonic separator of the wear particles that is incorporated in oil scavenge line. The separator's design allow the spiralled oil wash out the surface of the sensitive element providing the easy removal of particles used by the sensor.

The magnet field in the gap area of the sensor is created by the constant magnet which has relatively a low coercive force which allows to remagnitize it applying a small-size coil energised by the pulsating current. As the sensor most of it's time is being operated in the mode of wear particles accumulation the appliacation of the constant magnet instead of the electrical magnet decisively reduces the average consumed energy of the system. Within the period of demagnitisation-magnetisation which is 2,5 sec HWPS consumes about 3 wt.

The sensor's signals reach the registrator where they are being constantly registered and along with it they come to the unit of the measurement data processing which specifies the trend of the intensiviness criterion of the friction units wear and provides monitoring of it's changes. The trend is specified by the consequent processing of the reversed intervals values between the sensor's signals using the method of exponential smoothing sliding medians. In case of the constant tendency towards the trend growth the unit sends the alarm signal. The operation of HWPS is not interrupted during the entire period when the continous signal is being sent.

THE RESULTS OF THE PRELIMINARY SYSTEM'S TESTING

The operation of HWPS was tested in the laboratory of the drive's designer, the testing rig for GTE oil system sensors and fittings was used. The separator with the sensor was installed in the oil scavenge line.

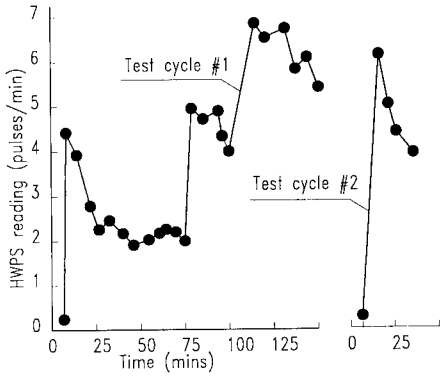
The main objective of the experiment was to estimate HWPS relative sensitiveness to small-size (less than 0,2 mcg) and medium-size (up to 2,5 mcg) ferromagnetic particles. Small size pieces of the steel chips being 20...50 mcm in size and about 10 mcm wide were used to simulate small size particles. Medium-size particles were represented by steel chips pieces being 100...150 mcm in size and about 15 mcm wide.

The experiment was conducted in two cycles. During the first cycle HWPS was operated applying small-size chips, during the second period - medium-size chips. There was no change of oil when transiting to the second cycle of the experiment. The oil was cleaned each time before starting the next cycle applying the removable filter and the removable magnetic catching device. The oil volume in the oil system was equal to 25 l, the oil consumption - 28 l/min. The pressure drop on the separator was equal to 0,3 kg/cm^2.

Chips were injected inside the oil tank in portions, 300 mg each.

Three portions of small-size chips were injected during the 1 cycle of the tests: the first portion on the 10-th minute of the cycle, the second one - 75-th minute, the third one - 100-th minute. On the 10-th minute of the second cycle the medium-size chips portion was enjected.

The picture presents the tests results.



Points used for drawing the curves corresponded to the average in time number of sensor's signals that were registerred on the consequent 5-minutes intervals on the register tape.

HWPS was monitoring and registering both the initial background of the residual ferromagnetic debris of the circulating oil as well as the resolution processes of the concetrated chips gradients injected in the oil tank.

So HWPS can be operated with a far more finer wear particles (0,2 mcg; 50 mcm) being compared with QDM wich can register particles the mass of which is more than 5 mcg and that are more than 250 mcm in size [3]. The results of the experiment do not allow to estimate the HWPS sensitivity threshold as the used chips portions were too big. The necessary data will be obtained in the process of the further study and the system development for the engine.

SUMMARY

The combination of self demagnetising contact sensor with the cyclonic separator allowed to develop a new system for monitoring wear particles in GTE oil. This system is less sensitive than QDM , but however it is good enough

for a more wide dynamic range and can be applied for the wear particles spectrum which is expanded finer sizes oriented.

The results of the preliminary tests give hope that after the system will be developed on the engine it will respond to any abnormal wear of the monitorred friction units which are characterised by the accelerated formation of ferromagnetic wear particles, being 20 - 50 mcm in size.

REFERENCES

1. F.M.Newman Automatic Self-Cleaning Ferromagnetic Metal Detector. U.S. Pat. № 4, 100, 491, Jul.11,1978

2. J.H.Magee, T.E.Tauber Lubricating Oil Debris Monitoring System U.S. Pat № 4, 219, 805, Aug. 26,1980

3. F.DiPassquale Field Experience with Quantative Debris Monitoring SAE paper № 871736,1987

EQUIPMENT FOR CARRYING OUT TECHNICAL MONITORING Of GAS PUMPING UNITS BY LUBRICATION OIL ANALYSIS

B. R. Matveyevsky, V. A. Petrosyanz
I.M. Gubkin Gang, Moscow, RUSSIA

Abstract: A portable on-line monitor is described which measures iron, copper, tin and aluminium. The device is recommended for use in combustion engines, compressors, turbines, gear boxes, oil and gas pumping systems. The device consists of a set sensors installed into the lubrication system. The sensors measure quantities of wear debris from 0,01 to 200 mg with an estimated error of 2%. As case study of the application of these sensors to a compressor pumping stating is quoted.

Key Words: Electron Devices; Friction Units; Line Debris Monitor; Lubrication System; Sensors.

Among the most important problems high-lighted by the development of the gas industry is that of ensuring that gas pumping units (GPUs) work reliably. This problem is being tackled from many angles. One approach is to improve monitoring methods and equipment.

The introduction of equipment capable of making a swift and precise evaluation of the condition of working GPUs enables correct decisions to be taken on the replacement or repair of any defective component parts, or of the unit as a whole, and also allows one to establish the optimum frequency and extent of preventative maintenance.

The on-line monitoring method of analyzing abrasion products in GPU lubrication systems while the unit is operating is an effective method for evaluating, controlling and predicting the condition of GPUs.

The most widely used method is spectral anaylsis of oil samples, which is characterised by rapidity, precision and sensitivity [1, 2]. However, taking periodic oil samples from GPU lubrication systems, sending them to diagnostic centres for analysis, and receiving back findings on the condition ofthe unit involves considerable expenditure of time. The most promising method of getting quick information on the condition of the GPU at the combustion chamber while it is actually working is by using primary monitoring equipment. The apparatus recommended in studies 3 and 4 for evaluating the condition of GPUs is the FS (FS-112; FS-

151) analyzer (meter) for carrying out dispersion analysis of oil samples from GPU lubrication systems.

Essentially this method involves periodic analysis of the dispersion structure of oil from GPUs directly at the combustion chamber and subsequent comparison of the readings with the maximum permitted values for particle concentration with a view to evaluating the condition of the unit. If the oil samples exceed the maximum permitted values this indicates that the GPU's friction assemblies are not functioning correctly.

As an example, figs 1 and 2 show the dependence of particle concentration of oil samples under dispersion analysis and the indicator elements under spectral analysis on the running time of centrifugal (STD-4000) and piston (10 GKNA) GPUs. Comparing these findings one can note the correlation between the dispersion and spectral analyses which confirmed the presence of three characteristic periods in the units' operation: the running in period, the period where deterioration sets in, and the period of intensive deterioration. The method developed was introduced into the Uzbektransgaz combustion chambers with a saving of R.129,000.

Recently, with combustion chambers going over to systems of non-duty maintenance, studies aimed at developing new methods of tribological monitoring of GPUs using built-in control equipment have acquired special importance. At the industry's laboratory attached to the I. M. Gubkin GANG Centre for Joint Analytical Research one such method has been developed which consists in the sedimentation of abrasion products (APs) on a special apparatus (sensor) where their mass is recorded by an electronic device. [5] The abrasion sensors are mounted in the oil feed line, in the crankcase or directly under the unit's friction assembly, and are connected to the recording device by a screened electric cable. This allows the quantity of both ferro-magnetic abrasion particles and of non-ferrous metal particles present in the sensors' working space to be established. With the help of such built-in control equipment it is possible to differentiate and determine the deterioration rate of the crank and connecting rod assembly, ie each bearing unit (babbit bush) individually, the reduction gear (cogged gearing), piston cylinder group etc.

The I. M. Gubkin GANG and the Gazpriboravtomatika Design Office have developed two design variants of built-in equipment for the tribological monitoring of GPUs: the first type is a small portable device with independent power supply and a set of abrasion sensors for work in the field; the second type is a fixed computerized system, mounted on the combustion chamber control panel, working round the clock, equipped with emergency alarms and carrying out constant monitoring of the sensors. (fig. 3)

The method developed was approved for use on Mostransgaz's Kasimov combustion chambers for Solar type GPU and for piston compressors (PC) type DR-12. Bearing in mind that the Solar plant contain 10 units, a special ten-channel computerized system designed for working with ten abrasion sensors was developed and manufactured.

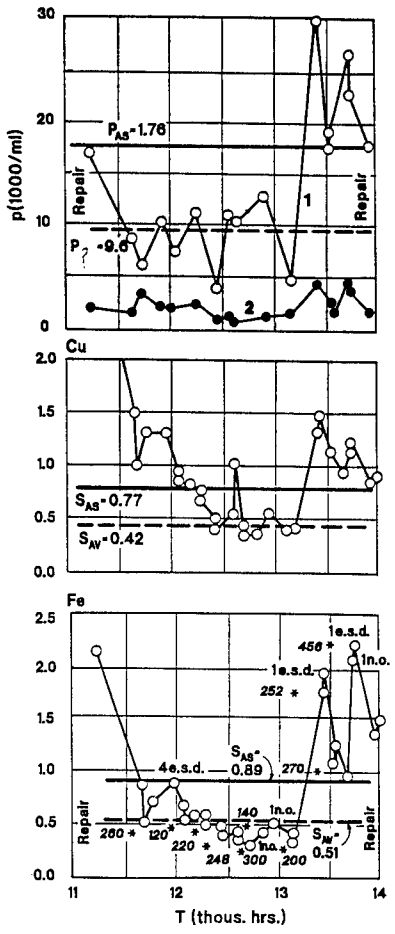


Fig. 1. Dependence of Particle Concentration p and Indicator Elements S in the Lubrication System of Uzbektransgaz's EGPU CTD-4000 No. 2 Combustion Chamber - 4a on Running Time T:

1 and *2* = particle size 5-10 and 10-25mkmk respectively;
*260 = oil added to the GPU, litres;
1 e.s.d. = one electricity shut down; AB]
1 n.o. = one normal stop.

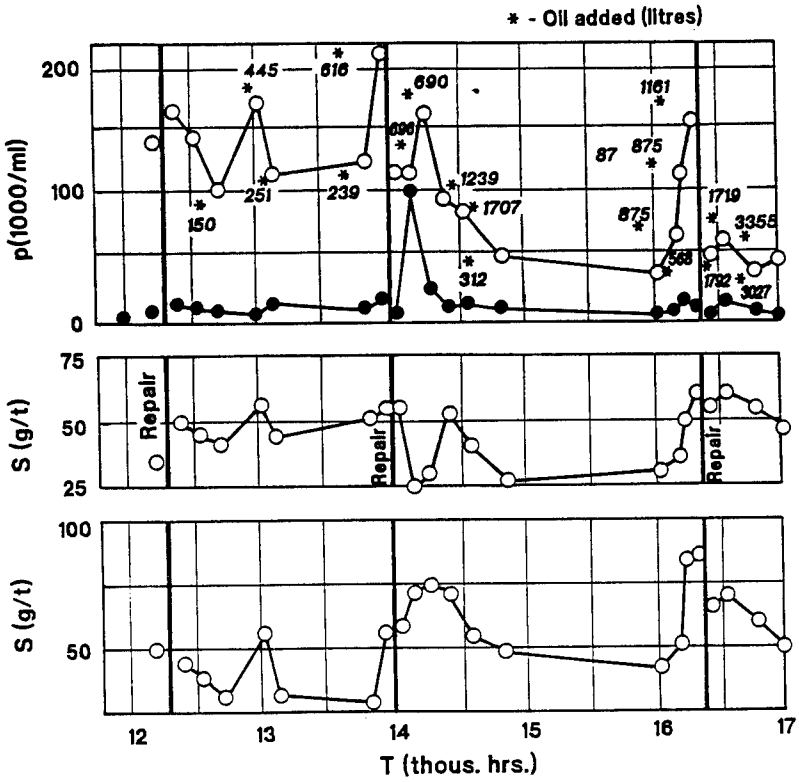


Fig. 2. Dependence of Particle Concentration p and Indicator Elements S in the Lubrication System of Uzbektransgaz's GML 10 GKNA Combustion Chamber-5 on Running Time T.

The sensors were mounted on a common discharge manifold in front of the oil tank on each Solar GPU and on the crankcase discharge on each DR-12 PC (the plant has five units). The computerized recording assemblies were mounted in the control room of each plant. At this particular stage the task was to evaluate the efficiency of the developed and manufactured computerized systems and to compare their findings with the actual condition of each plant. As an example, fig. 3 shows some practical results for DR-12 PC. As can be seen from the graph the friction assembly abrasion varies for each PC. The highest deterioration rate is that for MOGs nos 1 and 3. This example clearly shows the dynamics of defect development. The findings obtained showed that the method developed and the built-in control equipment were efficient and that there was a good correlation with the actual condition of the MOG, which means that they can be recommended for future introduction on other types of GPU.

Literature

1. Stepanenko, V. P., Practical Monitoring of Gas-turbine Aviation Engines, (Transport, Moscow, 1985), p. 102.

2. Petrosyants, A. A., Matveyevskii, B. R., Sorokin, I. A. et al., 'Development of a Method for Checking the Condition of Gas Pumping Units by Lubrication Oil Analysis', Vestnik mashinostroyeniya [Journal of Mechanical Engineering], 1982, no. 6, pp. 41-2.

3. Matveyevskii, B. R., Petrosyants, V. A., 'A Method for General Evaluation of the Condition of Centrifugal Gas Pumping Units by Oil Analysis', Vestnik mashinostroyeniya, 1985, no. 7, pp. 31-5.

4. A Method for Determining the Abrasion of Friction Pairs Working in the Presence of Liquid Lubrication Material, A. S. No. 1408296 USSR.

5. Denisov, V. G., Matveyevskii, B. R., 'Checking the Condition of Internal Combustion Engines Units while they are Running', Vestnik mashinostroyeniya, 1989, No. 9, pp. 20-1.

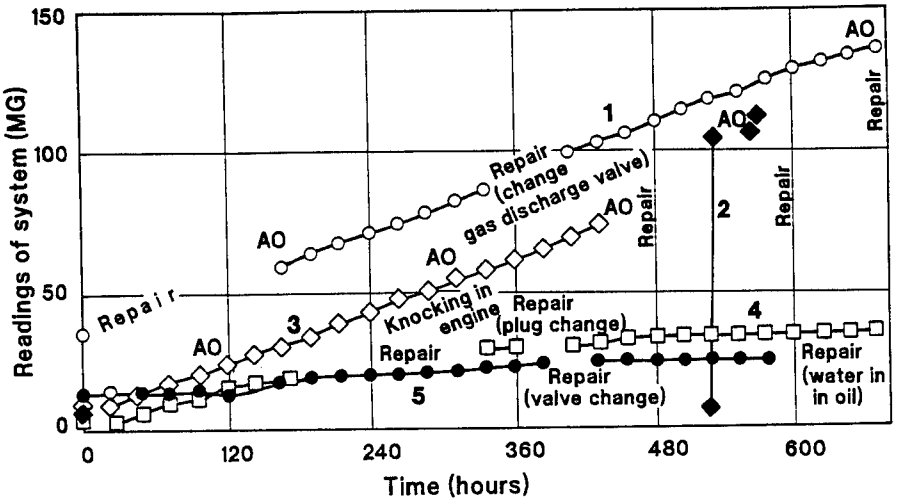


Fig. 3. Dependence of Computerized System Readings on Working Time of Mostransgaz's Kasimov PC Combustion Chamber DR-12.

1, 2, 3, 4, 5 = GMKs Nos 1, 2, 3, 4, 5 respectively.[4]
A0 - damage stop

BASIC TRENDS OF FRICTIONAL INTERACTION AND TRIBODIAGNOSTICS

Bronovets M.A.
Interdisciplinary Scientific Tribology Council
101, Prospect Vernadskogo, Moscow, 117526

Zaritski S.P.
Technical and Engineering Center "ORGTECHDIAGNOSTIKA"
13/1, Karamzina str., Moscow, 117463

Lopatin A.S.
JSC "ROS"
65, Leninskii prospect, Moscow, 117917

Abstract: This paper contain development the theory and practice of the tribology and tribo-diagnostics. It is includes the main directions in tribology, magnetofluid tribosystems, classification of the technical objects, tribomonitoring systems, thin layer activation, restoring of tribosystems without its disassembling, tribosystems with wear autocompensation, controlling tribochemical regime.

Key Words: Condition monitoring; corrosion; friction; lubrication; theory; tribo-diagnostic; wear.

INTRODUCTION: The most widespread cause of machinery operational performance deterioration is wear combined in some case with corrosion. It is wear that limits useful service life of any machine. If the wear of critical component parts exceeds some reasonable value, further operation of machine becomes non-effective, dangerous, and ecologically intolerable.

The main trends in **tribological developments** are given in table 1 [1, 2].

It is now believed that proper attention to tribology, especially in education, research and application, could lead to economic savings of between 1.3% and 1.6% of GNP [3]. Prof.P.Jost, President of the International Tribology Council(London), as the result of the developments in the world of the past 20 years, the subject of Tribology were to be divided into its main constituents, i.e. in its input areas, the division could be as follows (table II) [3]:

Condition Monitoring and Tribo-diagnostic could lead to economic savings near 0.1% of the GNP, it is, for example, for USA make 7 billion dollars.

Table I

The main trends in tribological developments

Theory of friction, wear and lubrication				
Theory of hydrodynamic and elastohydrodynamic lubrication	Physico-chemical processes in friction and wear	Contact mechanics	Theory of dry friction	Theory of wear
Study of friction, wear and lubrication by experimental methods				
Applied directions				
Wear-resistant materials	Technology of surface friction modification and reinforcement			Efficient lubricants and additives
Calculation, design and manufacture of friction units, machines and instrument with optimal properties				
Rolling and sliding bearings	Brakes and clutches	Sealing devices	Lubrication systems	Machines, equipment and instruments
Tribo-diagnostics, condition monitoring				
Friction units	Machines	Instruments	Other technical	arrangement

Table II

Main constituents in Tribology

a) Materials Science & Technology	40%
b) Mechanical Systems	30%
c) Lubrication and Lubricants	20%
d) Condition Monitoring, Tribo-diagnostics, Instrumentation, Tribo-information (Data Banks & Others)	10%

Classification of the technical objects: The technical objects may be divided in two groups.

The first one contains the objects which wear condition may be judged by variations of their operating performance characteristics, e.g. sensitive decreasing of speed of icebreaker cutting a passage through floating ice due to corrosion wear of her hull, drop of compression and variation of exhaust gas composition in internal combustion engines due to wear of cylinder-piston group, increasing of a gap in transmissions and gearboxes, significant rise of leakage in mechanical seals. Wear condition of engine may be judged by its vibrational characteristics, amount and composition of wear particles contained in lubricating oil, and a number of other means. An estimation of parameters which characterize the operational performance of an object is based on a measurement of an integral characteristic of object subsystems state which is related with joints wear.

The second group involves the objects which may be in operation up to the time instant when those operational capability loss occurs drastically as in the cases of magnetofluid seal hermetization loss or pipeline breakage. State variations of such objects, i.e. change of magnetic fluid composition and its magnetization or loss of material from inner surface of pipe resulting in wall thinning, does not affect practically on the operational performance. The development of methods for diagnostics and estimation of residual service life of such systems is very important problem. On the board of orbital space station "Mir" several dozens of magnetofluid seals are installed. Breakage any one of them sealing mechanisms transferring motion from inner compartment to the outer space causes loss of station hermetization. Breakage of gas or oil pipeline results to product delivery interruption and environment pollution with disastrous ecological consequences.

Magnetic fluids systems: Magnetic fluids represent a qualitatively new class of synthetic material which possess unique properties. The application magnetic fluids for tribotechnical purposes, such as seals, bearings, machine friction units with magnetic lubricants, was the result of the development of technology in space research [4]. Magnetic fluids are stable colloids of ultradisperse magnets in a neutral carrier fluid. The stability of such colloids is provided by using special high molecular weight substances with active surfaces. Water, hydrocarbons, oils, silicon fluids, esters and even liquid metal have been used as carrier fluids [5]. Such materials, retaining their fluid properties in the magnetic field, are magnetic controlled. A magnetic fluid, disposed on a flat glass plate and kept at a distance from a magnetic field, is illustrated in Fig. 1. In technical arrangements, the magnetic fluid is drawn into the clearance in the maximal magnetic induction area under the action of the magnetic field. This builds up an elevated pressure which provides hermetical sealing or lubrication. Of the magnetofluid tribotechnical systems, the most widely used application was seals (Fig. 2, Fig. 3). Such arrangements have been in use

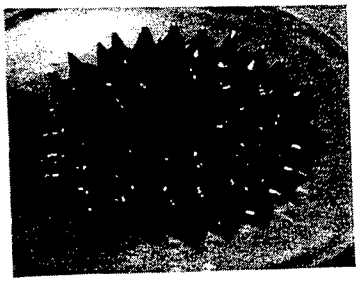
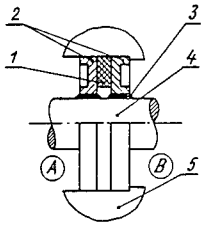


Fig 1. Magnetic fluid on a glass plate in a magnetic field.



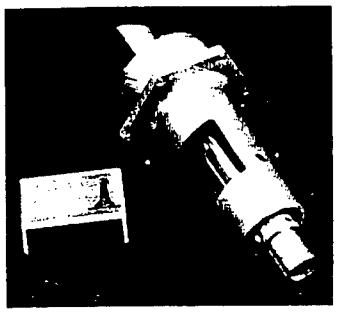


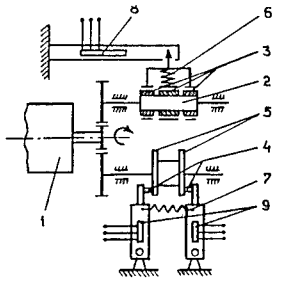
Fig. 2. Magnetofluid seal: 1, permanent magnet; 2, polar stands; 3, fluid; 4, shaft of magnetic material; 5, housing.

Fig. 3. Arrangement for the transmission of motion from the inside of a spacecraft to an outboard fixture: reverse rotatory, 0.16 Hz; rotation angle, ±30°; temperature interval, ±50°.

for a long time on some Russian spaceships. They are capable of maintaining their service properties in outer space for a period of more than 5 years. When the magnetic fluid used in sliding bearings, it ensures a spatial stabilization with a low rotation resistance. When it used in dampers, it give big possibilities to monitoring of the damper characteristic and to construct the system with back bond, as, for example, in the holder solar panel of the space station or suspension bracket of the perspective cars.

We would like to draw your attention to an unusual application of tribological development. At the present time, magnetic fluids are used in medicinal drugs and used for treating fistula and ulcers. The fluid is administered by any means to the patient at the required site, and remains therein with the help of a magnet fixed on the part of the body being treated.

Tribomonitoring systems: One of the important part of the tribo-diagnostics or tribo-monitoring systems there are measuring systems and systems for transfer the information. On the Fig.4 and Fig.5 are presented friction simulator which proved to operate successfully in



1 - drive
2,3 - shaft-bush
4,5 - pin-disk
6,7 - spring
8,9 - tensobeams

Fig. 4. Friction simulator: scheme.

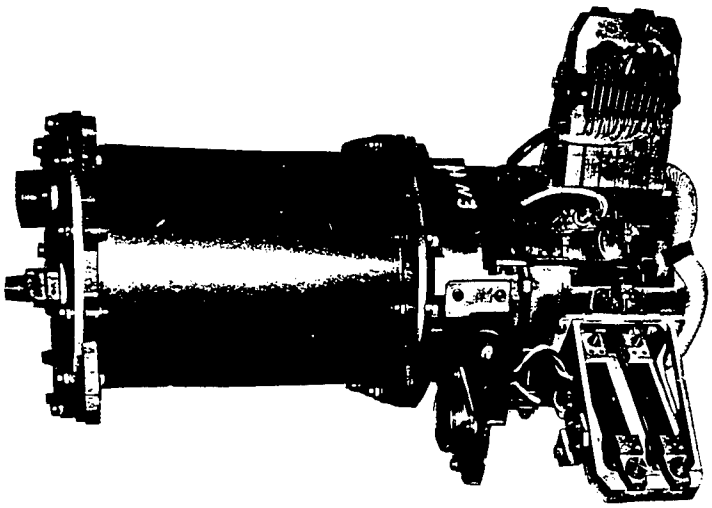


Fig. 5. Friction simulator: external view.

open space around moon. It was build on the Lavochkin`s plant. The friction simulator allows one to test materials and coatings for friction and wear simultaneously in nine friction pairs: three pairs by the "shaft-bush" scheme, six pairs by pin-on-disk rotary motion. The simulator is supplied with tensometric apparatus sending an output signal to a telemetry system and recording device. The same device may be used to measure friction in strong radiation, in aggressive media and another cases.

Thin layer activation: There are several methods allowing to measure local wear of machine components. Between of them the most remarkable one is the surface activation method (thin layer activation). This method, as a version of tracer technique, is very useful in a non-destructive wear and corrosion control. It is based on the irradiation of selected areas on rubbing surfaces by high-energy particle beam in order to create self-activity of areas mentioned. These prepared radioactive labels are used for wear measurements of irradiated components after their installation into the machine. Activity induced does not exceed 10 microcurie and this method may be used without any special protection. For example, it enables to measure cylinder insert wear of marine diesel without disassembling and provides continuous wear condition control during ship voyage using measurement equipment installed out of the engine. Numerous works with combustion engines carried out in different countries are well known and do not need additional illustrations. There were studied the choice of design and material, running-in and operation life, wear profile of the piston ring, influence lubricating oils, fuels and much other. The works on thin layer activation application for the wear control of cutting tool and different type of the

bearing are often described as well. There was also a large variety of technical objects studied - from precise ball bearings, recorder heads and artificial mitral valves to rotary excavators, mining scroll centrifuge and in many another cases.

The technique allows to solve a great variety of problems: a) on-line state control of important parts and units of unique and dangerous machines in operation; b) operative examination of different technical ideas concerning new technics; c) evaluation of operation life, i.e. mean-cycles between failure time in order to replace or to repair the equipment; d) the choice and optimization of operating conditions; e) qualify materials and media closely involved in the operating process, e.g. lubricants, fuel, gas mixture etc.

The main material of modern human civilization is yet iron and its alloys. Its charged particles irradiation generates the radionuclides of cobalt, being perfect indicators with a very convenient half-lives and gamma-ray energies. There is state standard on the application thin layer activation technique [6].

It seems now that one of the promising the surface activation method application is the wear and corrosion control of gas-oil pipe-lines and the equipment of oil-refining and chemical plants. The next schematic diagram shows a picture of a screw with the radioactive insert and the whole control system (Fig.6). Such screws are installed in the pipe-line walls and serve as wear and corrosion indicators. The results received are in good agreement with the gravimetric data of standards. This system are developments in the Institute of Physics and Power Engineering in Russia [7].

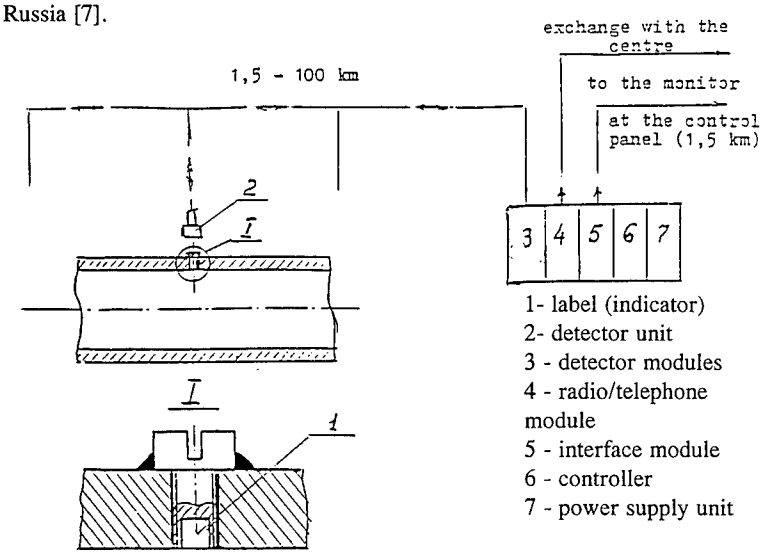


Fig. 6. Schematic diagram of pipe-line wear and corrosion control.

Two difficulties were met in this problem - a long control time (10 years and more) and large depth loss, reaching in some cases 3-5 mm. Their solution was found in the use of classical radiotracer technique. The corrosion indicators in the form of thin flat insert are manufactured from the same steel as objects under control with addition of cobalt (1-2%). Their activation by thermal neutrons in nuclear reactor gives 60 Co having very long half-life (5,24 years) and intensive high-energy gamma-radiation, convenient for remote monitoring through thick absorber. Error of the wear measuring is not more 5-15%. The laboratory studies of this material physical and chemical properties did not indicate any remarkable departures from those of initial metal.

The monitoring system for different technical objects is worked out and includes the following: *a) wear and corrosion indicators* placed inside without distortion of the flow and without any influence on physical and chemical properties of the surface under control; *b) detection units* mounted at the outward side of the pipe or facility; *c) system of automatic data collection and processing* which can also receive the information from a number of other monitors (temperature, pressure, flow rate and another) placed at the equipment.

So in many cases diagnostics allows to estimate machine state without any influence on rate of its operational capability loss.

RESTORING OF TRIBOSYSTEMS WITHOUT ITS DISASSEMBLING: There are more complex systems which diagnostics is related with tribosystems operation control, for example, by changing the composition of fuel supplied to engine or by influencing on lubricant applied providing adaptation of tribosystem to varying operational conditions. Restoring of worn cylinder-piston group of engine is possible without its disassembling by introduction of appropriate chemical species into combustion chamber, motor oil or engine fuel. Such means of nondestructive engine restoring are well approve now and widely used in practice. In table III are presented the results of the restoring old combustion engines with help lubrication composition PIK-02 [8].

Table III

Restoring of car engines without its disassembling

Type car	Run car before restoring, km	Compression on cylinders, kg/cm^2							
		before restoring				after restoring			
		1	2	3	4	1	2	3	4
Volga GAZ-3102	305074	7.0	7.0	7.5	6.0	10.0	10.0	10.0	9.0
Volga GAZ-2412	148154	9.5	9.0	8.0	9.5	10.5	10.5	9.5	10.0
Volga GAZ-2410	166688	9.5	9.0	9.0	9.0	10.5	10.5	10.5	10.5

Tribosystems with wear autocompensation: It is possible to create tribosystems with wear autocompensation and restoring of degraded lubricant properties. This possibility is based on deep investigations and discovery of means to control these processes. As an example of such approach one can regard the use of polymerforming additives or additives producing selective transfer phenomenon. These additives form self-restoring antifriction films on rubbing surfaces. Due to friction reduction these films enhance significantly the service life of tribological joints.

The selective transfer phenomenon is accomplished by using chemical compounds or dispersion of very small copper (or another metals) particles as additives in oils or greases, as well as if one of the elements of the rubbing pairs is made of a copper-containing alloy, for example steel-bronze pairs, while the lubricant medium contains corresponding surface-active substances [9]. In the case of friction in polyatomic alcohol (glycerol and xylitane) the diffusion process is accompanied by surface layer depletion of doping elements and the formation of a rich of the vacancies plastics copper film located on a support. Owing to the high degree of smoothing of the friction surface, substantial growth of the real contact area of the interaction bodies takes place, as a result the pressure peaks decrease at the contact surfaces and reduce the conditions of the formation of microcrack centers and reduce following formation of the wear particles.

Test: The experiments were carried out in condition of sliding by testing machine SMT-1, roller specimens with diameter 35 mm made from steel, bronze pads, synthetic oil "Shell Tivella" and the additive SURM-3 manufactured in Russia. Nominal pressure - 30Mpa; nominal area of contact - 100 mm^2; friction distance - 3000 m. Such parameters as friction moment, wear, bulk oil temperature and electric conductivity were continuously measured during the test. The experimental data were acquired through data acquisition card PCL 812 by the computer IBM PC, what allowed torealize measurement procedure and treatment of experimental results outomatically.

The results of testing: The dependence of friction coefficient on sliding velocity is showing Fig.7. Friction coefficient measured for oil "Shell Tivella" is between 0.018 and 0.021, what confirm high lubrication ability of the oil and allows to consider it as one of the best oils.

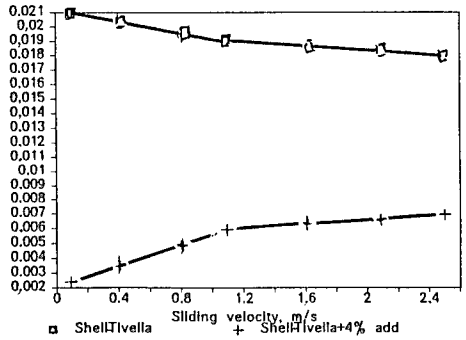


Fig. 7. Friction coefficient vs sliding velocity (m/s).

Application of the additive SURM-3 results in friction coefficient diminishing up to 10 times for the sliding velocity 0.1m/s and up to 2.5 times for sliding velocity 2.5 m/s (after sliding velocity 1.0 m/s perceptible influence hydrodynamic lubrication). The values of friction coefficient become equal to 0.0025 and 0.007 respectively what is close to the values of friction coefficient in hydrodynamic regime of lubrication. The dependence of the pad undimensioned wear ($\mathbf{I_h=h/L}$, where **h** is linear wear, **L** is friction distance) on sliding velocity is shown in Fig. 8.

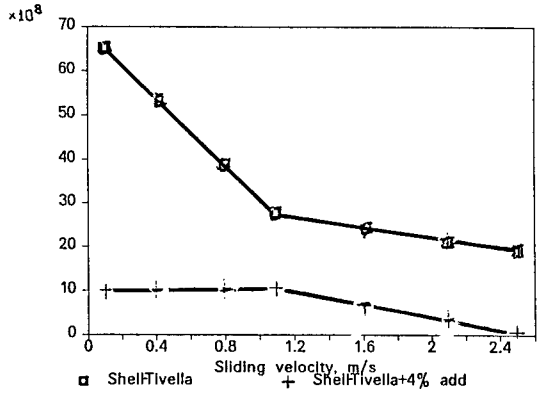


Fig. 8. Undimensioned wear vs sliding velocity (m/s).

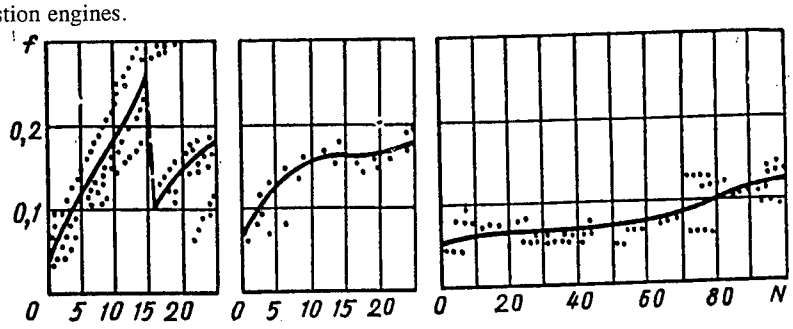
Undimensioned wear of the pad decreases as result of the additive SURM-3 application in 6.5 times for the sliding velocity 0.1 m/s. The influence of hydrodynamics effects increases with the rise of the velocity, however, the additive is effective in all investigated region of sliding velocities, undimensioned wear is decreased in 2.5 times as a result of the additive application for sliding velocity 2.5 m/s.

Analysis of the experimental results allows to conclude that application of the additive SURM-3 essentially changes the mechanism of friction interaction what results in tribotechnical characteristics improvement.

Show the example to using the greases, which development as realization selective transfer phenomena, in very stress joints of the changing geometry aircraft wing. On the Fig. 9 presented dependent of the friction coefficient to number of the using this joints with single lubrication. There are limit of the friction coefficient 0,2. Lifetime of the joints increase more than 10 times [9].

Controlling tribochemical regime: It is known that chemical reactions occurring in the zones of contact of rubbing bodies lead to a deterioration of the lubricant properties. The main reason for this are the temperature flashes at the contact points, contact strains as well as oxidation of lubricants. If we are concerned with friction units and machine aggregates, the wear products of the rubbing parts, the dust particles and the combustion products from the engines all contaminate the lubricant. All this leads to aging of the lubricant compositions,

deterioration of their serviceability and, as a result, to an increase in wear intensity. Therefore filters are used in these aggregates to remove solid inclusions and lubricants are employed which are more stable to chemical attack during the friction interaction processes. However, when traditional approaches are used, no permanent work of the oil can be reached in the case of overstressed running of the machine units and their aggregates, for example, of the combustion engines.



Lubricant TSIATIM - 201 Lubricant SVINTSOL - 01 Lubricant VNIINP - 254

Fig. 9. Friction coefficient vs to number of the using joints of the changing geometry aircraft wing.

One of the effective means in this area is the special device for combustion engine which contains chemical species affecting on the pumped-through motor oil [10]. These species restore lubricant properties and form on rubbing surfaces of cylinder-piston group, bearings and another moving joints the organometallic antifriction films containing Na-Sn alloys, metallic iodine, and hydrocarbons. The film thickness is about 200 Å. As the film possess poor mechanical properties, they can easily wear out, but can continuously be restored and it is possible to reach theoretically a dynamic equilibrium state at a definite chemical reaction rate. In practice, the application of this method leads to an increase in the capabilities of the combustion engine by a factor of several times with permanent oil work (Fig.10).

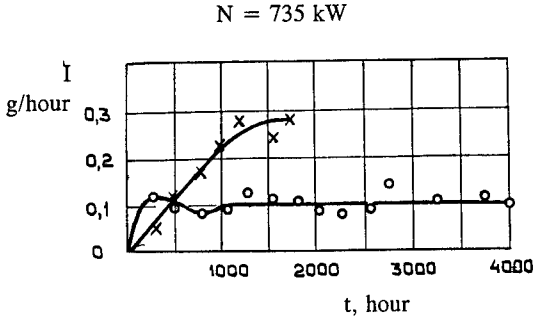


Fig. 10. Wear intensity vs. time: x, standard diesel; o, diesel with special device.

Reference

1. В.С. Авдуевский, М.А. Броновец. Трибология и машиностроение. Трение и износ. Т.11, №1,1990. Стр. 7-19.
2. В.С. Авдуевский, М.А. Броновец, Н.А. Буше, В.М. Школьников. Теоретические и прикладные аспекты современной трибологии. Первая международная конференция "Энергодиагностика". Сб. трудов. М. ИРЦ Газпром. Том. 1. Стр. 31-61.
3. H. Peter Jost. Tribology - Origin and Future. Wear. Intern. Journal on the Science and Technology of Friction, Lubrication and Wear. Vol. 136, № 1,1990, p.p.1-17.
4. V.S. Avduevsky, M.A. Bronovets. Main Trends in Tribological Development in the USSR Wear. Intern. Journal on the Science and Technology of Friction, Lubrication and Wear. Vol. 136, № 1, 1990, p.p. 47- 64.
5. Д.В. Орлов, Ю.О. Михалев, Н.К. Мышкин, В.В. Подгорнов, А.П.Сизов. Магнитные жидкости в машиностроении. М. Машиностроение. 1993. Стр. 8-11.
6. ГОСТ 23.209-79. Измерение износа деталей машин методом поверхностной активации. М. Изд-во стандартов. 1980.
7. I.O. Konstantinov. Development of Thin Layer Activation Technique in the Institute of Physics and Power Engineering. Paper. IAEA Experts Meeting on Nuclear Methods in Monitoring of Wear and Corrosion in Industry, Obninsk, Russia, 5-9 September 1994.
8. В.Ф. Пичугин, Д.В. Пичугин. Безразборное восстановление работоспособности двигателей внутреннего сгорания. Первая международная конференция "Энергодиагностика". Сб. трудов. М. ИРЦ Газпром. 1995. Том.3. Трибология. Стр. 64-72.
9. Д.Н. Гаркунов. Триботехника.М. Машиностроение. 1985.Стр. 273-286, Стр. 291-292.
10. B.P. Grigor'ev, USA Patent 3,298,952, 1967.

EFFECTIVE MAGNETIC METHOD FOR DETECTING LONGITUDINAL CRACKS IN THE PIPES OF PIPE-LINE TRANSPORT

G. Shelihov,
Doctor, Scientific & Technical Centre "EXPERT",
140003, Liubertsy-3, Moscov region, a/b 8, Russia
V. Usoshin,
142717, pos. Razvilka, Leninsky dst., Moscow region, Russia
V. Lozovsky,
Professor, Scientific & Technical Centre "EXPERT"
140003, Liubertsy-3, Moscow region, a/b 8, Russia
V. Rozov
Gazprom
117884, Moscow, Namiotkina 16, Russia

Abstract: The magnetic method based on sensing the magnetic field distribution in the vicinity of a defect with a special measuring device is intended for using in flaw-detector gears for intra-pipe inspection and allows detecting longitudinal cracks, corrosion affections on an external pipe surface. The instrument УПМД-1 has been developed which implements this method and features high noise immunity and tolerance for many external factors.

Key words: Magnetic method; longitudinal cracks; intra-pipe inspection

An analysis of break-downs of cross-country gas lines has shown that in many cases the cause of pipe-line damages in operation is the presence of longitudinal cracks arising due to high internal static stresses in pipe's material, corrosion affections and pressure fluctuations inside the pipe.

After originating the longitudinal cracks propagate for a long time. Critical size having been reached, the crack causes a pipe-line break-down.

Existing systems of intra-pipe inspection give a principal possibility of detecting only cross cracks and corrosion affections. These systems are based on measuring tangential or tangential and normal field strength components at the internal pipe surface being checked. The sensor moving along the pipe crosses a cross crack and indicates the field distribution, usually the tangential component of field strength in the vicinity of the defect, according to which the record signal is produced. In some systems of flaw-detector gears this signal is compared with defect signals recorded in the course of prolonged operation and stored in computer library. Also, such systems do not allow longitudinal cracks to be detected because the sensor of the moving flaw-detector gear travels along the crack. So the sensor signal carries insufficient information about the field distribution over a defect.

What is more, as the flaw-detector gear moves the gaps

between the sensor and pipe surface vary, which leads to changing signal amplitude. Disturbances can result from surface irregularities, various ferromagnetic impurities (generated in mounting the pipe-line), variations in gaps between the pipe and magnetizing system, discontinuities in magnetic properties of pipe's material. These as well as other inferring factors decrease the sensibility of a flaw-detecting system and hamper enhancing the detection signal, especially when its amplitude is well in excess of a noise level.

Because of this there is a need for a different way to solve the problem of detecting longitudinal cracks, the alternative method of magnetic inspection.

The foundation of the method developed for this purpose is an original principle of acquiring information and processing signals which consists in the fact that according to the information obtained from the measuring device a vector field is automatically constructed and a field pattern, that is, the distributiom of magnetic lines of force is determined. This makes it possible to obtain sufficient amount of characteristic parameters of the field in the vicinity of a defect which are directly associated with defect parameters. The quantitative values of these parameters give a complex of criteria for screening in accordance with technical specifications.

On this basis the instrument УПМД-1 was developed which includes matching device, electronic scanner system, microprocessor, and output device for storing information obtained. The instrument УПМД-1 features high noise immunity and allows reaching higher sensibility as compared to existing systems.

For example, variations in the gaps between the measuring element and the pipe surface under check do not influence substantially the inspection results because in this case of significance are not absolute signal magnitudes but their relative values. The field pattern constructed according to signals from the measuring element with different gaps are similar.

Random electric and magnetic noise also does not substantially affect the inspection results, it is easily filtered owing to being significantly different from detection signals (synthesis of random signal does not influence the standard field).The instrument УПМД-1 has a wide range of measurable fields. This allows solving a number of other problems with the use of specially designed sensors, for example, in determining the regions with elevated levels of internal stresses in pipe's material.

A pilot sample of the instrument has been fabricated which is now undergoing laboratory tasts.

LUBRICANT CONDITION MONITORING #3

A NEW ORBITAL BALL TYPE OF VISCOMETER

by Albert Abnett,[1] and Ray Garvey[2]

Abstract: This paper describes a new "orbital ball" type of digital viscometer designed for use in maintenance departments to measure viscosity of lubricants and hydraulics. A steel ball placed in a small cup containing the test oil is caused to rotate around the inner wall of the cup by electromagnetic action. The speed of the ball is measured with a microcomputer and viscosity of the fluid is computed. The operator has the option of easily cleaning the cup, or discarding the cup and ball for little cost. Other advantages are low cost, ruggedness, small sample volume, wide range, high repeatability. Equations governing the operation are developed, and experimental results are presented. Temperature of the fluid is measured and results can be extrapolated to standard conditions.

Key Words: Digital, electromagnet, lubricant, oil, hydraulic, temperature, viscosity, viscometer

INTRODUCTION: A digital viscometer was designed and developed for use in-plant by maintenance mechanics and technicians, enabling them to measure the viscosity of lubricants and hydraulic fluids. The digital viscometer is intended to be used in conjunction with other on-site instruments which measure fluid quality, contamination, and wear debris. The primary purposes of this device are

- to detect misapplication of lubricant (e.g., wrong or mixed oils),
- to detect fuel dilution in liquid fueled engine applications,
- to establish the lubricant viscosity for purposes of trending,
- and setting parameters for further tests which have viscosity dependent characteristics.

Misapplication of lubricants is a common problem that can normally be detected quite easily by measuring viscosity. The problem occurs whenever someone inadvertently selects the wrong lubricant when topping off or refilling a machine. Most often the difference between right and wrong lubricant is simply a matter of viscosity -- ISO grade 32 was used instead of the correct ISO grade 68. This oversight can lead to significant damage when the lubricated machinery reaches design loads at peak operating temperatures.

[1]Caliber Instruments, Inc., 208 N. Main St., Nevada, OH 44849, (614) 482-2197

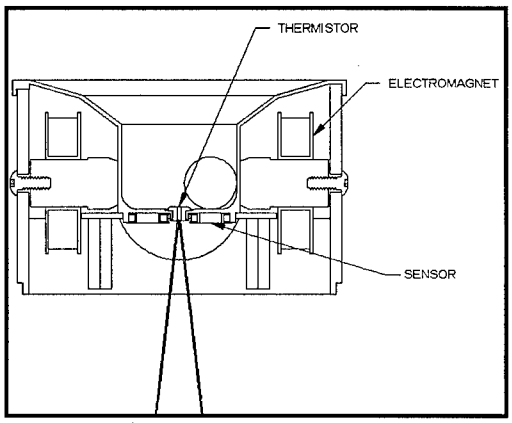
[2]Computational Systems, Inc., 835 Innovation Dr, Knoxville, TN 37932, (423) 675-2110

Fuel dilution can lead to rapid and severe wear problems for diesel and gasoline engines. The problem normally originates from a leaky fuel injector. Truck operators have termed this condition "making oil" because the dip stick may show more oil after operation than before. It does not take very much fuel to seriously compromise lubrication. Since lubricant viscosity is the primary factor affected by fuel dilution, viscosity can be a very simple and direct measure of the problem. In a blind test performed by the author with 300 diesel engine oil samples, the 7 samples exhibiting fuel dilution were correctly identified using viscosity measurement. All of those with fuel dilution showed in excess of 10% drop in viscosity as compared to a new oil sample. In all of these cases, the viscosity drop due to fuel dilution overpowered viscosity increases due to other factors such as oxidation or soot loading.

Third on the list of priorities for in-shop viscosity measurement is to measure and trend viscosity. It is well known that viscosity is one of the most important characteristics of a lubricant. As such it is a good idea to measure and trend this parameter, watching for adverse conditions to appear. Unfortunately, just about everything affects viscosity and as such, it becomes quite difficult to determine what, if anything, is wrong with a lubricant simply using viscosity. Temperature is possibly the greatest variable. For example the percent drop in viscosity due to heating from room temperature (20 to 25 C) to machine operating condition can be 70% if the machine operates around 40 C, or it can 95% if the machine operates around 100 C. Some other variables that affect lubricant viscosity include oxidation (typically increases viscosity, although some synthetics may drop in viscosity after chemical oxidation), soot loading (typically increases viscosity), and moisture contamination (this can increase viscosity until saturation occurs, then drop viscosity). For these reasons, viscosity trending can be a valuable confirmation of a suspected problem, detected using another instrument.

Finally, some of the other instruments used for on-site oil analysis suggest viscosity classification to determine flow rates, settling rates, or dilution requirements. In these cases, the digital viscometer is a convenient accessory.

DESCRIPTION: The principle of operation of this new "orbital ball" viscometer is simple and produces results consistent with the falling ball type of manual instrument. A small cup holds a sample of the test fluid. See Figure 1. Electromagnets spaced around the circumference of the cup create a rotating magnetic field, similar to a D.C. motor. A steel ball is immersed in the fluid and its position is reported by proximity sensors acting like the commutator on a motor. As each sensor detects the ball position the "next" magnet is turned on, pulling the



ball around the inner wall of the cup. The speed of this orbital motion is determined by the strength of the magnetic force and the viscosity of the fluid. For a constant field, higher rotation speeds result from lower viscosities. As will be shown, at low Reynolds numbers the motion is independent of extraneous effects and depends only upon fluid viscosity at the viscometer operating temperature.[3] Once it is calibrated, the digital viscometer can be used to measure viscosity of most fluids (10 cp or cSt to 1000 cp or cSt) at ambient temperature. Narrowing from "most fluids" to "most lubricants," the viscosity-temperature equation, ASTM D-341, is used to calculate viscosity values at 40 C.

The actual test volume is 12 ml. Tests typically take approximately 1 minute. Fluid is discarded at the end of a test. The orbital ball viscometer is convenient to use due to its ease of cleaning. The cup can be readily cleaned with a shop rag in the field.

EQUATIONS OF MOTION: The ball will rotate at a velocity such that the force due to the magnetic field is exactly balanced by the viscous drag force. The accepted formula for drag of an object moving through a fluid is:

$$F_d = C_d \bullet A \bullet p \bullet V^2 / 2 \qquad (1)$$

C_d is drag coefficient
A is cross section of ball
p is fluid density
V is relative velocity

The drag coefficient is a function of the Reynolds Number, N_R , and has been found empirically to follow the graph reported by Vennard .[4] Stokes has shown that for Reynolds Numbers less than 1, C_d is a linear function of N_R. In order to arrive at motion equations for a wide range of viscosities, this curve was further divided into three regions: Reynolds Numbers less than 1, 1 to 10, and 10 to 1000. Equations for drag coefficients in each region were derived by simple curve fitting.

Reynolds Number less than 1:

$$C_d = 24 / N_R \qquad (2)$$

Reynolds Number 1 to 10:

$$C_d \approx 24 / N_R^{3/4} \qquad (3)$$

---

[3]First disclosed in U. S. Patent 5,394,739, issued to Computational Systems, Inc.

[4]Elementary Fluid Mechanics, by Vennard, 2d edition published 1949 by John Wiley, graph of drag coefficients vs Reynolds numbers, page 297

Reynolds Number 10 to 1000:

$$C_d \approx 12 / N_R^{1/2} \quad (4)$$

The Reynolds Number can be calculated from:

$$N_R = pVD/\mu \quad (5)$$

p is fluid density (kg/m^3)
V is velocity (m/s)
D is ball diameter (m)
μ is fluid absolute viscosity (kg/m•sec)

Substituting equations 2 - 5 into equation 1, and plugging in actual dimensions of the 9/16 inch ball, gives three equations of motion for the orbital viscometer:

Reynolds Number less than 1:

$$T_{perRev} = (.135 / F_{mag}) \bullet \mu \quad (6)$$

Reynolds Number 1 to 10:

$$T_{perRev} = (.086 / F_{mag}^{4/5}) \bullet \mu^{3/5} \bullet p^{1/5} \quad (7)$$

Reynolds Number 10 to 1000:

$$T_{perRev} = (.040 / F_{mag}^{2/3}) \bullet \mu^{1/3} \bullet p^{1/3} \quad (8)$$

T_{perRev} is 1/V and is time in seconds for one revolution of ball
F_{mag} is magnetic force in Newtons to overcome drag
μ is viscosity in kg/m•sec (1000 centipoise)
p is fluid density in kg/m^3

Equation 6 shows that the time per revolution is linearly proportional to fluid viscosity. This linear relationship between revolution time and viscosity is born out by experimental data of higher viscosity fluids (from 50 to 5000 centipoise) if the ball velocity is kept low. For one specific design of viscometer with a 1.125" diameter cup, a 9/16 inch ball, and using .9 for fluid density, the Reynolds number is seen to be:

$$N_R = 640 \bullet \text{Revolutions per Second / Centipoise} \quad (9)$$

So for $N_R \leq 1$, Revs per second must be kept slower than Centipoise / 640. This is accomplished by the microprocessor controlling the electromagnet current. However, for low viscosity fluids (< 50 centipoise), it is not practical to run with the very low magnetic force. In this case it is necessary to accept the non-linear relationship, and to plug in fluid density; this is not considered a limitation in a computer controlled instrument.

The magnetic force on the ball is given by:

$$F_{mag} = \beta^2 A / \mu_0 \qquad (10)$$

The magnetic field, β, is determined primarily by exciting current and air gap, and is:

$$\beta = N \bullet I \bullet \mu_0 / l_{air\ gap} \qquad (11)$$

Substituting (11) into (10) and using specific constants for this viscometer:

$$F_{mag} = 32.2 \bullet 10^{-6} (I / l_{air\ gap})^2 \qquad (12)$$

F_{mag} is force, in Newtons
I is magnet current in amperes
$l_{air\ gap}$ is in meters

If all were simple, equation 12 could be substituted into equations 6, 7, or 8, and the revolution time of the ball could be calculated for a given fluid viscosity. There are additional factors that come into play when computing F_{mag}, however. First, the air gap is not constant; because of the geometry the air gap "tends" toward a fixed number because as the ball approaches one pole it is leaving the previous pole, so when between poles the air gap is nearly constant. But the magnets must switch before the ball gets between poles, so it starts with a larger gap. This means that the velocity is not constant but goes through a max and min every quarter revolution. So the average flux is between the extremes. Second, the magnetic flux is diverted from the exciting pole toward the other free poles as well as through the ball, another dilution. Both effects tend to reduce the useful magnetic flux. Only a fraction of the calculated peak energy contributes to ball motion.

Just to get a feel for the equations a specific example will be considered. In one experiment an oil of 525 cp and a magnet current of .65 amps produced a ball revolution time of 3.8 seconds. Equation 9 shows this is a Reynolds number of .32 so equation 6 can be used. The magnetic air gap was seen to vary from about .5 inches to .8 inches. According to 12, this should produce a force excursion of .028 to .072 Nt, or an average of .05 Nt. If all the flux went to ball motion, equation 6 shows a revolution time of:

$$T = (.135/.05)\ (525/1000) = 1.4 \text{ seconds}$$

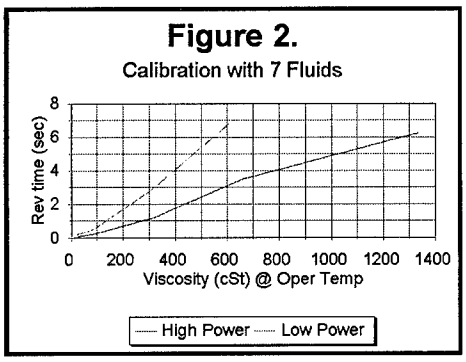
The "wasted" flux is proportional to the square root of the time ratio, or $(1.4/3.8)^{.5} = .61$. So 39% of the magnetic flux is not producing ball-viscous energy. This is a geometry constant so when the viscometer is calibrated with known fluids all later readings will produce correct results. A program in the host computer facilitates calibration.

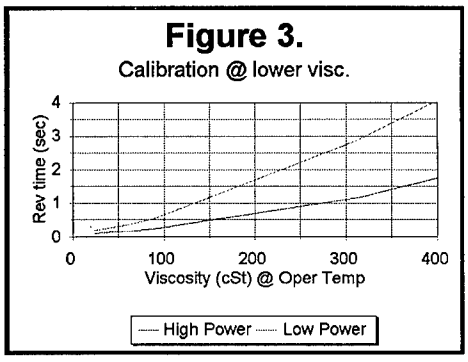
EXTRANEOUS EFFECTS: There are two obvious concerns of variables which might affect ball speed: rolling friction of the ball, and surface roughness effects of fluid flow. An accepted formula for rolling friction[5] is:

$$F_f = k \bullet W / r \qquad (13)$$

W is normal force
r is ball radius
k is coefficient of friction (approx .01 for steel on plastic)

The normal force is the vector sum of the weight of the ball in liquid, and centrifugal force. Because of low speeds and buoyancy of fluid the friction forces are less than 1% of the viscous forces down to 1 cp conditions. It has been shown[1] that fluid flowing through a pipe is unaffected by surface finish of the inner walls of the pipe so long as laminar flow conditions are maintained, i.e. the Reynolds number is kept below 2300. It seems consistent that the same conditions apply to a ball rolling around the wall of a cup filled with a fluid. Thus the speed of the ball's rotation is seen to be independent of surface finish of the cup wall and of the ball itself.





CALIBRATION: A piecewise linear interpolation curve, similar to that shown in Figure 2 and Figure 3, is used to calibrate the digital viscometer. These graphs show the calibration test results for various fluids representing the overall viscosity range. The viscosities shown reflect the actual viscosity at operating or ambient temperature.

VISCOSITY-TEMPERATURE COMPENSATION: Since the primary purpose of the digital viscometer is to measure viscosity for in-plant oil analysis, it is critical that the measurements be translated into units which are convenient for the user. The industry norm for lubricant viscosity measurement is to report results in centistokes (cSt) at 40 C and 100 C. The authors have found that although the digital viscometer physically measures centipoise, if it is calibrated with mineral based fluids in centistokes, the results can be reported in centistokes with reasonable accuracy as shown below. The following methodology has been developed and demonstrated to provide reasonable results in units of cSt at 40 C for a wide range of lubricants.

The temperature of the fluid is measured by means of a thermistor in the bottom of the cup. Using this thermistor, the viscosity of a *known* lubricant can be calculated at any temperature using the following equation from ASTM D-341 can be used down to 2 cSt [4]:

$$\log \log (\mu + .7) = A - B \bullet \log K \qquad (14)$$

μ is in centistokes (cSt)
K is temperature in Kelvin
A and B are constants to be determined

Rearranging terms, the constants for a specific lubricant are given by:

$$A = \log \log (\mu_1 + .7) + B \bullet \log K_1$$
$$B = [\log \log (\mu_2 + .7) - \log \log (\mu_1 + .7)] / [\log K_1 - \log K_2]$$

where K_1, K_2, μ_1, μ_2 are measured values

A convenient method for determining the "A" and "B" constants for a particular type of lubricant is to make use of the 40 C and 100 C viscosity values which are readily available from a lubricant data sheet or fluid analysis report. In this case K_1 and K_2 are 313.16 and 373.16 K respectively, and μ_1 and μ_2 are as reported on the lubricant data sheet or lab oil analysis report.

Once these constants have been determined then the viscosity for the *known* lubricant can be determined at any temperature from equation 15:

$$\mu|_{Temp} = 10^{10^{[A - B \bullet \log(K)]}} - .7 \qquad (15)$$

Equation 15 was used with *known* lubricants to calculate the true viscosity values at operating temperatures as shown in Figures 2 and 3 above. To measure the 40 C viscosity of an *unknown* oil, the following procedure can be used.

First, the operator the 40 C and 100 C viscosity values measured from a new sample of the particular lubricant. If the operator does not input these, then default selection is automatically

applied, sacrificing measurement accuracy while still giving good repeatability. These inputs are then used to compute the "A" and "B" constants for new or "reference" lubricant.

Next, approximately 12 ml of used oil is added to the sample cup and a test is run. The digital viscometer automatically operates through high, then low, power settings. If the viscosity is extremely high causing revolution time to be longer than about 4 seconds per revolution, the viscometer will stop after the high power setting.

The average revolution time for the orbital viscometer is converted directly to viscosity at operating temperature using data from Figures 2 or 3. This measurement is defined as V1

A separate calculation of viscosity at this particular temperature is performed using equation 15. The calculated viscosity for the reference lubricant at actual operating temperature is defined as V2.

Finally, the temperature compensated viscosity measurement is computed using the following equation:

$$V_{40}used = V_{40}reference * (\frac{V1}{V2}) \qquad (16)$$

where,

$V_{40}used$ = 40 C viscosity for used oil
$V_{40}reference$ = 40 C viscosity for reference
V1 = measured viscosity at operating temperature
V2 = reference viscosity at operating temperature

In an alternate method, the orbital viscometer may be used to measure the viscosity at two temperatures. Typically the lower temperature is at ambient and a second higher temperature is achieved through self-heating of the electro-magnets. Then the constants, A and B are calculated and used to calculate viscosity at any desired temperature. For instance, the viscosity at 40 C is computed by:

$$\mu|_{40_C} = 10^{10^{[A - B \bullet \log(313.16)]}} - .7 \qquad (17)$$

TEST RESULTS: Figures 4 and 5 show results using the digital viscometer to measure viscosity for multiple fluids at two different power settings. These 40 C viscosity results were obtained by applying measured revolution time to calibration curve to obtain viscosity at operating (ambient) temperature then compensating this viscosity value to 40 C using equations 15 and 16.

CONCLUSIONS: The "orbital" digital viscometer is designed for field use in measurement of lubricant viscosities. It is capable of identifying misapplication of lubricants, liquid fuel dilution, and viscosity measurement for the purpose of trending. It is also suitable for use with other in-shop oil analyzers which are use viscosity as a basis for setting flow rates, settling rates, and dilution parameters.

The digital viscometer uses a patented orbiting ball to measure viscosity. The speed of the ball is determined by the physical geometry of the cup, by the power settings for the electromagnets, and by the viscosity of the fluid in the sample cup. By controlling the geometry and electromagnet power, viscosity is directly related to ball speed. Testing is done using approximately 12 ml of fluid. Tests take approximately 1 minute.

To convert measured lubricant viscosity at a the test temperature to that which would be measured at 40 C, the standard viscosity-temperature equations from ASTM D-341 are used.

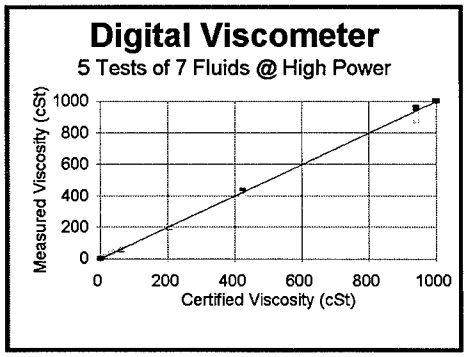


Figure 4

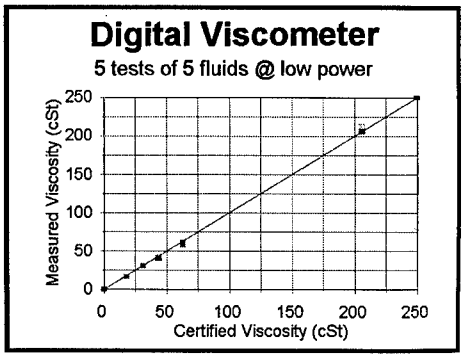


Figure 5

REFERENCES

1. Fundamentals of Fluid Mechanics, by Gerhart, Gross, and Hochstein
   2d edition published 1992 by Addison Wesley Publishing Co.

   a) surface roughness page 477

2. Elementary Fluid Mechanics, by Vennard,
   2d edition published 1949 by John Wiley

   a) Stokes drag law page 296
   b) graph of drag coefficients vs Reynolds numbers, page 297
   c) formula for drag force, page 285

3. Fluid Mechanics Source Book, by Sybil Parker, published 1988 by McGraw-Hill

4. F.A. Litt, "Viscosity - Temperature Relations, " Starting from Scratch, Tribology Basics
   Society of Tribologists and Lubrication Engineers, updated publication.

5. Baumeister and Marks, Standard Handbook for Mechanical Engineers
   1967 edition by McGraw-Hill

   a) rolling friction page 3-40

6. Elements of Physics, by Shortley and Williams; published by McGraw-Hill, 1955

7. Electric Machinery, by Peter F. Ryff, published 1988 by Prentice-Hall

   a) magnetic force page 38

ESTIMATING WATER CONTENT IN OILS: MOISTURE IN SOLUTION, EMULSIFIED WATER, AND FREE WATER.

by Ray Garvey and Grahame Fogel
Computational Systems Incorporated, 835 Innovation Dr, Knoxville, TN 37932

Abstract: This paper discusses the detrimental effects of moisture in lubricants, hydraulics, and transformer oils. It describes a means for automatic estimation of moisture content in lubricating oils, hydraulic fluids, and transformer oils. This is a field test method, used as an alternative when it is not convenient to measure water content by the Karl Fisher test or another analytical method. The total moisture content is found to be the sum of three separate measurements: moisture in solution, emulsified water, and free water. A single capacitive sensor is used to make all three measurements. Moisture in solution is instantaneously measured from a well mixed contaminated oil sample as compared to a similar measurement for a clean reference oil sample. Emulsified water is measured by determining the median-average rate of change in sensor output over time. Free water is measured by determining the accumulation of step changes in sensor output over time. The capacitive sensor is also used to identify the general oil type based upon the nominal dielectric value and upon the change in dielectric as a function of temperature change.

Key Words: Additive, dielectric, emulsified, moisture, oil, proactive, synthetic, water

Background: The measurement of moisture in lubricants, hydraulics, and transformer fluids using an automatic in-shop device is a valuable capability for industrial plant maintenance departments. Following is a list of some advantages for doing analysis on site:

- ownership and control;
- immediate results with immediate retest when needed;
- testing by people who know most about machinery;
- data is electronically stored with no transfer between lab and user; and
- test incoming fluids, with ability to fix problems, and verify correction.

The adage that "oil and water don't mix" is not entirely true. They do mix, to a greater or lesser extent, depending on oil base stock and additives. This paper addresses a method for evaluating water content, considering that water can either be in solution with the lubricant, emulsified with the lubricant, or carrying free water droplets in the lubricant. Before introducing the measurement methods, one should understand the answers to these questions.

- What are the functions of a lubricant? a hydraulic? a transformer oil?
- What are the effects of moisture in these fluids?
- In what forms does water exist in the oil?
- What are the effects of water on lubricant additives?

The answers to these important questions establishes the need to not only measure total water content, but to identify what amount is in solution, emulsion, or free state.

Functions of a Lubricant, Hydraulic, or Transformer: Lubricating oils serve multiple functions in mechanical systems. They reduce friction in each of the three lubrication regimes depending on application type: hydrodynamic, elestohydrodynamic (EHD), and boundary.

For applications with conformal bearings, such as Babbitt bearings, the lubricant transfers distributed loads, in the form of radial pressure, from shaft to sleeve. This regime is referred to as hydrodynamic lubrication and is similar to hydroplaning of an automobile on wet pavement

For highly loaded anti-friction bearings and gears, the viscosity of the lubricant must increase exponentially under the pressure due to lubricant being forced between the rolling element and the raceway. The lubricant viscosity increases until the oil has the bulk elastic modulus of a solid, and is able to transmit highly concentrated loads. Elastohydrodynamic lubrication practically eliminates metal to metal contact with extremely low frictional energy losses.

For traction type applications such as slides and wheels, the lubricant serves to reduce wear due to sliding contact and abrasion. In this case the physical loads are transmitted through-metal-to metal contact. The traction or friction in this type of system can be very high, with coefficient of sliding friction as high a 1/3 of the normal force that presses the two components together. The friction, wear, and traction can be greatly reduced by lubricating the surface(s) so that only a limited amount of shear can be transmitted between components. This can be done with either solid lubricants (molybdenum disulfide or graphite) or with liquid lubricants including extreme pressure and anti-wear additives.

In addition to these requirements for transferring loads, reducing friction, and minimizing wear, lubricants fulfill several other functions critical to machine life. These include sealing surfaces from corrosion and other forms of chemical attack, transmitting hydraulic power, cleaning away contaminants and wear debris, cooling hot surfaces, and isolating (electrically) dissimilar metals from galvanic interaction. In the special case of transformer applications, oil provides cooling and electrical isolation (i.e., a low dielectric medium).

Effects of Water on Lubricants, Hydraulics, and Transformer Oils: Moisture is deleterious to most mechanical systems, hydraulic systems, and transformers. It causes problems by rendering the oil additives ineffective, by corroding metal surfaces, and by incapacitating elastohydrodynamic (EHD) lubrication.

Moisture chemically reacts with many oil additives thereby disabling their ability to provide antiwear, anti-oxidation, extreme pressure (EP), anti-foaming, and detergency functions. Moisture also consumes all or part of the dispersant and emulsifier additives intended to prevent contaminants from coalescing or agglomerating.

It is well known that moisture promotes corrosion of metal surfaces. Water is a reactive species at room temperature, especially towards metals. Typically metals under ambient conditions are

passivated and contain an oxide coating. However, metals that are under stress or wear may produce a "fresh" surface that is highly reactive. For example, a scratch or crack exposes metal resulting in the following reaction:

$$M + H_2O \rightarrow MO + H_2$$

In this reaction the electron rich oxygen atom reacts with the electron deficient metal to form a new compound. The metal forms a new oxide coating and produces hydrogen gas. Depending on the metal (or alloy) present, the hydrogen may diffuse along grain boundaries, causing an effect known as hydrogen embrittlement, which accelerates wear rates.

The polar nature of the water molecule supports ionic mobility and encourages chemical attack to copper, lead, and other reactive metal surfaces, particularly in the presence of air and water. It also supports electrochemical interaction between galvanically dissimilar oil wetted metal parts leading to rapid corrosion.

As mentioned earlier, EHD lubrication depends on the unique pressure-viscosity property of mineral oils and synthetic lubricants. Free water droplets are unable to maintain physical separation between rolling elements or between gear teeth under these extremely high pressures. The resulting impacts can cause surface damage and shortened component life.

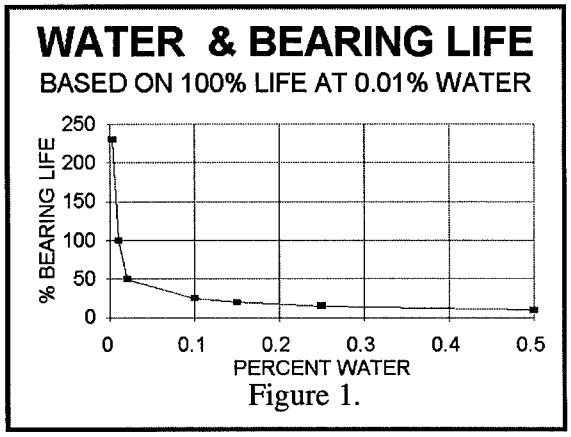


Figure 1.

Water contamination of the lube or hydraulic oils shortens component and machine life. Timken Bearing Company report on the effect of water on bearing life as shown on the attached graph. Notice that reducing water contamination from 0.01% (100 ppm) to 0.0025% or (25 ppm) increases bearing life by a factor of 2.5 times!

The Form(s) of Water in Oil: There are three forms or states in which water is found when combined with oil. The water can be dissolved in the oil in the sense that it is hydrated, dissolved or reacted with additives that are mixed with the oil. In another state, oil can be dispersed or emulsified with the oil. Finally, it can be independently stratified or mixed as free water droplets which are separate from the oil. Each of these damages the oil and the mechanical system in a different way.

Moisture dissolved in the oil consumes performance enhancing additives and promotes corrosion. "Water's reaction with oxidation inhibitors produces acids and precipitates. These water-reaction

products increase wear and interferences. At high operating temperature (above 60 C), water reacts with and destroys zinc type antiwear additives. For example, zinc dithiophosphate (ZDTP) is a popular boundary lubricant added to hydraulic fluid to reduce wear in high pressure pumps, gears, and bearings. When this type additive is depleted by reacting with water, abrasive wear accelerates rapidly. The depletion shows up as premature component failure, resulting from metal fatigue and other wear mechanisms."[1] Other reactions produce acidic compounds which secondarily react with metal compounds of the machine the lubricant is indented to protect.

Water emulsified in oil consumes nearly all remaining additives, increases corrosion, and changes fluid viscosity (normally increasing it). Systems designed to operate with water present must either *emulsify* or *demulsify* the water to extend life of mechanical parts. This is discussed in more detail in later sections of this paper. Automotive applications must emulsify the water until it can be driven off by heat. Steam turbines typically demulsify water, dropping it out in the oil compartment (sump) where it can be bled off. In either of these cases, water left in the system can lead to corrosion and even microbial growth. Under certain conditions, microbial bacteria can live and reproduce very rapidly when sufficient water is present in oil. These living organisms can ruin the oil system, clogging filters, changing water emulsion characteristics, increasing corrosion rates and producing acidic waste products.[2]

Free water droplets, which occur after the oil has been fully saturated, are desirable in some lubricating systems (those designed to demulsify water) and very detrimental in others. This is explained in the following paragraphs.

The Effects of Moisture on Oil Additives: In mineral oils and polyalphaolephin (PAO) synthetic hydrocarbon oils, additives are required for any significant amount of water either to be dissolved in oil or to be emulsified with oil. The polar nature of the water molecule and the non-polar nature of the hydrocarbon mineral oil make them immiscible. Pure mineral oils and transformer oils are saturated with as little as 5 ppm water, and turbine oils saturate with as little as 200 ppm water. Above saturation level, water coalesces into free water droplets eventually settling to the bottom of the oil compartment.

This is a very desirable characteristic for most steam turbines[3] and paper machines[4]. In these applications, water contamination is common. So to avoid the consequences of moisture retention in these machines, oils are designed to demuslsify, or drop water out, in the oil

---

[1]"Walking the Fluid Cleanliness Tightrope, Part 2," by Al Zingaro, *Hydraulics and Pneumatics,* December, 1994, pg. 26.

[2]"Microbial Problems in Metal Working Fluids," by Frederick Passman, Angus Chemical Company, Northbrook, IL, *Starting from Scratch: Tribology Basics,* STLE Publication.

[3]"Delivery Control of Lubricating Oil," K-Standard K-1171-7, dated 94-07, ABB Stal.

[4]"New Direction for Paper Machine Lubricants, Ashless vs. Ash Type Formulas," by Dr Louis R. Kray, et. al., Chevron Research and Technology Company, Richmond California, *Lubrication Engineering,* October, 1995, pg. 834-838.

compartment before returning to the machine. Then periodically, water is bled off from the bottom of the oil compartment (e.g., sump). Demulsibility of oil is good if water separates quickly at operating temperatures. It is bad if it separates slowly.

These oils which demulsify water are normally pure mineral oils or PAOs with very little additive. Mr. N. G. Nilsson of ABB Stal points out that one liter of automotive lubricant has enough additive to destroy the demulsification characteristics of 6000 liters of turbine oil, if one were to accidentally mix the two types. In fact, when additive concentrations > 1.0 ppm for Ca, Mg, or Ba are found, then emulsion characteristics of the oil must be validated by ASTM 1401 emulsion test to demonstrate clear separation of water from oil within 15 minutes at 54 C. This ensures that water picked up in the turbine will be dropped out in the oil compartment before returning to lubricate the machine again.

TABLE I

ADDITIVE ELEMENTS	RANGES
AUTOMOTIVE	
MAGNESIUM	50-400 PPM
SODIUM	0-200 PPM
BARIUM	0-150 PPM
PHOSPHOROUS	700-1300 PPM
CALCIUM	400-2000 PPM
ZINC	800-1400 PPM
TOTAL	≈ 4000 PPM
INDUSTRIAL GEAR OIL	
BARIUM	0-25 PPM
PHOSPHOROUS	300-750 PPM
CALCIUM	5-50 PPM
ZINC	500-1000 PPM
TOTAL	≈ 1500 PPM
E P ADDITIVE GEAR OIL	
SODIUM	75-500 PPM
BORON	998 PPM
PHOSPHORUS	175-425 PPM
CALCIUM	10-150 PPM
ZINC	300-550 PPM
TOTAL	≈ 1500 PPM
HYDRAULIC OILS	
BARIUM	0-200 PPM
PHOSPHORUS	225-325 PPM
CALCIUM	25-150 PPM
ZINC	350-500 PPM
TOTAL	≈ 1000 PPM
COMPRESSOR OILS	
BARIUM	0-200 PPM
PHOSPHORUS	0-35 PPM
CALCIUM	0-35 PPM
ZINC	0-50 PPM
TOTAL	≈ 200 PPM

Engine oils and many other lubricants will tend to *emulsify* rather than *demulsify* water. The additives in these serve to disperse water and prevent it from coalescing into free water droplets. Additives modify the solubility and emulsion character of mineral oils. Transformer oils saturate with 3 to 10 ppm water. Oil-base hydraulic fluids are typically saturated with 100 ppm (0.01%) to 1000 ppm (0.1%) water[5]. Industrial lubricants are typically saturated with 600 ppm (0.06%) to 5000 ppm (0.5%) water. Automotive lubricants are typically saturated with 1% to 5% water. Stern tube oils (those used to lubricate the aft bearing on a ship propulsion shaft) have been tested with 16% water with no evidence of saturation. Each of these lubricants has increasing amounts of additive, particularly increasing amounts of total dispersant + detergent + anti-oxidant + anti-wear + extreme pressure.

[5]"Walking the fluid Cleanliness Tightrope, Part 2," by Al Zingaro, *Hydraulics and Pneumatics,* December, 1994, pg. 25-26.

One can estimate the percentage of additive in a lubricant from the total ppm of additive elements including Zn, Ca, Mg, P, K, Ba, and sometimes Na. Table I displays typical ranges for elemental compositions of additive for various lubricant types.[6]

Traditional Water Measurement Methods used by Oil Labs: Several different methods are used in oil laboratories to measure moisture content of lubricants, hydraulics, and transformer oils. Three of the more common methods are Fourier Transform Infrared (FTIR), the "crackle" procedure, and Karl Fisher procedure.

The FTIR is a form of infrared absorption spectroscopy used to identify and quantify organic functional groups within the oil molecular structure. Water exhibits an absorption peak at 2.9 micron and 9.6 micron wavelengths. Results are typically reported with an index value commonly called, *"absorption units"* or *AU*[7]. Some industrial fluid analysis laboratories use the FTIR method to flag samples containing water for subsequent testing using the Karl Fisher method.

Crackle is a qualitative test in which a few drops of oil are placed on a hot surface (about 130 C) . If water is present, it quickly boils and sputters, giving audible indication. Although this is a very widely used test, "it is extremely subjective, and does not indicate the actual level of water present."[8] Since the crackling sound results when free water boils, it is not a good method for measuring water in solution and has limitations with respected to emulsified water measurement.

Water content by the Karl Fisher method (ASTM D1744) is a very accurate test, particularly for small amounts of water (<1000 ppm or 0.1%) although it can be used for greater levels of contamination as well. This procedure involves wet chemistry titration to an end point, reacting an oil sample with the Karl Fisher reagent. It is normally done in a ventilated laboratory.

Time-base Capacitance Measurement Apparatus: A time-base capacitance measurement apparatus[9] was evaluated with respect to water contamination measurement capabilities. This device is intended to be used in plant maintenance departments for the purpose of immediate testing of used lubricants. In addition to measuring moisture, the device also reports on wear condition and lube quality which was first reported in 1992.[10]

---

[6]"Oil Analysis and Wear Particle Analysis", by E. Zigler, DuPont Accession Report 17677, Dated April 1991, presented to the *1991 PPM Users Seminar,* May 14-16, 1991.

[7]"Tech Traks" by Lubricon, Lubricant Consultants, Inc., Box 51506, Indianapolis, IN.

[8]"Oil Analysis, Part 2 - Determining What Tests to Run," reprinted from *Fluid and Lubricant Ideas,* Amsoil Corporation, document G-286, pg. 6.

[9]The apparatus used is the OilView Analyzer available from Computational Systems Incorporated, Knoxville, Tennessee 37932, (423) 675-2110.

[10]"Portable Oil Analyzer for Predictive and Proactive Maintenance," by J. Kirkpatrick, Computational Systems, Incorporated, *Annual Meeting of the Vibration Institute,* 1992.

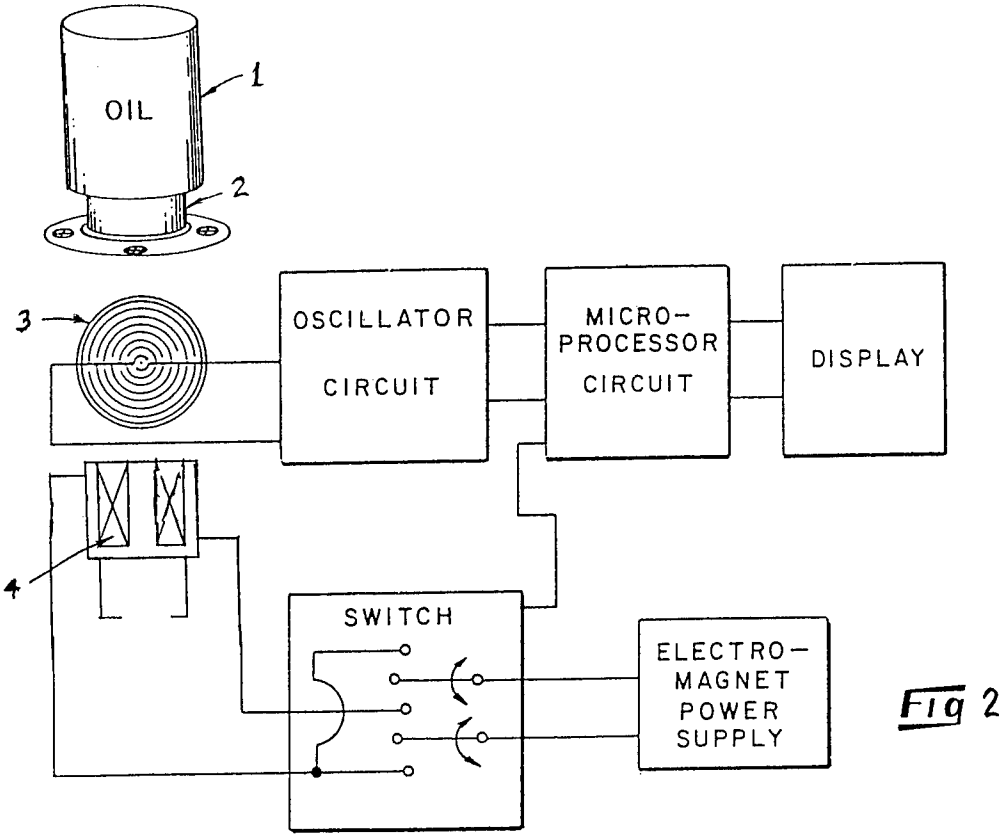

OIL
1
2
3
4
OSCILLATOR CIRCUIT
MICRO-PROCESSOR CIRCUIT
DISPLAY
SWITCH
ELECTRO-MAGNET POWER SUPPLY

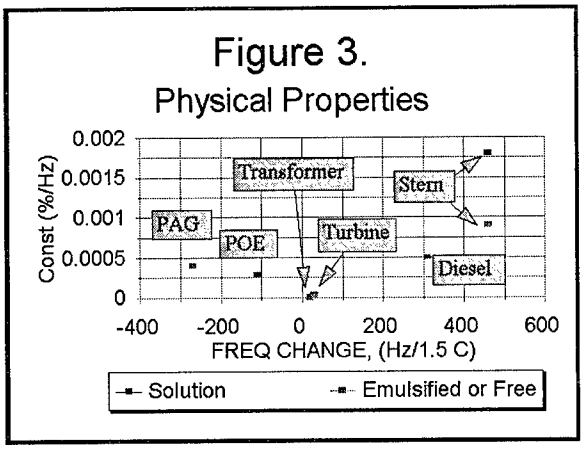
Fig 2

The apparatus which is used to estimate water in the oil employs a time-based measurement from a capacitive sensor at the bottom of an oil sample. See Figure 2. A sample of oil is poured into a wide mouth (item 1 on Figure 2) sample bottle (item 2). The sample is attached and sealed over a capacitive sensor (item 3) and then is shaken until the sample is homogenous. Then the sample is allowed to settle for a period of time, typically 500 seconds, while the microprocessor controls an oscillator circuit which tracks the natural frequency of the capacitive sensor.

The capacitive sensor (item 3) consists of two non-intersecting conductor traces. The substrate beneath the traces is fiberglass or another good dielectric material. The capacitance of the sensor (item 3) is a relative measure of the dielectric constant for the oil covering the space between the traces. A low relative dielectric constant for the oil produces a low capacitance for the sensor, which in turn causes a high natural frequency for the oscillator circuit. This is the case when new, uncontaminated oil is placed on the sensor - the frequency is relatively high because the oil is a very good insulator (e.g., it has a very low dielectric constant, <2.5). On the other hand, when water is present on the sensor, the capacitance increases and the frequency drops.

The electromagnet (item 4 in Figure 1), has a primary function assisting in the measurement of ferrous wear debris. This electromagnet also plays an important role with respect to water measurement -- it heats the reference oil approximately 1.5 C during the 500 second test time. This slight heating triggers a change in dielectric value as a function of temperature which can be used to identify the type of class of lubricant.

The abscissa (X-axis) on Figure 3 shows how the change in sensor frequency, labeled "frequency change," due to electromagnet heating (approximately 1.5 C) varies with oil type for mineral and synthetic oils. This fingerprint response is used to characterize reference oil samples with respect to general type or class of lubricant (e.g., PAG, POE, transformer, turbine/hydraulic, diesel engine, or stern oil).



Each class of lubricant has a different affinity for water due to its base oil and additive composition. The ordinate (Y-axis) on Figure 3 shows an empirically derived constants for water affinity of each oil type. This material constants are displayed with units of "Percent water per Hz" of sensor output. These constants, when multiplied by the measured frequency output, yield the water content in each of three physical states: solution, emulsified, and free.

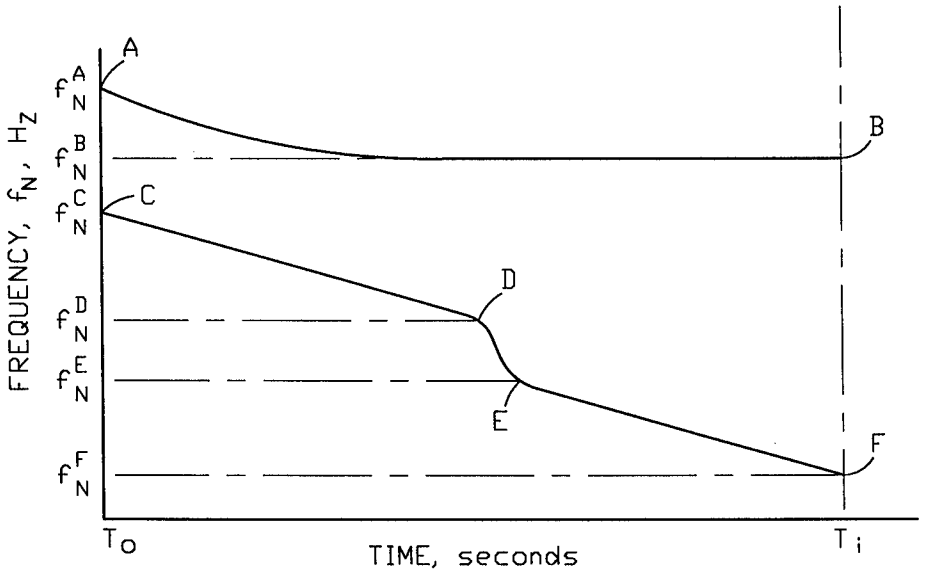

FREQUENCY, f_N, Hz
f_N^A
f_N^B
f_N^C
f_N^D
f_N^E
f_N^F
A
B
C
D
E
F
T_o
TIME, seconds
T_i

Fig. 4

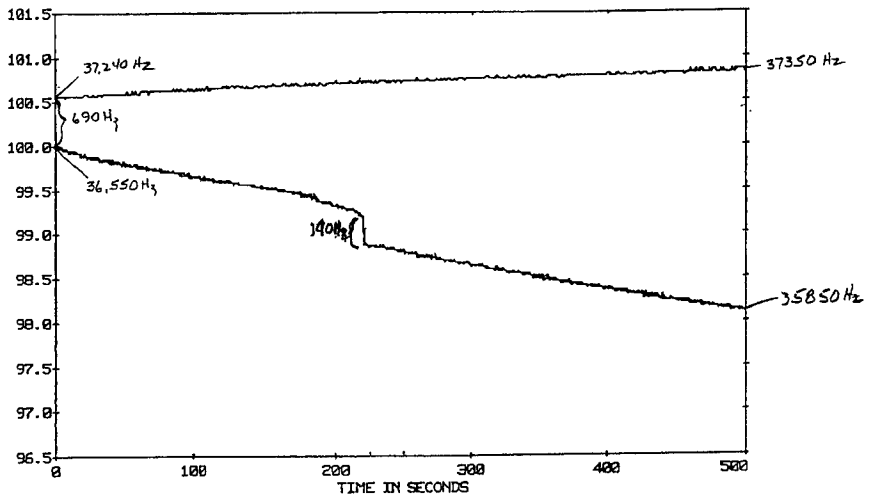

101.5
101.0
100.5
100.0
99.5
99.0
98.5
98.0
97.5
97.0
96.5
37,240 Hz
690 Hz
36,550 Hz
140 Hz
37350 Hz
35850 Hz
0
100
200
300
400
500
TIME IN SECONDS

Fig. 5

TABLE II

TEST	LUBRICANT	% H_2O		NATURAL FREQ. f, Hz		Δf, Hz	Δf, Hz	Δf, Hz	RATE % $H_2O/\Delta f$		ppm	H_2O, %	H_2O, %	H_2O, %	H_2O, %	Error H_2O, %
		Actual	Calc	Start	End	Step	Offset	Drop	Solution	Emul Free	Additive	Solution	Emul	Free	Sum	Diff
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	POE (1)	0	0	37240	37350	0	0	-110	28.6×10^{-5}	28.6×10^{-5}		0	0	0	0	0
2		0.04	0	37220	37240	0	-20	-20	28.6×10^{-5}	28.6×10^{-5}		0.00572	0.02574	0	0.03146	-0.00854
3		0.08	0	37200	37170	0	-40	30	28.6×10^{-5}	28.6×10^{-5}		0.01144	0.04004	0	0.05148	-0.02852
4		0.16	0.1	37010	36880	0	-230	130	28.6×10^{-5}	28.6×10^{-5}		0.06578	0.06864	0	0.13442	-0.02558
5		0.32	0.3	36550	35850	140	-690	700	28.6×10^{-5}	28.6×10^{-5}		0.19734	0.19162	0.04004	0.429	0.109
6		0.64	0.6	36110	34990	0	-1130	1120	28.6×10^{-5}	28.6×10^{-5}		0.32318	0.35178	0	0.67496	0.03496
7		1.28	1.2	35930	33130	0	-1310	2800	28.6×10^{-5}	28.6×10^{-5}		0.37466	0.83226	0	1.20692	-0.07308
8	PAG (2)	0	0	29920	30190	0	0	-270	40.0×10^{-5}	40.0×10^{-5}		0	0	0	0	0
9		0.04	0	29800	30070	0	-120	-270	40.0×10^{-5}	40.0×10^{-5}		0.048	0	0	0.048	0.008
10		0.08	0	29710	29990	0	-210	-280	40.0×10^{-5}	40.0×10^{-5}		0.084	-0.004	0	0.08	0
11		0.16	0.1	29540	29810	0	-380	-270	40.0×10^{-5}	40.0×10^{-5}		0.152	0	0	0.152	-0.008
12		0.32	0.2	29160	29410	0	-760	-250	40.0×10^{-5}	40.0×10^{-5}		0.304	0.008	0	0.312	-0.008
13		0.64	0.4	28520	28780	0	-1400	-260	40.0×10^{-5}	40.0×10^{-5}		0.56	0.004	0	0.564	-0.076
14		1.28	0.8	27070	27180	0	-2850	-110	40.0×10^{-5}	40.0×10^{-5}		1.14	0.064	0	1.204	-0.076
15	DIESEL	0	0	60960	60650	0	0	310	50.0×10^{-5}	50.0×10^{-5}	4000	0	0	0	0	0
16		0.1	0.05	60680	60300	0	-280	380	50.0×10^{-5}	50.0×10^{-5}	4000	0.14	0.035	0	0.175	0.075
17		0.5	0.41	60600	59530	0	-360	1070	50.0×10^{-5}	50.0×10^{-5}	4000	0.18	0.38	0	0.56	0.06
18		0.75	0.79	60050	58480	0	-910	1570	50.0×10^{-5}	50.0×10^{-5}	4000	0.455	0.63	0	1.085	0.335
19		1	0.89	60430	58690	0	-530	1740	50.0×10^{-5}	50.0×10^{-5}	4000	0.265	0.715	0	0.98	-0.02
20	TURBINE	0	0	59530	59500	0	0	30	2.0×10^{-5}	4.0×10^{-5}	1000	0	0	0	0	0
21		0.03	0.01	59240	58700	0	-290	540	2.0×10^{-5}	4.0×10^{-5}	1000	0.0058	0.0204	0	0.0262	-0.0038
22		0.13	0.13	59520	56110	0	-10	3410	2.0×10^{-5}	4.0×10^{-5}	1000	0.0002	0.1352	0	0.1354	0.0054
24	TRANSF	0	0	59720	59700	0	0	20	12.6×10^{-7}	13.0×10^{-7}	200	0	0	0	0	0
25		0.005	0.1	59690	55740	0	-30	3950	12.6×10^{-7}	13.0×10^{-7}	200	3.8×10^{-5}	0.004952	0	0.00499	-1.0×10^{-5}
26	STERN	0	0	50290	49830	0	0	460	18.0×10^{-4}	9.0×10^{-4}	6000	0	0	0	0	0
27		2.3	3	49090	47110	0	-1200	1980	18.0×10^{-4}	9.0×10^{-4}	6000	2.16	1.368	0	3.528	1.228
28		5.7	5.3	47970	45510	0	-2320	2460	18.0×10^{-4}	9.0×10^{-4}	6000	4.176	1.8	0	5.976	0.276
29		11	9.3	45400	42650	0	-4890	2750	18.0×10^{-4}	9.0×10^{-4}	6000	8.802	2.061	0	10.863	0.137
30		16	12.7	43180	39960	0	-7110	3220	18.0×10^{-4}	9.0×10^{-4}	6000	12.798	2.484	0	15.282	-0.718

Figures 4 and 5 show the sensor frequency vs. time plot for a clean reference oil sample (AB) and a contaminated used oil sample (CDEF). Water content is computed from the data reported in these curves in the following manner. Actual data which have been collected on a variety of oil types are reported in Table II. Values for starting frequencies, "A" and "C" are given in column D. Values for ending frequencies, "B" and "F" are given in column E. Step changes, "D-E", are given in column F. The example shown in Figures 4 and 5 is taken from test number 5 with POE (column A) having actual water content of 0.32% (column B). Actual water content was determined by using a precision micropipet to measure water added to reference oils for all data reported herein.

Measuring Moisture in Solution: Three elements are needed to compute moisture content in solution: (1) the identity of the lubricant type, (2) the water affinity constant for this type of lubricant, and (3) difference between the starting frequency of the new vs. used oil samples.

The identity of the lubricant type can be input by the individual doing the test, or it can be automatically determined using frequency difference, A-C, in conjunction with Figure 3. The water affinity constant for this lubricant type is then obtained from column "I" in table II (also graphically represented in Figure 3). Moisture in solution is then computed from the following:

Moisture in Solution = (A-C) x (Solution Constant)

Measuring Emulsified Water Content: Three elements are needed to compute emulsified water content: (1) the identity of the lubricant type, (2) the water affinity constant for this type of lubricant, and (3) the median-average slope for the curve CDEF.

The identity of the lubricant type can be input by the individual doing the test, or it can be automatically determined using frequency difference, A-C, in conjunction with Figure 3. The emulsified water affinity constant for this lubricant type is then obtained column "J" in Table II (also graphically represented in Figure 3).
Emulsified water content is then computed from the following:

Emulsified water content = Median Average Slope $\overline{\text{[CDEF]}}$ x (Emulsibility Constant)

Measuring Free Water Content: Three elements are needed to compute free water content: (1) the identity of the lubricant type, (2) the water affinity constant for this type of lubricant, and (3) cumulative step changes in the slope of the sensor output vs. time plot.

The identity of the lubricant type can be input by the individual doing the test, or it can be automatically determined using frequency difference, A-C, in conjunction with Figure 3. The free water affinity constant for this lubricant type is then obtained column "J" in Table II (also graphically represented in Figure 3). Moisture in solution is then computed from the following:

Free water content = (D-E) x (Free Water Constant)

Figure 6
STERN OIL - UP TO 16 % WATER

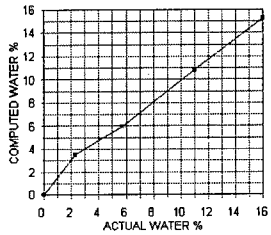


Figure 7
DIESEL OIL - UP TO 1% WATER

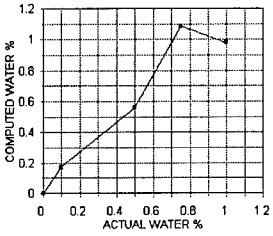


Figure 8
TURBINE OIL - UP TO 0.13% WATER

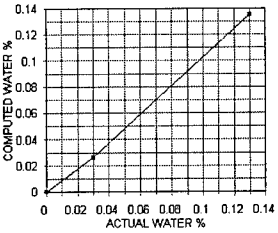


Figure 9
POE OIL - UP TO 1.28% WATER

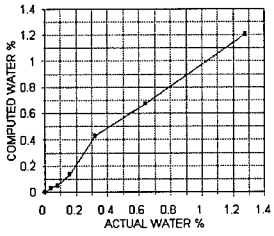


Figure 10
PAG OIL - UP TO 1.28% WATER

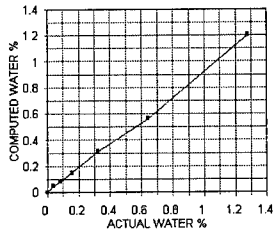
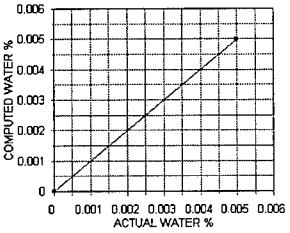


Figure 11
TRANSFORMERPAG OIL - UP TO 5 PPM WATER



Total Water Content: The total water content is simply the sum of the three parts: solution, emulsion, and free. Figures 6 to 11 show data measured on a variety of different oil types, comparing actual water content to calculated total water content. All of these data are reported in Table II.

Removal of water: There are a number of different actions one might take after detecting water in lubricated machinery:

- Do nothing
- Operate at high temperature
- Change the oil
- Centrifugal separation
- Vacuum dehydration
- Coalescing filtration.
- Bleed from bottom of oil compartment

There are advantages and disadvantages, costs and benefits to each of these approaches. The selection of the best approach depends on the application and form(s) of water to be removed.

The "do nothing" alternative may be the best short term action, at least until repair intended to prevent ingress of coolant or water can be accomplished. All too often, the first action after undesirable water is detected is to drain and replace the oil. This action can be a wasted if the cause of water contamination is not identified and corrected first.

Water is naturally driven off by heat. This method of water removal is only practical for high temperature operating systems with vented oil compartments. For example moisture, which is a product of combustion and condensation, is driven off when the engine operates at 100 C for a sustained period. The low vapor pressure of moisture in solution and emulsified water makes this makes this impractical for most industrial applications which operate below 50 C.

Changing the oil may be the lowest cost method to remove water, *after* the source of contamination has been eliminated. Naturally, this removes all three forms of water.

Centrifugal separation is an effective means for maintaining the cleanliness of the lubricant with respect to free water and most of the emulsified water. Centrifugal separators cannot remove moisture from solution which is chemically bound to additives or synthetic base stock.

Coalescing filters use a gel with high affinity to absorb water from the oil. These filters have limited water capacity and can be relatively expensive. They are best used to maintain low water contamination in systems which are not prone to ingress of very much water.

Finally, many lube systems are designed to demulsify free water in the oil compartment so that it can be bled off periodically. This approach works very well unless oil additives are used which support emulsified water. These oil compartments tend to be relatively large. The supply side of the oil compartment should be tested to verify absence of free and emulsified water going back into the lubricated machinery.

Conclusion: Moisture contamination of lubricants, hydraulics, and transformer oils lead to a variety of problems, all of which shorten machine life through corrosion, electrochemical interaction, and greatly accelerated wear rates. Water combines with the oils in the form of a solution, an emulsion, or a mixture of free water in oil.

The time-based capacitance apparatus was demonstrated to have the ability to measure the quantity of water in each of these forms. Moisture in solution is quantified by comparing the dielectric constant of a vigorously shaken contaminated oil sample with a clean oil sample. Emulsified water is quantified by determining the settling rate of moisture laden oil within this sample. Free water is quantified by monitoring the discrete step-changes in time as drops fall to the sensor surface.

By monitoring these three forms of water contamination in oil, industrial maintenance departments are able to take proactive measures to preclude the damaging effects water has on machinery.

RAPID, PORTABLE VOLTAMMETRIC TECHNIQUES FOR PERFORMING ANTIOXIDANT, TOTAL ACID NUMBER (TAN) AND TOTAL BASE NUMBER (TBN) MEASUREMENTS

R. E. Kauffman
University of Dayton Research Institute
300 College Park
Dayton, OH 45469-0161

Abstract: This paper describes rapid voltammetric techniques for determining the antioxidant, total acid number (TAN), and total base number (TBN) measurements of a wide variety of oils and fluids. The voltammetric techniques require less than 1 milliliter (mL) of oil, 5 mL of acetone or ethanol solution, and less than 2 minutes for each analysis and can are performed in vials using inexpensive, portable instruments. In contrast to currently used TAN (ASTM Method 664) and TBN (ASTM Method 2896 or 4739) techniques, the voltammetric techniques do not involve titrations, and consequently, do not rely on endpoints which are operator dependent. Once prepared, the oil/analysis solution mixture can be analyzed several times (20 seconds per repeat analysis) to ensure the accuracy of the voltammetric analyses. The current ASTM methods require resampling followed by a new titration (requiring several minutes per titration) for duplicate analyses.

This paper also presents results for the voltammetric and ASTM method (TAN and TBN) analyses of various new, laboratory stressed, and authentic used fluids obtained from different types of normally and abnormally operating equipment. The evaluated oils and fluids include diesel engine oils obtained from test stands, phosphate ester based hydraulic fluids obtained from commercial airliners, and ester based turbine engine lubricating oils obtained from military and commercial aircraft. The results presented herein demonstrate that combined knowledge of the antioxidant/TAN/TBN (if applicable) measurements improves the capability of the analyst or equipment operator to evaluate the remaining useful life of the tested fluid as well as the condition of the operating equipment. Interrelationships among the different voltammetric analyses and wear debris analyses of the used oils and fluids are demonstrated and discussed.

Key Words: Antioxidant, Oil Analysis, Condition Monitoring, Remaining Useful Life, Total Acid Number, Total Base Number, Voltammetric Analysis

INTRODUCTION: Organic fluids are used in a wide range of applications in which the fluid experiences thermal/oxidative stresses. Depending on the application, different types of compounds are added to the fluid to enhance its performance. One type of fluid uses esters of

carboxylic acids fortified by antioxidants (aromatic amines) and antiwear (tricresyl phosphate) additives to meet the lubrication requirements of aircraft gas turbine engines. Another type of fluid uses esters of phosphic acid fortified by corrosion inhibitors (amines) to meet the non-flammability, lubrication requirements of commercial aircraft and industrial hydraulic systems. However, the majority of fluids use long chain hydrocarbons (petroleum or synthetic) fortified by antioxidants (hindered phenols), multifunctional antiwear additives (metal and non-metal dithiophosphates) and corrosion inhibitors (amines) to meet the lubrication requirements of a wide range of industrial machinery and hydraulic systems. Compounds (calcium carbonate as well as overbased detergents and antiwear additives) are added to oils used in diesel and gasoline combustion engines (to neutralize sulfur and nitrogen acids caused by combustion of fuel) in addition to the antioxidant/antiwear/anti-corrosion additives.

During use in the operating equipment, the additives deplete at a rate dependent upon the environment experienced by the fluid [1]. For instance, the R&O (rust and oxidation inhibited) oils used in industrial gas and steam turbine engines are typically formulated with hindered phenol antioxidants and amine corrosion inhibitors. The hindered phenol antioxidants deplete at a faster rate than the corrosion inhibitors in the gas turbine system (thermal-oxidative environment) but deplete at a slower rate than the corrosion inhibitors in the steam turbine system (lower temperature, water present). Another example of competing depletion mechanisms occurs for the hindered phenol antioxidants and zinc dithiophosphate (ZDTP) antiwear additives used in hydrocarbon based hydraulic fluids. The phenol antioxidants deplete faster than the ZDTP antiwear additives in the thermal-oxidative environment of precision machining equipment (thermal-oxidative, low friction) but deplete slower than the ZDTP additive in high pressure pumps (thermal-oxidative, high friction).

Once the additive(s) decrease to a critical level, the lubricating/corrosion inhibiting/flowing capabilities of the fluid begin to decrease at an accelerated rate resulting in equipment damage, and possibly, equipment failure. To avoid equipment damage caused by fluid overusage, most condition monitoring programs detect fluid degradation using the total acid number (TAN) measurement technique (ASTM Method 664) which measures the accumulation of acids from fluid oxidation (fuel combustion) and hydrolysis. In the case of combustion engine oils, condition monitoring programs also employ total base number (TBN) measurements (ASTM Methods 664, 2896 or 4739) to determine the fluid's ability to neutralize corrosive combustion acids. In addition to fluid degradation, programs use analytical techniques to measure trace metal contents (additives, wear metal debris, dissolved metal from corrosion, etc.) of oil samples using atomic absorption (AA) or atomic emission (DCP or ICP) spectrometric techniques. [The discussion of metal analyses of fluids will be minimal in this paper except for wear metal debris increases caused by additive depletions.]

The main drawback of the TAN technique is its inability to predict the operating time from when the analyzed fluid was sampled until a fluid change will become necessary due to additive depletion, i.e., TAN techniques cannot predict the remaining useful life of a fluid. In addition to being non-predictive, the TAN analyses are affected by the operating conditions of the equipment (carboxylic acids from accelerated fluid oxidation volatilize) or by the additives. For instance, new fluids containing ZDTP antiwear additive have TAN values of 1-2 mg KOH/g of oil which

decrease (additive depletion) upon initial use, level off during further use (additive depletion compensated by oxidation acid increase), and then increase rapidly (oxidation acids accumulating). Consequently, measured increases in the TAN of a sampled oil may be due to fluid degradation (fluid being overused) or fluid replenishment (fluid being underused).

In contrast to fluid degradation techniques, analytical techniques which determine the additive (antioxidant, antiwear, anticorrosion) concentration of used oil samples are able to predict the operating time (assuming minimal fluid make-up) at which the additive will be totally depleted. However, the concentration at which the additive becomes ineffective (acid accumulation, wear debris or trace metal increase, etc.) will depend on the operating conditions of the fluid application. Historical relationships between additive concentration and fluid degradation need to be established so that percent remaining useful life (% RUL) plots can be established for each fluid application.

If fluid degradation does not occur until the additive is totally depleted, then the % RUL will be equal to the percent remaining antioxidant level. However, the 0% RUL (rapid degradation occurs) usually corresponds to 20-40% remaining antioxidant. In most cases the higher the temperature application, the larger the difference between 0% RUL and total antioxidant depletion. Also, other mechanisms such as hydrolysis, depletion of overbased additive, overheating (thermal) and so on may cause fluid degradation without affecting the antioxidant concentration.

Therefore, an oil analysis procedure which measures additive levels (% RUL) as well as fluid degradation [TAN and TBN (if applicable)] will afford the most accurate assessment of new and used fluids (and in numerous cases provide information on equipment health) regardless of the application. The combined additive/degradation measurement has the following capabilities:

(1) Quality assurance of incoming batches of lubricants - check type and concentration of additives, compare antioxidant packages of different manufacturers, etc.

(2) Determine correct oil change intervals - determine the depletion rate of the antioxidants and the antioxidant concentration at which the fluid degradation rate dramatically increases to predict correct oil change intervals for each piece of operating equipment.

(3) Extend oil change intervals - determine the depletion rate of individual antioxidants so that the selected antioxidants can be replenished to extend the useful life of the monitored lubricant.

(4) Detect abnormally operating equipment - once the normal depletion rates of the antioxidants have been established, accelerated antioxidant depletion rates can be used to detect abnormally operating equipment.

(5) Monitor non-oxidative degradation mechanisms such as hydrolysis, overheating, and acid accumulation from combustion gases.

TECHNOLOGY DESCRIPTION: Although, numerous long-term oxidation tests are available to perform antioxidant depletion (% RUL) measurements of different type fluids, they are unsuitable for routine use. Voltammetric techniques [1-4] are capable of performing rapid (less than 1 minute) antioxidant analyses but are unable to perform TAN or TBN measurements. Therefore, research was performed to develop TAN and TBN methods that could be performed on small, portable voltammetric instruments designed to perform antioxidant analyses.

The main objections to performing TAN (ASTM 664) and TBN (ASTM 2896 and 4739) methods are the methods require time consuming titrations with hazardous solvents (toluene, chlorobenzene, acetic acid, etc.). The TAN and TBN titrations are performed until an endpoint (rapid change in pH or color change with single drop of titrant) or predetermined pH value (established with buffers) is reached. Since the endpoint or predetermined pH value is not always distinct or stable, the TAN and TBN values tend to be highly operator dependent, especially TBN measurements of used oils. Also, in the case of ester based fluids, chemical reactions may occur between the basestock and titrant leading to superficially high TAN and TBN numbers.

The main basis for the voltammetric TAN technique is the use of electroactive (detected by voltammetry) compounds that become non-electroactive (not detected by voltammetry) in the presence of accumulated acids (TAN method). For the TBN technique, insoluble compounds which become soluble and electroactive in the presence of hydrochloric acid (TBN titrant) were used. In the voltammetric TAN and TBN techniques, the concentration of the electroactive species is determined before and after adding, then shaking the oil sample with the voltammetric solution. The difference in electroactive species concentration and the amount of oil added is then used to calculate the TAN or TBN of the oil sample. Since the voltammetric analysis detects (does not titrate) the concentration of the electroactive species, multiple analyses can be made on the same prepared solution to increase the accuracy of the voltammetric TAN and TBN results. The ASTM methods require resampling of the oil to perform duplicate analyses increasing the time requirements and waste of the TAN and TBN methods. In addition to other advantages, the voltammetric TAN and TBN techniques use smaller quantities (5 mL vs up to 100 mL) of less toxic solvents (ethanol or acetone vs toluene, chlorobenzene, acetic acid) compared to the ASTM methods.

This paper also presents results for the voltammetric and ASTM method (TAN and TBN) analyses of various new, laboratory stressed, and authentic used fluids obtained from different types of normally and abnormally operating equipment. The evaluated oils and fluids include diesel engine oils obtained from test stands, phosphate ester based hydraulic fluids obtained from commercial airliners, and ester based turbine engine lubricating oils obtained from military and commercial aircraft. The results presented herein demonstrate that combined knowledge of the antioxidant/TAN/TBN (if applicable) measurements improves the capability of the analyst or equipment operator to evaluate the remaining useful life of the tested fluid as well as the condition of the operating equipment. Interrelationships among the different voltammetric analyses and wear debris analyses of the used oils and fluids are demonstrated and discussed.

EXPERIMENTAL: Lubricating Oils. The used lubricating oils and fluids analyzed in this paper were obtained from normally and abnormally operating equipment as well as laboratory stressing tests and diesel engine test stands. Several new oils and fluids were also analyzed. The types of oils and fluids analyzed included: aircraft gas turbine engine oils (carboxylate ester based), aircraft hydraulic fluids (phosphate ester based) and diesel engine oil (hydrocarbon based).

ASTM TAN and TBN Methods. The ASTM TAN and TBN results reported here in were performed by industrial oil analysis laboratories from which the used lubricants and fluids were obtained.

Voltammetric Additive Analyses. The voltammetric method was performed with a commercially available voltammograph equipped with a digital readout, stripchart output and electrode system (glassy carbon working electrode, a platinum wire reference electrode, and a platinum wire auxiliary electrode). A fresh oil typical of the application (100% standard) and the solvent system (0% standard) were used to calibrate the voltammetric instrument for % remaining antioxidant determinations.

The oil samples (100 - 500 µL) were diluted with acetone or ethanol containing a dissolved electrolyte and a suspended solid substrate. For aromatic amine type antioxidants the solvent system was acetone and the electrolyte was a neutral salt. For hindered phenol-type antioxidants and phenol degradation products the solvent system was water/ethanol and the electrolyte was an inorganic base. For zinc dialkyldithiophosphate type additives the solvent system was a water/acetone solution and the electrolyte was a neutral salt. When the oil/solvent/solid substrate mixture was shaken, the insoluble oil coated the solid substrate and upon standing the agglomerated particulates quickly settled out to produce a clear solution for analysis. The voltage of the auxiliary electrode was scanned from 0.0 to 1.0 V at a rate of 0.5 V/second. The height of the peaks produced by the voltammetric method were then used to evaluate the remaining additives of the used oil samples.

Voltammetric TAN Analysis. The voltammetric method was performed with the same voltammetric analyzer described for voltammetric additive analysis. The system was calibrated for TAN measurements as shown in Figure 1. The blank water/ethanol solution containing overbased phenol or phenyl phenol) was analyzed with the described voltammetric analyzer to produce the first peak (due to overbased phenol) in Figure 1. This peak height (analyzer reading = 520) represents a TAN value of 0.0 mg KOH/g of oil. To show the repeatability of the test, the same blank was reanalyzed to produce the second peak in Figure 1 (analyzer reading = 515). To the same blank solution, 50 µL of a 0.1N HCl in water solution was added and shaken for 3 seconds. The third peak in Figure 1 (analyzer reading = 140) represents the overbased phenol unreacted by the added acid and corresponds to a value of 0.28 mg of KOH. When the same experiment was repeated with a second blank and 50 µL of a 0.1N octanoic acid in water solution, the vial had to be shaken for 10 seconds to get complete reaction (peak decreased on second analysis after additional shaking). The peak produced by the octanoic acid was similar to the third peak in Figure 1 and had an analyzer reading of 130.

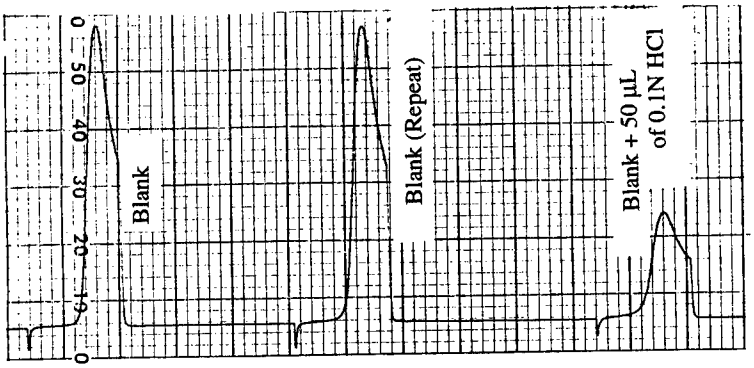


Figure 1. Voltammograms produced by analysis of the blank, repeat analysis of the blank, and analysis of the blank after addition of 50 µL of 0.1N HCl in water solution.

Therefore, the voltammetric TAN detects strong and weak acids equally well and can be used for qualitative comparisons of acid strength, i.e., if peak height obtained after 3 seconds of shaking does not change with additional shaking all of the detected acid is strong (mineral acid), if the peak height decreases substantially upon additional shaking a weak acid (carboxylic acid) is present. To calculate the TAN of a fluid sample using the data from Figure 1 the following equations are used:

$$\text{Calibration} = \frac{50\mu\text{L of 0.1 N acid}}{\text{Blank Reading - Acid Reading}} = \frac{0.28 \text{ mg of KOH}}{520-140} = 7.4 \times 10^{-4} \frac{\text{mg of KOH}}{\text{readout unit}}$$

$$\text{TAN (mg of KOH / g of oil)} = \frac{\text{Blank Reading - Sample Reading}}{\text{Weight of Sample}} \times 7.4 \times 10^{-4} \frac{\text{mg of KOH}}{\text{reading unit}}$$

Voltammetric TBN Analysis. The voltammetric method was performed with the same voltammetric analyzer described for voltammetric additive analysis. The system was calibrated for TBN measurements in a similar manner to the TAN method. The blank was prepared by dispensing the reaction solution (water/ethanol solution containing HCl acid) into a vial containing copper powder. The solution/powder mixture was shaken for 10 seconds and allowed to sit undisturbed for 5 seconds to allow the powder to settle. The solution was then analyzed with the described voltammetric analyzer. The produced peak(analyzer reading = 1550) represents a TBN value = 0 mg of KOH/g of oil. The TBN standard was then prepared by dispensing the reaction solution and an oil (200 µL) of known TBN (usually TBN = 13) into a vial. The reaction solution/oil mixture was then shaken for 20 - 30 seconds and allowed to sit undisturbed for 5 seconds. Since ZDTP species produce peaks in the same area of the TBN method, the solution is analyzed for additive content and to establish the baseline for the TBN calculations. The reaction solution is then poured into a vial containing copper powder and

shaken for an additional 10 seconds to react any remaining HCl. The vial is allowed to sit undisturbed for 5 seconds then analyzed to produce the peak representative of a TBN value = 13 mg of KOH/g of oil (standard).

The oil samples to be analyzed for TBN are then analyzed in a manner similar to the calibration standard provided the samples' TBN values are below 13. For samples with TBN values between 13 and 52 mg of KOH/g, 50 µL samples are used. The TBN value of the oil sample is then calculated in the following manner:

$$\frac{\text{TBN of Standard}}{\text{Blank reading - Standard reading}} \times \frac{200\ \mu L}{\text{Oil sample size}(\mu L)} = \frac{\text{TBN Value}}{\text{Unit Reading}}$$

$$\text{TBN value of oil sample (mg KOH/g of oil)} = (\text{Sample reading - Baseline reading}) \times \frac{\text{TBN Value}}{\text{Reading Unit}}$$

RESULTS AND DISCUSSION: To initially evaluate the capabilities of the combined voltammetric additive/TAN/TBN measurements a wide range of new and used oils and fluids were analyzed. The used analyzed oils and fluids were obtained from laboratory engine test stands and stressing tests as well as normally and abnormally operating equipment.

Diesel Engine Oils. Of the numerous applications evaluated for this paper, the combined voltammetric analyses have the greatest potential benefit for diesel engine oils. As opposed to other analytical techniques, voltammetric analyses are not limited by the soot levels found in most used diesel engine oils. To evaluate the voltammetric additive/TAN/TBN potential for analyzing diesel engine oils, used engine oils were obtained from an extended IIIe engine test (test usually terminated at 48 hours). The voltammetric additive analyses and the ASTM TAN/ TBN measurements were made by the laboratory personnel performing the engine test. The voltammetric TAN/TBN measurements were performed by the author. The resulting voltammetric and ASTM measurements of the new and used diesel engine oils were plotted versus engine operating time as shown in Figure 2. The results in Figure 2 show that the voltammetric and ASTM TAN measurements are in good agreement with the voltammetric values being slightly lower during the useful life of the oil (TAN values stable). The lower voltammetric TAN values are attributed to the fact that the voltammetric test detects the overbased additives (ZDTP,detergents,etc) to a lesser extent than the ASTM method. The results in Figure 2 also show that the voltammetric and ASTM TBN measurements are also in good agreement. The difference between the TBN tests after 48 hours is attributed to the fact that the voltammetric test is unaffected by the organic acids produced by oxidation but the ASTM method is buffered by the produced organic acids resulting in superficially high values.

In addition to the TAN/TBN measurements, the results in Figure 2 show that the voltammetric ZDTP analyses are the only analyses capable of determining the remaining useful life (RUL) of the oil being tested in the engine test. The oxidative RUL of the oil ends (TAN increases) when the ZDTP depletes to approximately 15% of its original concentration. A dramatic increase in the Fe concentration occurred (not plotted) at a ZDTP level of 10%, indicating the initiation of accelerated wear in a diesel engine can be predicted by the voltammetric ZDTP analyses.

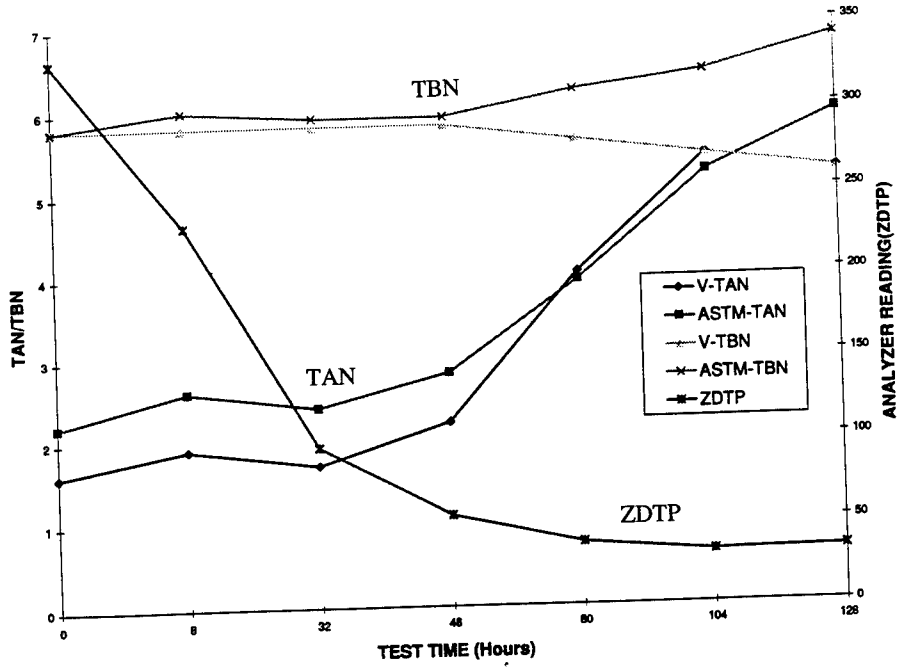


Figure 2. Plots of voltammetric and ASTM analyses versus engine test time for used diesel engine oils obtained from an extended IIIe engine test.

Therefore, the results in Figure 2 as well as additional evaluations [5] of used oils obtained from normally operating diesel and gasoline engines demonstrate that the combined voltammetric additive/TAN/TBN measurements provide the user with on-site capabilities to monitor both the RUL and condition of used diesel engine oils with minimal time and cost requirements. The results in Figure 2 also indicate that the voltammetric TBN values are more accurate than the ASTM TBN values for used diesel engine oils. The accuracy of voltammetric TAN results can be improved by subtracting out the additive concentrations detected by the voltammetric additive analyses.

Aircraft Turbine Engine Oils. As opposed to the diesel engine oils, aircraft turbine engine oils operate in a clean environment and are not subject to combustion acids, coolant leaks, soot, and so on. Therefore, the aircraft oils are not formulated with overbased detergents and calcium carbonates, and consequently, do not have a meaningful TBN value.

Previous research [2-4] has shown that voltammetric additive (aromatic amine antioxidants) analyses are able to detect abnormally operating aircraft engines. In particular, C-130 transport and commercial airliner engines experiencing severe oxidation due to cracked seals were detected by voltammetric additive analysis prior to accelerated wear and component failure. In

fact, the results from the prevoius studies indicated that oil degradation was the main failure mechanism for C-130 engines explaining the inability of spectrometric oil analysis to detect wear prior to engine failure [6].

In contrast to C-130 transport oil degradation mechanisms, oil samples from military fighters and commercial airliners have shown indications of overheating (e.g., F-16 black oil samples). The overheating at a hot spot causes coke formation with only minor decreases in the antioxidant level of the bulk oil. Consequently, additive analysis is ineffective in detecting the coking process. If allowed to continue without detection, the coke buildup will eventually break lose clogging oil jets and other small orifices resulting in decreased oil flow and lubricant starvation of the bearings. Previous studies [2] have shown that the conductivity of the oil increases dramatically during the coking process indicating polar species, possibly acidic species, are produced by the coking process.

Therefore, twenty oil samples obtained from normally and abnormally operating commercial airliner engines were analyzed for additive and TAN using the voltammetric analyzer. The results in the table below indicate that the majority (15 samples) of the oil samples have antioxidant levels greater than 90% and TAN values below 0.3 mg of KOH/g. The high oil makeup rates of the aircraft engines account for the high additive, low TAN measurements. Of the other 5 samples, two oils show significant additive depletion with minimal TAN values; these samples are in the early stages of accelerated oxidation (cracked seal) but have not reached rapid acid accumulation, i.e., % RUL > 0 (Figure 2) [records were not available for confirmation]. The other three oils show significant TAN value increases without significant antioxidant depletion; these samples are experiencing varying degrees of overheating. Conductivity [2] measurements detected overheating in samples 18-20 and airline maintenance personnel confirmed maintenance action had been required.

COMPARISON OF VOLTAMMETRIC ADDITIVE
AND TAN MEASUREMENTS FOR USED COMMERCIAL AIRLINE ENGINE OILS

Sample No.	% Antioxidant Remaining	TAN Value (mg KOH/g oil)
1-15	92-98	0.13 - 0.28
16-17	56-70	0.38, 0.31
18-20	81-85	1.1 - 2.8

Therefore, the results in the above discussion indicate that combined voltammetric additive/TAN measurements enables the user to better evaluate the RUL and condition of used aircraft oils by detecting two kinds of oil degradation mechanisms; accelerated oxidation (additive depletion) and coking (TAN increase). Long-term storage of aircraft oils can cause hydrolysis of the ester basestock which produces organic acids, and consequently, can also be detected by voltammetric TAN measurements.

Phosphate Ester Based Hydraulic Fluids: In contrast to hydrocarbon based hydraulic fluids, the fluids used in commercial aircraft hydraulic systems have phosphate ester basestocks. The

phosphate esters are used in aircraft hydraulic systems because they are fire resistant in comparison to hydrocarbon oils. Also, the phosphate esters don't require antiwear additives (e.g., ZDTP) due to their inherent lubricity. However, the phosphate esters undergo hydrolysis, oxidation, and thermal degradation to produce corrosive acids and phenolic compounds. Contamination of the hydraulic fluids with turbine engine oils is also a concern for operators of aircraft.

To evaluate the capabilities of the voltammetric analyses for phosphate ester hydraulic fluids, fresh fluids and authentic used fluids obtained from commercial aircraft hydraulic systems were analyzed. The voltammograms produced by two widely used phosphate ester hydraulic fluids and a suspect used hydraulic fluid are shown in Figure 3. The results in Figure 3 show that the two fresh fluids produce distinctly different spectra and that one of the used hydraulic fluids appears to be aircraft oil. The voltammetric analyses of the used phosphate esters showed that the amines decreased, and the phenolic degradation compounds increased dramatically, with use. The voltammetric analyses were in good agreement with the total acid number (ASTM Method D-664) and trace metal (e.g., Fe, Cu, etc.) measurements of the used fluids.

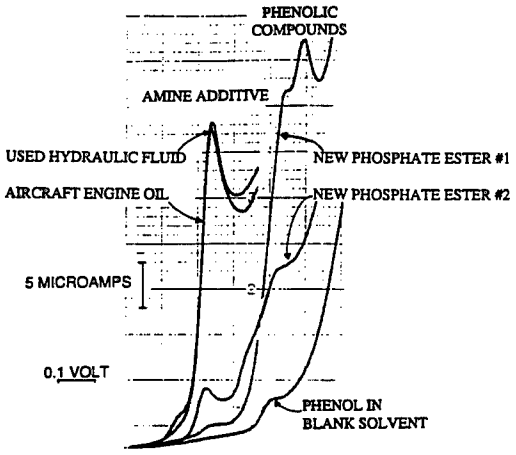


Figure 3. Voltammograms of two fresh and one suspect used phosphate ester hydraulic fluid in comparison to fresh aircraft engine oil.

To evaluate the voltammetric TAN method, the used fluids were analyzed and the results are plotted in Figure 4 versus the ASTM TAN values. The results in Figure 4 show that the voltammetric TAN values are in good agreement with the ASTM TAN values once an offset of 0.6 mg KOH is employed. The higher voltammetric TAN values for the phosphate ester fluids are caused by the phenolic species (initially present and produced during use) detected in the fluids (Figure 3) and can be compensated using the voltammetric additive analyses.

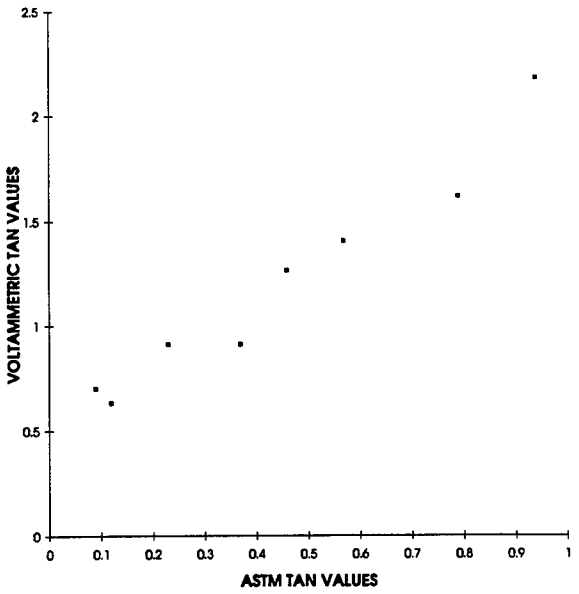


Figure 4. Plots of voltammetric TAN value versus ASTM TAN value for fresh and used phosphate ester hydraulic fluids obtained from a normally operating aircraft hydraulic system.

Therefore, these results indicate voltammetric additive/TAN measurements have great potential for monitoring the RUL and condition of phosphate ester hydraulic fluids. Whether monitoring phenol degradation products detected by voltammetry are of importance to operating efficiency (responsible for deposits on servo values) or for adjusting the voltammetric TAN values has yet to be determined.

SUMMARY: The initial results presented in this paper demonstrate that combined knowledge of the additive/TAN/TBN (if applicable) measurements improves the capability of the analyst or equipment operator to evaluate the RUL and condition of the tested fluid as well as the condition of the operating equipment. The results indicate that for diesel engine oils depletion of the ZDTP additive and for aircraft engine oils (e.g., C-130 aircraft oils) depletion of the antioxidant (and resulting basestock degradation) allow accelerated wear to occur. Consequently, voltammetric analyses can be used to aid in the determination of the wear mechanisms detected by spectrometric oil analysis programs. Present work with the combined additive/TAN/TBN analysis procedure includes industrial steam and gas turbine oils, mineral oil based hydraulic fluids, automotive engine oils and transmission fluids, water based metal working fluids and fuels (jet, diesel engine, and gasoline).

References

1. Kauffman, R. E., "Remaining Useful Life Measurements of Diesel Engine Oils, Automotive Engine Oils, Hydraulic Fluids, and Greases Using Cyclic Voltammetric Methods," Lub. Eng. 51(3), pp 223-229 (1995).

2. Kauffman, R. E., "On-Line and Off-Line Techniques for Monitoring the Thermal and Oxidative Degradation of Aircraft Turbine Engine Oils - Part 1: Laboratory Evaluations," Lub. Eng., 51(11), pp 914-920 (1995).

3. CRC Handbook of Lubrication and Tribology, Volume III Monitoring, Materials, Synthetic Lubricants, and Applications," Rapid Determination of Remaining Useful Lubricant Life, pp 89-100, Editor E. Booser, CRC Press, Ann Arbor MI (1994).

4. Kauffman, R. E. and Rhine, W. E., "Assessment of Remaining Lubricant Life," Report No. AFWAL-TR-86-2024, November 1986.

5. Kirkpatrick, J. F., "Diagnostic Tools for Reciprocating Engine Cogeneration Systems, Phase I," Gas Research Institute Report No. GRI-90/0022, January 1990.

6. Hancock, D. O. and Synovec, R. E., "Early Detection of C-130 Aircraft Engine Malfunction by Principal Component Analysis of the Wear Metals in C-130 Engine Oil," Appl. Spect. 43(2), pp 202-208 (1989).

A NEW ELECTRONIC VISCOMETER BASED ON RAYLEIGH WAVE MECHANICS

James J. Kauzlarich, Robert A. Ross, & D. Sharif Abdallah
University of Virginia
School of Engineering & Applied Science
Thornton Hall
Charlottesville, VA 22903-2442

Abstract: The operation of a new viscometer is shown to be based on Rayleigh wave mechanics, where a thin walled tube of 5 mm diameter is immersed to a fixed depth of 60 mm in a fluid and is caused to undergo axial vibrations due to a vibrating piezoelectric plate which supports the tube. The motion of the tube is damped by the fluid, and the piezoelectric crystal feedback circuit gives a voltage inversely proportional to the square root of the viscosityxdensity product. The design and performance characteristics of the viscometer are described. The device is under patent application.

The instrument is portable and gives a measurement of viscosityxdensity within 1 minute over a wide range, and can be modified to permit continuous measurements as a function of time. The results are accurate for a Newtonian fluid and give an apparent viscosityxdensity for other fluids. A probe measures the fluid temperature during the test. Only a small amount of fluid is necessary for measurement. The Rayleigh theory is shown to be obeyed by the instrument.

The electronic circuit uses a phase-locked loop to drive the thin walled tube at resonance, about 1.5 kHz. The processed feedback signal is large, varying from 2 volts down to about 0.5 volts DC when measuring SAE 10 to 60 weight motor oils. There is an onboard computer to process the data. The computer gives a prediction of viscosity at 40 °C, or can be modified to predict viscosity at any reference temperature, for a paraffinic lubricating oil. For a particular fluid it is possible to build in data reduction in the computer. For example, equations for density as a function of temperature can be stored in the computer to reduce the viscosityxdensity product to viscosity and to density.

Key Words: Viscometer; Rayleigh Wave Mechanics; Electronic Phase-Locked Loops.

Background: The viscosity of a fluid is the most important property when considering lubricated bearings. It is the viscosity which makes it possible for the rotating shaft to drag the lubricant into the narrowing gap of the journal bearing. As the gap becomes narrower the pressure must increase, and this effect is used to support the rotating journal of the journal bearing to carry the load without metal to metal contact. In 1687 Sir Isaac Newton formalized the concept of viscosity. His postulate can be described mathematically, as the

resistance to shear stress (τ) of a fluid is proportional to the velocity gradient (du/dy; also called shear rate since it has dimensions of sec^{-1}), thus

$$\tau = \eta \frac{du}{dy} \qquad (1)$$

Fluids for which the coefficient of viscosity, η (N-s/m^2), is essentially independent of the shear rate are called newtonian fluids, but since the coefficient of viscosity is a truly thermodynamic property of a fluid, it is a function of the pressure and temperature. For many hydrodynamic bearings the pressure does not change the coefficient of viscosity (usually just called the viscosity) much, but changes in temperature have a large effect on the viscosity. For mineral oils, at atmospheric pressure, the viscosity changes by an order of magnitude for a change in temperature of 40 °C around room temperature. Although mineral oils are essentially newtonian fluids, greases, paints, and vegetable oils may have limited ranges of shear rate where they act newtonian.

There are a great number of viscometers on the market (see Van Wazer, et.al. [1]), but most of them can be classified by operation into a few categories. One of the earliest devices was a Falling Ball Viscometer. This device predicts the viscosity by measuring the terminal velocity of the ball in the fluid, where the gravitational force is balanced by buoyant force (Archimedes 285-212 B.C.) and viscous drag force (Stokes 1901).

The Saybolt Viscometer is another very simple device. The flow of the sample fluid through an orifice at the bottom of a cup is established under gravity, and the time to collect 60 ml of fluid is called Saybolt seconds. Saybolt seconds can be converted to kinematic viscosity (absolute viscosity divided by density) using an empirical equation. The original design was by H. G. Saybolt, Standard Oil Co., and the first published description was given by W. H. Herschel in 1918. This method is no longer supported by ASTM Standards, and a glass capillary viscometer has succeeded in its place.

The Capillary Tube Viscometer is supported by ASTM Standards D445 and D446. This system operates by measuring the time a known quantity of fluid passes through a capillary tube, and the result can be related to kinematic viscosity. A scientific understanding of the flow in a capillary was first presented by Hagen in 1839.

The Rotating Drum Viscometer gives the absolute viscosity of a fluid by measuring the drag of the fluid on the drum. This is based on Couette's viscometer devised in 1890.

All of these devices usually operate at or near atmospheric pressure, and measure or control the temperature of the test fluid.

In 1947 a new class of viscometer was constructed based on Rayleigh wave mechanics. Mason [2] has shown that a high-frequency torsionally oscillating crystal in a test fluid will generate viscous waves. Both a reactive and resistive loading occurs in the crystal which lowers the frequency and raises the resistance at resonance. The viscosity times density

product of a fluid can be determined by measuring the changes in the properties of the crystal with respect to an essentially inviscid fluid (air). Since the shear rate varies during the cyclic oscillation, correct values of viscosity are only obtained for newtonian fluids. An apparent viscosity results for non-newtonian fluids. Apparent viscosity is especially useful to follow chemical changes in a fluid, such as a polymerization process.

Woodward [3] introduced a vibrating plate viscometer in 1953, where a thin plate immersed in a fluid is vibrated in its plane. Shear waves are set up in the liquid, and the liquid exerts a retarding force on the plate consisting of a resistive and reactive component. The retarding force is inversely proportional to the square root of the absolute viscosity times density product of the fluid and is measured in comparison to an essentially inviscid fluid (air).

The subject of this paper is a new device to measure viscosity times density operating on the same principle as the Woodward device, but using an axially vibrating thin wall tube immersed in the fluid.

Piezoelectric Viscometer: A stainless steel piezoelectric audio bender is used to vibrate a thin wall tube immersed in a test fluid, and it is found that the viscosityxdensity product is inversely proportional to the square of the velocity amplitude of the tube in the fluid. A brief introduction to piezoelectric elements is given by Singmin [4].

Figure 1 is a drawing of the viscometer sensor. The viscosity probe is a thin wall stainless steel tube which is inserted in a sample of fluid to be measured. The end of the tube is sharpened so as not to present a blunt edge which can act as an apparent additional mass when vibrated, in accordance with the theory of Lamb [5]. The tube is immersed in the test fluid to the center of the hole through the tube wall, which also allows the test fluid to rise up inside the tube to the same height as on the outside. The sensitivity of the probe (see governing equations section) is inversely proportional to the surface area of the tube in the fluid, and at the depth of 60 mm it is on the order of 10 mV per mm. The viscosity probe is attached to the piezo ceramic audio transducer stainless steel plate. The piezo ceramic is sintered to the top side of the plate and has sections which are used for the application of a drive signal and for feedback signal monitoring. The piezo ceramic is driven by a square wave using a phase-locked loop to maintain axial mode resonance of the viscosity probe. The feedback circuit generates a voltage in proportion to the deflection of the ceramic which is used to determine the effect of the test fluid on the viscosity probe. The design of the mechanical spring (transducer), mass (viscosity probe), base support, and isolation neoprene spring is such that the sensor operates only at its axial resonance mode. The housing is isolated from the vibrating probe-transducer system so that forces on the housing do not affect the measurement. The base-and-pins support is designed to minimize forces being transmitted to the housing and to avoid radial forces on the transducer due to differential thermal expansion. The Neoprene isolation spring is very soft to minimize force transmission to the housing. An integrated circuit is used to measure the fluid temperature with a resolution of 0.1 °C. The sensor head is connected to the electronics by a shielded cable.

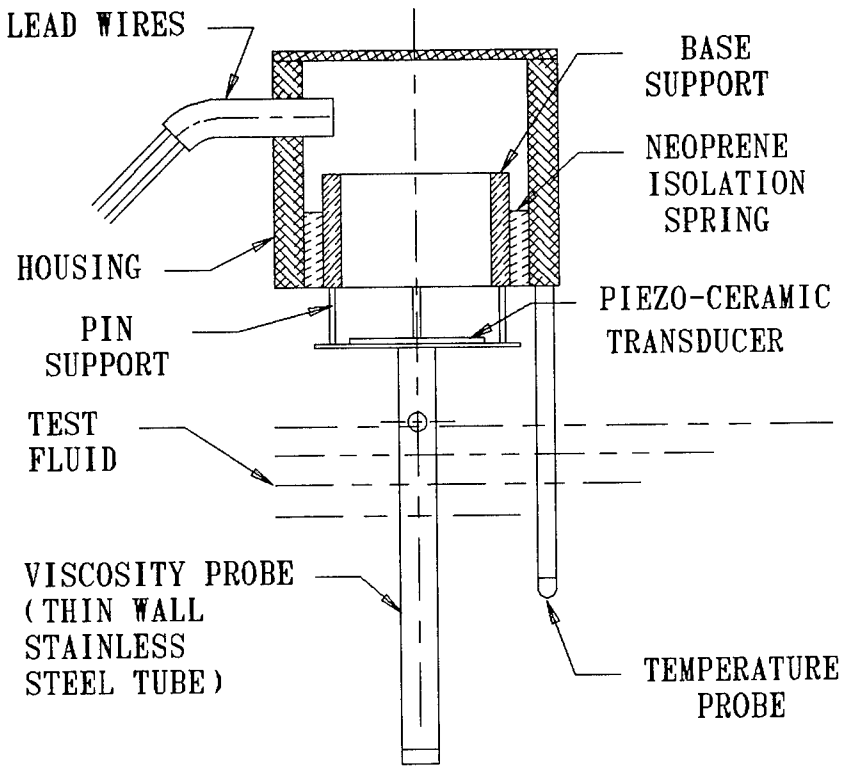


Fig. 1 PIEZOELECTRIC VISCOMETER SENSOR

Theory: The theory for a vibrating plate immersed in a fluid is based on Rayleigh wave mechanics, and was initially given by Woodward [3]. More recently, the theory of Woodward was reexamined by Oka [6], and the results of his work will be presented here. In the Oka derivation, it is assumed that the system is operated at resonance and the driving force remains constant. The resulting equation is

$$\eta\rho = \left(\frac{R_o^2}{2A^2}\sqrt{\frac{m}{k}}\right)\left[\frac{V_{oa}}{V_o} - 1\right]^2 \qquad (2)$$

Where R_o is internal friction, A is wetted area of tubular probe, m is probe mass, k is spring rate of transducer, v_{oa} is velocity amplitude of probe in air, and v_o is velocity amplitude of probe in test fluid.

Since the velocity amplitude can be replaced with the frequency times displacement amplitude, and the displacement amplitude is proportional to the piezo ceramic feedback voltage amplitude, Eq.(2) can be modified for application to the piezo ceramic transducer as follows.

$$\eta\rho = \left(\frac{R_o^2}{2A^2}\sqrt{\frac{m}{k}}\right)\left[\frac{f_a V_{oa}}{f V_o} - 1\right]^2 \qquad (3)$$

Where V_{oa} is the ceramic feedback voltage amplitude with the probe in air, and V_o is the ceramic feedback voltage amplitude with the probe in the test fluid. It is usual to assume the frequency ratio, f_a/f, is equal to unity.

In Eq.(3) the terms in parenthesis are constant for an instrument, as well as the feedback voltage amplitude in air in the bracketed term; requiring two measurements for calibration. The device can be calibrated in air and in a calibration fluid, or by using two different calibration fluids. The latter method of calibration is the method that has been found to be most useful.

Electronics: The microcontroller-based piezoelectric viscometer has been designed as a battery powered portable unit with an onboard computer. The electronics are built on a printed circuit board in a separate housing from the sensor, and connected by cable to the sensor head. Figure 2 shows the block diagram of the electronic circuit.

As currently implemented, the viscometer's phase-locked loop, shown in Fig. 2, has a running frequency of 1467 Hz. It will acquire and track the resonant frequency for measuring heads that resonate within a range of approximately +/- 300 Hz. This frequency range is free of nonaxial mode vibration, therefore only the axial mode vibration of the tubular probe is activated. Also, there is a small increase in resonance frequency of the measuring head with more viscous oils, and this range accommodates this effect well.

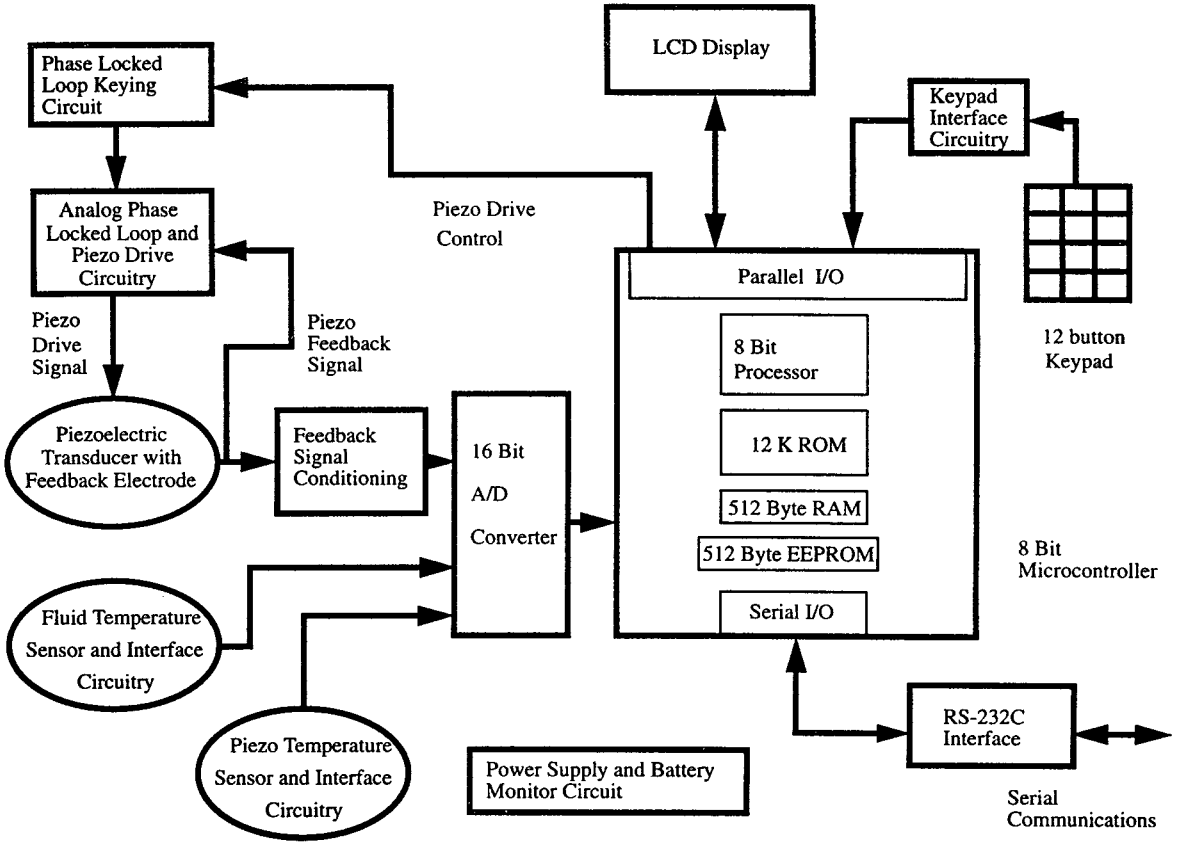


Fig. 2 Block Diagram of Microprocessor Controlled Piezoelectric Viscometer

A 14.4 volt nickel cadmium battery is used to power the device, but both positive and negative 12 volts as well as 5 volts are required for operation. A switching power invertor circuit was used to develop the +/- 12 volt sources. These circuits generate electrical noise, and special shielding was designed to eliminate this problem.

The battery life tests show that the battery lasts 8 hours between recharge cycles. A recharge time of 1.5 hours is required during which time the viscometer may be operated from an external power supply (wall cube).

The Motorola MC68HC711E9 microcontroller used by the device has 12 kbytes of onboard read only memory, which limits the software to fit within these constraints. It has been necessary to restrict the use of floating point real numbers to a few inputs in favor of integer input from the keypad.

The viscometer will measure, calculate, and display the viscosityxdensity product and the temperature of the test fluid. In the case of a specific type of oil, such as a paraffin based mineral oil with viscosity index near 100, the software contains curvefit data which relates the viscosityxdensity product to the ISO number for the oil at the temperature of measurement. Thus, for a paraffinic oil the computer can reduce and present the following data: 1. test temperature, 2. viscosityxdensity at test temperature, 3. ISO Number, 4. viscosity at test temperature, 5. density at test temperature, and 6. viscosity or density at any other reference temperature. The viscometer can display the results of the measurements through the RS-232C interface to a personal computer equipped with communications software, such as Pro-Comm Plus, or the simple TERMINAL program provided with the Microsoft Windows package. The link operates at 9600 bits per second with one stop bit, eight data bits, no parity bit, and no flow control.

Experimental Results: Test results using Cannon Instrument Co. reference oils are shown on Figure 3. According to the theory given by Eq. (3), a log log plot of $[V_{oa}/V_o - 1]$ versus viscosityxdensity should give a straight line with a slope of 1/2. The data is in agreement with the theory line as shown on Fig. 3. Test results using mineral oils give a range of application from 10 to 2000 cPoise (a higher range is possible but was not tested) with an accuracy of +/- 1.5% of full scale viscosity, but typically within +/- 5% absolute error. The operating temperature range is +/- 5 °C around room temperature, with a repeatability of 2%. There are 3 factory adjustments required per unit. The chemical resistance of the stainless steel sensor probe is satisfactory for many fluids, and the tube material could be modified for special case fluids such as acids.

Conclusions: The piezoelectronic viscometer is a useful instrument for measuring viscosityxdensity at the test temperature of a newtonian fluid where the viscosity is greater than 10 cP. The viscosity can be calculated by making a separate measurement of density at temperature. High pressure measurements can be made by inserting the measuring head into a high pressure chamber, but the test fluid must have a free surface for proper immersion by the tubular probe. For a specific fluid where density versus temperature data is known, the computer software can include a calculation to reduce the data automatically during a measurement. In the case of a non-newtonian fluid the device will give a value

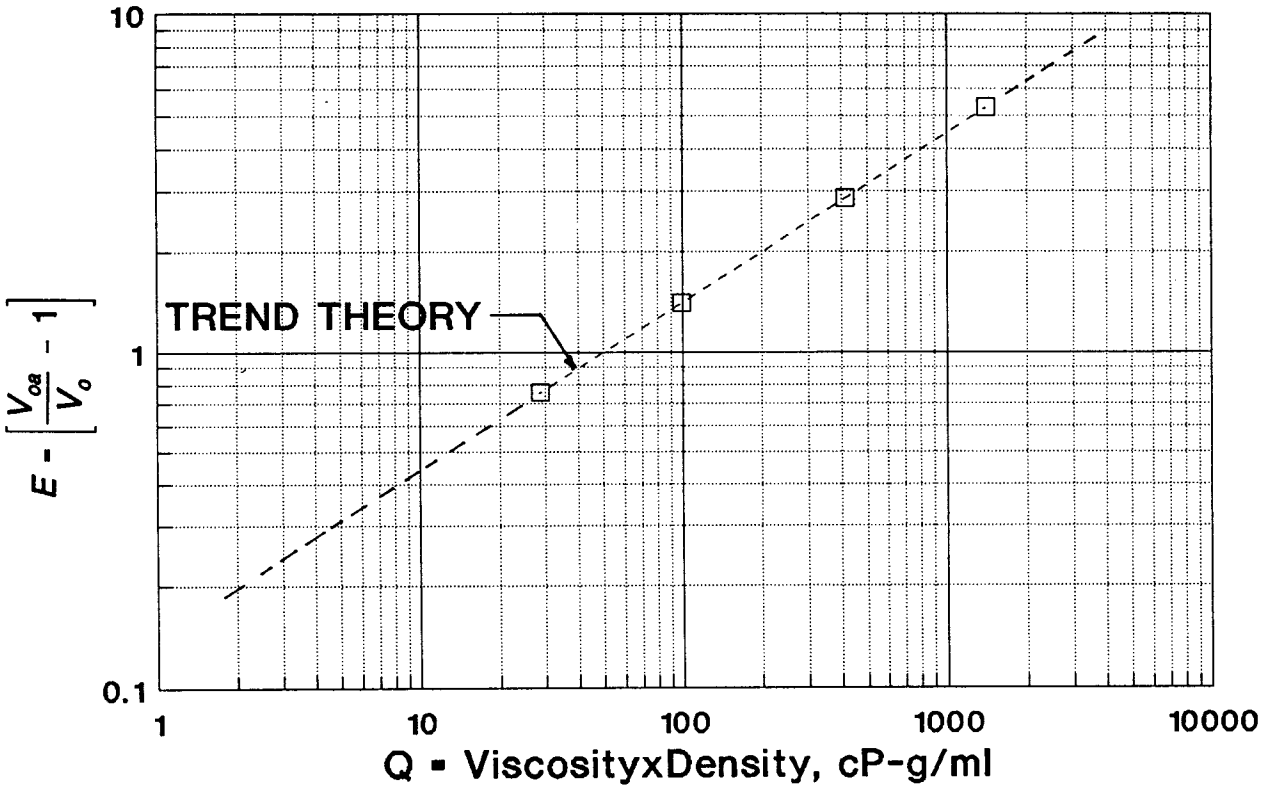


Fig 3 VISCOMETER TEST DATA

of the apparent viscosityxdensity product, but this can be related to an operational requirement of the fluid such as a mixing criteria, for example.

References:

1. J. R. Van Wazer, J. W. Lyons, K. Y. Kim, and R. E. Colwell, "Viscosity and Flow Measurement," Interscience, 1963.

2. W. P. Mason, "Measurement of the Viscosity and Shear of Liquids by Means of a Torsionally Vibrating Crystal," Transactions of the American Society of Mechanical Engineers, V. 69, May 1947, pp. 359-370.

3. J. G. Woodward, "A Vibrating Plate Viscometer," Journal of the Acoustical Society of America, V. 25, January 1955, pp. 147-151.

4. A. Singman, "All About Piezoelectric Elements," Popular Mechanics, March 1994, pp. 69, 70, and 96.

5. H. Lamb, "On the Vibrations of an Elastic Plate in Contact with Water," Proceedings of the Royal Society A, V. 98, 1921, pp. 205-216.

6. S. Oka, "Oscillating Plate Viscometers," Rheology Theory and Applications, F. R. Eirich Editor, Vol. 3, Academic Press, 1960, pp. 65-82.

PARTICULATE/WEAR DEBRIS ANALYSIS #2

A PHENOMENOLOGICAL APPROACH TO WEAR DEBRIS ANALYSIS

Unchung Cho and John A. Tichy
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

Abstract: In wear debris analysis, phenomenological observation and characterization of wear debris can be systematically exercised to offer clues about the underlying wearing conditions. While there is terminology which describes phenomena occurring in worn surfaces and subsurfaces, phenomenological observations of wear debris have not been coined in technical terms. In this paper, the concept of wear debris phenomena is proposed to ascertain the phenomenological aspect of wear debris analysis. Two terms are defined to describe phenomena which can occur in relation to wear debris: cutting and severe sliding. An extended classification of wear is made to accommodate wear debris phenomena.

Key Words: Ferrography; oil analysis; wear debris analysis

Introduction: Wear debris analysis is an important subject in maintenance, especially condition based maintenance. It represents one of the very limited ways by which the wear life of mechanical components can be predicted without significant interference with machinery operation. In wear debris analysis, a broad spectrum of techniques is applied to extract information. Among these, optical examination of wear debris has long been used as an effective method to find clues to the progress of wear that occurs in machinery. Its origin as a diagnostic tool can be traced to Ferrography developed in the early 1970's. Westcott and Seifert [1] state the heart and soul of Ferrography, or optical debris monitoring, as follows.

> *The key to Ferrography or optical examination of wear debris is to find marks or features on wear debris which suggest likely wearing conditions from which they were generated. When changes in wear rates are detected, it is important to predict whether equipment will go back to a normal condition or get worse and fail. Since most equipment is designed for normal wear, abnormal wear may quickly lead to catastrophic failure. Therefore, it is essential to detect machine faults, even though it is hard to predict the exact time of failure.*

Wear processes produce an amazingly complex array of types of wear debris from surfaces. The interpretation of the appearance of the wear debris is subject to the individual judgment and optical means employed. However, human expertise is not usually discussed in the open literature nor is it readily available in the present tribology community.

Since the advent of Ferrography, much effort has been devoted to correlating the visual assessment of wear debris with the relevant wear modes. There exists internationally understood knowledge [2] which bridges the gap between wear debris and its origin. However,

phenomenological observation of wear debris has not been solidified in technical terms. The lack of terminology describing wear debris might be one of the reasons why wear debris analysis is not usually recognized in the conventional framework of wear. In this paper, the phenomenological aspect of wear debris analysis is addressed. The concept of wear debris phenomena is proposed so that the outcome of phenomenological observation and characterization of wear debris is effectively accumulated in one domain. Terminology is coined to represent two significant phenomena, the effects of which can be visually observed on wear debris, namely cutting and severe sliding. An extended classification of wear is undertaken to accommodate wear debris phenomena, in addition to worn surface phenomena and wear processes.

Identification of wear modes based on wear debris: In maintenance, the motivation to investigate wear modes arises from the fact that the wear life of mechanical components can be predicted by the identification of wear modes occurring in a machine. The wear mode-the way surface damage proceeds or a manifestation of wear processes-can be studied by examining the evidence of wear, i.e. worn surfaces, wear debris and subsurface deformation (Figure 1). Of these three results of wear, visual assessment of worn surfaces and wear debris is adequate for routine maintenance practice. However, the examination of wear debris is more cost-effective.

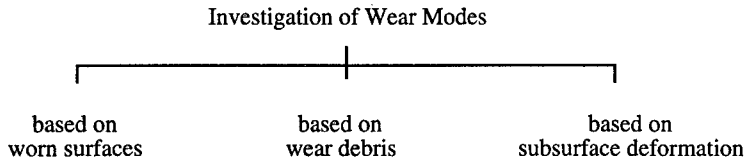


Figure 1. Investigation of wear modes.

In wear debris analysis, the wear mode is usually concluded by examination of wear debris using optical means. Before examination, wear debris are usually mounted on a glass slide or trapped in filter paper. It is not conceivable to rotate debris to see the exterior appearance all round [3]. Since observations are usually done from above, some shapes, e.g., concave shapes, are misleading and overall shape is hard to tell. Nevertheless, visual assessment of wear debris by light microscopy can be correlated to the wear mechanisms by which they were removed.

With technological advances in the field of computer vision, attention has focused on an automated means for wear debris analysis. To date, results have been limited by the 2-dimensional nature of conventional light microscopy [4]. Recently, it has been suggested that the high resolution 3-dimensional imaging capabilities of laser scanning confocal microscopy may provide a more complete analysis of wear debris [5].

Wear debris phenomena: Whenever there is a tribological interaction, the surface is worn and deformed, the subsurface is deformed, and wear debris are generated. In the course of wear

processes, several phenomena occur in the surface, subsurface and wear debris. Phenomenological observations and characterizations of wear are attempts to pinpoint significant wear processes. There exists terminology which describes meaningful phenomena that occur on worn surface (pitting, scuffing, plowing, scoring, etc.) and subsurface deformation (microcracking and recrystallization, etc. [6]). However, phenomenological observations related to wear debris have not been coined into technical terms. The concept of wear debris phenomena is proposed with the hope of embracing the efforts of phenomenological observations and characterizations. The meaning of wear debris phenomena is self-explanatory. In the field of wear, terminology assessing the results of wear is highly descriptive of the appearance, and in a sense is indicative of possible wear mechanisms [7]. So terminology of wear debris phenomena should also be capable of such descriptions. Due to their small size, phenomenological observation of debris is highly subject to the use of optical means. For example, wear debris phenomena can be defined as follows.

Curling: A process by which overall shape of debris becomes like machining swarf (Figure 2, left). This phenomena can be observed when surface is cut either by a sharp projection or by an abrasive particle.

Scratching: Fine-scale deformation resulting in a multitude of lines which run parallel to the sliding direction [6] (Figure 2, right). Scratching results in striated surfaces on wear debris, which can be indicative of severe sliding.

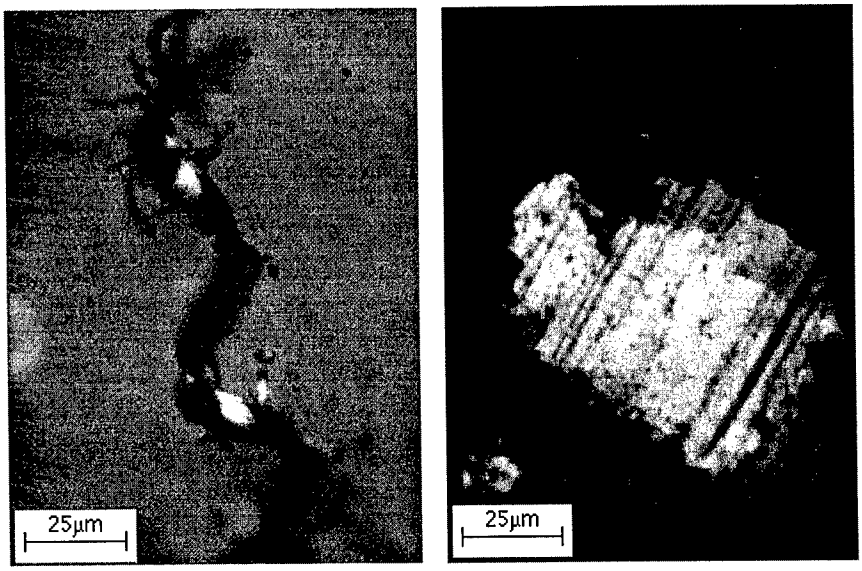


Figure 2. Wire-like aluminum debris cut by abrasive grains (left), aluminum debris with striated surface generated in a severe sliding condition (right).

An extended classification of wear: Wear is generally regarded as the "unwanted removal of material by chemical or mechanical action" [8]. Another definition [9] is the "progressive loss of substance from the operating surface of a body occurring as a result of relative motion at the surface." Laws, mechanisms and theories have been developed to understand fundamentals of wear; and the accumulated knowledge of wear has been systematically arranged according to numerous classification schemes. In general, wear can be broadly classified based on the results of wear and the nature of the underlying processes [7, 10]. For example, in Dowson's classification [7], one category is highly descriptive of worn surfaces as well as somewhat indicative of the wear mechanisms. The other is based on the wear processes (Figure 3). However, the conventional classification of wear does not include any wear debris phenomena.

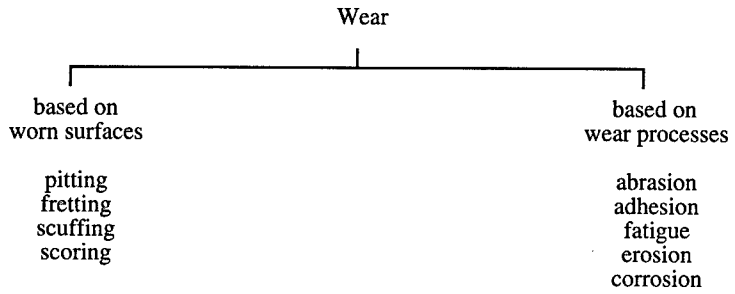


Figure 3. Classification of wear.

While wear processes can be investigated by the examination of worn surfaces, wear debris features can also provide indications of the tribological interactions by which they were formed. For many years, wear debris analysis has been used to provide cost-effective and continuous monitoring of the degradation of machinery components without dismantling the machines. As the number of studies of wear debris has been increased, the amount of knowledge about it has taken a more significant position in the area of wear. A better understanding of the relationship between wear debris features and the relevant wear modes will enrich our knowledge of wear. Because of their importance, it is fair to say that the phenomena occurring in relation to wear debris should be included in the conventional classification of wear.

In addition to worn surface phenomena and the wear processes, traditional classification of wear is extended to accommodate wear debris phenomena (Figure 4). In both conventional and extended classification, there are two backbones: one is the results of wear and the other is the underlying processes. Between two classifications, the difference is that: in the extended classification of wear, the phenomena of wear debris as well as worn surface are considered as the results of wear. The motivation for the extended classification is two-fold: first, wear debris phenomena should be included for the completeness of classification of wear. Second, identifying processes of wear modes based on the analysis of wear debris should be recognized in conventional framework of wear-so that more attempts can be made to discriminate the

validity of claims and counter-claims in wear debris analysis. In this way, the interpretation of wear debris would become more reliable.

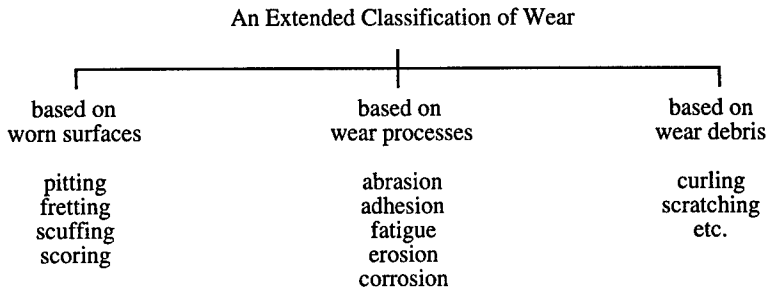


Figure 4. An extended classification of wear.

Summary: The phenomenological approach to wear debris analysis can be a useful tool in digging out significant correlations which can bridge the gap between wear debris and the relevant wearing conditions. The concept of wear debris phenomena and the extended classification of wear are proposed not only to make a systematic effort, but also to make wear debris analysis recognized in conventional framework of wear-so that the acquired knowledge can be easily screened.

Acknowledgments: The authors would like to thank Mr. S. F. Murray and Mr. S. J. Calabrese of Rensselaer Polytechnic Institute for helpful discussions. The authors are also grateful to Dr. Simon C. Tung and Mr. Kevin Brogan of GM Research & Development Center for wear tests and the supply of oil samples.

References:

1. Westcott, V. C. and Seifert, W. W., Private communication, August 1995.
2. Anderson, D. P., Wear particle atlas (revised), Naval Air Eng. Center, Report NAEC-92-163, 1982.
3. Hunt, T. M., Handbook of wear debris analysis and particle detection in liquids, Elsevier Applied Science, London and New York, 1993.
4. Roylance, B. J. and Raadnui, S., The morphological attributes of wear particles - their role in identifying wear mechanisms, Wear 175 (1994) 115-121.
5. Kirk, T. B., Panzera, D., Anamalay, R. V. and Xu, Z. L., Computer image analysis of wear debris for machine condition monitoring and fault diagnosis, Wear 181-183 (1995) 717 - 722.
6. Blau, P. J., Friction and wear transitions of materials, Noyes Publications, Park Ridge, 1989.
7. Dowson, D., History of Tribology, Longman, New York, 1979.
8. Peterson, M. B., Wear testing objectives and approaches, Selection and Use of Wear Tests for Metals: ASTM Special Technical Publication 615, R. G. Bayer, editor, Philadelphia, 1976.
9. Organization for Economic Cooperation and Development(OECD), Glossary of Terms and Definitions in the Field of Friction, Wear and Lubrication (Tribology), 1969.
10. S. Jahanmir, On the wear mechanisms and the wear equations, Fundamentals of tribology: proceedings of the International Conference on the Fundamentals of Tribology, MIT, June 1978, pp 455 - 467.

MONITORING FLUID SYSTEM DEBRIS *VIA* DIAGNOSTIC FILTERS

Puliyur Madhavan
Pall Corporation
Scientific & Laboratory Services Department
25 Harbor Park Drive
Port Washington, NY 11050, U.S.A.

Abstract: The utility of the filter element as a diagnostic tool to augment existing diagnostic techniques for monitoring fluid system debris is discussed. Two specific ways of utilizing the filter element for the above purpose are presented:
1) The incorporation of a removable(pull-out) diagnostic layer of porous medium, in conjunction with the regular filter element medium pack, allowing for on-site inspection of the collected debris and hence, delineation of 'abnormal' fluid system component wear and/or impending component failure. The above can assist operators in preventive maintenance and 'trouble shooting', and augment existing diagnostic techniques including chip detection and spectrometric fluid analysis.
2) The on-line monitoring of the rate of contaminant loading of the filter element *via* measurement of the filter element differential pressure in conjunction with the fluid temperature and flow rate. This will assist operators in identifying 'abnormal' contaminant loading conditions that may be associated with excessive contaminant ingression or accelerated fluid system component wear once a 'normal' baseline trend for filter element differential pressure build-up in a fluid system is established. This will also assist operators in optimizing filter element service life.
It is anticipated that on-line monitoring of filter elements incorporating diagnostic layers will significantly augment existing diagnostic techniques for fluid systems.

Key Words: Diagnostics; fluid systems; component wear; debris monitoring; diagnostic filter element; diagnostic layer; on-line differential pressure monitoring; failure prediction.

Introduction: The fluid is a repository for wear debris generated by the rolling/sliding component surfaces in the fluid system, particulate contamination ingressed from the operating environment into the fluid system, and the products of fluid breakdown such as carbonaceous material resulting from the thermal break down of engine lubricants. Monitoring of the particulate contamination present in the fluid can provide valuable information about the condition of the fluid system[1] and allow for the characterization of 'normal' and 'atypical' or 'accelerated' component wear and hence, the possibility of preventive maintenance prior to component malfunction and/or system failure. Two widely utilized debris monitoring techniques are detection/evaluation of metallic/magnetic debris (chip detection) and spectrometric fluid analysis. A useful technique that augments the above diagnostic procedures is monitoring of debris collected in the filter element[2-5].

The coherent surface and 'full-flow' characteristics of filter elements are ideal for efficiently capturing particulate contamination in the fluid, including non-magnetic and non-metallic debris of interest, such as material from seals, mineral compounds that may be ingressed from the operating environment into the system, and particulates associated with fluid break down. In addition, the monitoring of the filter element differential pressure, in conjunction with fluid flow rate and temperature, provides an estimate of the particulate contaminant loading of the filter element and allows for the identification of 'abnormal' contaminant loading conditions, and prediction of impending requirement for filter element replacement[5,6], once a baseline trend has been established.

The use of filter elements with convenient diagnostic layers for monitoring fluid system debris, and filter element differential pressure monitoring as an on-line diagnostic tool, are discussed below.

Filter Elements with Diagnostic Layers: The contamination collected in filter elements is analyzed at varying levels of sophistication by many equipment operators, often to verify the results of spectrometric fluid analysis or chip detector indications, and by equipment manufacturers during 'Acceptance' testing, such as the 'Green Run' testing of aircraft engines. The use of conventional filter elements for this purpose has several drawbacks. A primary difficulty is in accessing the particles trapped within the matrix of fibers constituting the filtration media of disposable filter elements. Typically, this is time consuming and involves cutting the filter element medium pack along the circumference of the end caps, removing the pack, and separating the filtration medium from the support meshes, usually metal meshes. The process can introduce metal debris from the cutting tool as well as from the support meshes. Separating the relevant wear debris from the large background of contaminant collected on the filtration medium may also present a problem.

The above problems may be resolved by the incorporation of a removable(pull-out) layer of porous medium(diagnostic layer) in conjunction with the regular filter element medium pack. The removable diagnostic layer is pleated in with the regular filter element medium pack, above the upstream support mesh, and can be readily removed on-site for examination of collected debris, Figure 1.

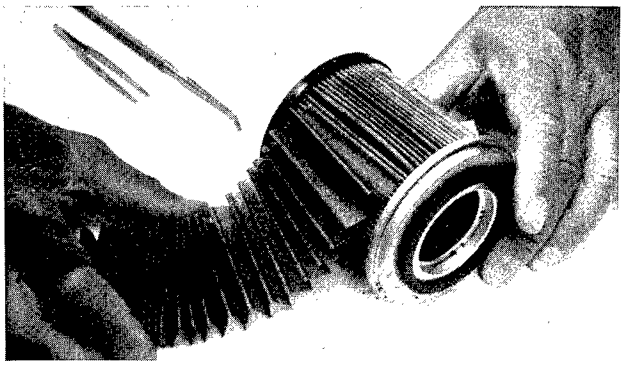


Figure 1. Filter element with removable diagnostic layer.

Important characteristics relevant to the diagnostic capability include the minimal thickness and appropriate porosity of the medium comprising the diagnostic layer. The minimal thickness ensures that the captured debris are trapped predominantly on the surface, allowing for convenient inspection, while the appropriate porosity is necessary to control the size range of particles retained on the diagnostic layer. As an example, one version of the diagnostic layer, utilized in several applications, exhibits a particle removal efficiency of ~ 95% for particles in the ≥ 75 μm size range, and a lower efficiency (~ 60%) in retaining debris in the ≥ 20 μm size range.

Filter elements with diagnostic layers have been evaluated in several bearing failure studies carried out by aircraft engine manufacturers. In these studies, the filter elements were incorporated in test stands, downstream of magnetic/metallic debris monitors. The bearings were deliberately damaged and allowed to operate to failure. The filter elements were replaced periodically, and the diagnostic layers were inspected for bearing wear debris build-up. Magnetic/metallic chip counts from the debris monitors were also recorded throughout the tests. The diagnostic layers of the filter elements, removed on 'chip indication' by the debris monitors, showed the presence of significant amounts of bearing wear debris, indicative of the onset of bearing failure. As expected, the quantity of bearing wear debris observed on the diagnostic layers, and the chip counts, increased significantly during the tests due to accelerated bearing wear prior to bearing failure, thus demonstrating the ability of the technique to augment on-board magnetic/metallic debris detectors. It should be noted that examination of the wear debris on the diagnostic layer could also provide the ability to distinguish between 'nuisance chip detector warnings' and more significant wear debris.

At present, a principal area of application of filter elements with diagnostic layers is in aircraft engine lubrication systems, both during 'Green Run' testing of engines and during regular flight operation. Numerous engine manufacturers and overhaul facilities are utilizing lubricant filter elements with diagnostic layers to examine built-in debris in engine lubrication systems during 'Green Run' testing of newly manufactured and overhauled engines. Figure 2 depicts photomicrographs of debris removed from diagnostic layers during the 'Green Run' testing of overhauled engines.

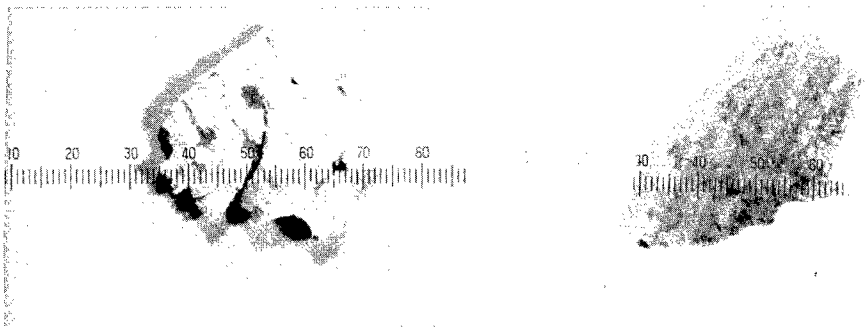


Figure 2. Photomicrographs of a titanium chip(left) and a magnesium chip(right) removed from diagnostic layers during 'Green Run' testing of overhauled engines; magnification : 50X.

Examination of the debris on the diagnostic layer during 'Acceptance/Green Run' testing can provide information about the debris built into the fluid system, during overhaul and manufacturing processes, which can contribute to accelerated component wear and failure. Characterization of the origin of the debris can provide information about the various stages in the overhaul and manufacturing processes that may contribute to built-in debris and hence, allow for process improvement.

Table I summarizes results from the evaluation of diagnostic layers during post production testing of two aircraft engines, designated Engine #1 and Engine # 2, at an engine manufacturer's test facilities. Filter elements with diagnostic layers were utilized in place of the standard lubricant filter element in the engine as well as in the test stand, downstream of the engine, corresponding to the lubrication system scavenge section. Based on the standard criteria employed by the engine manufacturer, it was determined from the visual observation of the diagnostic layers that the diagnostic layers corresponding to the test stand filter elements were heavily contaminated. The diagnostic layers corresponding to the engine filter elements were not as contaminated due to the fact that the engine filter is located downstream of the test stand filter. The mass of contaminant collected on the diagnostic layers and the chemical elemental composition of the contaminant are summarized in Table I. The results confirm the conclusions of the visual examination of the diagnostic layer, with the mass of contaminant collected on the diagnostic layers corresponding to the test stand filter elements being significantly higher than the mass of contaminant collected on the diagnostic layers corresponding to the engine filter elements. Further, the chemical elemental composition of the contamination proved to be of value in determining the origin of the contaminant. The above illustrates the utility of the diagnostic layer during 'Acceptance/Green Run' testing.

Table I. Evaluation of contamination collected on diagnostic layers of lubricant filter elements during post production testing of aircraft engines.

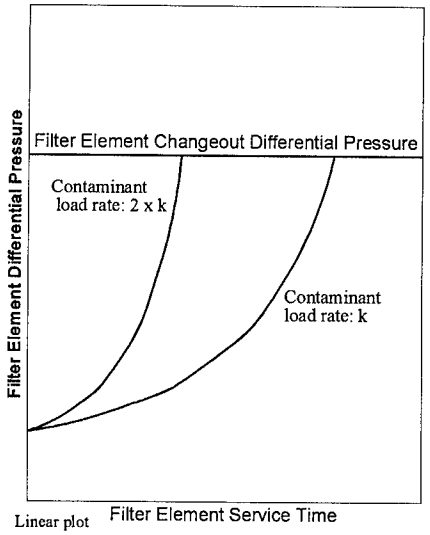
Engine #	Filter Location	Contaminant Weight(mg)	Chemical Elemental Composition[1] (XES Analysis)
Engine # 1	Engine (primary)	4.8	Major: Si, Cl Moderate: Al, Ag
	Test Stand (scavenge)	36.7	Major: Si, Ag Moderate: Fe, Ti, S
Engine # 2	Engine (primary)	2.6	Major: Si, Cl Moderate: Fe
	Test Stand (scavenge)	39.1	Major: Si Moderate: Fe, Ag

[1] Al = Aluminum; Ag = silver; Cl = chlorine; Fe = iron; S = sulfur; Si = silicon; Ti = titanium

The wear debris collected on the diagnostic layer can also provide valuable information about component wear during regular fluid system operation. Once a baseline is established, the presence of 'abnormal' debris can signal component malfunction and/or impending component failure. Lubricant filter elements with diagnostic layers are currently utilized in one series of engines, during regular flight operation, by a major airline in North America; over 1.5 million hours of flight service have been accrued to date. Experience with the diagnostic layer has been positive, and specific protocols and baseline trends have been developed by the airline to characterize debris collected on the diagnostic layer, and to take appropriate maintenance actions. In several reported instances, corrective maintenance actions, based on the evaluation of the contaminant on the diagnostic layer, have resulted in avoiding delays/in-flight engine shutdowns.

Filter Element Differential Pressure Monitoring: Since the differential pressure across a filter element is a function of the fluid flow rate, viscosity(temperature), and the contaminant loading, monitoring of the filter element differential pressure, in conjunction with information about the fluid temperature and flow rate, can provide information about the contaminant loading(load rate) of the filter element.

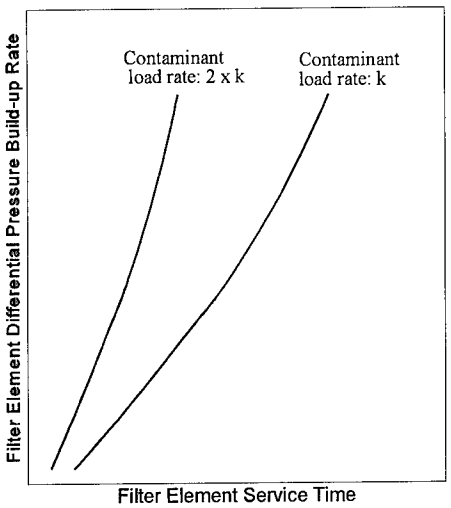
Figure 3 illustrates typical curves of filter element differential pressure build-up with service time, based on accelerated contaminant loading tests in the laboratory at constant fluid temperature and



Note: Constant fluid flow rate and temperature

Figure 3. Typical curves of filter element differential pressure build-up *vs* service time.

flow rate, for a baseline contaminant load rate designated k(mass/time) and for a contaminant load rate corresponding to twice the baseline load rate(2 x k). In Figure 4, the first derivatives of the pressure build-up curves in Figure 3 are plotted to illustrate the rate of change of filter element differential pressure for the two contaminant load rates. As expected, the curve corresponding to the



Log - linear plot

Note: Constant fluid flow rate and temperature

Figure 4. Filter element differential pressure build-up rate *vs* service time.

higher contaminant load rate(2 x k) exhibits a steeper slope. Thus, on-line monitoring of the filter element differential pressure build-up rate, under 'rated' fluid flow and temperature conditions, would allow for the identification of 'abnormal' contaminant loading conditions that may arise due to accelerated component wear, excessive contaminant ingression from the environment, or particulates associated with fluid break down, once a baseline trend has been established. The use of a filter element with a diagnostic layer would also permit operators to examine/identify the contaminant on-site in the event that 'abnormal' contaminant loading conditions are detected. In conjunction with other diagnostic information, this can assist in the identification of accelerated fluid system component wear and hence, allow for preventive maintenance.

The development of the technology for on-line monitoring of filter element differential pressure is at the 'Proof of Concept' stage. An electronic sensor, designed to measure filter element differential pressure and temperature, has been developed. It is comprised of a fused strain gauge circuit in a pie-

zoresistive differential pressure transducer, and a precision thermocouple to measure temperature. The sensor provides continuous differential pressure and temperature output signals(analog or digital) and can be interfaced with electronic control systems such as the ECU or FADEC systems in aircraft engines. A flow sensor may be incorporated, if required, or flow rate information obtained by other means could be interfaced with the control system.

In addition to identifying 'abnormal' contaminant loading conditions, the sensor can also assist operators in optimizing filter element service life. Consequently, sufficient advance warning of impending filter element by-pass will be provided, allowing operators to take appropriate measures to prevent filter element by-pass.

References

1. Aerospace Information Report(AIR) 1828, 'Guide to Oil System Monitoring in Aircraft Gas Turbine Engines', SAE, March 1984.

2. P. Madhavan, 'Monitoring Aircraft Power Plant and Transmission Systems via Diagnostic Filters', presented at the 43rd Meeting of the Mechanical Failures Prevention Group, October 3-6, San Diego, CA, 1988.

3. P. Madhavan, 'Advances in Filtration Technology for Reducing In-Flight Engine Shutdowns and Maintaining ETOPS Ratings', presented at the Second Annual Aero-Engine Cost Management Conference, January 25-26, San Francisco, CA, 1994.

4. K. K. Yeung, A. J. McKenzie and D. Liew, and G. A. Luoma, 'Development of Computer-Aided Image Analysis for Filter Debris Analysis', Lubrication Engineering, Vol. 50, No. 4, pp 293-299, 1994.

5. P. Madhavan, 'Monitoring Gas Turbine Engine Lube System Debris *via* Diagnostic Filters', presented at the SAE E34 Aviation Propulsion Lubricants Committee Technical Symposium, 28-29 September, Munich, Germany, 1994.

6. F. E. Bishop, 'In-System Sensors for On-Condition Monitoring', presented at the SAE A6 Fluid Power Actuation and Control Technologies Meeting, 2-7 May, New Orleans, LA, 1993.

DEVELOPING AN INTEGRATED CONDITION MONITORING SYSTEM

Jason Tranter
Predict/DLI, DLI Engineering Division
253 Winslow Way West,
Bainbridge Island, WA, 98110

Abstract: This paper discusses some of the background of condition monitoring systems, the need for integration between the technologies, and some of the approaches already taken in the industry. This paper also discusses some of the issues faced when designing an integrated condition monitoring system. The issues include: database structure, similarities between test types, the difference between data and information, access to data and information, and where to draw the line between condition monitoring and maintenance management systems.

Key Words: Condition monitoring; data collectors; integration; oil analysis; performance analysis; thermography; vibration analysis; wear particle analysis;

BACKGROUND: For many years it has been broadly recognized that condition monitoring (predictive maintenance) can reduce maintenance costs. By monitoring the vibration, lubricant, wear particles, performance parameters, electrical parameters, and many other indicators, it is possible to determine the state of a piece of equipment. The different technologies are used to test for different conditions, and are appropriate to different kinds of equipment. In many cases there is overlap between the technologies, for example vibration and wear particle analysis can be successfully applied to rotating machinery. It is also true that some technologies can reveal different information about the same equipment, for example electrical analysis can tell you about eccentricity and broken rotor bars whereas wear particle analysis can only tell you about the components in the path of the lubricant.

In many industries, multiple technologies are implemented, often by a combination of in-house and external services. However in almost all cases, the results of the various tests are reviewed by different people, often in different departments, using different software packages. Rarely can one user look at the wear particle analysis results and the vibration results related to the same equipment at the same time, even though together they paint the clearest picture of the condition. Instead, two people are forced to make individual assessments and hope to meet and discuss their results. As a result there has long been a desire to create one system capable of tracking all of these data types, with the appropriate analysis and reporting tools.

There are other challenges and problems that can arise due to this separation of technologies:

- Pockets of expertise form. It is rare to find someone expert in more than one field. Thus the organization is exposed to that person leaving.
- Poor decisions can be made due to that lack of information.
- Problems can be missed altogether. Individually the vibration, oil, etc. data may not clearly show that a problem exists. However when combined, a clear picture of an incipient problem may be revealed.
- It is difficult to perform root cause analysis. Only with the history of all the data is it possible to see what may have lead to a problem. Rather than simply fixing the problem, it is always best to understand and resolve the cause of the problem.
- Ownership of data and results can occur, resulting in inefficient and non-effective communications between parties.

A DESIRE FOR THE FUTURE: The ultimate solution to this problem, in the author's opinion, is as follows:

- The system should have one database which holds all of the key equipment and test data, and all of the recommendations and reports.
- The system should deal equally well with all major technologies, and should be able to grow to accept others.
- The system should be centered on the equipment rather than the test data, after all it is the equipment we are maintaining.
- The system should allow three basic levels of access to data and information: the technician who performs the test and performs basic analysis, specialists and systems engineers who are required to solve the problem and make a recommendation, and upper management (maintenance managers, plant managers, etc.) who need fast access to the recommendations without being confronted by raw data.
- The data and information should be available at any workstation, and should be integrated with other data management systems, for example maintenance management (CMMS) and distributed control systems (DCS).
- The system should support condition monitoring products from a variety of vendors, rather than limiting the user's choice.

- The final recommendation should be derived from all the information available, as shown in Figure 1.

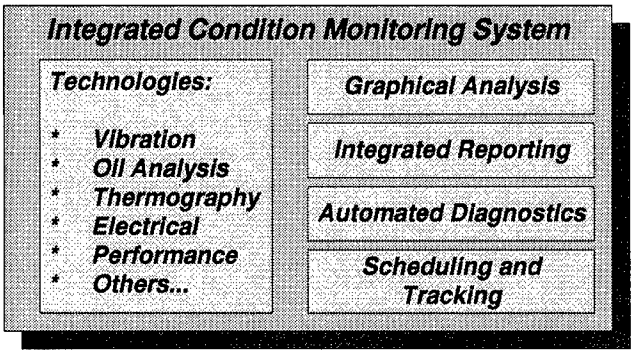


Figure 1: Structure of integrated condition monitoring system

As with any software package, additional requirements include:

- Ease of use and quick access to commonly used functions, as these systems have to be used by people unfamiliar with computers and software.

- Open database structure, for example SQL, to allow the user ease of movement of data in and out of the system, and for flexible reporting using external reporting tools.

- Support for client-server networking to reduce computer requirements and improve network performance.

- Use of off-the-shelf tools, and support for the integration of other tools via DDE and OLE, where ever possible to allow for non-vendor-specific growth.

- Flexible report writing to enable a clear and professional presentation of results, and to reveal hidden facts in the data (for example, do we have a bad pump manufacturer?).

- Support for financial analysis, first to justify the continued growth (or existence as the case may be) of the condition monitoring program, and second to justify individual recommendations.

WHY MULTIPLE TECHNOLOGIES? It has long been known that no one technology can give all the answers. And even though each specialist will have his or her own favorite technology, only a combination of information will lead to an accurate diagnosis. Your confidence level will grow considerably when a fault is supported by more than one source of information. And technologies have their sphere of influence: vibration cannot tell you about contamination occurring in the lubricant, and oil analysis cannot diagnose imbalance problems.

Remember, when performing a diagnosis we are attempting to learn four things: what is the nature of the fault, what is the severity of the fault, how soon must a repair be performed, and what is the root cause of the fault. When making a final recommendation for a piece of equipment, we need to know about the machine as a whole - not just the bearings or the gearbox. Only a combination of technologies can reveal this combination of information.

As a simple example, vibration readings may indicate a moderate bearing problem on a pump. Wear particle analysis may reveal particles of lead and tin also supporting, and adding confidence to the bearing call. Just knowing this you may recommend that the machine be left in service with increased monitoring. However if you were also collecting performance information on the pump, and saw that the efficiency was down, you may choose to overhaul the pump earlier, replacing the bearings and resolving the efficiency problem, as a reduction in efficiency could affect the process at large, and increase the energy costs.

WHY SHOULD THE SYSTEM BE INTEGRATED? With information held in separate systems, a user must be familiar with the operation and nomenclature used in each system - and naturally must have access to those systems - before an integrated conclusion can be made. Otherwise there will have to be frequent meetings where all of the specialists will come together to discuss their individual findings and arrive at a consensus. This can often mean going back to their respective systems to perform additional analysis.

If all of the information is held (or available from) one system, all users will be able to make these kinds of decisions in a much more timely fashion. It assumes of course, that these users are familiar with all technologies. However, the ultimate situation is one where all users have a level of competency in each technology. In any case, individual experts (or computerized expert systems) should first verify and process the data, turning the raw numbers into easily understood information, before it is used in a decision making process.

Other benefits to using an integrated system include:

- The level of cross-competency will grow with time as users inevitably become exposed to the different technologies.
- The system should be easier to learn, use and maintain given that just one software package would be used.
- All reports and analysis methods would be more consistent, reducing the chance of miss-communication.
- All users would know one place to turn for the key information.

ALTERNATIVE APPROACHES: There are a number of ways an integrated system could be developed. Each have their strengths and weaknesses.

Combination of separate systems: One approach, as depicted in Figure 2, is to create a number of independent condition monitoring systems: vibration, oil, thermal imaging, expert system, etc. with a simple menu system to jump between them. A user of such a system would be required to make a conscious decision which technology was necessary for a given task and select the appropriate package. Ideally there would be a consistent user interface, and there should be some kind of link between the data so that comparing data would be possible.

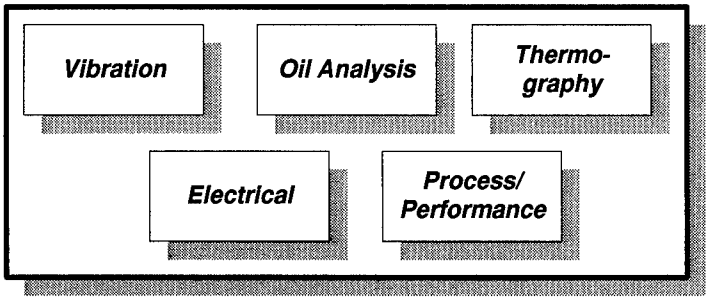


Figure 2: Combined system

The strengths are:

- Users focused in one technology area would have a dedicated package to learn.
- The individual packages are probably less resource intensive than a fully integrated system.
- It would be possible to choose packages from different vendors (although the level of integration may be reduced).

The weaknesses are:

- It is likely that the user interface is not consistent between packages.
- It is likely that there would be more than one database.
- It is therefore likely that equipment would have to be described many times.
- It is likely that there will not be a clear link between tests and equipment in the different packages. The odds are high that users will describe their equipment differently.
- It is likely that most users would not actually utilize more than one technology.
- It is likely that the entire system would be more difficult to learn.
- It is likely, therefore, that the full benefits of integrated condition monitoring would not be realized.

Vibration centered system: Another approach, as depicted in Figure 3, is to take a vibration system and add the required fields and analysis tools to support other technologies like oil analysis, thermography, etc.

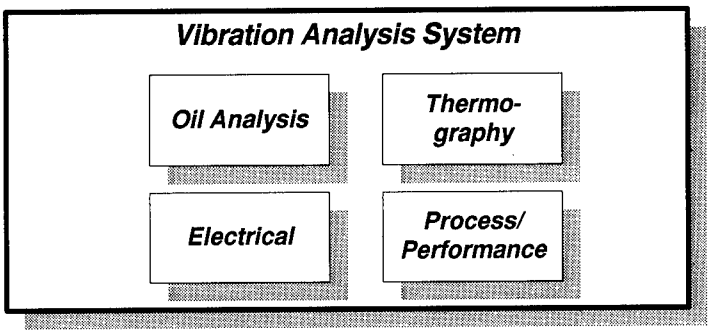


Figure 3: The vibration centered approach

The strengths are:

- If you believe that vibration is the key technology and other technologies simply support vibration, this structure makes it easy to compare data.
- Users familiar with vibration may find it easier to migrate to other technologies.

The weaknesses are:

- Users unfamiliar with vibration will have to become experienced. They will see this as a kludge.
- It is likely that the non-vibration data will be forced to fit the vibration model. For example, unlike vibration, not all oil samples are associated with just one machine. This subject will be discussed in greater depth later in this paper.
- It is unlikely that the system will be terribly flexible. It may be difficult/tedious to describe and access the data.

Separate systems, centralized integration: Another method, as depicted in Figure 4, would have separate condition monitoring systems (potentially from different vendors) for oil, vibration, etc., with a central software system that is able to read data from the other systems on demand. The central system would enable the user to view a variety of data, thus enabling an integrated decision.

The strengths are:

- This system would be independent of any vendor. The user could purchase the strongest packages in the different fields.
- The system could potentially be expanded to take data from other systems, thus extending the range of analysis.

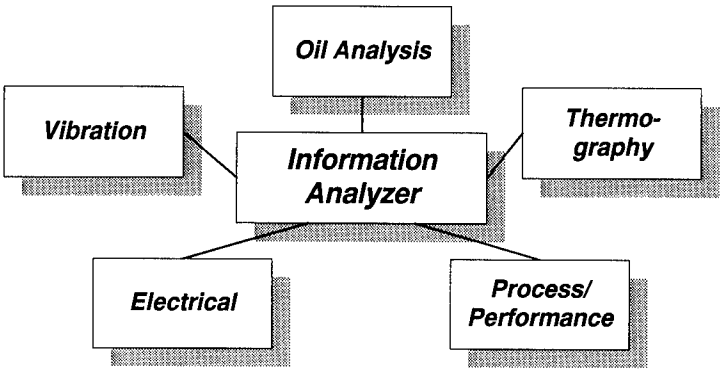


Figure 4: Centralized information analyzer

The weaknesses are:

- Special modules (currently unavailable) would be required to integrate the different packages. They would need to know how to read the stored data.
- Significant integration and customization efforts would be required before the system could be used. Issues such as networking, communications, and access would complicate the set-up.
- It is likely that the individual condition monitoring systems would be used for the detailed analysis, while the central system is used for an overview of the data.

One integrated, extensible system: The final method, the one preferred by the author, is to create one software system, with one database, one user interface, one analysis tool, one reporting tool, and one place to go for summary information about the condition of a piece of equipment. This system is depicted in Figure 5. The system would support a number of different technologies - as many as possible from multiple vendors. And the system would support the flexible creation of 'special test' modules that would support practically any maintenance or performance test performed in the plant.

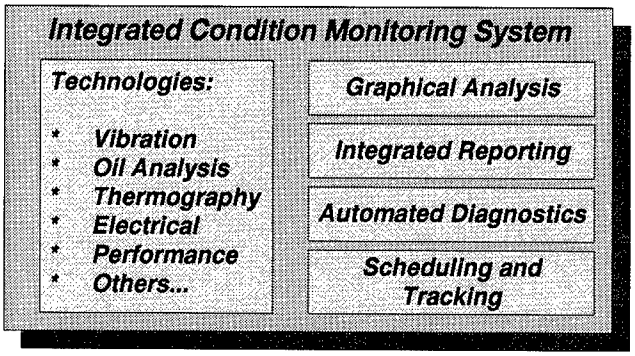


Figure 5: Integrated system approach

The strengths are:

- All users would have one system to turn to for up-to-date information.
- All users would have one software package to learn.
- The system would support growth into different technologies.
- The user would benefit from all the strengths of integrated condition monitoring, as described earlier in this paper.

The weaknesses are:

- Given the level of integration, it is possible that any individual technology would be more involved to use than if it were a single technology package.
- It would be perceived to be more vendor specific than the other alternatives mentioned. If the vendor of this system did not choose to support a particular technology, the user would be unable to utilize it in an integrated sense. Instead the user would be forced to adopt the first method described, where that one technology would be used separately.

CHALLENGES: There are a number of challenges to overcome before this final method can be made to work. These are not only issues that the developer of the system had to overcome, to a large extent they are issues that the user must be aware of as well. The flowing discusses just a few of the challenges.

How Far Should You Go? The first question to ask is how far to take the concept of integrated condition monitoring. For example, one could look at condition monitoring as just vibration, oil, thermography and motor current analysis. But as you delve more deeply, other questions and opportunities arise:

- When you consider oil analysis, most people think of Used Oil Analysis (UOA), where the user is interested in the condition of the lubricant, and Wear Particle Analysis (WPA), where the user is interested in the nature of the particles found in the lubricant. First, should the system support an on-site laboratory, with an interface to the various instruments and batch processing of samples, or just support the access of data from centralized laboratories? Second, should the system stop at UOA and WPA? What about other laboratory tests: transformer oils, fuel oils, soil tests, to name but a few. Why shouldn't the system support other tests types?

- What about process/performance analysis? Many vibration collection systems will allow a user to collect temperatures and pressures from the machine under test, but in general this is restricted to just a few readings, if any, and certainly not enough to tell you about the operating characteristics of the machine. And what about the rest of the process? If the data collected via the Distributed Control System (DCS), and the data collected via the routine 'logs' was available to the system, a much more complete picture of machine and plant condition could be made. Perhaps the system should also support hand held loggers that could be used to collect the logs. The question then arises of what should be done with all

that data. First, it is usually more useful to compute some more informative parameters, such as efficiency, from the raw data. And second, rather than having to review the levels and alarm status of all the collected data, a simple expert/advisory system should be used to reduce the mass of data into useful information.

- What about the variety of other tests performed in the plant: megger readings, diesel performance tests, crank web deflection measurements, and others? This information can tell you about equipment condition, so it should be part of an integrated system.

The tests and technologies mentioned above all play an important role in the assessment of equipment and system condition. It is up to the individual user and corporation to decide which tests should be performed, and what to do with the data. But as you can see, the problem is not truly solved by simply combining the 'traditional' condition monitoring technologies.

Database Design And Data Access: One of the keys to a successful system is in making it easy to define all of these tests, while offering all of the flexibility described thus far. When you sit down with experts in each of the individual fields and ask how they would describe their database hierarchy, you will get different answers. For example, the vibration expert will describe a machine with vibration test points, as shown in Figure 6. The oil expert will describe oil sample points that relate to individual components. Thermography users are concerned with items around the plant that have been identified as having a problem: air ducts, switch gear, motor bearings.

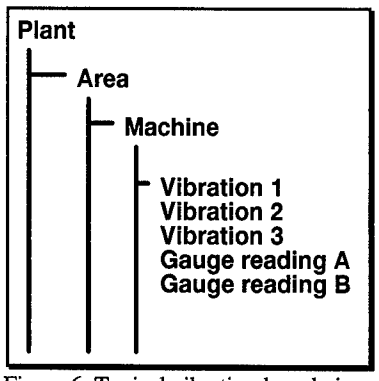


Figure 6: Typical vibration-based view.

If the system is to have one database, it needs to have a consistent hierarchy that supports these different views, while at the same time allowing the user to access the data in the most convenient manner. At the end of the day, the key is to be able to quickly access data related to the equipment under consideration - that is the only way an integrated decision can be made.

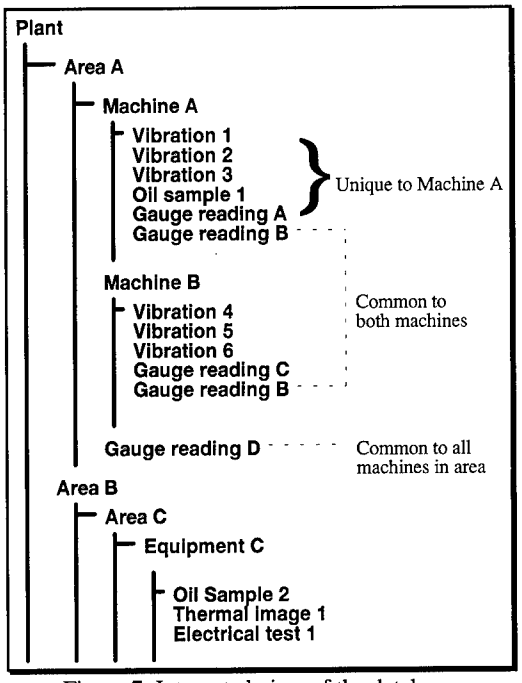


Figure 7: Integrated view of the database.

The second hierarchical diagram in Figure 7 illustrates how the database can be structured to support this variety of test data. It also illustrates a number of points:

- In some cases it is convenient to simply describe one or two levels in the hierarchy. In other cases it is desirable to add additional layers.
- It is necessary to show that some measurements are related to more than one piece of equipment. A measurement such as Vibration 1 is related to just Machine A. Gauge Reading B (a header pressure for example) is related to two machines: A and B. And Gauge Reading D is related to all machines in Area A (an ambient temperature for example).
- Although not illustrated in the diagram, it is also desirable to have more than one way to get to the same item of equipment. In the most simple case, one user may like to access a pump via its physical location (select the plant then the building). Another user, the Systems Engineer for example, may also like to describe equipment according to the system it belongs to. So the Systems Engineer may first select the circulating water system to find the same pump.

Test Synchronization: Another important issue when attempting to compare test results from different technologies is the timing of those tests. Ideally all tests will be taken on the same day,

under the same conditions. In reality this does not happen. Even if tests were taken on the same day, test results are not always immediately available. In the worst case, oil samples may be taken before an overhaul and vibration readings after. The user must be sensitive to this issue when performing comparisons.

User Education: A much larger issue is whether individual users will have the expertise to understand the data available in the different technologies. One thing is certain, when a system like this is first installed, it is important that inexperienced users are not looking at totally new data and expected to make decisions. Instead, the people most familiar with the technology should first review the data, ensure that it is valid (not taken from the wrong location for example), and make comments/recommendations on what is being revealed. As each specialist performs this 'filtering', an overall picture of equipment health will be formed and made available for general consumption. Any user can now look at any of the supporting data, and over time, become more experienced.

Over time, with the right tools, users will look at information rather than data - this is the ultimate goal. For example, rather than looking at vibration spectra, non specialists should look at information describing the severity of individual faults in the equipment, with the final conclusions. All technologies should be handled this way.

MAINTENANCE MANAGEMENT SYSTEMS: One of the trickiest questions is where to draw the line between the condition monitoring system and the computerized maintenance management systems (CMMS). The information flow is illustrated in Figure 7. This is a huge subject all by itself, but some of the issues are described below:

- Which system should be responsible for scheduling the tests?
- Which system should be the main source of machine condition information?
- At what point should information be handed to the CMMS in order to generate a work/inspection order?
- How should you get overhaul, oil change and other key information from the CMMS to the condition monitoring system?

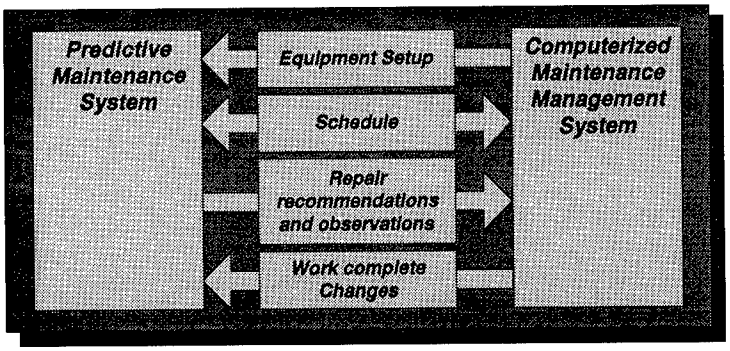


Figure 8: The link between Predictive Maintenance and CMMS

There is no clear answer to these questions. It seems every organization uses their CMMS in different ways. At one extreme users will state that their CMMS is inaccessible and unusable, thus the condition monitoring system should do as much as possible. Other users have excellent CMMS systems, and feel that the condition monitoring system should only be used to collect and diagnose data. The solution, therefore, is to support two-way links between the condition monitoring and CMMS system, and allow the end-user to utilize those links as appropriate.

WHAT DOES THE FUTURE HOLD? As far as the author is concerned, the future is in open, integrated systems. The days of proprietary systems with closed databases and limited functionality are numbered. The days of walking into a condition monitoring group office to find five computers, each exclusively used for a single technology, are numbered. This will not happen over night. Most users are loath to simply replace their existing system with the newest system on sale. So the first step is to begin the migration to an integrated system with just two technologies. Once some of the benefits of the change are realized, addition changes can be made. With cooperation within the industry (all condition monitoring users and vendors), all systems will be able to work together, either to support the transition of a site from one vendor's system to another, or to support the wide spread use of the key data.

With the move to open, integrated systems will come greater success in condition monitoring. With greater success comes greater visibility, viability and acceptance - something that is lacking in condition monitoring today.

VIBRATION ANALYSIS/MONITORING

A FUZZY-NEURO SCHEME FOR FAULT DIAGNOSIS AND LIFE CONSUMPTION OF ROTORDYNAMIC SYSTEMS

Michael J. Roemer

Stress Technology Incorporated
1800 Brighton-Henrietta Town Line Rd.
Rochester, New York 14623
Phone: (716) 424-2010
Fax: (716) 272-7201

Abstract: An intelligent health monitor for rotating machinery is presented that integrates proven neural network and fuzzy logic technologies with rotordynamic, finite-element modeling. A rotor demonstration rig is used as a proof of concept tool. The approach integrates rotor shaft vibration measurements with detailed, rotordynamic, finite-element models through a fuzzy-neuro scheme which is specifically developed to respond to the rotor system being monitored. The advantage of this approach over current methods lies in the use of a neural network classifier and fuzzy logic reasoning algorithms. The real-time neural network is trained to contain the knowledge of a detailed finite-element model whose results are integrated with system measurements to produce accurate machine fault diagnostics and component stress predictions. The availabilty of these real-time stresses allows for critical component life estimates to be calculated during machine operation. Fuzzy logic is implemented to overcome system measurements uncertainties, provide machine fault severity information, and make informed decisions about maintenance actions that should be performed based on operator experience.

Key Words: Artificial Intelligence, Diagnostics, Condition Monitoring, Neural Networks, Fuzzy Logic, Rotor Dynamics, Finite Element Models, Real-Time, Health Monitoring.

Introduction: Over the last several years, the practical strengths and weaknesses of applying neural networks and fuzzy logic in real-world condition monitoring systems have become more clear. Namely, pattern recognition based neural networks do not work very well on inconsistent or changing failure modes that many different types of machines exhibit (turbo-machinery in particular). In addition, the system specific training data required for training a neural network to recognize common machinery faults is typically not available. Also, fuzzy logic schemes do not lend themselves easily to modeling complex, nonlinear systems due to the large rulebases that would need to be generated for proper accuracy. However, by recognizing the *practical* capabilities of each of these technologies, a system developer can judiciously implement the individual technical benefits of each to meet the demands of more reliable and accurate condition health monitoring.

Utilizing the system modeling strengths of neural networks, machinery component finite element models can be represented with neural networks. By training a neural network from parametric analysis performed with a FE model, critical component stresses can be predicted in real time. These real-time, monitored stresses can then be used in LCF/HCF fatigue life algorithms to monitor critical component life damage as the machine is operating. In addition, the strengths of fuzzy logic can be utilized to minimize the effects of measurement uncertainties and system non-linearities. Confirmed fault diagnosis (based on operator experiences) and fault severity monitoring is also easily accomplished with fuzzy logic algorithms.

This paper demonstrates a **fuzzy-neuro** system (i.e. a system the implements both fuzzy logic and neural networks) for improving the present state-of-art in machinery health monitoring by increasing the effectiveness and reliability of mechanical diagnostics and component life monitoring. Specifically, a trained network is used to process relevant system sensor data in order to make informed decisions on a rotor's mechanical health and monitor critical rotor stresses. The real-time cyclic stresses are then analyzed by a damage accumulation algorithm to report a remaining component life estimate. A fuzzy logic scheme is developed to monitor the severity of the diagnosed fault, check the diagnosis performed by the neural network, and report on the maintenance action required to correct the fault.

Rotor Demonstration Rig: A rotor rig was constructed to demonstrate the concepts proposed in this paper on actual hardware. The demonstration rig was designed to be versatile enough to duplicate various vibration-producing phenomena found in all type of rotating systems. Many different types of vibration related characteristics were created and measured by changing rotor speed, degree of unbalance, degree of misalignment, shaft rub, and rotor bearing clearances. The resulting dynamic characteristics are measured with proximity probes and/or accelerometers and are processed with a multi-channel dynamic signal analyzer. The rotor configuration utilized in this paper is shown in Figure 1.

Two roller bearings support the motor armature, while four, oil impregnated, bronze sleeve bearings are positioned between the various couplings and disks. A solid 36" aluminum base with adjustable bearing pedestal locations and rubber isolation feet provide sufficient rigidity to the rotor configuration. Motor speed control is maintained with a proportional speed feedback algorithm, with speed sensed by a dedicated proximity probe and toothed wheel.

Seeded faults were introduced into the rotor demonstration system by applying mass unbalances to the disks, misalignment across the rigid coupling, loosening the bearing pedestals, and installing pre-worn bearings. Under each of these conditions, measurements were obtained from each of four proximity probes to determine the magnitude and phase of each transducer with respect to the reference key phaser. The specific magnitude and phase measurements were logged into a database and used in the neural network training procedure.

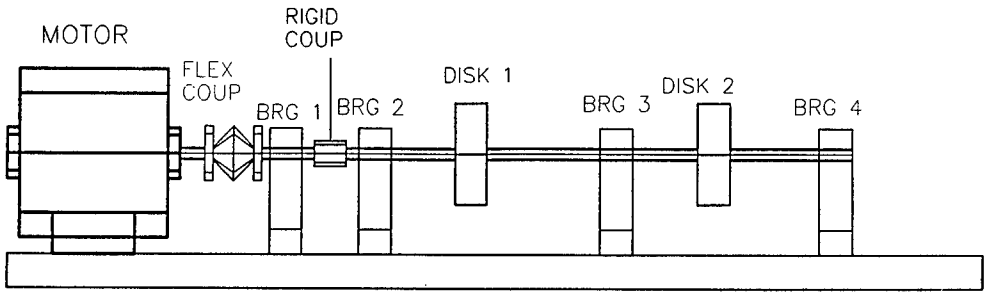


Figure 1 Rotor Demonstration Rig

Rotor Dynamics Finite Element Model: A detailed finite element model of the rotor demonstration system was developed and correlated with measured experimental data. This computer model was used to simulate rotor operation and to train the neural network classifiers. In particular, the network was trained from the model to determine dynamic stresses and forces in critical mechanical components. Figure 2 illustrates the first critical mode associated with the finite element model. In addition, the model predicts overall rotor vibratory characteristics as well as local vibratory stress levels. The real value in having a finite element model based diagnostic system is that it provides a very accurate picture of the rotor stress distribution and reaction forces. These stresses and forces are the cause of many of the component failures in the rotor, bearings, seals, etc. With the rotating shaft component stresses predicted, an automated life analysis algorithm will be able to determine what the expected component life will be with any damage condition.

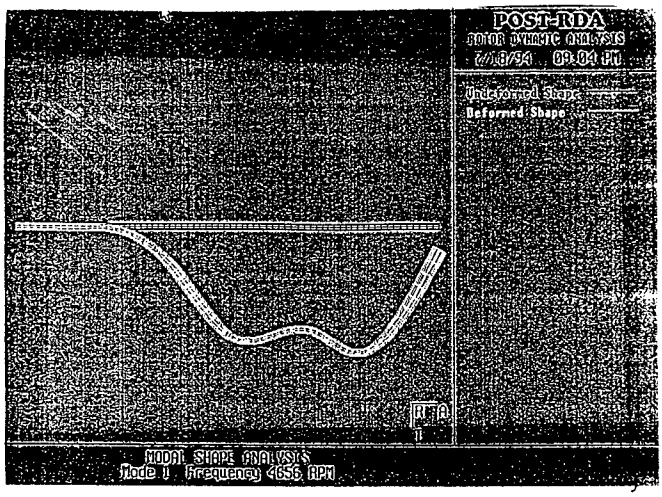


Figure 2 Calculated First Critical Rotor Mode

Fuzzy-Neuro System Development: The **fuzzy-neuro** system architecture developed in this paper provided for efficient measurement data processing and resulted in confirmed diagnoses of the rotor mechanical faults and rotor shaft component life estimation. A block diagram of the developed system architecture is given in Figure 3. In this figure, the "Diagnostics" block represents the neural network classifier, the "Fault Severity" and "Maintenance" blocks represent the fuzzy logic membership functions and rulebases, and the "Life Module" block contains the life accumulation algorithm. The system was developed as a stand-alone application comprising multiple functions and a single executable file to assist system testing and verification.

The complete diagnostic system shown in Figure 3 consisted of 6 transducer inputs and 4 diagnostic related outputs. The system inputs included four bearing vibration signals, relative phase across the shaft coupling, and rotor speed. A neural network was used to map specific one per-rev rotor responses to corresponding unbalances or misalignments as well as predict shaft stresses based on FE model training. The outputs of the "Diagnostic" neural network predict the probability of having an unbalance in balancing plane 1, balancing plane 2, or a misalignment across the coupling. The "virtual sensor" outputs of the network predict the stresses at two critical locations on the shaft. These real-time shaft stress are transferred simultaneously with the rotor speed measurements to the component "Life" module. The component life module continuously computes and updates the remaining fatigue life associated with the maximum cyclic shaft stress monitored during the current start/stop cycle.

The bearing vibration measurements are also processed by a "Fault Severity" fuzzy logic module. Processing the raw measurement data allows the fault severity module to perform a check on the network diagnosis as well as performing its standard function of determining the severity of a diagnosed fault. The outputs of this "fuzzy" module indicate the severity of an unbalance condition or coupling misalignment in linguistic terms such as "severity is minor", "severity is moderate", etc.

The final module entitled "Maintenance" accepts data from the output of "Diagnostic" neural network and "Fault Severity" fuzzy logic. By examining the diagnosed fault and associated fault severity, this module recommends specific maintenance actions that should be performed by the operator in linguistic terms. For example, if a fault is diagnosed as an unbalance condition in plane 2 with a severity level of minor, the "Maintenance" module would prompt the operator to monitor this situation over the next several days to see if the condition is worsening. If time passes and the unbalance severity worsens to a moderate level, then the maintenance action might be to "monitor condition very closely". Finally, if the severity continues to worsen, then a "shut down and balance plane 2" message would be given to the operator. The complete fuzzy-neuro system comprised of the diagnostic, fault severity, life, and maintenance modules is described next.

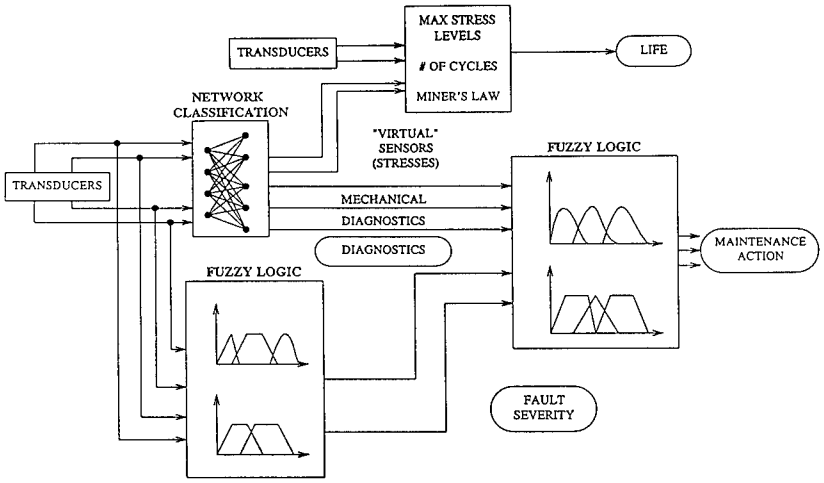


Figure 3 System Block Diagram

Neural Network Module: The neural network configuration illustrated in Figure 4 utilizes four bearing vibration input measurements and relative phase information across the flexible coupling to yield 5 input nodes to the network. One hidden layer, consisting of 10 nodes, is used to increase the "flexibility" of the network. Hidden layers, when used properly, can provide more accurate correlation between complex, linear and nonlinear training patterns. The output layer of the network consists of 5 nodes.

The first three output nodes of this network configuration diagnose the gross fault condition as either; 1.) an unbalance in balance plane 1, 2.) an unbalance in balance plane 2, or 3.) a misalignment across the coupling. The remaining two output nodes give important "virtual sensing" information about shaft stresses at two critical locations.

Virtual sensing refers to indirectly measuring a parameter such as shaft stress by matching patterns of directly sensed data (such as bearing displacement) with a finite element model to yield an accurate measurement of the unmeasured parameter. For the demonstration rotor system, the shaft bending stresses are calculated using a detailed finite-element model of the rotor for particular rotor conditions. The neural network is then trained to recognize the sensed patterns and relate them to the values calculated from the model. The result is a neural network (trained from measurements and FE model) that is capable of "virtually" sensing stresses on particular components in real time without actually having installed strain gages on-board.

Training the neural network involved evaluating the weights and thresholds of the numerous interconnections between the input and output layers. This was conducted utilizing a supervised training procedure. The supervised training technique specifies what target outputs should result from an input pattern. The neural network variables (weights and thresholds) are then self adjusted to generate that target output. The training procedure utilized a back propagation least-square error approach to achieve the desirable network accuracy. Network training was based on experimental

case histories and analytically derived input/output pairs resulting from the rotor dynamics computer model.

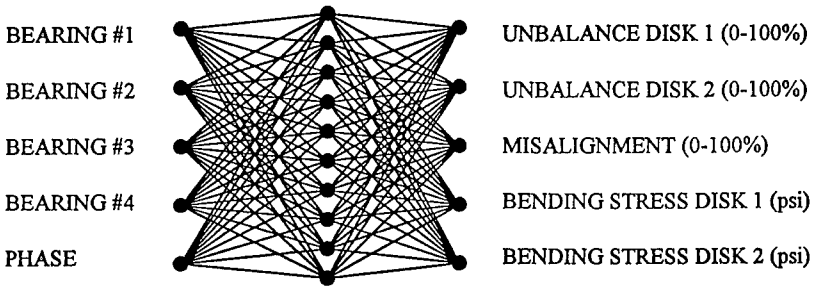


Figure 4 Neural Network Architecture

The network architecture was trained with 50 input/output training patterns devised from both experimental measurements and the finite element model analysis. The training patterns of the network database focused on diagnosing mass unbalance conditions, coupling misalignment, and shaft stresses. As an example, experimental data was collected from the rig to train the neural network to distinguish the differences between misalignment and an unbalance condition. Both of these conditions exhibit similar one/rev vibration characteristics. Phase angle measurements were obviously very important for the network to make this distinction.

The rotordynamics finite element model was exercised extensively with numerous unbalance force and shaft misalignment conditions. The results from each run of the finite element model yielded steady-state shaft bending stresses for each of these forcing conditions. These results were then used in conjunction with the measured data to build the training pattern database.

Fuzzy Logic Modules: Two fuzzy logic modules were developed for the rotor health monitoring system. First, a "Fault Severity" fuzzy logic scheme was introduced to monitor the level of risk a particular fault is producing. This severity level is calculated from the raw transducer measurements and therefore acts as a check for the neural network diagnosis. The outputs of the fault severity module are combined with the outputs of the diagnostic network to form the six inputs to the "Maintenance" module. The maintenance module examines any diagnosed fault and corresponding level of severity to determine the best action to be performed by the machine operators.

The "Fault Severity" module utilizes four bearing vibration inputs to determine three fault severity outputs. Fuzzy logic membership functions and corresponding rulebase were developed for each input and output variable. An example of an output variable membership function is given in Figure 5.

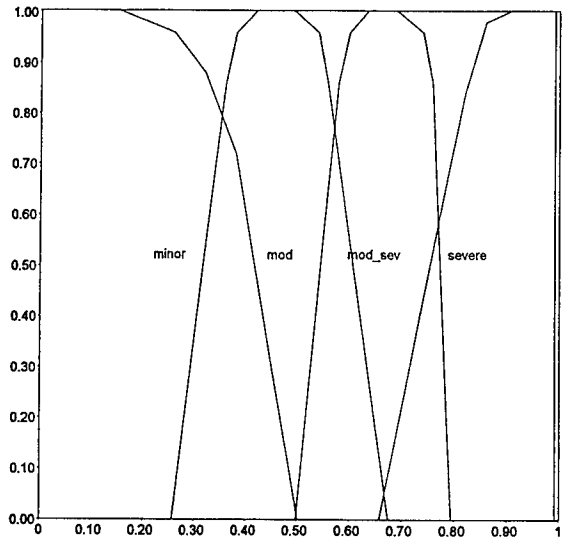


Figure 5 Fault Severity Membership Function

The fault severity membership functions were developed based on the experienced gained from operating the rotor demonstration rig. For example, the vibration amplitudes measured near bearing 1 were consistently higher than the levels measured near bearing 4. Therefore, the corresponding membership functions must represent this knowledge accurately. The rulebase developed from the vibration patterns exhibited from the unbalance and misalignment seeded faults were accurately represented with only 16 rules. A representative rule would be; (IF bearing_1 IS high) AND (bearing_2 IS low OR bearing_2 IS medium_low), THEN coupling misalignment severity is severe. The combination of the input/output membership functions and rulebase make up the knowledge of the fault severity module. A product-sum fuzzy inference method was used to scale and combine the membership functions, while the centroid technique was used for defuzzification.

The maintenance fuzzy logic module was developed similarly to the fault severity module except six input and three output membership functions were needed. The rulebase consisted of 22 rules that primarily examined the relationship between a diagnosed fault and corresponding severity level. An example rule is; IF unbalance_plane_2 IS high AND unbalance_2_severity is severe, THEN shut down balance plane 2. Maintenance module outputs for a plane 1 unbalance, plane 2 unbalance, or coupling misalignment are described as either "operational", "to be monitored", "to be monitored closely", or "shut down".

Component Life Module: A fatigue life algorithm was developed that utilizes the virtually sensed shaft stresses as a basis for computing fatigue initiation life. The algorithm estimates the amount of time to crack initiation, with crack propagation not being considered. Neuber's Rule is used to compute the true stress and strain in the crack initiation region. Morrow's Method is used to

incorporate the mean stress effects in the life calculations, which are based on strain-amplitude and the number of reversals. Miner's Law computes the cumulative fatigue damage.

Strain-Life Equation: The local strain approach was used to calculate the total strain, ε, including elastic and plastic components, from the given stress state and the fatigue properties of material:

$$\varepsilon = (\sigma'_f - \sigma_o)\,(2N_f)^b / E + \varepsilon'_f (2N_f)^c$$

where E is the elastic modulus, σ_o is the true mean stress or the true steady stress, and N_f is the number of cycles required for crack initiation. In the right hand side of the equation, the first term represents an elastic strain and the second term represents a plastic strain. This equation is the foundation for the cyclic strain-based approach to fatigue prediction and is usually called the strain-life equation.

Cycle Counting and Cumulative Damage: Under spectral loading, the dynamic strain conditions at critical locations of a component may have very complex waveforms. Several procedures exist to deal with this situation, of which, the Rainflow cycle counting procedure is well known. Simply stated, this procedure consists of dividing the complex waveform into a sequence of simple cycles, and then counting the number of strain cycles within a given strain range. The resulting number is then compared with the tested fatigue life of the material at this strain level to determine the degree of incremental damage. The best known cumulative damage assessment procedure is Miner's Law, which states that the cumulative damage is equal to the sum of the incremental damage at the various strain ranges.

$$\sum \frac{n_i}{N_i} = 1;$$

this procedure is utilized in this fatigue life algorithm. The number of cycles n_i occurring at a given strain level is first computed from the Rainflow cycle counting procedure. The number of cycles to failure at each strain level, N_i , is based on test sample data and adjusted for mean stress effects. This is obtained from the strain-life equation. The portion of damage at this strain level becomes n_i/N_i. The summation sign in the Miner's Law equation indicates that the cumulative damage is the sum of damage portion due to all existing strain levels. Accordingly, the crack initiation is expected to occur when the cumulative damage is equal to or greater than unity.

System Verification and Test Results: Several parametric tests were conducted involving seeded fault conditions applied to the demonstration rotor system. Accuracy of the neural network diagnostic outputs, fuzzy logic module outputs, and overall system outputs were compared with the "true" seeded fault condition. As an example, if a "minor" mass unbalance was imposed on disk #1, then the network diagnostic output should ideally respond "100 percent chance of unbalance on disk #1", the fault severity module respond as "severity is minor", and the overall system output should respond "monitor disk #1 unbalance.

To convey the diagnostic and life monitoring testing results most completely, a worsening unbalance condition was subjected to balancing plane 1 of the demo rotor rig. At first, sensor measurements were acquired from the rig under the absence of the unbalance condition. Next, data was acquired after a "minor" unbalance of 0.05 ounce-inches was applied to disk 1. Consecutive, unbalance magnitude increases were applied to disk 1 until a "severe" unbalance of 0.25 ounce-inches was reached. A sample of the output data file showing four discrete testing results is given in Figure 6.

The four discrete system outputs resulted from four sets of measurement data acquired from the rotor rig over a large time frame. During actual monitoring system operation, a system output describing unbalance, misalignment, and shaft life status is provided for every set of transducer measurements acquired. Reported system monitoring results can be provided at user specified intervals.

Several seeded fault conditions were examined similar to the one described above. The rotor monitoring system was capable of accurately tracking a worsening rotor fault condition as well as monitor remaining life of the rotor shaft. The system "fuzzy" outputs describe to a machine operator what steps he/she should take as a condition worsens.

```
----- Life Module Results:  Stress 1 = 6.787332   Stress 2 = 7.724948
-----
Coupling Alignment Is Operational
Balancing Plane 1 Is Operational
Balancing Plane 2 Is Operational
Remaining Shaft Life = 0.000000e+00  years
-----
----- Life Module Results:  Stress 1 = 86.611595   Stress 2 = 18.802980
-----
Coupling Alignment Is Operational
Balancing Plane 1 Should Be Monitored
Balancing Plane 2 Is Operational
Remaining Shaft Life = 3.784470e+31  years
-----
----- Life Module Results:  Stress 1 = 159.930435   Stress 2 = 38.713123
-----
Coupling Alignment Is Operational
Balancing Plane 1 Should Be Monitored Closely
Balancing Plane 2 Is Operational
Remaining Shaft Life = 7.918761e+16  years
-----
----- Life Module Results:  Stress 1 = 263.891205   Stress 2 = 61.637596
-----
Coupling Alignment Is Operational
Balancing Plane 1 Should Be Re-Balanced ASAP
Balancing Plane 2 Is Operational
Remaining Shaft Life = 2.503154e+15  years
```

Figure 6 System Testing Results

Conclusions: A fuzzy-neuro machine health monitor that performs both mechanical diagnostics and component life prediction was demonstrated with the use of a miniaturized rotor system. The rotor system was subjected to mechanical fault scenarios including; mass unbalances and coupling misalignments in order to examine the benefits of utilizing fuzzy-neuro technologies for monitoring rotating machinery. A neural network classifier was able to accurately diagnose mechanical faults based on the associated vibration signatures measured from the desktop system. Fuzzy logic was used to determine fault severity levels and make decisions on required maintenance action. In addition, component life accumulation was monitored based on the diagnostic outputs provided by the neural network

The concept of training neural network classifiers with both rotor system measurements and detailed finite element models is highlighted as a significant advancement in condition monitoring applications. The rotor dynamics finite element model was used to train the diagnostic network to recognize fault patterns and their resulting effect on shaft stresses. This real-time, "virtual" sensing of shaft stresses allows for component life monitoring to be achieved in real-time. In other words, the ability of the neural networks to recognize particular vibration signatures and correlate them with associated shaft stresses is of particular significance.

Finally, through the use of fuzzy logic, the diagnostic system provided the necessary robustness to measurement noise and changing failure mode vibration patterns. In addition, by processing the raw transducer inputs in both the diagnostic neural network and fuzzy logic scheme simultaneously, increased accuracy of the diagnosed faults and corresponding corrective action recommendations was accomplished while preserving the real-time requirement of the system.

References

1. Kosko, B., "Neural Networks and Fuzzy Systems", Prentice-Hall, NJ, 1992.

2. Hong, C. A., "Development of Advanced Rotordynamics Software for Root Cause Investigations", STI Report PA791, 1993.

3. Widrow, B., "DARPA Neural Network Study", AFCEA International Press, 1989.

4. Aleksander, I., and Morton, H., "An Introduction to Neural Computing", Chapman and Hall Publishing, 1990

5. Dietz, W.E., Kiech, E.L., and Ali, M., "Jet and Rocket Engine Fault Diagnosis in Real Time", Journal of Neural Network Computing, 1989.

6. Kandel, A., Langholz, G., "Hybrid Architectures for Intelligent Systems", CRC Press, Boca Raton, FL, 1992.

7. Zadeh, L. A., "Fuzzy Sets", Information and Control, vol. 8, pgs. 338-353, 1965.

SYNC-PERIOD FREQUENCY ANALYSIS AND ITS APPLICATION TO THE DIAGNOSIS OF MULTIPLE ELEMENT DEFECTS OF ROLLING BEARINGS

Wen-Yi Wang Michael J. Harrap

Acoustics and Vibration Center
Department of Aerospace and Mechanical Engineering
University College of UNSW
Australian Defence Force Academy
Campbell, ACT 2600
Australia

Abstract: A method for diagnosing multiple element defects of rolling bearings has been investigated. The method combines the time synchronous averaging and envelope spectral analysis techniques to produce spectra of synchronously averaged envelope signals with a range of synchronous frequencies. The spectra are displayed in the domains of synchronous period versus frequency to result in the sync-period frequency distribution. The distribution separates the characteristic defect frequencies and their associated sidebands in the synchronous period axis. This analysis technique makes it possible to detect and diagnose multiple defects appearing on different elements of rolling bearings. Another main benefit of the method is the significant noise reduction by both enveloping and synchronous averaging process. Analysis results from computer simulated data and experimental simulated data are presented.

Key Words: Defects diagnosis; Multiple defects; Rolling bearings; Synchronous average; Sync-period frequency analysis; Vibration analysis.

1. Introduction

Time synchronous averaging is a powerful tool to extract periodic waveforms from noisy signals. It has been used extensively in diagnosing gear defects. However, this technique has not found wide acceptance for diagnosing defects of rolling bearings, because of the asynchronous rotation [1] between the shaft and the rolling elements or cage of the rolling bearing, and the approximation of bearing characteristic defect frequencies [2]. On the other hand, the high frequency resonance technique (HFRT) is widely accepted in bearing defect diagnosis. For single defect cases, it has proven successful under reasonable signal to noise ratio (*SNR*) conditions, provided that the shaft is running at a constant speed and there exists no skidding of the rolling elements. But this

technique has proven less than satisfactory in detecting bearing defects under poor *SNR* conditions [3]. Also, to the authors knowledge, the successful application of this technique to the diagnosis of multiple element defects in bearings has not been reported.

In diagnosing multiple faults, Igarashi and Kato [4] investigated the vibration of a ball bearing with multiple defects on the race surface of **either the inner or outer ring** under given thrust loads. Using an FFT analyzer and examining the effect of the measuring position of the outer ring on the vibration, they clarified the fundamental characteristics of ball bearings with multiple faults, and they also established an approach to locate the faults and indicate the number and size of the defects. Mcfadden and Smith [5] described an investigation of a bearing with multiple point defects based on their model [6] of a single point defect. The vibration envelope spectra produced by a bearing with **multiple inner race defects** were illustrated by the reinforcement and cancellation of spectral lines due to the different phase angle of the vibrations induced by different faults. Braun and Datner [7] discussed a bearing fault diagnostic method based on time averaging. The RMS level of the averaged signal was utilized to indicate the relevant defects according to their characteristic defect frequencies. This method effectively extracts the periodic signals associated with bearing faults from noisy structural vibrations and is less sensitive to structure parameters than general spectral techniques. However, the RMS approach seems not capable of identifying which frequency components contribute the most to the RMS value, because one averaging period could involve several frequency components. Moreover, due to the low energy of the vibration caused by bearing faults, the characteristic defect frequencies may be very difficult to be discerned when the shaft rotation related frequency components (high energy) are closely adjacent to those defect frequencies.

The high frequency resonance technique is effective in excluding the shaft rotation related frequency components, but it has no advantage in separating multiple defect frequencies and eliminating high frequency noise disturbances. In cases where bearings have multiple defects located in different elements, each defect may have different shape and size. Thus each defect may excite a different resonance (structural and/or transducer's) to a different extent. Using the ordinary high frequency resonance technique, only one resonance is isolated by a bandpass filter for envelope detection. Therefore, where multiple faults exist, multiple defect frequency components are likely to be mixed together, provided all defects excite the same resonance to a comparable extent. Alternatively, if a particular resonance is dominated by one fault excitation, then only that fault will be distinguishable.

In this paper, we investigate a technique which introduces an averaging process into the high frequency resonance technique and expands the bandwidth of the bandpass (or even high pass) filter to include several high frequency resonances rather than just a single isolated resonance. The time synchronous averaging process is performed after the envelope detection. Due to the uncertainty of the bearing's characteristic defect frequencies, the synchronous frequency is swept around the nominal defect frequency until the true frequency is identified. Spectral analysis is then carried out for the resulting averaged envelope signal. The spectra are finally displayed in the so-called *syncperiod - frequency domain* to form the *syncperiod - frequency (SPF) distribution*. In SPF distributions, the individual characteristic defect frequencies and their associated modulation sidebands are well separated along the syncperiod axis. The exact frequency content

can then be acquired so as to diagnose the condition of the bearing with multiple element defects. It is shown that the enveloping and synchronous averaging processes used in this method significantly improve the signal to noise ratio.

2. Sync-Period Frequency (SPF) Analysis

In this section, a method which combines the time synchronous averaging technique and the high frequency resonance technique is described. From the high frequency resonance technique, we know that the envelope signal generated by localized bearing defects can be featured as a series of exponentially decayed impulses with the periods related to various defects [6, 7]. For the remaining part of this section, we are only dealing with the envelope signals.

2.1. Time Synchronous Averaging of a Bearing Signal and Its Spectrum

Using standard envelope detector, the raw signal is firstly bandpass filtered in the high frequency resonance region. The resulting signal consists of the bearing defect induced signal embedded in band-limited Gaussian noise. Suppose $x(t)$ is such an envelope signal, and it contains frequency components corresponding to the harmonics of the characteristic defect frequencies and their associated sidebands. It is decomposed into N segments, each of same length T_s. Each segment is separated by a variable period τ ($\tau \geq T_s$ – ie. no overlap). Summing up all these segments yields:

$$y(t,\tau) = \sum_{n=0}^{N-1} x(t+n\tau), \qquad t \in [0, T_s] \tag{1}$$

where τ is the averaging period, T_s is the segment length. According to the time-shift theorem of Fourier transform, the spectrum of (1) $Y(f,\tau)$ will be [8,9]

$$Y(f,\tau) = X(f) \cdot \sum_{n=0}^{N-1} e^{j2\pi f n\tau} = X(f) \cdot \frac{\sin N\pi f\tau}{\sin \pi f\tau} \cdot e^{j(N-1)\pi f\tau} \tag{2}$$

the amplitude spectrum will then yield

$$|Y(f,\tau)| = |X(f)| \cdot \left| \frac{\sin N\pi f\tau}{\sin \pi f\tau} \right| \tag{3}$$

which is the original amplitude spectrum $|X(f)|$ multiplied by the transfer function of a comb filter $|H(f)| = |\sin(N\pi f\tau)/\sin(\pi f\tau)|$. By examining the nature of the function $|H(f)|$ shown in Figure 1 [8, 9], we can see that **a)** it is a periodic function with the period of $f = 1/\tau$; **b)** for large N the function appears like a train of sinc functions and it becomes an impulse train when N goes to infinity, with period $f = 1/\tau$; **c)** its main lobes peak amplitude is N, they are located at $f = n/\tau$ ($n=0,1,\ldots,N-1$), its first zero is at $f = 1/N\tau$ and thus the main lobes have the width of $B=2/N\tau$; **d)** its main lobe $3dB$ point is at $0.45/N\tau$, and thus $3dB$ bandwidth is $B_{3dB} = 0.9/N\tau$; **e)** its side lobe peaks decrease with increasing N, the maximum side lobe peak is $12dB$ down from the main lobe peak when $N=5$, and $13dB$ down when $N=20$, it approaches the limit value of $-13.3dB$ as N increases.

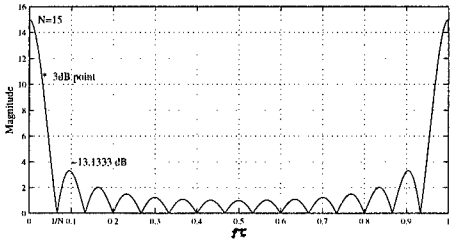


Figure 1. One period of function $|\sin(N\pi f\tau)/\sin(\pi f\tau)|$ [9]

As we mentioned early, $|X(f)|$ has the frequency components of mf_o, $mf_i \pm kf_r$, $mf_b \pm kf_c$ (m,k=0,1, 2,...) where f_o, f_i, f_b denote the characteristic defect frequencies of outer race, inner race and ball/roller respectively, f_r denotes shaft rotation frequency, f_c cage frequency. The relative amplitude of these components depend upon the damping factor of the bearing system, the load distribution and the variation of transmission path [1]. When the synchronous period τ equals an integer multiple M of the interval corresponding to one of the above frequencies, for instance $\tau=MT_o=M/f_o$, the frequency components from both signal and noise (except f_o and its harmonics) are rejected and a periodic signature with a period of T_o is extracted by the averaging process of (1) (see [10]). Therefore, the spectrum $|Y(f,\tau)|$ in (3) will be dominated by f_o and its harmonics.

2.2. Signal Processing Considerations

Due to the uncertainty of the characteristic defect frequencies of the bearings, it is necessary to identify the true synchronous period by searching around the nominal frequencies. The signal processing scheme that is used to find the true synchronous period is based on a search strategy (Figure 2).

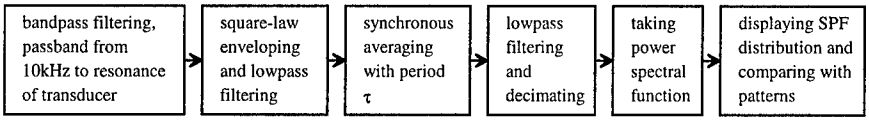


Figure 2. The flowchart of Sync-Period Frequency Analysis

In implementing the above scheme digitally, several factors have to be considered. In this section, we will discuss the effect of sampling rate f_s, data segment length T_s, synchronous period τ, the number of averages N and the search spacing $\Delta\tau$ on the performance of SPF analysis. For clarification, we tabulate all factors concerned in SPF analysis as following:

f_s	sampling frequency, $\Delta\tau_{min} = 1/f_s$
T	total data length
T_s	segmental data length
B	main lobes bandwidth of the comb filter, $B=2/N\tau$, $\Delta f \geqslant B$
B_{3dB}	main lobes 3dB bandwidth of the comb filter, $B_{3dB} = 0.9/N\tau$

f^*	defect frequency components to be detect
Δf	frequency accuracy, defined as $\Delta f = M[1/\tau - 1/(\tau+\Delta\tau)]$, $(\Delta f \simeq f\,\Delta\tau/\tau$, if $\Delta\tau << \tau)$
df	frequency resolution from the Fourier transform, $df = 1/T_s$
M	integer multiples of the defect periods
τ	synchronous period, $\tau = M/f^*$, $\tau \geqslant T_s$ (τ and T_s are in an equivalent order)
$\Delta\tau$	sync-period search spacing, $\Delta\tau/\tau = \Delta f/f^*$ (if $\Delta\tau << \tau$), $\Delta\tau \geqslant 1/f_s$
N	number of averages, $N = \text{mod}\,(T, \tau)$

To ensure the correct detection of frequency components, we are particularly concerning about the search spacing $\Delta\tau$ with respect to the frequency accuracy Δf (frequency resolution obtained from the syncperiod). Large $\Delta\tau$ may fail to identify the bearing defect frequency, whereas too small $\Delta\tau$ may cause a tremendous increase of computing time. In general, $\Delta\tau$ is determined such that the frequency accuracy Δf is larger than (or equal to) the bandwidth [10] of comb filter's main lobes shown in Figure 1. According to the requirement of the SPF analysis, factors like f_s, f^*, $\Delta\tau$, Δf and B are primarily selected, and other factors are determined by the formulas in the above table. Figure 3 depicts how non-dimensional measures $\Delta f/f^*$, $\Delta\tau/\tau$, $\Delta\tau f$, $\Delta f\tau$, $\Delta f/B$, N and M are related. The calculated factors could be verified by these non-dimensional graphs.

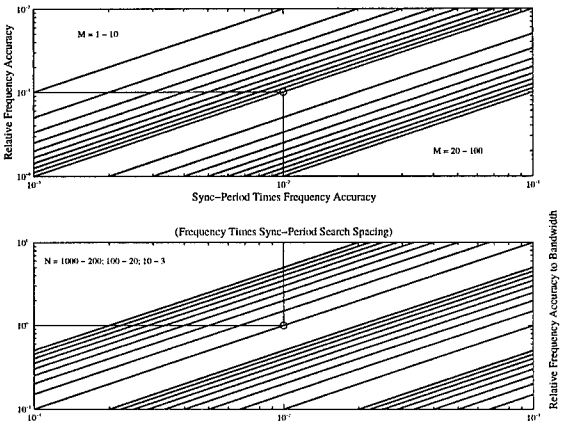


Figure 3. The relationship of $\Delta f/f^*$, $\Delta\tau/\tau$, $\Delta\tau f^*$, $\Delta f\tau$, $\Delta f/B$, N and M. (**a**) $\Delta f/f^*$ vs. $\Delta\tau f^*$(or $\tau\Delta f$), both in logarithm scale, $\Delta f/f^* = \Delta\tau f^*/M$; (**b**) $\Delta f/B$ vs. $\Delta\tau f^*$ (or $\tau\Delta f$), both in logarithm scale, $\Delta f/B = \Delta\tau f^* N/2$.

For example, assuming the sampling rate is 100kHz which leads to a $\Delta\tau_{min}$ of 0.01ms, if we select $\Delta\tau$ = 0.04ms and Δf = 0.25Hz, then to detect a frequency component of 250Hz would require M to be 10, and τ to be 40ms. Furthermore, if we select $B = \Delta f$ = 0.25Hz, N would be 200 and T would be 8 s (800000 samples).

In detecting multiple frequency components, we need to sweep the synchronous period τ in a wide range. We can keep $\Delta\tau$ unchanged during the sweeping while using different M 's to associate with multiple f^* 's. In the above example, if there are three frequency components to be detected, eg. 152Hz, 198Hz, 247Hz, M's would be 4, 6 and 10 respectively, and the sweeping range of τ would be 26.32ms ~ 40.5ms. To ensure a successful search, the range could be expanded to 24ms ~ 43ms. If a fixed averaging number N (eg. 200) is selected, the frequency bandwidth B will vary from 0.42Hz to 0.23Hz, and Δf from 0.23Hz to 0.24Hz. Provided the range of $\Delta f/B$ is acceptable, fixed N is more convenient for sweeping. In this regard, the frequency resolution for low frequency components will be poorer due to the smaller $\Delta f/B$.

Following the averaging process, the frequency spectrum of the averaged envelope signal is estimated. The FFT associated frequency resolution of the spectrum df is therefore another concern. However, during sweep synchronous averaging the sync-period τ ($=M/f$) can give a much higher frequency resolution than what could be achieved through spectral analysis. The required df should be high enough to allow individual frequency components to be distinguished. For the above example, the spectral resolution df is about 38Hz ~ 25Hz ($1/\tau$).

3. SPF Pattern of Multiple Bearing Defects

The SPF patterns provide a graphical exhibition of the frequency composition of a signal in the frequency-syncperiod domain. In this section, we present a SPF pattern for a combination of different types of element defects using nominal characteristic defects frequencies [7, 11] of a bearing. The main uncertainty in estimating the characteristic frequencies is the contact angle between rolling elements of the races, which is dependent on the type of bearing and loading conditions. Another source of uncertainty is the diameter variations of the balls.

In the cases where multiple defects in different bearing elements exist and/or the *SNR* is poor, envelope spectra [6] prove unsatisfactory for the identification of individual bearing defect frequency components. SPF patterns indicate where the characteristic defect frequencies and their sidebands are located in a two dimensional domain (notice that the relative intensity is not shown), referred here as *'syncperoid - frequency domain'*. Syncperiod axis represents the sweeping range for the synchronous period τ which is an integer multiple of the characteristic defect periods. the frequency axis is used to define the frequency contents corresponding to a given the syncperiod. In fact, with small $\Delta\tau$ the syncperiod τ contains much finer frequency resolution than frequency axis. However, without the frequency axis, the syncperiod along would not be able to distinguish which frequency components this period multiple ($\tau=M/f^*$) is associated with. For instance, a syncperiod of 24ms could be either 6 times 4ms (250Hz) or 4 times 6ms (166.67Hz).

Now we consider a rolling bearing (NSK EN202) with different types of defects and present the SPF pattern associated with a combination of various defects. This type of bearing is deliberately chosen for the near harmonics of various defect frequencies, ie. $3f_r$ and f_o, f_i-f_r and f_b, $2f_o$ and f_i +f_r etc. are closely adjacent, which makes the diagnosis of defects more complex. Dimensions and defect frequencies of the bearing are listed below. The SPF pattern is shown in Figure 4.

ball number	ball diameter	pitch diameter	contact angle	shaft speed f_r	outer race frequency f_o	inner race frequency f_i	ball frequency f_b	cage frequnecy f_c
8	6.54mm	27.5mm	0°	50Hz	152.47Hz	247.53Hz	198.52	19.06Hz

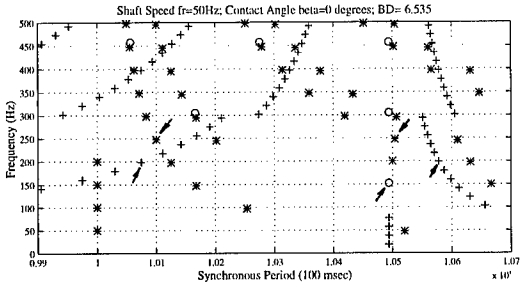


Figure 4. SPF pattern for a combination of outer race, inner race and roller/ball defects ('**o**' – outer race defect related components; '*' – inner race defect related components; '+' – ball/roller defect related components; all arrows are pointing at the fundamental defect frequencies)

4. Diagnosis of Multiple Defects Located in Different Bearing Elements Using SPF Analysis

In this section, we will discuss the application of the SPF analysis to detect multiple defects in different bearing elements. Results from both computer synthesised and experimental simulated data are described.

4.1. Diagnosis of Multiple Defects Generated by Models

The synthetic vibration signal used in this section is a bearing signal generated via Braun's model [2, 7] plus other vibration sources (eg. unbalanced shaft, etc.) and noise. The following example presents the SPF distribution from synthetic signal generated by the model. It consists of an impulse series induced by all three types of defects, shaft rotation related vibrations (narrow band but high power) and broadband white noise. The maximum impact amplitudes are 35, 90, 100 for the outer race defect, inner race defect and ball defects respectively. The noise standard deviation σ_n is 30. It is also assumed in our synthetic data that three different types of defects excite three different resonances, ie. f_1=25070Hz, f_2=18003Hz, f_3=21110Hz, and with different time constants of exponential decay, α_1=1/15, α_2=α_3=2/15, respectively. Using the SPF analysis described in section 2, the SPF distribution of the synthetic signal plus broadband white noise is generated and displayed in Figure 5. All factors used are tabulated below:

df	f_s	$\Delta f <$	$B <$	$T_s = \tau_{min}$	N	T	$\Delta\tau = \Delta\tau_{min}$	τ	M
10Hz	100kHz	0.03Hz	0.2Hz	100 ms	100	10.48576 s	0.01 ms	100.01~107 ms	16(f_o); 25,26(f_i); 20,21(f_b)

Comparing this SPF distribution with the SPF pattern shown in Figure 4, it is found that the characteristic defect frequency components and their associated modulation sidebands are well separated in syncperiod - frequency domain. For a given synchronous period, only one frequency component and its harmonics are exposed. It can be seen from the SPF distribution that the lower frequency components are more widely spreaded over the syncperiod domain. The noise level

has been significantly attenuated through enveloping (3dB improvement) [12] and averaging (a factor of $\sqrt{N}$) [10,13].

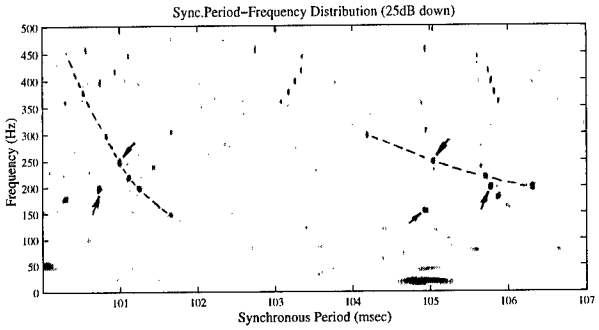


Figure 5. SPF distributions from synthetic signal plus broadband noise with σ_n=30

In fact, the original signal for Figure 5 has a very poor signal to noise ratio (-15.5 dB). From time domain, there is no evidence of any characteristic defect frequency related components because of high level noise disturbance and the intertwinement of multiple frequency components. Using the envelope spectral technique (Figure 6), it is also very difficult to identify individual characteristic defect frequencies.

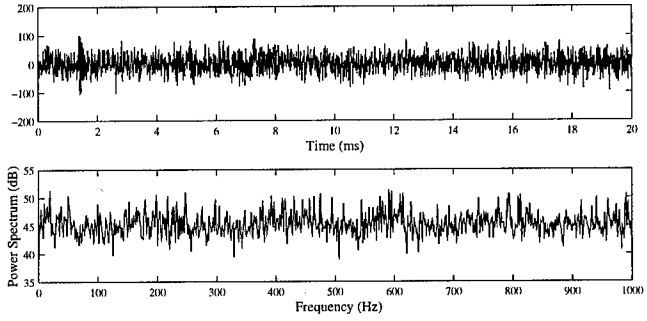


Figure 6. **(a)** Original time signal for Figure 5 and **(b)** the envelope spectrum (1024 averages) of (a)

4.2. Diagnosis of Multiple Defects With Experimental Data

Two NSK EN202 bearings were used for this experimental investigation. The artificial defects were implanted by either nitric acid corrosion or electric discharge erode. Bearing #1 has one defect on **both outer race and inner race** (nitric acid corrosion - about 2~3 mm in diameter, depth unknown but very shallow). Bearing #2 has one ball defect implanted (electric erode - diameter 0.38mm, depth 1mm), and its cage was dismantled and re-riveted after the introduction of the defect. The analysis results are shown in Figure 7 and 8. All SPF analysis parameters are identical to the examples in section 4.1 except the range of syncperiod τ for bearing #2 (from 99.01~107 ms). The experimental conditions are listed in the following table:

shaft speed	transducer	load applied	sampling rate	f_o, f_i, f_h, f_c
50 Hz	B&K 8307 (#1) B&K 4393 (#2)	240N radial (#1) 260N radial (#2)	100 kHz	152.47, 247.53, 198.52, 19.06 Hz (nominal defect frequencies)

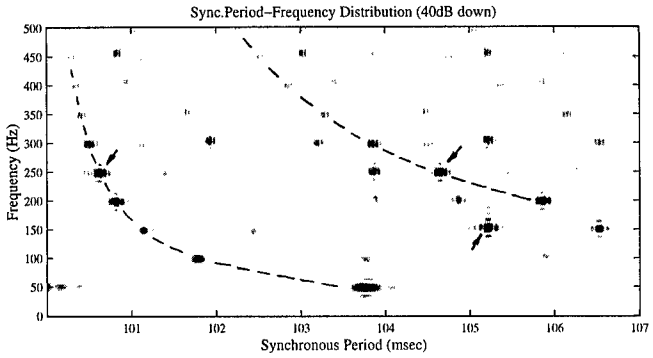


Figure 7. SPF distribution of Bearing #1, the signal was highpassed at 10kHz before enveloping.

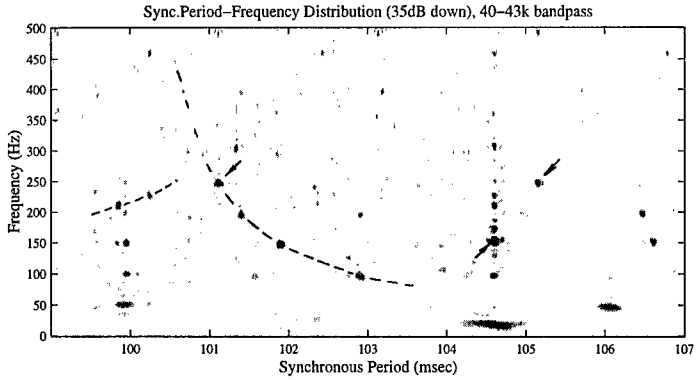


Figure 8. SPF distribution of Bearing #2, the signal was bandpassed at 40~43kHz before enveloping

5. Discussions and conclusion

By baseband spectrum analysis of vibration signal of bearing #1, it is found that a series of high frequency resonances have been excited. Envelope demodulated spectra at several resonances show both inner race and outer race defective frequencies and associated shaft rotation frequency, but all those frequency components are intertwined together. The SPF distribution shown in Figure 7 separates those frequencies along the syncperiod axis and shows the defective frequencies and the modulation sidebands explicitly. The outer race defect frequency is derived from the point at 105.23ms on syncperiod axis and about 150Hz on frequency axis, which corresponds to 152.05Hz (M is 16 here). The inner race defect frequency (248.43Hz, M is 25) is corresponding to the point at 100.63ms and 250Hz. The SPF distribution of bearing #2 displays a multiple defect pattern rather than a single ball defect pattern. The indication of inner race and

outer race defects are very strong. This curious phenomenon suggests that inner race and outer race defects may have been produced during dismantling and re-assembly of the bearing. The envelope spectrum from a low frequency resonance (2kHz ~ 4kHz) reveals inner race fault related components. The outer race defect frequency dominates the envelope spectrum demodulated about a 26k~28kHz resonance. Ball defect components could be seen from the envelope spectra of 2kHz highpassed and 17k~23kHz bandpassed signals. The SPF distribution from the 40~43kHz bandpassed signal seems to give the best presentation of all three types of defects. By inspecting Figure 8 closely, the ball defect frequency and its sidebands are only discernible with ease around 100msec region due to the small size of the defect. The cage component at around 104.65ms, which corresponds to 19.11Hz (M=2), is an important indication of ball damage.

In Figure 5,7,8, we notice that the low frequency components tend to spread over a wide syncperiod region. The reason for that may be explained by *uncertainty principle* [13]. Because the syncperiod domain is virtually a frequency scale, a certain length of data segment contains less cycles for low frequencies than those for high frequencies, which results in poorer frequency resolution.

The above discussion suggests that the syncperiod frequency analysis combines the benefits of both high frequency resonance technique and time synchronous averaging technique in overcoming problems like low-duty bearing related vibration and asynchronous rotation of rolling element bearings. It is therefore concluded that the syncperiod frequency analysis is potentially an effective method for the diagnostics of multiple defects presented in different bearing elements.

References

1.P.D. Mcfadden. "Condition Monitoring of Rolling Element Bearings by Vibration Analysis". Machine Condition Monitoring (Papers presented at a Seminar) 1990, p49-53. Mechanical Engineering Publications Limited for the Institution of Mechanical Engineers, London.

2.S. Braun. "The Signature Analysis of Sonic Bearing Vibrations". IEEE Transactions on Sonics and Ultrasonics SU-27 (5) 1980, p317-328.

3.W.Y. Wang and M.J. Harrap. "Fault Diagnosis of Ball Bearings Using Autocorrelation Technique". Proceedings of Vibration and Noise '95 (Venice, Italy), p36-44. Edited by M.J. Doogwin, Published by Staffordshire University.

4.T. Igarashi and H. Hamada. "Studies on the Vibration and Sound of Defective Rolling Bearings (Third Report: Vibration of Ball Bearings with Multiple Defect)". Bulletin of the JSME 28 (237) 1985, p492-499.

5.P.D. Mcfadden and J.D. Smith. "The Vibration Produced by Multiple Point Defects in a Rolling Element Bearing". Journal of Sound and Vibration 98, 1986, 263-273.

6.P.D. Mcfadden and J.D. Smith. "Model for the Vibration Produced by a Single Point Defect in a Rolling Element Bearing". Journal of Sound and Vibration 96, 1984, p69-82.

7.S. Braun and B. Datner. "Analysis of Roller/Ball Bearing Vibrations". Journal of Mechanical Design 101, 1979, p118-125.

8.W. L. Stutzman and G. A. Thiele. Antenna Theory and Design, pp124-133. New York: John Wiley & Sons, Inc. 1981.

9.W.S. Burdic. Radar Signal Analysis, pp92-128. Englewood Cliffs, N.J.: Prentice-Hall Inc. 1968.

10.S. Braun. "Extraction of Periodic Waveforms by Time Domain Averaging". Acustica 32, 1975, p69-77.

11.T.A. Harris Rolling Bearing Analysis, pp229-252. John Wiley & Sons. 1984.

12.J.A. Betts Signal Processing, Modulation and Noise, pp97-115. London: The English University Press Limited. 1971.

13.F.D. Coulon. Signal Theory and Processing, pp80-121,162-178. Dedham, M.A.: Artech House Inc. 1986.

Coherent Phase Wide Band Demodulation Technique for Turbomachinery Cavitation Detection and Monitoring

Jen Jong
AI Signal Research, Inc.
3322 South Memorial Parkway, Suite-67
Huntsville, AL 35801

Tom Nesman, Wayne Bordelon, Jess Jones, Tom Zoladz
NASA, Marshall Space Flight Center, ED31
MSFC, AL 35812

ABSTRACT: This paper discusses a Coherent Phase Wide Band Demodulation (CPWBD) technique for turbopump cavitation detection and monitoring. The principle of cavitation detection is based on the unique nonlinear modulation phenomenon associated with pump cavitation in which the periodic shaft rotational motion (and/or its harmonic motion) amplitude modulates the wide-band noise generated from the collapse of cavitation bubbles. However, these periodicities associated with shaft rotation are hidden during the physical wide band modulation (WBM) process and cannot be identified within the conventional power spectral density (PSD) of the monitored dynamic signal. Such periodicity, hidden well within wide band noise, thus provides a unique signature and serves as the basis for effective cavitation detection. Existing techniques for cavitation detection utilize envelope analysis or full-wave rectification (FWR) spectral analysis to demodulate dynamic measurement signals over an isolated band of high frequency wide band noise in order to recover hidden cavitation-generated periodicity. However, the conventional FWR technique is subject to a practical limitation in that any discrete components present in the high frequency regions of interest, due to other vibration sources (mechanical, structural, etc.), will erroneously generate discrete peaks in the resulting demodulated signal appearing as genuine recovered hidden periodicities. The CPWBD technique provides an effective alternative to avoid such discrete interference. The method is based on an inherent signal property associated with the WBM signal, where a unique coherent phase relationship exists among all interacting spectral components associated with the cavitation process. The CPWBD technique can thus identify hidden periodicity by searching for such a coherent phase relationship. An inherent characteristic of the CPWBD is that it is not affected by any linearly superimposed discrete components, and as a result, overcomes the limitation due to discrete interference. Application examples with Space Shuttle Main Engine (SSME) turbopump test data are utilized to demonstrate the effectiveness of the CPWBD for cavitation detection.

KEY WORDS: Cavitation Detection, Demodulation, Impeller, Inducer, Modulation, Turbopump

INTRODUCTION: Cavitation within turbopumps can degrade pump performance and cause excessive vibration along with pump blade pitting. This phenomenon occurs when local fluid pressures in the accelerated flow areas around and within rotating pump blades drop below the vapor pressure. The damaging effects of cavitation are associated with the pressure shock waves generated by the collapse of the vapor bubbles, which impact the adjacent metal surfaces to cause pitting and erosion. The collapse of cavitation bubbles also generates broad band noise in the surrounding fluid and structure, with noise levels proportional to intensity of the cavitation. The principle for cavitation detection is based on the nonlinear modulation phenomenon between the bubble collapse wide-band noise with some synchronous (RPM or harmonics) and/or non-synchronous periodic components. However, the periodicities modulating with the wide band noise become hidden in the resulting WBM signal and cannot be identified with the conventional power spectral density (PSD) function. Such hidden periodicity thus provides a unique signature in the wide band noise conducive to cavitation detection.

Existing techniques for cavitation detection utilize envelope analysis or full-wave rectification (FWR) spectral analysis [1] to demodulate the dynamic measurement signals over an isolated band of high frequency wide band noise in order to recover any cavitation-generated hidden periodicity. The magnitude of the recovered hidden periodicity is then used to quantify cavitation intensities; while the frequency of the recovered hidden periodicity is used to identify various types of cavitation under different pumping conditions. Other experimental work has suggested that [2, 3] cavitation erosion results from the collapse of swirling transient vortices, originating at the leading edge of the hydrofoil. These vortices were found to shed periodically according to a constant Strouhal number. Also observed was that when erosion levels were high a strong modulating frequency was recovered at the periodic vortex shedding rate.

Experimental study of cavitation in hydroturbines [4] indicates that the recovered modulating frequency is found to be the rotating turbine blade/wicket-gate passing frequency when cavitation develops on all of the turbine blades (full blade cavitation). A recent Alternate Turbopump Development (ATD) High Pressure Oxygen Turbopump (HPOTP) inducer flow test series using water conducted at NASA/MSFC to investigate cavitation-induced vibration [5, 6] observed a fairly repeatable sequence of cavitation. At relatively high inlet pressure, small cavitation clouds are present behind each of the four inducer blades (full blade cavitation). The second type of cavitation occurred when the inlet pressure was lowered until two symmetric cavitation clouds appeared behind alternate blades (alternate blade cavitation). A third type of cavitation occurred just before head fall-off when two alternate blade cavitation clouds of different size formed in the inducer. In each case, distinct pressure oscillation frequencies are associated with each type of cavitation. Dominant hidden periodicities recovered from the demodulated

signal were found to be 4N (4th harmonic of RPM), 2N and 1N frequencies respectively for each of these three cavitation types [7, 8]. Another recent SSME ATD (High Pressure Fuel Turbopump) HPFTP impeller flow test series at NASA/MSFC produced cavitation-induced oscillations at synchronous frequency multiples, as well as at non-synchronous anomalous frequencies, as narrow band random signals, and as broadband random noise. Furthermore, modulation between the non-synchronous components and the synchronous frequency multiples, which generated sideband signatures, was also observed.

These experimental observations reported by various researchers indicate that cavitation-generated modulation phenomenon provides a unique signature to monitor the cavitation process within a wide variety of pumping systems. In addition, the detection of modulating periodic frequency or hidden periodicity within the wide band modulation signal provides a way to identity the type of cavitation under various pumping conditions. Recovery via demodulation of such hidden periodicity thus forms the basis of signal processing for cavitation detection and monitoring. Envelope or full-wave rectification spectral analysis is subject to a practical limitation in that any discrete components present in the high frequency regions of interest, due to other vibration sources (mechanical, structural, etc.), will erroneously generate discrete peaks in the resulting demodulated signal appearing as genuine recovered hidden periodicities. This limitation is critical in assessing the performance of rocket engines under severe operational environments in which other vibration sources contribute many high frequency components which can corrupt the demodulation signal.

The Coherent Phase Wide Band Demodulation method discussed in this paper provides an alternative for wide band demodulation which can avoid such discrete interference. The CPWBD method is based on the signal property within an ordinary (linear) wide band noise signal, spectral components at different frequencies are statistically independent. However, within a signal composed of wide band modulation (WBM) noise coupled with a hidden periodicity, a unique coherent phase relationship exists among all interacting components. The CPWBD technique can thus identify hidden periodicity by searching for such a coherent phase relationship. An inherent characteristic of the CPWBD is that it is not affected by any linearly superimposed discrete components since only phase information is utilized in searching for cavitation signatures. As a result, CPWBD provides an effective way to overcome the critical limitation of discrete interference suffered by the conventional FWR method. The CPWBD technique utilizes a Phase-Only (PO) filter coupled with an envelope detector to search for the unique cavitation-generated coherent phase relationship. Therefore, the CPWBD method offers the unique ability to accurately detect the cavitation condition without being subject to high frequency discrete interference. Application examples with the Space Shuttle Main Engine (SSME) turbopump test data are utilized to demonstrate the effectiveness of the CPWBD for cavitation detection.

WIDE-BAND MODULATION (WBM) SIGNAL: The signal processing techniques, upon which effective cavitation detection and monitoring is based, rely a unique phenomenon associated with cavitation physics:

When cavitation occurs in a rotating system, some periodic components related to the shaft rotational motion will amplitude modulate the wide-band noise generated from the collapse of cavitation bubbles.

A simplified signal model for a Wide-Band Modulation signal can be formulated as a periodic wave multiplied by a wide-band noise signal N(t):

$$x(t) = [\ 1 + r \cos(\omega_r t)]\ N(t) \qquad (1)$$

where $\cos(\omega_r t)$ represents some periodic motion associated with shaft rotational motion, and N(t) is zero-mean Gaussian White or Color Noise. A special kind of periodicity exists in such a WBM signal but is well hidden. An ordinary power spectral density (PSD) will not show a discrete peak at the frequency of the periodic component. This can be easily deduced from studying its signal model. The operation between N(t) and $\cos(\omega_r t)$ in the time domain is multiplication, but this multiplication becomes a convolution in the frequency domain. Since the PSD of noise is flat and that of a sine wave is a delta function, the convolution of these two PSD functions remains flat without any discrete peak. For this reason, a conventional PSD is unable to identify periodic components hidden within a Wide-Band Modulation signal.

FULL-WAVE RECTIFICATION WIDE-BAND DEMODULATION: The Full-Wave Rectification Wide-Band Demodulation (FWRWBD) technique for demodulating a WBM signal can be better understood from the tri-spectral approach. Dwyer [9] proposed a special tri-spectral function to identify the existence of hidden periodicity in underwater sonar signal processing. This tri-spectrum is based on the Fourier Transform of a special tri-correlation function $R_{XXXX}(\tau)$ of a measured signal x(t), which is defined as:

$$R_{XXXX}(\tau) = E[\ x(t)\ x(t)\ x(t+\tau)\ x(t+\tau)\ ]$$

$$= E[\ x^2(t)\ x^2(t+\tau)\ ] = R_{yy}(\tau) \qquad (2)$$

Where $y(t) = x^2(t)$

The special tri-spectrum $T(\omega)$ is defined as the Fourier Transform of $R_{XXXX}(\tau)$. Notice that, the function $R_{XXXX}(\tau)$ reduces to the ordinary auto-correlation of y(t), where y(t) is simply the square of the original signal x(t). In other words, this special tri-spectrum is equal to the ordinary PSD of the square of the original signal. By examining the WBM

signal model, it can be easily seen why such a simple squaring operation can recover the hidden periodicity in a WBM signal. The square of x(t) can be written as:

$$x(t)^2 = [\,1+ r\,\mathrm{Cos}(\omega_r t)\,]^2\ [\,N(t)\,]^2$$
$$=[\,1.5 + 2r\,\mathrm{Cos}(\omega_r t) + 0.5\,r^2\,\mathrm{Cos}(2\,\omega_r t)\,]\,[DC + N'(t)\,] \quad (3)$$

Where N'(t) is defined by the following relationship:

$$[\,N(t)\,]^2 = [DC + N'(t)\,] \quad (4)$$

When squaring a zero-mean noise signal, an additional bias, or DC (mean value) component is introduced, and this DC component is multiplied by the periodic component Cos(ωrt). The resulting component then becomes superpositioned on the new noise component N'(t). It is this new DC-introduced superposition term that allows the recovery of the periodic component hidden in the original WBM signal. To maintain the dynamic range of the demodulated signal, the FWRWBD method replaces the squaring operation by an absolute value operation since both operators are effective in recovering the hidden periodicity.

CAVITATION GENERATED WBM SIGNAL: The signal generated by cavitation can be modelled as the multiplication of two separate components p(t) and N(t):

$$x(t) = p(t)\,N(t) \quad (5)$$

Here, N(t) represents the wide-band high frequency noise generated from the collapse of cavitation bubbles, while p(t) represents some periodic pressure fluctuation associated with the shaft rotational motion. This pressure signal p(t) contains a DC component due to its static pressure component P_{static},

$$p(t) = [\,\underbrace{P_{static}}_{(DC)} + P(t)\mathrm{dynamic}\,]. \quad (6)$$

Therefore, a cavitation generated pressure signal is a typical WBM signal and is a good candidate for WBD processing. However, in most operational environments, the spectra in the low frequency region always contains the fundamental rotor synchronous (Sync) frequency component along with its harmonics (2N, 3N, 4N...), which are generated from the combination of all the effects of rotor dynamics, structural dynamics and hydrodynamics. These combined effects make it difficult to isolate sources of vibration problems. However, when cavitation occurs, its signature should be contained in the noise floor of the high frequency region due to the unique phenomenon of wide-band modulation. The significance of this phenomenon is that a hydrodynamic source of vibration can now be isolated due to its unique information contained in the high frequency noise floor.

With cavitation detection using the FWRWBD method, the raw high frequency signal must first be high-pass filtered to remove low-frequency discrete components. Demodulation is then performed on the remaining high-frequency noise floor in order to recover any existing low frequency periodic component which is modulating the cavitation generated noise signal. As a result, a new low frequency WBD PSD is generated in addition to the original raw data PSD. However, unlike the raw data PSD which includes multiple contributions from rotordynamics, structure dynamics and hydrodynamics, this new WBD PSD only reflects the hydrodynamic ones. If cavitation does not exist, the WBD PSD should reduce to regular broadband noise. However, if cavitation does occur, the WBD PSD should show discrete peaks corresponding to the low frequency periodic rotational processes modulating collapsing bubble noise.

DEMONSTRATION OF THE FWRWBD WITH ENGINE TEST DATA: Real test data from the Space Shuttle Main Engine (SSME) Alternate Turbopump Development (ATD) E8 component test stand is used to demonstrate the effectiveness of the FWRWBD technique in detecting cavitation. Figure 1 shows the ordinary raw PSDs taken from four different accelerometers and pressure measurements across ATD LOX pump unit 3-1A during test E8-162 with a maximum frequency of 5 KHz. The inducer of this unit has four blades. These PSDs all show the fundamental Sync frequency component and its harmonics. Figures 2 shows the ordinary raw PSDs of these same measurements with the maximum frequency increased to 50 KHz. For this engine component test, discrete peaks are concentrated in the low frequency region with only limited amount of high frequency line noise discrete peaks present in the spectra. Notice that the signal energy in the wide-band high frequency noise floor is the prospective information to be used in effective cavitation detection. This high frequency energy may be made up of just regular noise, or may have the very unique wide-band modulation phenomenon hidden within it. Using just the PSDs of both figures 1 and 2, one would not be able to distinguish such a subtle difference.

The high frequency signals shown in figure 2 are first high-pass filtered at 20 KHz to remove all the discrete components, after which demodulation is performed using the FWRWBD method. Figure 3 shows the resulting WBD PSDs of these four measurements. For the first three measurements, which include two accelerometers and one inducer outlet high frequency pressure measurement, no periodic component is recovered in the WBD signal. This indicates that the original wide-band high-frequency noise components of these five measurements are just ordinary noise signals with no modulation phenomenon present. However, in the last plot of figure 3 (figure 3-d), corresponding to the Three Quarter Chord Inducer Inlet Kistler high frequency pressure measurement, several strong discrete components show up in the WBD PSD. This indicates that wide-band modulation phenomenon indeed exists in the high frequency noise floor. In other words, cavitation is present in this test, and the rotational periodic

components modulate with collapsing bubble noise generating a unique WBM high-frequency noise floor. The demodulated WBD signal thus recovers these hidden discrete components.

Moreover, an interesting phenomenon is present in this wide-band demodulation result. Notice that this WBD PSD of figure 3 has a strong 2N component and relatively weaker 1N, 3N and 4N components. However, the Raw PSD of the inlet pressure measurement in figure 1 shows a strong 4N component. If a judgement based on this raw PSD had to be made about the cavitation condition, full-bladed (4-blade) cavitation would be the likely candidate. However, the WBD PSD recovers a strong 2N component rather than a 4N component therefore indicating an alternate-blade rather than a full 4-blade cavitation condition, The alternate blade cavitation has been confirmed using high speed video taken during the test for flow visualization. This situation is typical in an actual engine operational environment, since the raw PSD will pick up many other vibration effects in addition to cavitation. The WBD process effectively isolates cavitation signature and provides more reliable information about the true operational condition.

COHERENT PHASE WIDE-BAND DEMODULATION (CPWBD) METHOD: Cavitation detection using the FWRWBD technique is powerful; however it suffers from a commonly encountered limitation. Since the hidden periodicity is recovered from the high frequency noise floor by demodulation, any discrete components present in the high frequency region of the original raw signal will generate false discrete peaks in the WBD signal which appear to be recovered hidden periodicities. These false peaks are generated due to the newly introduced cross coupling components when performing FWRWBD since demodulation represents a nonlinear operation. This limitation may not be so critical in an isolated environment such as during laboratory testing where no discrete components show up in the high frequency region. However, it is a very critical limitation in the analysis of rocket engine static firing or flight data in which many other vibration sources contribute all kinds of high frequency components which corrupt the WBD signal.

Due to its frequency domain formulation of coherence phase information, the Coherent Phase Wide Band Demodulation technique provides an effective method to avoid such discrete interference. As discussed previously, the FWRWBD method is directly related to the special tri-spectrum $T(\omega)$ for hidden periodicity detection in a signal $x(t)$, where $T(\omega)$ is simply equal to the ordinary PSD of the square of $x(t)$. It can be shown that [10], the Discrete Fourier Transform (DFT) of $y(t)=x^2(t)$ can be expressed as:

$$Y(k)=\frac{1}{N}[\sum_{i=0}^{N-1}X((i))_N X((k-i))_N]R_N(k) \qquad (7)$$

Where X(k) is the DFT of x(t)

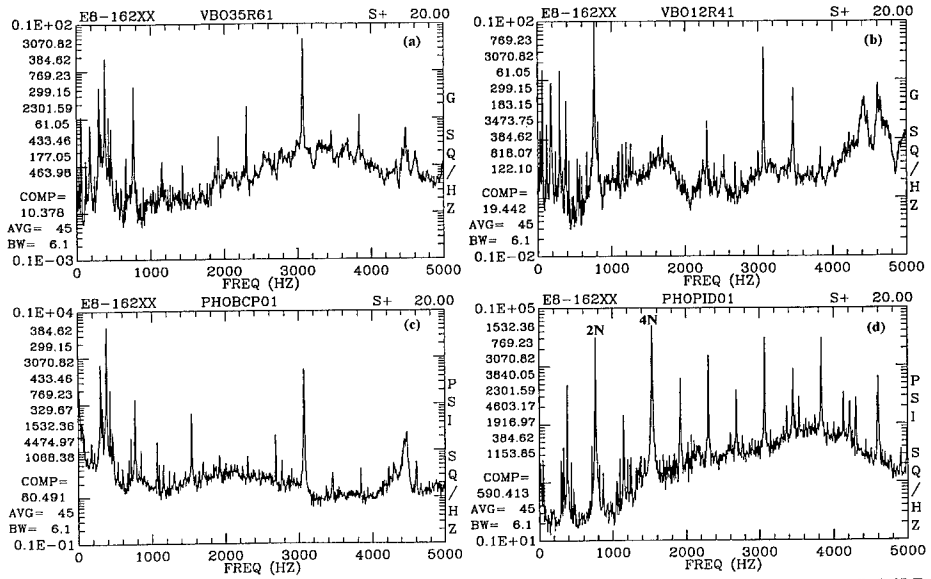


Figure 1: 5 Khz Raw PSDs Taken From Four Different Measurements Across ATD LOX Pump Unit 3-1A During Test E8-162

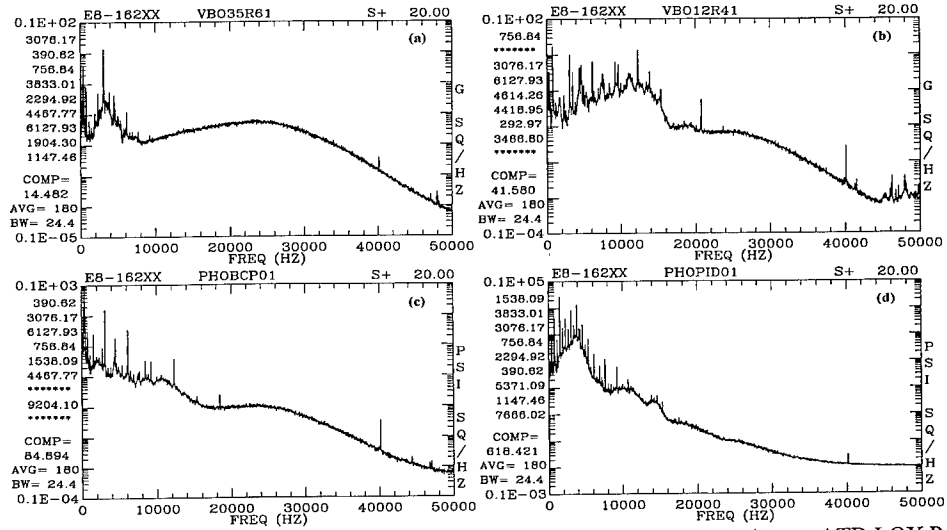


Figure 2: 50 Khz Raw PSDs Taken From Four Different Measurements Across ATD LOX Pump Unit 3-1A During Test E8-162

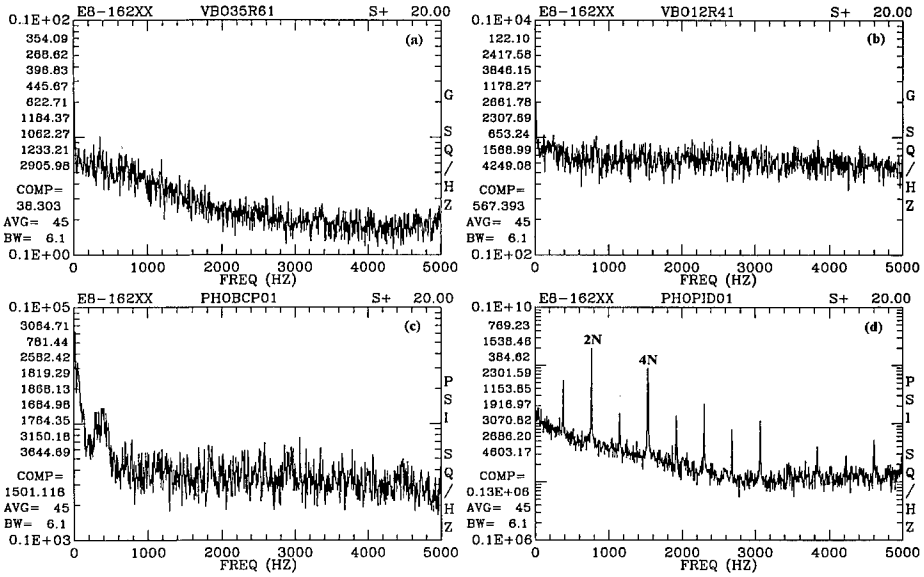


Figure 3: FWRWBD PSDs Obtained From Four Different Measurements In Figure 2

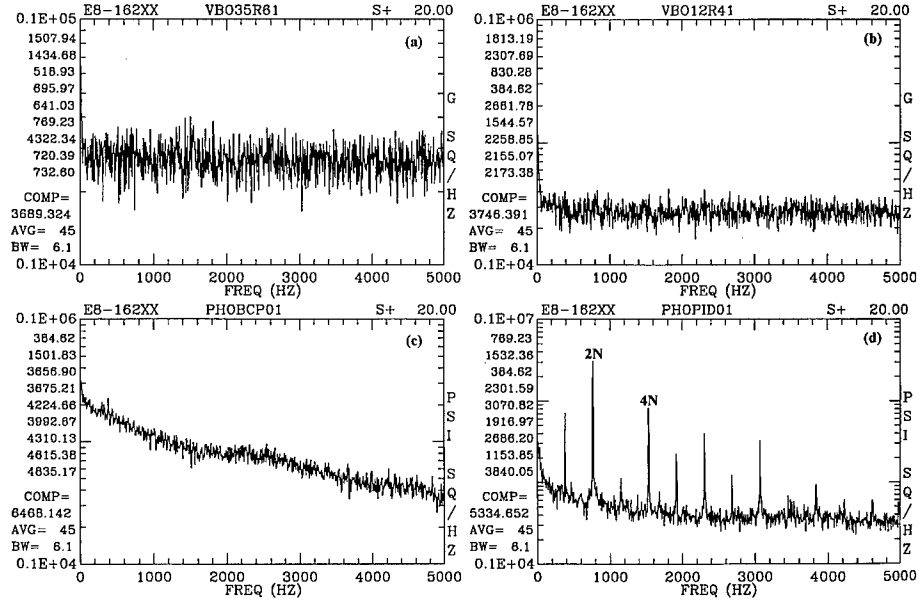


Figure 4: CPWBD PSDs Obtained From Four Different Measurements In Figure 1

Y(k) is the DFT of y(t) (or $x^2(t)$)

$R_N(k)$ is the rectangular function

$((n))_N$ is the (n modulo N) function

In other words, the DFT, Y(k), is equal to the convolution of X(k) with X(k) in the frequency domain. An important observation can be made from this equation; for ordinary independent wide band noise, the phases of $X((i))_N$ $X((k-i))_N$ will be independent random phases for i=0 to N-1, therefore, the superposition of all the terms of $X((i))_N$ $X((k-i))_N$ during convolution will be a random walk process and vector construction will not occur to form a discrete component in Y(k). On the other hand, if x(t) is a true WBM signal, the coherent phase relationship among all the terms of $X((i))_N$ $X((k-i))_N$ will make each individual vector point in a common direction. As a result, the superposition of $X((i))_N$ $X((k-i))_N$ during convolution will lead to vector construction, and a hidden periodicity will be recovered in Y(k) at the frequency of the WBM modulating frequency. The significance of this observation is that, the hidden periodicity within a WBM signal can be recovered solely from its unique coherent phase relationship among all interacting components. This unique property thus provides an effective way to solve the discrete interference problem. When an independent discrete component is present within a signal, it will introduce a large step vector in the random walk process during convolution, which would in turn erroneously generate a strong discrete component in the resulting WBD signal pretending to be a recovered hidden periodicity.

To avoid such an unwanted large step random walk due to discrete interference, the CPWBD technique thus performs WBD by introducing a Phase-Only (PO) filter prior to the convolution process. This PO filter will retain the coherent phase relationship associated with a WBM signal to allow hidden periodicity recovery. Furthermore, due to the amplitude normalization effect generated from the PO filter, the CPWBD would no longer be subjected to any independent discrete interference. Due to such a normalization effect, the contribution of unwanted discrete interference is now reduced to the same level of other spectral components. Notice that the amplitude of the hidden periodicity recovered from the CPWBD is also subjected to some normalization. Therefore, the CPWBD should be viewed as a normalized "coherence" function for recovering true hidden periodicity from a WBM signal. It should also be pointed out that, for a noise-free or deterministic signal, both the FWRWBD and CPWBD techniques could produce erroneous results since phase information within the wide band noise floor is heavily corrupted by the side-lobes emanating from the most dominant spectral components, commonly referred to as leakage. Therefore, these techniques should not operate on diagnostic signals with suspiciously "smooth" noise floors corrupted by such leakage.

DEMONSTRATION OF THE FWRWBD WITH ENGINE TEST DATA: The test data from the SSME/ATD E8 component test stand is used to demonstrate the effectiveness of CPWBD method for cavitation detection by only utilizing the low frequency (5 KHz) signal. The objective here is to demonstrate that (1) if the WBM phenomenon also exists at lower frequencies, the hidden periodicity associated with cavitation can also be recovered from the low frequency (5 Khz) signal rather than having to perform high frequency analysis which requires more processing time and larger data volume; (2) the apparent discrete components (e.g. Sync and Harmonics) in the low frequency region will not affect the CPWBD technique in detecting hidden periodicity (Notice that, a direct application of FWRWBD to the 5 KHz frequency signal will unavoidably introduce erroneous peaks during the demodulation process due to these apparent discrete components).

Wide-band demodulation is performed on the 5 KHz frequency data by the CPWBD method with its resulting WBD PSDs shown in figures 4. The WBD PSD in figure 4-d clearly shows that strong hidden periodicity at 2N due to alternate blade cavitation is identified. The more significant result is that, erroneous peaks are no longer present in the other three measurements in figures 4-a, 4-b, and 4-c, as would be with the FWRWBD method. This is because the CPWBD technique effectively identifies hidden periodicity by searching for a coherent phase relationship and is not be affected by linearly superimposed discrete components. The original discrete components, unrelated to cavitation, in the raw PSDs are therefore rejected during the demodulation due to the lack of phase correlation.

CONCLUSION: Cavitation occurs in many types of turbomachinery, causing performance degradation, excessive structural vibration, material erosion and pitting damage. The principle for cavitation detection is based on the nonlinear modulation phenomenon between the bubble collapse generated wide-band noise with some synchronous (RPM or harmonics) and/or non-synchronous periodic components. However, the periodicities modulating the wide band noise become hidden in the resulting WBM signal and cannot be identified with the conventional power spectral density function. Such hidden periodicity thus provides a unique signature in the wide band noise conducive to cavitation detection. When a measured diagnostic signal is not subjected to discrete components interference due to other vibration sources, the full-wave rectification technique is an effective method to detect the existence and identify the intensity of cavitation.

The CPWBD method utilizes a Phase-Only filter coupled with envelope analysis to search for a unique coherent phase relationship associated with the cavitation-generated wideband modulation phenomenon. As a result, the cavitation analysis/prediction limitation due to discrete interference is eliminated. Analysis results from actual SSME ATD engine test data indicate that the CPWBD can effectively detect cavitation signatures by searching for such coherent phase relationships. In addition, the signal

processing associated with the CPWBD can take advantage of the computational efficiency of Fast Fourier Transform (FFT) making it highly suitable for real-time implementation in an on-line cavitation monitoring system.

REFERENCE:

1) Abbot, P.A., Arndt, R.E.A, and Shanahan, T.B., "Modulation Noise Analysis of Cavitating Hydrofoils," ASME Winter Annual Meeting, International Symposium on Bubble Noise Cavitation Erosion in Fluid Systems, New Orleans, LA, December 1993.

2) Avellan, F. and Dupont, E., "Cavitation Erosion of Hydraulic Machines: Generation and Dynamics of Erosive Cavities," International Association for Hydraulic Research Proceedings From 14th Symposium, Trondheim Norway, June 1988.

3) Bourdon, P., Simoneau, R., Avellan, F. and Farhat, M., "Vibratory Characteristics of Erosive Cavitation Vortices Downstream of a Fixed Leading Edge Cavity," International Association For Hydraulic Research, Proceedings From 15th Symposium, Belgrade Yugoslavia, September 1990.

4) Bourdon, P., Simoneau, R., and Avellan, F., "Erosion Vibratory Fingerprint of Leading Edge Cavitation on a NACA Profile and of a Francis Model and Prototype Hydroturbine," ASME Winter Annual Meeting, International Symposium on Bubble Noise Cavitation Erosion in Fluid Systems, New Orleans, LA, December 1993.

5) W. J. Bordelon, Jr., J. L. Minor, and T. E. Nesman, "Cavitation Testing of the ATD 14.6 Degrees High Pressure Oxygen Turbopump Inducer"., NASA 1994 Conference on Advanced Earth-to-Orbit Propulsion Technology, May 1994

6) T. E. Nesman, W. J. Bordelon, Jr., and J. Y. Jong, "Dynamic Water Flow Tests Using Four-Bladed Axial Flow Inducers", Workshop for CFD Applications in Rocket Propulsion and Launch Vehicle Technology, 1995 .

7) Jen-Yi Jong "Coherent Phase Cavitation Monitoring System for Turbomachinery" NASA-93 SBIR Phase I Final Report. July, 1994.

8) Jen-Yi Jong. "Anomaly Identification for Space Shuttle main Engine Diagnostics"., 48th Meeting of the Mechanical Failure Prevention Group (MFPG). April, 1994.

9) Dwyer, R. S., "Fourth-order spectra of Gaussian Amplitude-Modulated Sinusoidals," J. Acoustical Society of America, 90(2), Pt. 1, August, 1991.

10). A. V. Oppenheim and R. W. Schafer. Digital Signal Processing. Prentice Hall. 1975.

Self–aligning Roller Bearings Fault Detection Using Asynchronous Adaptive Noise Cancelling

Kikuo Nezu Yimin Shao
Department of Mechanical Engineering
Gunma University

1–5–1 Teinjin–cho kiryu City
Gunma Prefecture Japan

Abstract: A new method is presented for increasing the signal to noise ratio from measurements of bearing housing vibration. The method is based on the conventional ANC technique principle, it makes some variation on the main signal input and reference signal input. Because the signal of the main and reference input are collected at the different time, this method is called asynchronous adaptive noise cancelling (AANC). It can improve the signal to noise ratio as to the case of the detecting bearing faults when vibration signal of machinery are collected under low frequency noise condition. Experiment result has shown that these techniques can be made more effective only after AANC has reduced the background noise from the diagnostic signals. A new approach is proposed to the traditional fault detecting.

Key Words: Bearing; fault detection; asynchronous adaptive noise cancelling;

INTRODUCTION: Methods of machine condition monitoring are now receiving considerable attention, on the premise that financial benefits can be obtained by reducing planned maintenance costs, by avoiding catastrophic failure and by increasing plants availability. The successful operation of a machine relies on the performance of each component element, e.g., bearings, shafts, foundation, and so an effective approach to overall machine monitoring requires a detailed understanding of the behavior of these elements and their interaction[9]. Condition monitoring and failure diagnosis of rolling element bearings is one of the most fundamental and important components in the mechanical system failure diagnostic technology.

Vibration monitoring is recognized as being an effective tool when applying a programme of machine condition monitoring. A lot of methods have been developed for localized faults of bearings. For example, statistical methods[9], time and frequency field analysis methods[3][10], analysis of bearings noise[14], adaptive noise cancelling technique[4], fuzzy and neural networks identification methods[2][12][15][16], etc.. But there is still not an ideal way to diagnose the early failure of the self–aligning roller bearing. Especially, the detection fault is fraught with difficulties, when the self–aligning bearing run in the condition of low speed. The signal to noise ratio is lower, and the impulse signal is more difficult to extract when the bearings are running at the high shaft speed condition.[16][7][1]

Adaptive noise cancelling (ANC), as this is widely known, is a method of estimating signals corrupted by additive noise. It has been applied to speech signals, electrocardiography, and adaptive antenna arrays.

It has also been applied to fault detection of bearings[4]. This method makes use of two signals input, a primary input, which contains the corrupted signals, and a reference input, containing noise correlated with the primary noise. Generally, the signals of primary input and signals of reference input are collected at the same time. The reference input may be derived from a sensor located at the point in the noise field where the signal is very weak. The signal to noise ratio will be affected if reference measurement point was unsuitably chosen. So, the correctly selected position of reference measurement point is very important, and also is difficult. The method can only cancel part of the noise resulted from the near component of machine or the machine, it can't completely cancel the heavy noise resulted from the normal vibration of bearings.

In this paper, a new diagnostic method is proposed to solve the problem that the success rate of fault detection is affected by the heavy low frequency background noise. Based on the idea of signal relative change, the abnormal running condition of bearings is an relative change as to normal running condition of bearings. The measurement points of primary and reference input are chosen at the same place (same sensor). The primary and reference signals are collected at difference time. Due to the reference input was collected at the early running time (normal running condition of bearing), and the primary input was collected at the present, this method is called asynchronous adaptive noise cancelling (AANC). It can eliminate unwanted normal background noise. The background noise includes normal vibration noise of bearings and noise machine vibration. Subsequent to applying AANC, enveloping spectrum analysis techniques were used for the purpose of detection and diagnosis of faults.

THEORY ANALYSIS: The Characteristic of Vibration Signals of The Self–aligning Roller Bearing: Due to the structure features of the self–aligning roller bearing, the rolling elements contact the inner race and outer race in line. Although the defect's pit of bearing is the same either in the ball bearing or in the self–aligning roller bearing, the degree of collision between rolling elements and defect's pit is not the same bigness, when each time a rolling element passes over a defect. The impulse's duration of the ball bearing is shorter than that of the self–aligning roller bearing. So, there are two main features in the vibration signals of self–aligning roller bearing. First, there are no short and sharp impulse signals enough in the vibration signals of self–aligning roller bearing. Second, the vibration signals of self–aligning roller bearing contain heavy low–frequency noise. Based on the these two features of vibration signals of the self–aligning roller bearing, the AANC technique can be used to improve the signal to noise ratio under the condition of heavy background noise of low frequecy.

The Method of Asynchronous Adaptive Noise Cancelling: The special principle of AANC is shown in Fig. 1. The signal S is corrupted by a noise N_0 and received at the primary sensor when the bearing was running in abnormal condition. Where S is fault signal of bearings, N_0 is noise of machine that it contain normal vibration of bearings. A reference noise N_1, which is related to the noise N_0 in some unknown way but uncorrelated with the signal S_0, is received at the reference sensor. The vibration signals of self–aligning roller bearing is mainly relative to the shaft speed. Under the shaft speed is a steady constant, collected primary input and reference input at different time can satisfy the three assumption conditions of adaptive noise cancelling. The three assumption conditions have been shown in the paper[4]. Its were:

(1). S, N_0, N_1, and y are statistically stationary and zero means.

(2). S is uncorrelated with N_0 and N_1.

$$E[SN_0] = 0 \qquad (1)$$
$$E[SN_1] = 0$$

(3). N_1 is correlated with N_0.

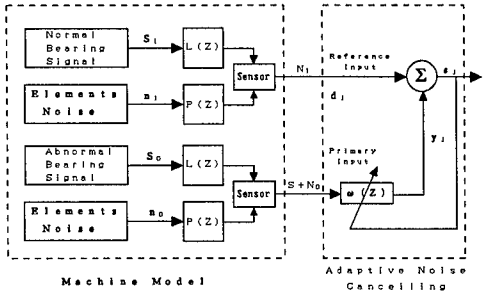


Fig. 1 The Model of An Asynchronous Adaptive Noise Cancelling

The output of the ANC system is given by

$$C=S+N_0-y \qquad (2)$$

then

$$C^2=S^2+2S(N_0-y)+(N_0-y)^2 \qquad (3)$$
$$\because y=W_j^TX_j \qquad (4)$$

Where W_j is weight vector at the jth instant of time, X_j is reference input vector.

$$\therefore C^2=S^2+2S(N_0-W^TX)+(N_0-W^TX)^2 \qquad (5)$$

Taking the expectation of both sides of equation

$$E[C^2]=E[S^2]+E[(N_0-W^TX)^2] \qquad (6)$$

The signal power $E[S^2]$ will be unaffected as filter weights are adjusted to minimize $E(C^2)$

$$\min E[C^2]=E[S^2]+\min E[(N_0-W^TX)^2] \qquad (7)$$

For an optimal set of filter output y is then the best least square estimate of primary noise N_0. From equation:

$$(C-S)=(N_0-y) \qquad (8)$$

Adjusting the filter weights to minimize the output power thus causes C to be a best least square estimate of the signal S.

EXPERIMENTAL SET-UP AND DATA ACQUISITION: The experimental self-aligning bearings are a bearing type of NSK 22207. These bearing could be artificially localized defects induced by electric-discharge machine. The bearing failure occurs with defects in the outer race. The experimental data include two failure size types of self-aligning bearing under two kinds of shaft speed and one load. The test conditions are shown in Table 1.

Fig. 2 shows the bearing, housing, and the structure supporting them as well as the drive. The motor was connected to the driven shaft on which the aid bearings and test bearings were housed by a belt and pulley system. The motor was controlled by an alternating governor system (FUJI Frenic 5000 G9S/P9S). The vibration was measured with conventional, piezoelectric accelerometers. The signals passed through a multi-path A/D board and then the computer, where signals were analyzed. The

software not only contains time and frequency domain signals analysis system, but also the digital adaptive filter of asynchronous and a digital low-pass filter. A time domain data set contains 8 samples, a sample has 1024 points. Spectra are averaged over 8 samples with a HANNING window. A figure of enveloping spectrum was drawn by using 512 points. The enveloping spectrum of AANC (after the application of AANC) was drawn by using 256 points.

Table 1 The Experiment Conditions

Name of set up condition	Value of set up	Frequency of outer race
Defect Type 1	2.6x1x0.5 (LxWxH mm)	
Defect Type 2	3x1x0.5 (LxWxH mm)	
Shaft speed 1	417 r.p.m.	44 Hz
Shaft speed 2	1667 r.p.m.	180 Hz
Load	5 kN	

Generally, frequencies of outer race and inner race or roller malfunction were used on identification faults of bearings. For feature frequency of the malfunction of inner race, outer race, or roller, there are formulas for calculating the frequency regions in which features may appear in the spectra[10].

$$F_i, F_o = \frac{n}{2} f_r \left[1 \pm \frac{d}{D} \cos \alpha\right] \quad (9)$$

$$F_b = \frac{D}{d} f_r \left[1 - \left(\frac{d}{D}\right)^2 \cos^2 \alpha\right] \quad (10)$$

where f_r is shaft rotation frequency, F_i and F_o are frequenies for inner and outer race malfunction, respectively, F_b is frequencies for ball malfunction, d is the rolling element diameter, D is pitch diameter, n is the number of rolling elements, α is contact angle.

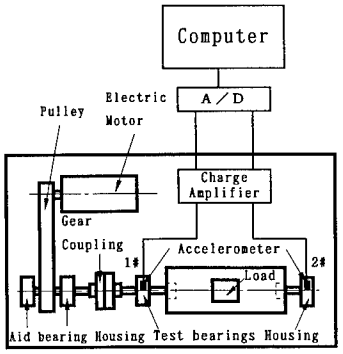


Fig. 2 Experimental Set-up

DISCUSSION OF RESULTS: This paper presents the results of asynchronous adaptive noise cancelling for the cases referred to in Table 1, the results of the time waveform, enveloping spectrum (before and after the application of AANC) are also given to show the effectiveness of AANC in eliminating the unwanted noise.

Fig. 3A and Fig. 3B show the normal signal of bearings when the self–aligning bearings were running at the 417 r.p.m and 1667 r.p.m.. They are also the reference signal of AANC. The capital letter A flag of figures show defect type 1. The capital letter B flag of figures show defect type 2. From Fig. 5 to Fig. 8 show two defect types at a shaft speed of 417 r.p.m. From Fig. 9 to Fig. 12 show two defect types at a speed of 1667 r.p.m.

Detecting the faults of bearing are difficulty as to signals of vibration when the bearings are running at the low speed and the small size of defect is small. There are heavy low frequency background noise when the bearings are running at the low (417 r.p.m.)or high shaft speed (1667 r.p.m.). Fig. 4 and Fig.6 also show the heavy disturbance of low frequency whether the bearings have defect or no defect. For example, near 228 Hz and 487 Hz have bigger value of amplitude.

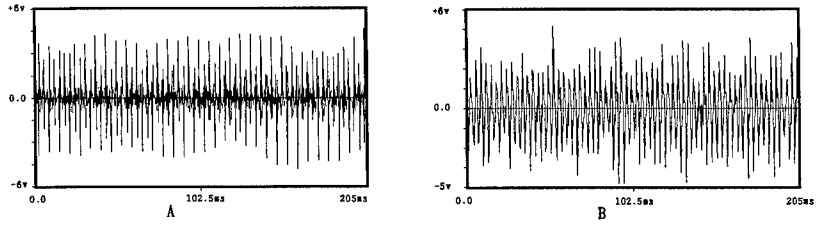


Fig. 3 Normal Time Waveform of Vibration Signal of Self–aligning Bearings

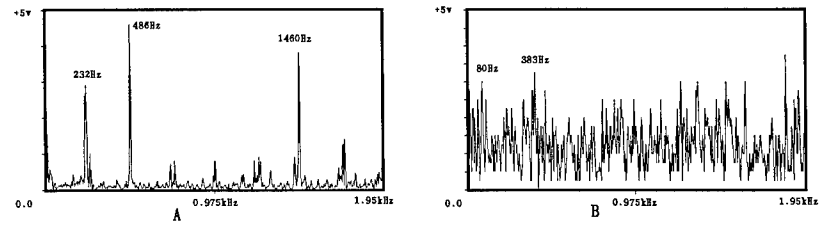


Fig. 4 The Enveloping Spectrums of Normal Signals

A comparison of the estimated bearing signal waveform Fig. 11 with the raw bearing signal Fig. 5 shows these two waveforms are similar. In Fig. 7 and Fig. 11 , the impact waveform is more clear than the raw bearing signal waveform. The heavy low frequency background noise has been eliminated, but the faults recognition still is difficulty by time waveform.

In Fig. 8 and Fig. 12 the defective bearing waveform is shown along with its enveloping spectrum (before and after the application of AANC). The enveloping spectrum clearly shows the amplitude of the fault bearing waveform after the application of AANC. The failure frequency of outer race can be found at the 42Hz and 46Hz in the Fig. 8, but it is not very clear to see from spectrum. The case may be resulted from resolution of spectrum, the resolution is 10Hz in Fig. 8 and Fig.12.

In contrast to this, Fig. 6 and Fig. 10 can find out about four times frequency of outer race, but it is difficulty to recognize faults of bearings. However, the failure frequency of outer race can be found clearly from Fig. 8 and Fig. 12. It has been shown that the signal noise ratio can be well improved by ANNC method under heavy low frequecy noise condition.

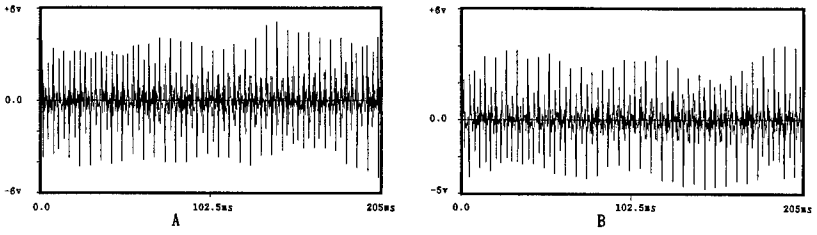


Fig. 5 Defect Type 1 Time Waveform

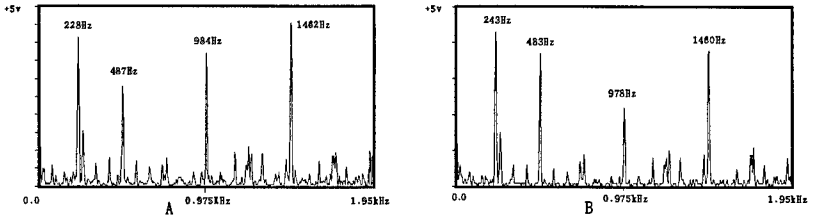


Fig. 6 Defect Type 1 Enveloping Spectrum

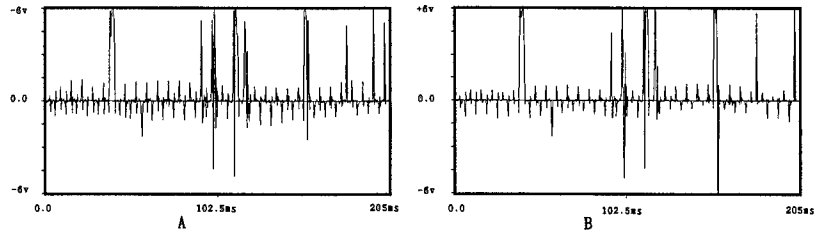


Fig. 7 Defect Type 1 Time Waveform After AANC

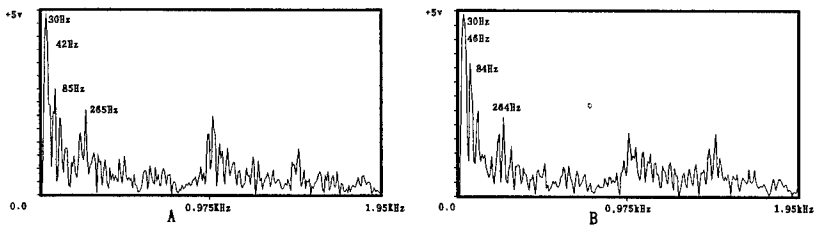


Fig. 8 Defect Type 1 Enveloping Spectrum After AANC

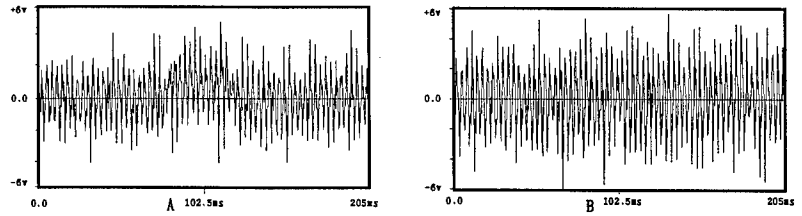


Fig. 9 Defect Type 2 Time Waveform

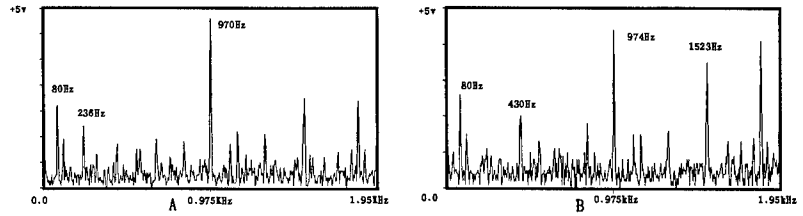


Fig. 10 Defect Type 2 Enveloping Spectrum

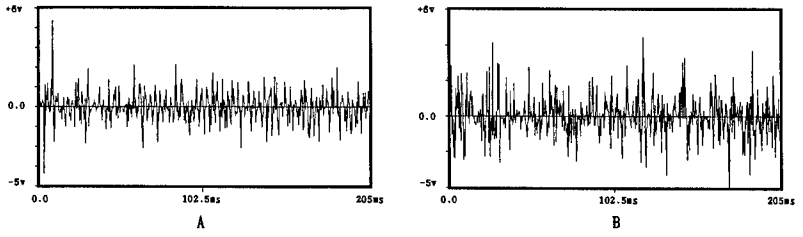


Fig. 11 Defect Type 2 Time Waveform After AANC

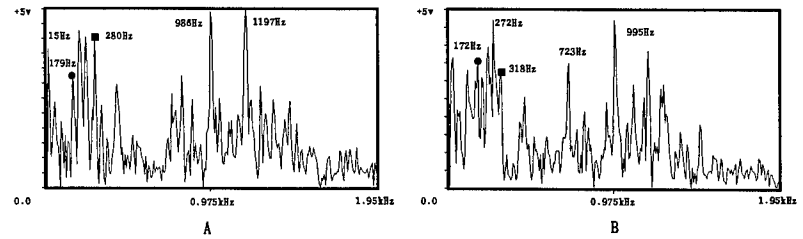


Fig. 12 Defect Type 1 Enveloping Spectrum After AANC

CONCLUSION: It has been shown that AANC method is a potentially powerful method to identify faults when the bearing was running at low speed or defect size of bearing is small. It is more effective for self–aligning bearing fault detection and diagnosis when combining AANC with enveloping spectrum

analysis technique.

This paper proposes a variation of ANC for those situations where the main input signal and reference input signal are collected at the same sensor from bearing's housings. It can not only increase the signal to noise ratio under heavy low background noise, but also simplify measurement location choice. The technique enhances traditional fault detecting of self-aligning bearing.

REFERENCES

[1] Yimin Shao, Kikuo Nezu, " An On-Line Monitoring And Diagnostic Method Of Rolling Element Bearing With AI ", Proceedings of SICE '95, 1543-1548, Sapporo, Japan, July 26-28, 1995.

[2] Yimin Shao, Kikuo Nezu, " Feature Extraction of Machinery Diagnosis Using Neural Network ", Proceedings of IEEE Interbational Conference on Neural Networks, Vol 1,459-464, Perth, Austrilia, November, 1995.

[3] Karl Barthel, " The Shock Pulse Method for Measuring the Condition of Antifriction Bearing ", Tappi,Vol. 60, NO.8, Aug. 1977.

[4] G. K. Chaturvedi, D. W. Thomas, "Bearing fault detection using adaptive noise cancelling", Transactions of the ASME Journal of Mechanical Design, Vol. 104, 280-289, 1982.

[5] C.J.LI and S.M.Wu, "On-line detection of localized defects in bearings by pattern recognition analysis", ASME Journal of Engineering for Industry, 111,331-336,November, 1989.

[6] J. Mathew, R. J. Alfredson, "The condition monitoring of rolling element bearings using vibration analysis", Journal of Vibration, Acoustics, Stress, and Reliability in Design, Vol. 106, 447-453, July 1984.

[7] S. Braun, B. Datner, "Analysis of roller / ball bearing vibrations", Transactions of the ASME Jouranal of Mechanical Design, Vol. 101, 118-125, January,1979.

[8] M. Serridge, "Fault detection techniques for reliable machine condition monitoring", sound and vibration, 18-23, May, 1989.

[9] D.Dyer, R. M. Stewart, "Detection of rolling element bearing damage by statistical vibration analysis", Transactions of the ASME Journal of Mechanical Design, Vol. 100, 229-235, April, 1978.

[10] J. I. Trylor, "Identification of bearing defects by spectra analysis", Trans. of the ASME Journal of Mechanical Design, Vol. 102, 198-204, April, 1980.

[11] Sherman S. Wang, "Diagnostic Expert System for Industry", Proceedings of the 2nd International Machinery Monitoring & Diagnostics Conference &Exhibit, xvii-xxv, Losangeles, California, 1990.

[12] J. Xu, H. Peeken, "Failure diagnosis system using Fuzzy logic", The 1989 ASME design technical conferences-12th biennial conference on mechanical vibration and noise, De-Vol. 18-3, 18-4, 18-5, 93-99, Montreal Quebec Canada, 17-21,1989.

[13] J. Sandy, "Monitoring and diagnostic for rolling element bearings", Sound and vibration, 16-22, June 1988.

[14] Simon G. Braun, "The signature analysis of sonic bearing vibrations", IEEE Trans. on sonics and ultrasonics, Vol. Su-27, No. 6, November 1980.

[15] T. I. Liu, N. R. Iyer, "Diagnosis of roller bearing defects using neural networks", The International Journal of Advanced Manufacturing Technology, Vol. 8, No. 4, 210-215, 1993.

[16] I. E. Alguindigue, A. L. Buczak, R. E. Uhrig, "Monitoring and diagnosis of rolling element bearings using artificial neural networks", IEEE Trans. on industrial electronics, Vlo. 40, No. 2, April, 1993.

NON-DESTRUCTIVE INSPECTION

EDDY-CURRENT EXAMINATION OF LARGE-DIAMETER INSULATED PIPES

John C. Griffith
Framatome Technologies, Inc.
155 Mill Ridge Road
Lynchburg, VA 24502

Abstract: A new concept eddy-current technique has been developed to examine large-diameter insulated pipes without removing the insulation, and in many cases, without securing system operation. The new concept allows the in-place examination of piping systems using an encircling eddy-current coil that is a proven eddy-current technology. Eliminating the requirement to secure or drain the system or to remove the insulation from the pipe provides significant cost savings. The new concept also provides a significantly higher percentage of inspection volume coverage than ultrasonic techniques despite limitations of pipe brackets and other physical obstructions. The technique has been demonstrated on copper, copper nickel and limited carbon steel pipe ranging from 4 to 8 inches in diameter with insulation thicknesses from 1 to 2 inches.

Key Words: Condition monitoring; diagnostics; encircling coil; eddy current; pipe examination;

INTRODUCTION: Operational piping systems in general are subject to a number of degradation effects such as corrosion, erosion, fatigue, pitting, and wear. Non-destructive examination (NDE) of piping systems is essential in cases where a failure of the system could result in a loss of product, damaged equipment, or injured personnel. The NDE of large-diameter piping systems has typically been done by manual ultrasonic testing (UT) of expected degradation locations. The UT examination normally requires removing pipe insulation or lagging, extensive surface preparation, and time consuming manual or automated scanning of a grid pattern at the suspect location followed by system restoration. The UT examination provides a costly but very accurate characterization of expected degradation locations, but provides no information on other unexpected degradation locations in the piping system. Other NDE techniques could require system shutdown, cooldown, draining and insulation removal.

Framatome Technologies, Inc. (FTI) has developed a unique method of eddy-current examination for large-diameter piping systems that does not require system shutdown or insulation removal. The proven technique of using an encircling eddy-current coil had been adapted for use on an installed and insulated pipe. The method provides a rapid examination with good sensitivity. The technique has been proven in laboratory testing as well as in field examination of operational piping systems.

SYSTEM DESCRIPTION: The Framatome Technologies encircling coil system consists of several components. A rail system or a movement fixture is used to provide a smooth movement of the coil relative to the pipe surface. Coil forms are used to define the coil geometry and minimize any unwanted movement of the coils during the examination. The coils are designed to provide good defect sensitivity and allow rapid installation and removal. Figure 1 shows a typical setup for an encircling coil examination. Each of the components will be discussed below in more detail.

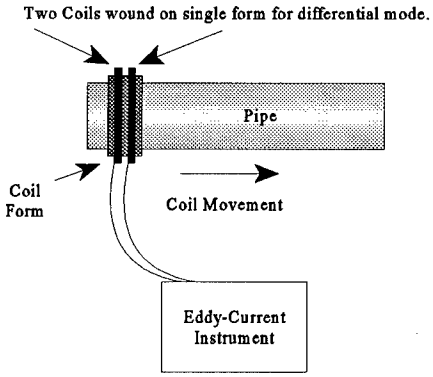
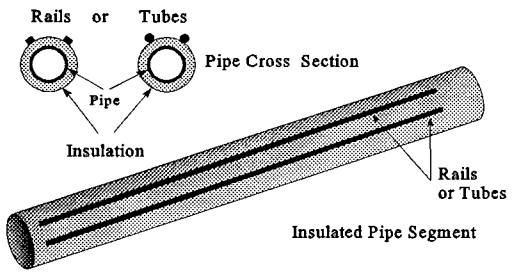


Fig. 1 - Eddy-Current Encircling Coil

Movement Fixtures: The first critical requirement for the application of an encircling coil for pipe examination is that the coil movement be coaxial with the pipe centerline. Side to side movement of the coil produces undesirable signals which complicate data interpretation. In conventional encircling coil examinations of tubing, the coil form contacts the tube outside diameter (O.D.). This direct contact with the pipe O.D. is not possible in the case of an insulated pipe. Instead, a system of rails or a mechanical movement fixture is used to maintain the coil alignment during the data scan.

Figure 2 shows a system of rails or thin tubular forms which can be placed on the top or side of a straight insulated pipe segment to provide a reference surface for the coil movement. The insulation surface is often rough or uneven and would add significant liftoff noise to the eddy current data. The rails provide a smooth surface for the coil travel and closely approximate the pipe wall contour.



With the proper selection or design of rail material to conform to the pipe geometry, the rails could be used for examination of pipe bends or elbows.

Fig. 2 - Straight Movement Fixtures

Examination of bends and elbows provides a more significant challenge. Figure 3 shows two possible solutions. A mechanical movement fixture could be used to define the coil path by setting a pivot point at the center of curvature for the bend or elbow. Alternately, a set of flexible rails could be installed on the bend or elbow. The actual geometry and access restrictions would determine the best method for this case.

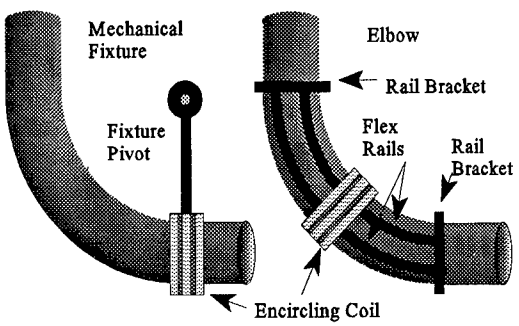


Fig. 3 - Elbow Movement Fixtures

Coil Forms: The second requirement for the use of an encircling coil in examining insulated piping is that the coil geometry must remain stable during the examination. In the case of a differential coil, the relative positions of the two windings must remain constant. A special design coil form is used to provide the stable structural geometry of the coils during a pipe examination. The coil form is constructed of a rectangular sheet of nylon or delrin with a hole pattern for attachments and guide posts. Typical coil form top, side and end views are shown in Figure 4.

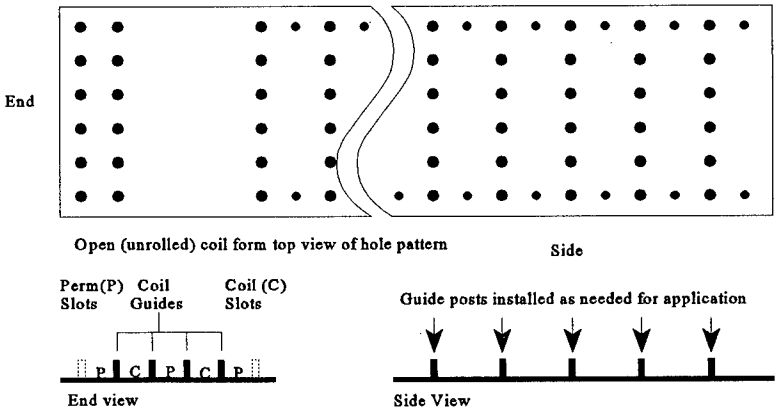


Fig. 4 - Coil Form

The coil form is wrapped loosely around the insulated pipe after the rails have been installed. The start of the form is secured to the end to create a rigid right circular cylinder around the pipe. Figure 5 shows a cross-sectional view of the installed coil form.

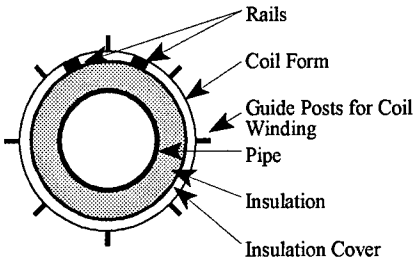


Fig. 5 - Coil Form Wrapped Around Insulated Pipe

Encircling Coils: A typical encircling coil is composed of 50 (or more) lengths of insulated copper magnet wire. The length and wire gage are determined by the pipe and insulation diameter and the number of turns required for the coil sensitivity and proper frequency range. The conductors are grouped together in heat-shrink tubing to provide structure and protect against abrasion. One end of the bundle has a male connector and the opposite end has the mating female connector as shown in Figure 6.

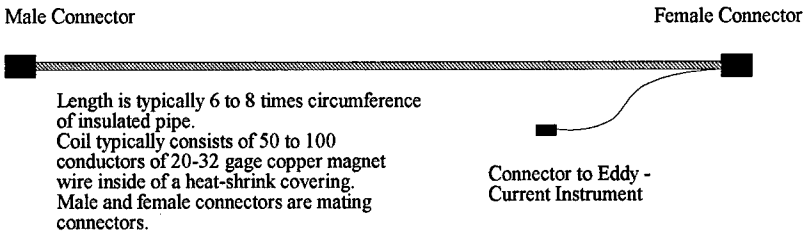


Fig. 6 - Coil Winding Design

The pinout of one connector is shifted one pin such that connection of the two ends provides a continuous single loop of the 50 (or more) conductors as shown in Figure 7. When this long coil of 50 conductors is wrapped around the pipe 6 times, the resulting encircling eddy current coil has 6 * 50 = 300 turns, which can be installed in about 4 minutes. This is one of the more unique features of this technique. **The coil design allows a sensitive encircling coil to be easily installed on an intact and operational piping system without system shutdown or insulation removal.**

The time and ease of coil installation can be improved significantly with minor design changes.

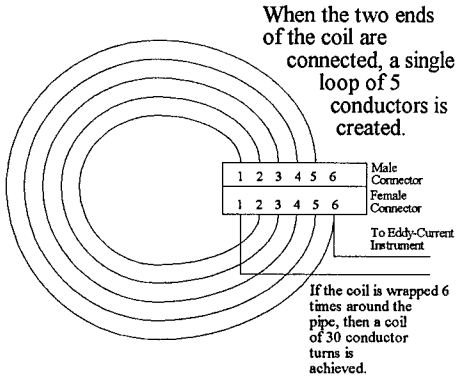


Fig. 7 - Encircling Coil Concept

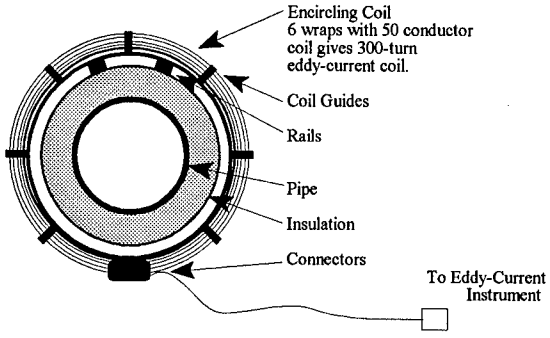


Fig. 8 - Coil Winding on Coil Form

Once the coils are wrapped in place on the coil form as shown in Figure 8, the ends are connected to complete the circuit and the wraps are secured in place. The coil form provides structure and uniformity for each coil winding.

The coils are connected to a standard reflected impedance eddy-current instrument and an electrical balance or nulling is performed as in a typical differential eddy-current application.

Examination Procedure: After the coil has been installed on a section of pipe, appropriate locating and identification information is recorded and the coil is moved to the start of the section. Eddy-current data is recorded as the coil is slowly pulled past the examination region. As shown in Figure 9, data may be recorded in either a differential or absolute mode for reflected impedance operation. It is also possible to operate the coils in a driver pickup (or remote field) configuration if required for examination of ferromagnetic piping.

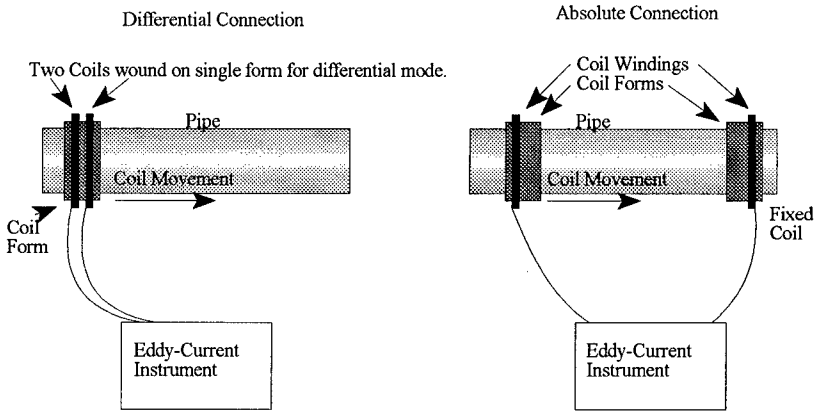


Fig. 9 - Eddy-Current Differential and Absolute Mode Connections

Disposition of Indications: Eddy-current responses can result from several conditions. The examination procedure must determine which signal responses should be eliminated and which represent pipe degradation. Signals are first correlated with brackets and fixtures external to the pipe. Eddy-current responses clearly resulting from factors external to the pipe are eliminated. Some signal response locations may require examination with an eddy-current driver pickup pancake coil for better localization / characterization. For example, a pipe weld would be detectable by the pancake coil for a full 360° around the pipe circumference.

Indications that are not eliminated as external influences would then require localized removal of the pipe insulation to determine the cause of the eddy-current indication. Visual examination of the pipe O.D. would show any denting or significant O.D. wall loss. After the insulation has been removed, the encircling coil sensitivity could be increased by wrapping the coil in a smaller diameter (closer to the pipe O.D.). If needed, ultrasonic examination techniques could then be applied to measure any inside diameter (I.D.) wall loss or pitting.

LABORATORY TEST RESULTS: Framatome Technologies, Inc.'s encircling coil concept was proven in laboratory tests prior to the first field examination. Samples of actual piping material for the following sizes were obtained for test purposes:

Pipe Diameter, (in.)	Pipe Wall Thickness, (in.)	Pipe Material	Pipe Insulation Thickness (in.)
8	0.18	CuNi	1 to 2
5	0.13	CuNi	1.3 to 1.8
4	0.12	Copper	1 to 1.5

Defects of 100 % through wall, 50 % O.D. and 50 % I.D. were machined or cut into the pipe standards to determine if the encircling coil could detect the target wall loss of 50 %. The diameter of the encircling coil was set to the maximum insulation diameter plus about 0.5-inch of clearance.

The FTI encircling coil technique was used to examine each standard. In every case, the installed defects were detected with a very good eddy-current signal response. The typical response from the 8-inch copper / nickel pipe is shown in Figure 10.

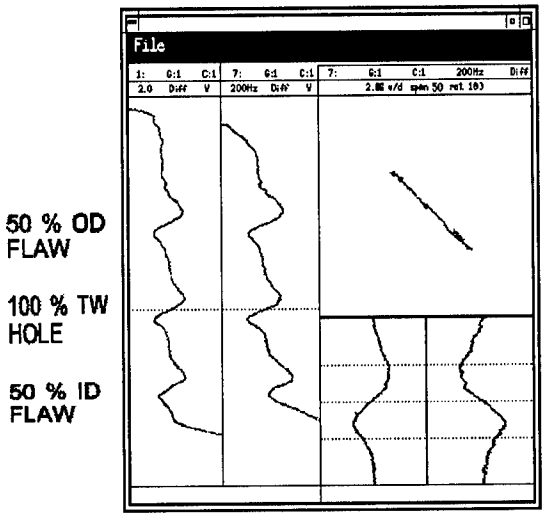


Fig. 10 - Eddy-Current Response for 8-Inch Pipe

FIELD TEST RESULTS: In August of 1995, the FTI encircling coil was used to examine shipboard piping in Norfolk, VA. The system clearly identified piping welds, some denting, and pipe O.D. degradation. At least one actual I.D. wall loss pit with a depth of approximately 40 % was detected and confirmed in an operating system pipe. The pit was confirmed by ultrasonic and visual examination.

DEVELOPMENT PLANS: The encircling coil technique has been proven in the laboratory and in a field examination. Planned and underway improvements will expand the scope of the technique, improve the examination speed and simplify the technique. The specific development plans cover the following areas:

* More work on carbon steel pipe examination
* Better methods / fixtures for elbow examination
* Improved coil installation methods and speed
* Better rejection of noise and external signal influences.

SUMMARY: FTI has developed a unique new examination concept that allows a proven encircling coil eddy-current technique to be applied to insulated and operational piping system. **The new concept offers the following significant advantages for any industry using large piping systems which require NDE:**

* **REDUCED COSTS FOR PIPE EXAMINATION**
* **IMPROVED EXAMINATION COVERAGE**
* **IMPROVED EXAMINATION SPEED**
* **MINIMAL INSULATION REMOVAL / REINSTALLATION**
* **MINIMAL IMPACT ON SYSTEM OPERATION**
* **MORE FLEXIBLE AND COST EFFECTIVE NDE SCHEDULING.**

OFF-LINE AND ON-LINE MOTOR ELECTRICAL MONITORING AND CONDITION ANALYSIS: PAYOFFS AND PROBLEMS

Jack R. Nicholas, Jr.,
PdMA Corporation
2448 Holly Avenue, Suite 200
Annapolis, MD 21401-3148

Abstract: This paper provides insight into and describes motor electrical condition monitoring and analysis methods in use for decades and provides a comparison with state of the art methods developed and introduced in the 1990's. Nine methods for condition monitoring of motors are described. In particular, the technology generically called "motor circuit analysis" available from an increasing number of commercial sources (and now used in many different countries) is compared in a balanced manner to three other off-line and five on-line test methods. Five techniques for analysis of motor electrical data from the nine technology methods are tabulated for comparison. A sixth, very powerful, analysis technique is discussed. Advantages and disadvantages of off-line and on-line methods are tabulated and tied to specific technologies where applicable. One case history involving two methods from a commercial activity is presented to illustrate a problem and a payoff. The difficulty of gaining acceptance of standards developed to support the analysis using newly introduced technologies is described. The status of industry consortiums established to set standards for new and refurbished motor purchases and for ensuring easy integration of data from any predictive technology used for monitoring of any machinery is briefly discussed.

Key words: Motor condition monitoring, motor circuit analysis, electrical predictive maintenance, motor diagnostics

INTRODUCTION: Motor electrical condition monitoring of a limited nature has been possible for decades using a number of qualitative and quantitative analysis methods. However, the process of arriving at a definitive, non-subjective conclusion regarding comprehensive motor electrical health has been difficult and time consuming until recent years. The availability of powerful, field portable computers has made it possible to quickly and definitively assess the major parameters which may be used to characterize condition. The computers, combined with an increasing number of software programs and small electronic packages for applying innovative measurement methods, enable technicians to quickly gather, store, recall, analyze and present the mass of data needed to perform electrical predictive maintenance and diagnosis.

This paper describes some of the most common "traditional" condition monitoring techniques, such as:
-resistance to ground (RTG) testing,
-surge comparison (Surge) testing,
-high potential (HiPot) testing,
-motor current balance analysis (MCBA).

It then contrasts those listed above to some newer methods, such as:
-motor circuit analysis (MCrA),
-motor current signature analysis (MCSA),
-motor power or electrical signature analysis (MPA),
-motor flux analysis (MFA),
-motor normalized temperature analysis (MNTA).

Payoffs using all of these methods are presented in the form of actual examples or results which may be reasonably expected based on the nature of the measurements taken. Problems most commonly encountered employing or interpreting data from the test methods are also described.

Analysis techniques used for each of the methods listed above are tabulated. Analysis techniques most commonly used in predictive condition monitoring include the following:
-trend analysis,
-pattern recognition,
-correlation analysis,
-tests against limits and ranges,
-relative comparisons,
-statistical analysis.

Correlation of data from various technologies can greatly strengthen a predictive maintenance program. This analysis technique can become much more widely used if vendors of predictive and diagnostic software adhere to "open architecture" and common communications and protocol standards being developed and discussed by an industry consortium which has been working on this subject for over a year.

Difficulty gaining acceptance by maintenance professionals of newer motor condition monitoring methods is described. Part of the difficulty comes from the fact that it takes a long time for organizations which set standards to reach consensus and publish criteria for employment of new monitoring methods. This is being overcome somewhat by individual and at least one group of companies which have established standards for purchase of motors. The standards require measurements based on the more advanced monitoring methods described in this paper.

CONDITION MONITORING METHODS: Following are descriptions of the condition monitoring methods listed above. It is recognized that other methods exist which have been proven useful in monitoring motor electrical conditions of various types. However, the methods selected for presentation in this paper are those which are commercially available and applicable to the broadest range of motor types and outputs.

-Resistance to Ground (RTG) Testing - As applied to motors, the most commonly used method of resistance to ground (RTG) testing measures the leakage current flowing to and through an insulation system to ground under the pressure of a controlled (known) electromotive force (voltage). The test result, in ohms, is derived by dividing the measured current into the known voltage. This relationship is known as Ohm's Law. National electrical codes, industrial and professional standards institutes and associations provide regulatory standard minimums for RTG values on circuits required to carry electric power. The condition being monitored is the integrity of the insulation system isolating the circuit from ground. For larger motor circuits, a variation of this test method involves calculation and analysis of ratios of RTG values taken seconds or minutes apart on the same motor circuit. These variations eliminate any need for temperature correction of RTG measurements and make apparent the combined effects of circuit capacitance and insulation polarization on total current flow caused by application of voltage to the circuit under test. Condition criteria exist for certain commonly used RTG ratios, such as polarization index (ratio of RTG after 10 minutes to RTG after 1 minute of continuously applied, constant voltage) and dielectric absorption ratio (RTG value taken at 1 minute divided by the RTG value taken after 30 seconds of continuously applied, constant voltage). Almost all maintenance electricians have conducted resistance to ground tests at some point in their careers. However, few know what is happening during the testing, which can be costly when readings approach minimum allowable values.

-Surge Comparison (Surge) Testing - Surge or surge comparison testing involves insertion of controlled electrical pulse(s) into a motor from (one or twin) capacitor or capacitor-like circuit(s). The return pulses, which have been "damped" and may exhibit instability caused by the effects of changing inductive reactance of the motor coils, are evaluated to assess the condition of winding coil turn-to-turn and ground insulation in all motors. Surge testing also reveals phase-to-phase insulation and coil orientation (erroneous connection) problems in polyphase motors. Evaluation is done using oscilloscope trace(s) of return pulse(s) created by effects of any variation in impedances of part(s) of the motor circuit on the inserted (DC) electrical pulses. Comparison is most often performed with simultaneously inserted pulses into polyphase motors. Modern surge testers include a computer for storage and later recall of oscilloscope traces of return pulses for comparison with those from the same circuits at widely separated times. This allows more reliable evaluation of changes in motor electrical condition to be performed by comparison of traces taken months or years apart on DC and single phase AC motor circuits. It also allows an analyst to subjectively evaluate health and degradation of polyphase motors over time. Many motor manufacturers use surge testing as a quality tool. Motor rewind shops use the method for diagnosing incoming motors. Many facilities use the technique for diagnosis, periodic testing and more recently as a predictive condition monitoring tool for motors and small generators.

-High Potential (HiPot) Testing - HiPot testing involves application of either AC or DC voltage higher in value than that for which a motor circuit is rated. The test is used to evaluate the integrity or margin of the ground insulation system against its breakdown under electromotive forces. Guideline evaluation criteria are available from various codes, standards and texts on maximum values to use in conducting the testing. Voltage is controlled by the test unit operator. Current flowing into the motor circuit and thence through and over the ground insulation system under test is measured and recorded. If no indication of insulation breakdown occurs at the maximum voltage established by the test criteria, the test is terminated and declared to be successful. Many motor rewind shops and facility maintenance teams use this method routinely.

-Motor Current Balance Analysis (MCBA) - For polyphase motors a simple set of two or three on-line current measurements, their mathematical comparison and calculation of percentage unbalance is an excellent indication of motor condition. Unbalance usually results from impedance mismatch between phases of a motor circuit. Impedance changes occur because of motor winding degradation (turn-to-turn or phase -to-phase shorts) and/or the development of high resistance connections anywhere in the motor circuit from motor control center into the motor. Current unbalance can also be caused by voltage unbalance from the generation, transmission or distribution supply systems, although not as frequently. In the latter case, although the cause may be external, motor degradation and failure of one or more of its insulation systems will occur (due to overheating and accelerated aging) if the unbalance is not corrected.

-Motor Circuit Analysis (MCrA) - Motor circuit analysis (MCrA) involves measurement of four "natural" electrical parameters using closely controlled AC and DC input signals from a test unit. These are resistance in the conductor path, inductance, capacitance to ground and resistance to ground. Analysis involves combinations of these parameters, calculations of unbalance and graphical diagnosis. Results may be compared to well accepted standards (e.g., for RTG) or empirically derived guidelines.The guidelines are based on experience analyzing tens of thousands of motors. Motor predictive maintenance personnel can then quantitatively characterize motor circuit electrical conditions. This method is growing in popularity and is used internationally. The most advanced application of this test method indicates the following problems:

- Resistive unbalance in AC three phase motor circuits (causing overheating and premature and uneven stator winding failure)
- Excessive resistance in DC motor circuits (which may interfere with control)
- Inductive unbalance in AC three phase motor stators (indicating quantitatively the extent of shorted turns)
- Loss of inductance in AC synchronous or wound rotors (indicating quantitatively the extent of shorted turns)
- Loss of inductance in DC field or armature winding circuits (indicating quantitatively the extent of shorted turns)
- Presence of broken or cracked rotor bars or end rings and electrical or mechanical eccentricity
- DC armature shorts, opens or grounds (through quantitative and graphically presented

bar-to-bar readings and profiles)
- Temperature corrected and time consistent resistance to ground (RTG) conditions.
- Build up or presence of dirt or moisture on the outside of winding insulation systems (through correlation with capacitance to ground measurements over time)
- Polarization Index, Dielectric Absorption or other ratios from RTG readings

-Motor Current Signature Analysis (MCSA) -This method is known by several different commercial and generic names, many of which abbreviate to the same set of initials, MCSA. The most common applications involve analysis of two AC motor line current spectra. The first is in the frequency domain around the power supply line frequency (60Hz in North America, 50 Hz in most other parts of the world). The second is in the frequency domain around the center frequency calculated by multiplying the number of rotor bars or stator slots times the motor rotational frequency. Frequency spikes (sidebands) in the spectrum around the line frequency spike indicate the presence and influence on current amplitude of various faults in motors as well as dominant mechanical characteristics of both the motor and device(s) driven by it. A set of empirically derived numerical relationships between amplitudes of the line frequency and sideband frequencies indicate the severity of the problem being analyzed. Sidebands around the slot times rotational frequency and their relative amplitudes indicate the presence and relative severity of eccentricity problems. The most common faults detected (but not all distinguishable from each other) by MCSA are:
- Broken or cracked rotor bars
- High resistance joints in rotor bars or wound rotor conductors
- Broken or cracked end rings in squirrel cage rotors
- Casting porosity affecting current flow in die cast rotors
- Static and dynamic eccentricity conditions between rotor and stator
- Mechanical defects associated with the rotating element (e.g., bearing degradation)

-Motor Power (or Electrical Signature) Analysis (MPA) - This method involves measurement, conditioning, instantaneous recording and further processing for analysis in time and frequency domains of all phase currents and voltages associated with a motor. From the analysis and various easily performed calculations, power factor, real, reactive and apparent power can be derived, tabulated and graphically presented. Instantaneous variations of these and the basic, measured values in each phase, one phase compared to the others and power in the overall circuit may be used to assess conditions of the motor, the power system supplying and controlling it and the device(s) driven by it. Power or electrical signature analysis provides a potentially huge amount of data and opportunities to detect degraded conditions at very early stages of their development. The method has been used to monitor not only for motor defects but also for analysis of mechanical performance of valves, pumps, couplings and other devices being driven by or through them. Full capabilities have yet to be realized due to the fact that the MPA method has gotten attention only in the past few years. Most of that attention has been in the nuclear utility industry, which has some limitations on extent to which relatively new methods can be applied.

-Motor Flux or Leakage Flux Analysis (MFA) - Magnetic flux created by currents in motors is concentrated largely inside the enclosure or shell. However, some of the magnetic field may be detected in the space near the outside of most motors. This "leakage flux" varies with conditions found inside the motor and the condition of power supplied to it. Over the years, various investigations have resulted in some correlation between trended variations in the leakage flux field and defects such as broken rotor bars and stator turn-to-turn or phase-to-phase shorts. Analysis involves comparison of conditioned output signals of a flux detection coil consistently mounted external to the motor. The signals are presented on a Fast Fourier Transformed frequency spectrum which typically runs from 0 Hertz to 10 or 20 Hertz above two times the line frequency. Specific frequency lines in the spectrum of flux amplitudes can be related to particular types of defects. As the amplitude of a particular line increases, it indicates that the condition identified as related to it is growing worse. Sideband analysis similar to that described in motor current signature analysis is also used, although there is no direct correlation established (or revealed) yet between the numbers used in each method.

-Motor Normalized Temperature Analysis (MNTA) - Most motor defects are accompanied at some point in their development with increased temperature. In applying MNTA, temperatures are taken at specific points, such as on the outside of an end bell closest to where a bearing is located, an outlet air vent or the outer skin of the casing. The measured temperatures must be "normalized" to account for load and ambient conditions and are trended over time. Increasing trends may indicate the onset of degradation. Effects of location relative to direct sunlight, of design (such as outer skin thickness and material used, internal airflow patterns, frame design, etc.,), and surface paint color must all be accounted for or mitigated. Many critical motors are equipped by the manufacturer with temperature sensors at key locations, making the application of this method of motor condition monitoring redundant. However, there are many critical motors without this capability which are candidates for monitoring with relatively low cost, hand held temperature measurement devices. Typical conditions manifested by increased temperatures in motors include:

- Degrading bearings
- Rotor faults (broken or cracked bars and end rings)
- Clogged ventilation filters or screens
- Stator winding faults (turn-to-turn and phase-to-phase)
- Couplings which are misaligned or need lubrication
- Unbalanced currents which are caused by high resistance in the motor circuit

PAYOFFS: Each of the methods described above provides either a subjective (non-quantitative) or objective (quantitative) measure of condition of motors. Some provide much more, indicating condition of the entire motor circuit, the power supplied to the circuit, condition of the motor drive in the circuit and in some cases the condition of the device being driven. The payoff is the capability to assess condition(s) before actual failure. This allows orderly planning and repair or replacement to occur at minimum overall cost, including the cost of lost production, which is

often the dominant factor in the payoff total. Table 1 summarizes the way (subjective and objective) conditions detectable from the methods described in this paper are characterized.

Table 1. Motor Electrical Condition Monitoring Capabilities of Various Methods
Condition Monitored O=Objectively S=Subjectively

Method	Insulation System Integrity	Contam-ination	Coils/ Wind-ings	Sq. Cage Rotor	Armature & Sync Rotor	Circuit and/or Drive	Bearings and Coupling	Driven End
RTG	O	O						
SURGE	S		S		S			
HiPot	O							
MCrA	O	O	O	O	O	O		
MCBA			O			O		
MCSA				O				O
MPA			O	O		O	O	O
MFA			O	O		O		
MNTA		O	O	O			O	

Motor circuit analysis, an off-line method, has the capability to monitor 6 of the 8 categories in the table. Motor power analysis, an on-line method, has the capability to monitor 5 of the 8 categories, including the two not monitored by motor circuit analysis. If both were applied to the same motor driven system, they would appear from the table to have combined capabilities to provide the most comprehensive condition data upon which to base overall assessments. No other combination of methods covers all of the categories tabulated. However, within each column of Table 1., the presence of an entry for a method (O or S) indicates a possibility of correlation of its data with the data from any other method which has an entry in the same column.

Correlation of data from predictive condition monitoring methods, such as vibration analysis, infra red thermography and wear particle analysis, with the methods described in this paper is strongly recommended and easily accomplished. Use of the correlation technique pays off in greatly strengthened predictive analysis conclusions and more compelling reports. Further discussion on this subject may be found in references 1 and 5 of the bibliography at the end of this paper.

Other analysis techniques, besides correlation, are used to assess present and predict future conditions in motor driven systems. Table 2. summarizes the techniques used for each method covered in this paper.

As illustrated in Table 2., only motor circuit analysis and motor power analysis provide data of sufficient quantity and variety to permit application of all five predictive analysis methods. This is another reason for selecting this combination of motor, predictive condition monitoring methods.

Table 2. Analysis Techniques Used On Data from Various Motor Electrical Monitoring Methods

	Analysis Technique				
Method	Trend Analysis	Pattern Recognition	Relative Comparison	Test Against Limit or Range	Statistical Analysis Technique
RTG	X			X	
SURGE		X	X	X	
HiPot				X	
MCrA	X	X	X	X	X
MCBA	X		X		
MCSA	X	X	X		
MPA	X	X	X	X	X
MFA	X	X	X		
MNTA	X	X	X		

PROBLEMS: Problems may be encountered in application of any condition monitoring method. Two particular problems concerning resistance to ground readings are managed effectively by one version of MCrA technology. They are consistent temperature correction and control of the time of measurement after application of the DC ground test voltage. These two criteria become very important when ground readings in circuits are approaching marginal conditions.

In a 1994 incident, excessive voltage was applied to 220 run out table motors in steel plant hot rolling mill. To determine if any were damaged as a result of the incident, it was decided that a "quick" check of resistance-to-ground (RTG) readings on these DC motor circuits was prudent. Electricians were dispatched with hand cranked and battery powered meters to take the readings as fast as possible. Readings taken indicated 138 out of 220 motors below acceptable values. The mill had only 10 spares available at the time. After questioning the results, the hot mill electrical foreman realized that the electricians had recorded the RTG readings without considering the time factor. RTG readings must be taken consistently at the same instant after application of the test voltage. Otherwise, circuit capacitive and polarizing influences will not be accounted for equally. Readings taken without consideration of the time factor are not repeatable nor comparable. The

foreman ordered the tests redone using the motor circuit analysis methodology and equipment. Because of the accuracy of the time controlled method, he was also able to lower the RTG acceptance values used for the plant. Results of the retests indicated only five motors had RTG readings low enough to need replacement under the revised standards. The accuracy and repeatability of the time controlled RTG test method allowed reliance on RTG readings closer to the alarm value where these motors were determined to be unacceptable for continued service.

Fifteen other run out table motors in the retest (which provided all four motor circuit analysis parameters described in this paper) were found to have high conductor path resistance. Most of these problems were located in the DC power connection plug at the motor and were corrected and confirmed to be fixed during testing. High conductor path resistance conditions, which could affect control of the motor speed, are not detectable with an RTG test.

Other problems with individual methods relate to perceptions and actual statements in authoritative documents which imply that use of certain methods such as Surge and HiPot may damage motors. In spite of compelling test results and arguments by users and condition monitoring vendors to the contrary, the perceptions persist.

Another way of looking at the nature of the problems is to review advantages and disadvantages of off-line and on-line methods. The first four methods listed in the above tables are all off-line methods. The last five are all on-line. Table 3 summarizes advantages and disadvantages.

TABLE 3. Advantages and Disadvantages of Off-line and On-line Motor Condition Monitoring Methods

	Off-line Monitoring	On-line Monitoring
Advantages	1. Active, known and controlled test signals used. 2. Fewer extraneous signals on de-energized motor circuits to mask test results.	1. Passive - No test signals needed. 2. Monitoring can be done without interfering with production. 3. Hookup and/or data collection may be achieved for some methods and installations without special electrical safety considerations (MFA, MNTA).
Disadvantages	1. Motor must be shut down to perform testing, which may interfere with production. 2. Circuit must be proven safely deenergized before hookup of ALL off-line methods. 3. Test signals used are not always the same for different vendors. 4. Residual magnetism may influence some results (MCrA)	1. Many normal conditions provide test indications which mask fault. indications (MCSA, MPA, MFA). 2.Test hookup may be difficult to achieve safely without motor shutdown at some installations for some on-line methods (MCBA, MCSA, MPA). 3. Local conditions may influence test results (MNTA).

CONDITION MONITORING STANDARDS: Organizations which establish standards are typically slow to take up the subject of new monitoring methods. In the eyes of some potential users, the lack of national or international criteria for acceptance or rejection of motors based on condition data from new monitoring methods is justification for not applying them.

At least one ad hoc group of companies which buy large numbers of new and rebuilt motors each year has established acceptance criteria derived from motor circuit analysis experience for their purchasing agents to use. Criteria such as resistive and inductive unbalance are included in the consensus standards. Motor vendors were slow to accept these new requirements at first but have started to respond positively after threat of disqualification from bidding was raised by buyers.

Another ad hoc industry group, The Machinery Information Management Open Systems Alliance (MIMOSA), has begun an effort to develop conventions to achieve cost effective data exchange between dissimilar condition assessment, maintenance information and process control systems. The alliance teams are providing recommendations which will allow connecting proven, but fragmented, asset management technologies into high value, comprehensive providers of essential information. Such efforts will greatly aid establishment of central data bases and enable correlation analysis to be applied more easily, among many other benefits.

BIBLIOGRAPHY:

1. J. Flude and J. Nicholas, Predictive Maintenance Primer, April 1991, Electric Power Research Institute Nuclear Maintenance Applications Center, Palo Alto, CA.
2. J. Nicholas, Motor Electrical Predictive Maintenance and Testing, 1995/96 Edition, PdMA Corporation, Tampa, FL.
3. J. Nicholas, "Evaluating Motor Circuits", Maintenance Technology Magazine, November 1992, Chicago, IL.
4. J. Nicholas, "Predictive Maintenance of DC Motors", P/PM Technology Magazine October 1993, Minden, NV.
5. J. Nicholas, "Predictive Condition Monitoring of Electric Motors", P/PM Technology Magazine, August 1993, Minden, NV
6. J. Nicholas, "Motor Circuit and Motor Current Analysis", AIPE Facilities Magazine, July/August 1994, Cincinnati, OH.
7. B. Roberts, "Power Signature Analysis for Motor Driven Device Monitoring and Diagnosis", Duke Power Company, Huntersville, NC.
8. G. Lang, "Of Cages, Induction, Deduction, Bars, Vars and Squirrels" Sound and Vibration Magazine, December 1994.
9. S. Bowers and K. Piety, "Proactive Motor Monitoring" Maintenance Technology Magazine, June 1995, Chicago, IL
10. D. Brooks, V. Morgan, J. Nicholas, "Motor Analysis at Dofasco" Paper presented at the Conference on Advanced Technologies for the Steel Industry, October 1995, Cleveland, OH.

THE USE OF THE MOTOR AS A TRANSDUCER TO MONITOR SYSTEM CONDITIONS

D. A. Casada and S. L. Bunch
Lockheed Martin Energy Systems, Inc.
P. O. Box 2009
Oak Ridge, TN 37888

Abstract: Motor current and power analysis methods have been developed to assist in the condition monitoring of a variety of motor-driven devices. The early work in this area was conducted at Oak Ridge National Laboratory (ORNL) on motor-operated valves in the mid-to-late 1980's in support of the U.S. Nuclear Regulatory Commission's Nuclear Plant Aging Research Program. The successful implementation of motor current signature analysis (MCSA) as a diagnostic for valves led to its application to other devices and to refinements in the methodologies used.

In the last few years, work for the Nuclear Regulatory Commission and independent activities at ORNL and the Y-12 defense plant have resulted in additional, more thorough consideration of the merits of using motors as transducers for understanding system conditions. A variety of applications, ranging from fractional to over 1200 horsepower have been analyzed.

Motor current and power analysis have been found to provide information that is complementary to that available from conventional diagnostics, such as vibration and pressure pulsation analysis. Inherent signal filtering associated with rotor to stator magnetic field coupling does limit the high frequency response capability of the motor as a transducer; as a result, certain phenomena, such as pump or fan vane pass energy, is not readily apparent in the motor electrical signals. On the other hand, the motor-monitored parameters have often been found to be much more sensitive than vibration transducers in detecting the effects of unsteady process conditions resulting from both system and process specific sources.

Key Words: Condition analysis; current; MCSA; motor; power; process conditions; system analysis.

Background: Almost all motor-driven devices experience fluctuating loads. The types of load fluctuations are dependent upon the specific design features of the device and the system within which it is operated. The load fluctuations can be periodic or chaotic in nature. For those loads which are periodic, the period can vary from fractions of milliseconds to months or years.

The relatively straightforward recognition that the motor speed (for induction motors), current, power and power factor change in response to load changes led investigators studying potential means of diagnosing motor-operated valve conditions to evaluate the usefulness of the motor as a

transducer in the 1980's [1]. The ability to identify a variety of load-related phenomena, such as lubricant degradation and gear tooth wear, by use of demodulated motor current was demonstrated. The researchers dubbed the methodology of analyzing the time and frequency domain current signals to understand the driven device conditions motor current signature analysis (MCSA). The application of this term has since been applied rather broadly to a variety of waveform and spectral analysis techniques, and has been used when characterizing both the motor and the driven device.

Probably the most common use of current monitoring is by operators simply observing permanently installed ammeter readings to verify that the motor is energized and that its current draw is consistent with expectations. In a diagnostic sense, spectral analysis of the motor current has been used to help understand the condition of the motor, with primary focus on rotor eccentricity and rotor conductor condition.

While the use of motor current data is helpful in understanding motor electrical and magnetic field conditions, this paper is focused on understanding other conditions in the drive train. There are both some relatively straightforward as well as more complex applications of motor current or power data. This paper will not discuss in detail either the theory behind the approaches used to extract useful diagnostic information nor the various types of instrumentation used, but will rather identify some examples of effects that can be identified.

In no way is it the intent of this paper to suggest that motor current or power analysis be considered as a replacement diagnostic for technologies such as vibration. There are clearly many defects and conditions detectable through vibration measurements, for example, that are not detectable through the motor data. However, there are some conditions for which motor data is more sensitive than most other diagnostic methods. There are also situations where it is easier to implement than many other techniques. Finally, it can be useful in confirming indications developed from other diagnostic techniques.

Typical motor performance characteristics: The early applications of motor current signature analysis focused on current specifically because of the ability to acquire current data in a relatively non-intrusive fashion (using a clamp-on current transformer, for example), as opposed to the somewhat more intrusive nature of power and power factor measurements which require direct connection to the motor supply leads.

A typical motor performance curve for a 4-pole 150 hp motor is shown in Figure 1. Note that the input power is very linear with output power across the full range of motor load; over the range from 50 to 150 bhp, the current is essentially linear as well. The power factor is relatively insensitive to load variations within this region, particularly near rated load. At very light loads, the current response to load changes is small, while the power factor changes dramatically with relatively small changes in motor load. Motor performance curves do vary somewhat for different designs, but the general performance characteristics are not dramatically different in the normal operating range.

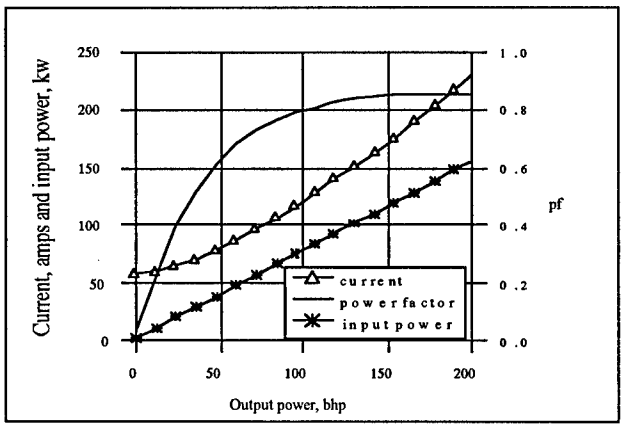


Figure 1. Typical Motor Performance Characteristics

There are complicating factors, such as deviation of actual bus voltage from design and power quality-related issues, that limit the accuracy of generic, or even motor-specific curves. However, a variety of motor models have been developed recently which have been found to be capable of predicting motor output power accurately to within three percent or better [2].

It is important to recognize that motor performance curves such as that shown in Figure 1 are typically based on measurements averaged over periods of seconds or minutes. On an instantaneous basis, significant departure from these average values may occur. For example, the head and flow developed by a pump at low-flow conditions are often less stable than when the pump is operated near its best efficiency point. The instability may consist of both some relatively well-defined frequency components, such as vane-passing as well as more chaotic instabilities that are manifested as broad band noise. When such instabilities affect shaft torque (and as a result, motor load), it becomes possible to detect both the existence and magnitude of the instabilities, at least in a relative context.

To a great extent, vibration monitoring is a useful diagnostic technique because of similar considerations. Accelerometers, for example, do not measure absolute forces for rotating equipment; rather, they sense the acceleration occurring in response to fluctuations in forces. The long history of vibration monitoring has provided a basis for not only rules of thumb, but industry standards.

At this point in time, there is minimal experience in the use of motor-transduced data as a system diagnostic. At ORNL and at the Y-12 defense plant, motor current and power data for a variety of equipment have been collected over the last few years. Motor current or power has been found to provide insights into both the monitored component specifically and to overall system conditions, as well. The balance of this paper will discuss examples of the types of results that have been found.

Comparison of the sensitivity of motor input power and vibration measurements to common sources of vibration: From test data acquired on a variety of types of equipment over several years, fluctuations in motor spectral data that are related to some load component have been observed. In many cases, researchers and test personnel have been content to observe the existence of some load-related peak in the spectral data, and perhaps trend the amplitude with time. It was noted during certain tests that there were characteristic peaks in the motor current or power spectra that also existed in vibration spectra. Motor power and pump vibration spectra for a belt-driven positive displacement vacuum pump are shown in Figures 2 and 3. The loads associated with belt pass, pump stroke, and motor running speed can be clearly seen in the motor power spectrum. Some, but not all of the spectral peaks also appear in the vibration spectrum; the most noticeable difference is the absence of most of the belt pass frequency components from the vibration spectrum. This is typical for belt-driven devices, and is due to the torsional load fluctuations associated with both belt irregularities and belt looseness. While loose belts can result in elevated vibration in the radial direction, the radial effect is not nearly as dramatic as the torsional effect. As a result, radial vibration is often minimal, as was the case for the vacuum pump.

Positive displacement pumps, by their very nature, involve considerable torque fluctuations at the drive shaft during individual compression sequences. These torque fluctuations are readily transduced by the motor; as a result, the spectral amplitudes associated with pump compression frequencies are usually distinct (e.g., Figure 2), and are very sensitive to flow and head changes.

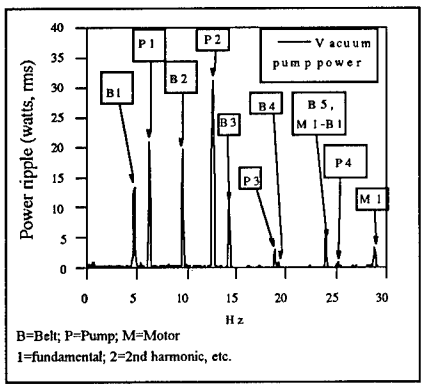


Figure 2. Motor Input Power Spectrum for Belt-Driven Vacuum Pump

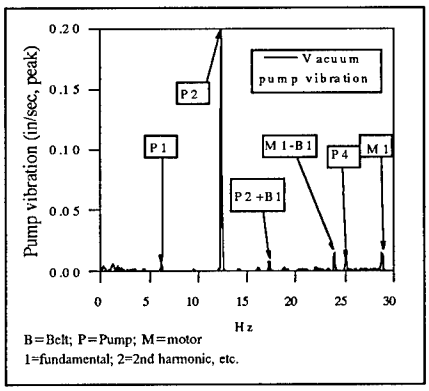


Figure 3. Vacuum Pump Vibration Spectrum

In general, motor-based measurements are more sensitive to torsionally-related loading than radial or axial vibration data. On the other hand, the motor is, generally speaking, a relatively poor transducer of radial and axial vibration that is not accompanied by torsional loading. Fundamentally, in order for a radial or axial load fluctuation to be seen in motor data, the radial or axial load must either result in relative motion between the motor rotor and stator and/or cause a change in torsional load.

Sensitivity to mechanical unbalance: A comparison of the results of vibration and motor data measured during tests conducted to measure the effects of unbalance and alignment have validated the observation that the motor, as a transducer, is primarily sensitive to torsional loads. In a controlled balance test performed at Oak Ridge, measurements of motor input and output data and vibration amplitude were made as the mechanical balance of a motor was changed. For this test, a 4-pole, 50-hp motor was connected to an eddy-current brake dynamometer, and the mechanical balance was changed by using a dynamic balancing device provided by Balance Dynamics, Inc. The dynamic balancer allows for on-line balancing of rotating equipment by moving fluid between chambers in a wheel attached to the shaft. The primary purpose of the test was to determine if mechanical balance had any effect on motor efficiency. While there was marginal, if any, effect on motor efficiency, a relatively well defined correlation was found between radial vibration and motor input power spectra. A plot of the energy of a running speed-related frequency from the motor power spectrum vs. motor inboard horizontal vibration velocity is provided in Figure 4. This almost linear relationship is not necessarily reflective of all types of unbalance. In this particular test, the balance wheel was located between the motor and the coupling. In field testing conducted on a 6000-hp fan motor, where two balancing wheels were used on the inboard and outboard ends of the fan, there was no observable change in motor spectra as the balance was changed. In the case of the 50-hp motor, it is likely that the unbalance forces resulted in slight dynamic changes in rotor to stator air gap, and thus was reflected in the motor data, while for the 6000 hp motor, the unbalance location (on the fan side of the coupling) affected neither the torsional shaft load nor the rotor to stator air gap.

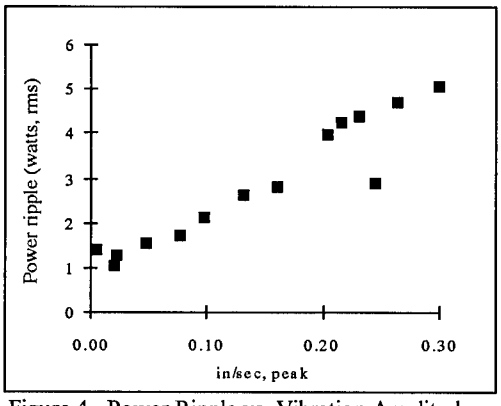


Figure 4. Power Ripple vs. Vibration Amplitude - Running Speed-Related Components for an Unbalanced Rotor

In general, motor data is very insensitive to mechanical unbalance. Note that even for the data shown in Figure 4, only a few watts of energy are involved. The motor was being run at its rated condition of 50 hp (output). The motor input power was about 41000 watts. About 0.01 percent or less of the motor energy was associated with the unbalance. On the other hand, the vibration data was dominated by the mechanical unbalance. Thus, although mechanical unbalance may be

detectable from careful analysis of motor data, it is neither a reliable nor sensitive indicator of unbalance, particularly when the unbalance exists on the driven device side of the coupling.

Sensitivity to misalignment: Because misalignment fundamentally results in not only radial and axial vibration, but torsional loading, the motor has been found to be a very sensitive transducer of misalignment. Testing at the Y-12 plant is presently assessing the effects of alignment on motor system efficiency. In conjunction with this testing, the sensitivity of motor and vibration data to alignment changes are being evaluated.

Motor inboard vibration, output torque, and input power spectral data for a 10 hp, 4-pole motor after laser alignment to the dynamometer are shown in Figures 5-7. Spectra for the same parameters with the motor outboard feet elevated by 20 and 30 mils are shown in Figures 8-10 and 11-13, respectively.

The running speed and two times running speed motor power (input) ripple and vibration velocity vs. output torque ripple for these and several other misalignment conditions are shown in Figure 14 and 15. Note that the amplitude of vibration is very low, even for the most severely misaligned condition (motor inboard feet elevated by 40 mils).

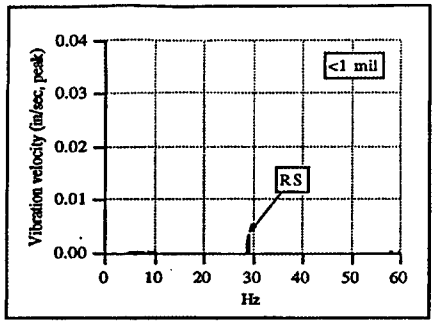


Figure 5. Well-Aligned Vibration Spectrum

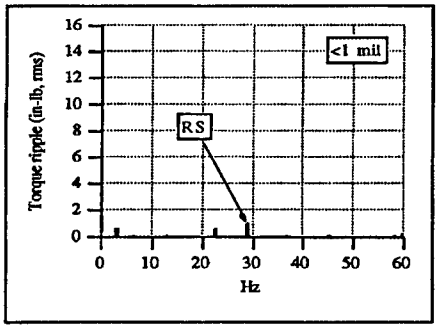


Figure 6. Well-Aligned Torque Ripple Spectrum

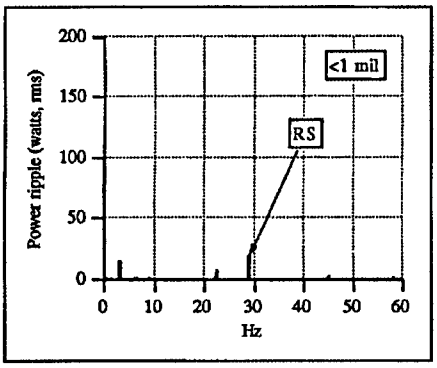


Figure 7. Well-Aligned Power Ripple Spectrum

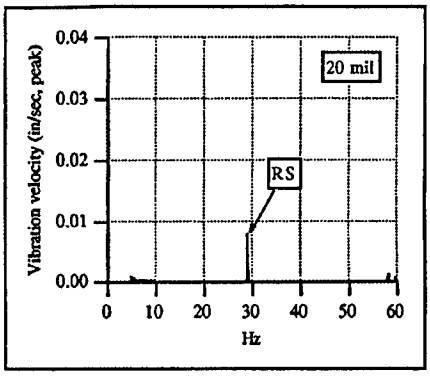


Figure 8. Vibration Spectrum with 20 Mil Shim

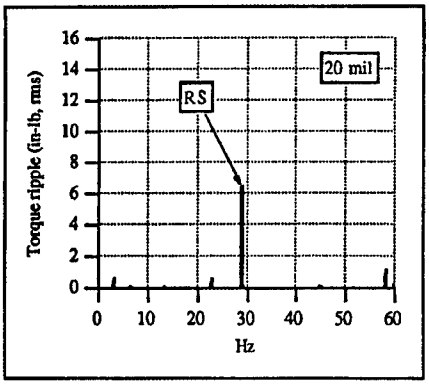


Figure 9. Torque Ripple Spectrum with 20 Mil Shim

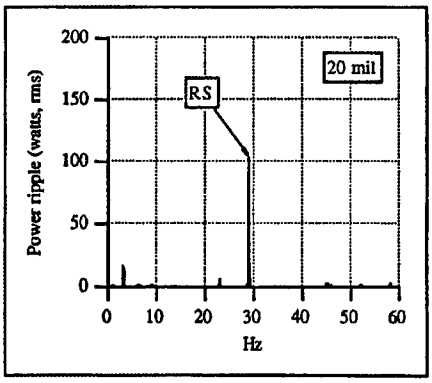


Figure 10. Power Ripple Spectrum with 20 Mil Shim

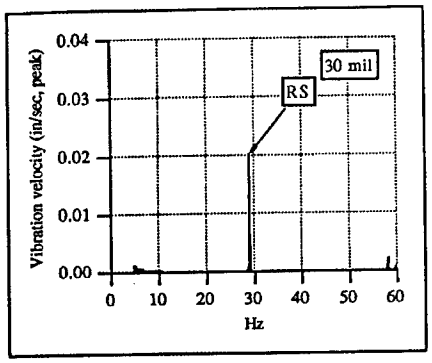


Figure 11. Vibration Spectrum with 30 Mil Shim

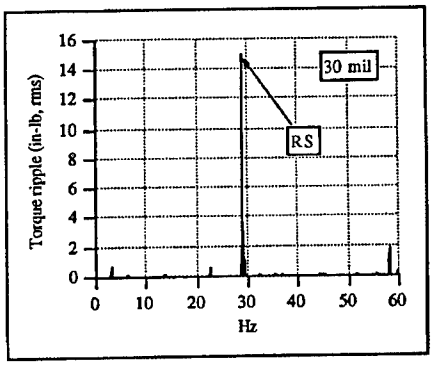


Figure 12. Torque Ripple Spectrum with 30 Mil Shim

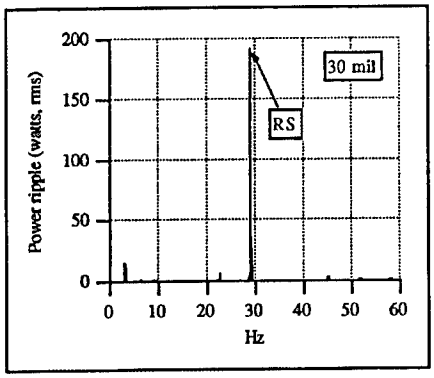


Figure 13. Power Ripple Spectrum with 30 Mil Shim

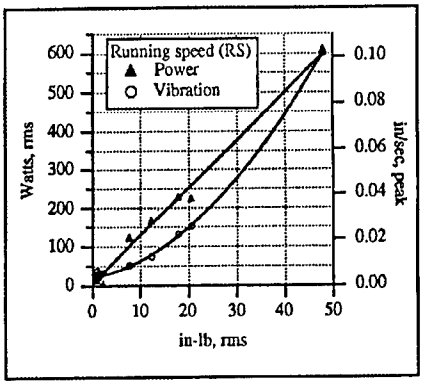


Figure 14. Running Speed Power Ripple and Vibration vs. Torque Ripple at Various Levels of Misalignment

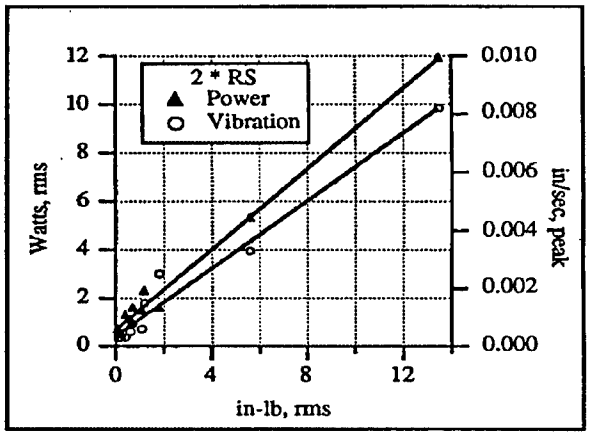


Figure 15. Two Times Running Speed Power Ripple and Vibration vs. Torque Ripple at Various Levels of Misalignment

Results of on-line monitoring: The Y-12 plant is in the process of implementing on-line motor current monitoring for motors used in a variety of processes. Both overall amplitude and spectral data are periodically acquired and logged. The data are archived and trended in the same database used for periodic vibration monitoring.

This system is still not fully implemented, but it has proven useful in detecting degraded component conditions. Two examples are provided below which are illustrative of both the capability and limitations of motor current/power monitoring.

Clogged suction strainer: Motor current for several pumps used in a chilled water facility are periodically monitored. A trend plot of motor current for two identical pumps operating in parallel over the course of 35 days is shown in Figure 16. In Figure 17, the motor power spectrum for the J102 pump at the end of the trend period is shown. Power spectral data for the same pump with a clear suction strainer is shown in Figure 18. From the combination of overall current trend and spectral data, as well as knowledge of previous experience for these pumps, it was deduced that the J102 pump suction strainer had become clogged. Subsequent removal of the suction strainer showed considerable clogging (Figure 19). Vibration data acquired for these pumps (not shown) failed to indicate significant differences.

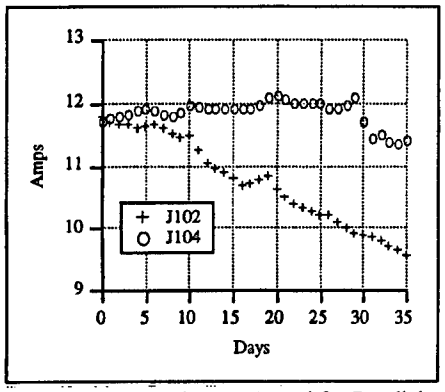


Figure 16. Motor Current Trend for Parallel Pumps

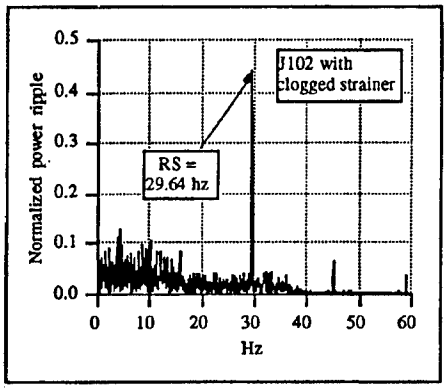


Figure 17. Motor Power Spectrum for Pump with Clogged Suction Strainer

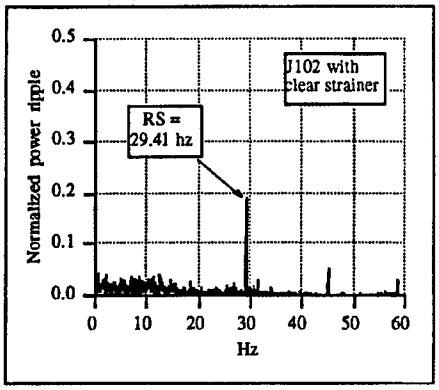


Figure 18. Motor Power Spectrum for Pump with Clean Suction Strainer

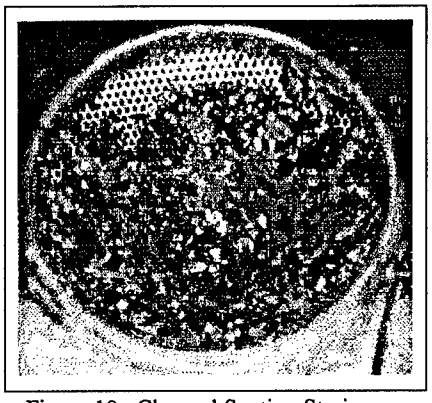


Figure 19. Clogged Suction Strainer

Damaged motor bearing: Spectral vibration data from a belt driven exhaust fan motor indicated the presence of a severe bearing defect (Figure 20). Since this motor had been recently put on the on-line current monitoring system, the motor current spectrum was evaluated for confirmatory indication. As shown in Figure 21, the fault frequency was also seen in the current spectrum. The amplitude shown for the current spectrum in this case is relative amplitude. Note that the fault here was in a motor bearing, and it was severe; even then, the amplitude of the fault frequency in the motor data was extremely small. There have been no examples at the Oak Ridge facilities where bearing faults in the driven equipment have been detected by motor spectral data. Also, many cases have been found where less severe motor bearing flaws have been detected by vibration data with no corresponding indication in the motor spectral data. Two important conclusions from the vibration and current data analysis experience relative to bearing fault

detection can be drawn:

- Spectral vibration data is much more sensitive to bearing fault conditions than motor data.
- Although severely degraded motor bearing faults may be manifest in the motor spectral data, it is judged to be unlikely that anything less than a severe fault would result in detectable levels in the current or power spectrum. Degradation in driven device bearings would be even less likely to be sensed by motor current or power spectral data.

These observations relative to bearing degradation are not based on controlled research studies, but rather on observations from field data. They are certainly not conclusive, but represent experience from a variety of motors and driven equipment.

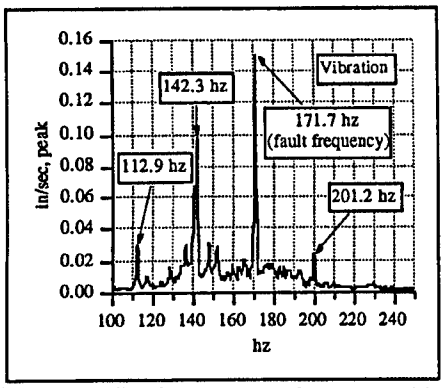


Figure 20. Vibration Spectrum for Motor with Damaged Bearing

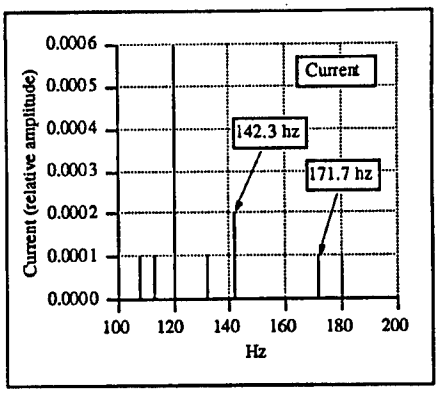


Figure 21. Motor Current Spectrum for Motor with Damaged Bearing

Resources: The use of motor current or power data in analyzing load behavior is not a well-developed field. For those that are interested in exploring its potential, the following may be of interest:

- The technologies developed at ORNL and at Y-12 to assess equipment performance from motor data have been licensed by several domestic diagnostic service companies, including ITI MOVATS, Inc. (a Westinghouse company), and Babcock and Wilcox, Inc. These organizations provide both diagnostic systems and field service based on the motor-monitoring technologies.

- ORNL and the Y-12 plant, both of which are Department of Energy facilities, have ongoing motor-related activities in both the research and facility operation support arenas. Motor data from pumps and other equipment are being used and analyzed in a variety of ways to improve operational efficiency and reliability, minimize unplanned maintenance, and reduce operational costs, as well as to support research objectives. Inquiries about technical aspects of these activities may be addressed to the authors.

Conclusion: Monitoring important system performance parameters by using the motor as a transducer has been found to be both feasible and cost effective. Fundamentally, the motor is effective at transducing torsionally-related load phenomena.

The motor can be effectively used as a transducer to aid the understanding of both alignment and process conditions. In contrast, the motor has not been found to be a particularly effective transducer of mechanical unbalance or bearing fault conditions.

Motor current and power analysis offer an alternative diagnostic approach that may be used in conjunction with existing diagnostics, or in some cases, as a primary diagnostic. As more experience is gained in its field use, it is expected that it may become possible to establish experientially-based indications of the acceptability of certain characteristic loads, such as those associated with misalignment or overall load stability.

References:

[1]. Haynes, H. D., *Aging and Service Wear of Electric Motor-Operated Valves Used in Engineered Safety-Feature Systems of Nuclear Power Plants,* U. S. Nuclear Regulatory Commission Report NUREG/CR-4234, Oak Ridge National Laboratory, August, 1989.

[2]. Kueck, J. D., Gray, J. R., Driver, R. C., and Hsu, J. S., *Assessment of Available Methods for Evaluating In-Service Motor Efficiency*, Bonneville Power Administration, expected publication March, 1996.

SUBSURFACE DEFECT DETECTION IN CERAMIC MATERIALS USING LOW COHERENCE OPTICAL SCATTER REFLECTOMETER

Mark Bashkansky, Michael D. Duncan, Manfred Kahn*, David Lewis III[†] and J. F. Reintjes[‡]

Laser Physics Branch, Code 5640
Naval Research Laboratory
4555 Overlook Ave. SW
Washington, DC. 20375
Tel: (202) 767-2512 Fax: (202) 404-7530

Abstract: We demonstrate the use of optical gating techniques for determining the size and location of subsurface defects in advanced ceramic materials. Various silicon nitride based ceramic materials are probed non-destructively using an optical gated reflectometer based on a low-coherence fiber interferometer. This device is capable of depth and lateral resolutions of 10 μm and 4 μm, respectively. Experimental results indicate that the size and position of small subsurface defects can be determined as deep as 500 μm below the surface.

Key Words: Ceramic; nondestructive evaluation; optical gating

Introduction: Advanced ceramic materials have great potential in industrial and military applications because of their superior thermomechanical and chemical properties. However, their widespread use is limited in part due to the presence of defects in both the bulk and on the surface of the ceramic. Defects are manifested as voids and cracks, or as inclusions which are caused by inherent powder defects or poorly distributed second-phase material. Such defects can be introduced at various stages of the manufacturing process. Potentially, these defects can reduce structural integrity of the ceramic components and lead to variability in their material properties, both of which may be unacceptable in critical applications. Because the potential market for ceramic components is so large, a considerable amount of effort has been put into developing techniques to detect flaws at all stages of the manufacturing process[1-6].

Defects that exist near the surface are particularly critical in many applications, since operational stresses in this region of the ceramic component can be greatest. Therefore, flaws in the vicinity of the surface can compromise the strength, degrade performance and shorten lifetime of a ceramic component. These flaws may be intrinsic to the bulk material, introduced in the final

stages of fabrication (e.g. machining, grinding and polishing) or appear during the operation of the device.

Recently, several ultrasonically based non-destructive evaluation (NDE) techniques have been applied to the detection of surface and subsurface damage in ceramic materials[3-5]. In photoacoustic microscopy an amplitude-modulated laser beam is used to excite the acoustic modes in the sample[4]. A transducer, attached to the sample, is used to detect the frequencies and the amplitudes of the acoustic modes. Even though, in principle, this measurement can yield information about the flaws in the surface and near surface regions, this technique has had difficulty in detecting smaller flaws and adequately shielding the device from other sources of noise. In pulse-echo NDE, the acoustic modes are excited by a highly focused ultrasonic pulse[5]. By focusing the ultrasound to a spot below the surface some of the pulse energy goes into exciting a surface mode of the ceramic. To date, to our knowledge, no distinguishing signatures have been found in the return echo which can be conclusively associated with subsurface damage.

An alternative to acoustic techniques has recently been developed where the polarization properties of the reflected light are utilized for defect detection in ceramics[6]. As light propagates through the ceramic it is both absorbed and scattered. The scattering occurs at the grain boundaries and interfaces with the second-phase material, where the index of refraction changes discontinuously. To distinguish between surface and subsurface reflections, both of which have different light scattering characteristics, a NDE technique using a cross-polarization method was developed[6]. If the surface of the ceramic is optically polished then the light that reflects from the surface will, for the most part, retain the polarization of the incident beam. In this case, an analyzing polarizer placed before the detector can be oriented to block most of the reflected light from the surface. However, the light that is scattered by the irregular defects below the surface is most likely depolarized and, therefore, is partially transmitted through the polarizer. Depending on the orientation of the analyzing polarizer, the light is detected, for the most part, from either the surface or the subsurface region. An advantage of optical techniques over ultrasonic techniques is that no contact with a sample is required. However, the limitation of the cross-polarization method is that it cannot determine directly how far below the surface a defect is located, or accurately measure how large the defect is. This information is important because defects that are either far below the surface or smaller in size may result in a fewer failures than large defects or defects near the surface.

Recently, a variety of ultra-fast optical range-gating techniques have been developed for potential biomedical applications[7-10]. In general, these techniques utilize the fact that photons which scatter below the surface travel farther to reach the detector than photons that scatter from the surface. Therefore, when a short optical pulse scatters from a sample, the detected return will be spread out in time. There will be a near one-to-one correspondence between the time the light arrives at the detector and the depth from which it came. Thus, by temporally resolving the intensity of the scattered light it is possible to determine the depth location of defects in the sample. Furthermore, it is possible to utilize either very short light pulses or long pulses with very short coherence length. These techniques have demonstrated depth resolutions in air $\leq 10\ \mu m$.

In this work, we applied an optically-gated reflectometer that is based on a low coherence fiber interferometer[11,12] to the detection of both the location and the size of defects in the subsurface region of ceramic components. This device has advantages of simplicity and low cost.

Experimental Setup: The experimental set up for detecting subsurface defects in ceramics is shown schematically in Fig. 1. Optical radiation at 1.3 μm is produced by a light emitting diode (LED). The full-width-at-half-maximum (FWHM) of the spectral content of the radiation was measured to be ~ 40 nm at an output power of 130 μW. The FWHM of the correlation length of this LED in air (equivalent to the depth resolution) is a function of the spectral content of the radiation and for a Gaussian spectrum is given by[9] $\Delta L = \ln(2)\frac{2}{\pi}\frac{\lambda^2}{\Delta\lambda} \approx 20\mu m$.

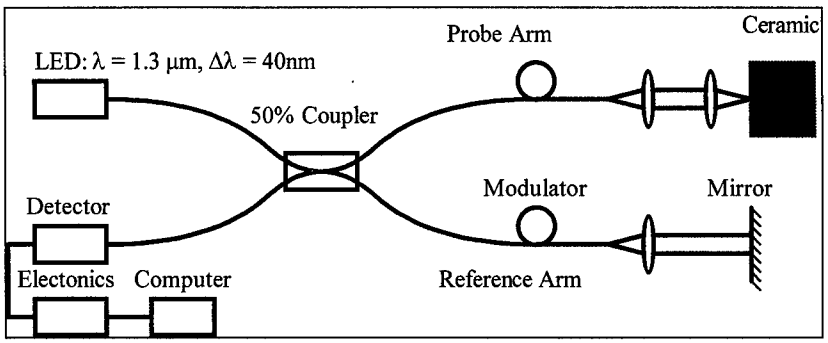


Figure 1

Experimental schematic of the low-coherence fiber interferometer system used to detect subsurface defects in ceramic material.

The LED radiation is coupled to a fiber and split into reference and probe beams in a 50 % fiber coupler. The phase of the reference beam is modulated by stretching the fiber with a piezoelectric transducer. The reference beam is retro-reflected from the mirror back into the fiber and toward the 50 % splitter. The probe beam is focused on the ceramic sample. The probe radiation scattered from the ceramic retraces its path through the fiber and recombines with the back-propagating reference beam in the 50 % splitter. This combined radiation is detected by a fiber-coupled photodiode. The spatial extent of the probe beam's focal spot was measured to be ~7 μm (FWHM). This gives resolution roughly 1/5 of that allowed by the diffraction limit at 1.3 μm.

When the pathlength difference between phase-modulated reference and probe beams is within the coherence length of the radiation, the interference between them produces an alternating current (AC) component on the photodiode signal. The frequency of this AC component depends both on the frequency and the depth of the reference beam modulation. This AC

component represents the range-gated signal and its magnitude is proportional to the amount of light reflected from the ceramic. The photodiode signal is AC coupled, and rectified such that a direct current (DC) signal is produced proportional to the root-mean-square (RMS) of the AC signal. The logarithm of this DC signal is stored in a computer after analog-to-digital conversion.

By adjusting the relative depth (Z) position of the ceramic, the light that is scattered from different depths of the ceramic will become correlated with the reference beam. Only the correlated component of the scattered probe beam contributes to the AC signal on the photodiode. Thus, by varying the X-Y-Z position of the ceramic at the focus of the probe beam, the sizes and locations of the defects in the ceramic material can be mapped in three dimensions.

RESULTS: The first ceramic material used in this study was a hot, isostatically pressed (HIPed) silicon nitride, designated NCX 5102, which was obtained from Saint-Gobain/Norton Industrial Ceramics Corp. This ceramic is primarily silicon nitride (Si_3N_4) with a small percentage of Y_2O_3, a second phase material.

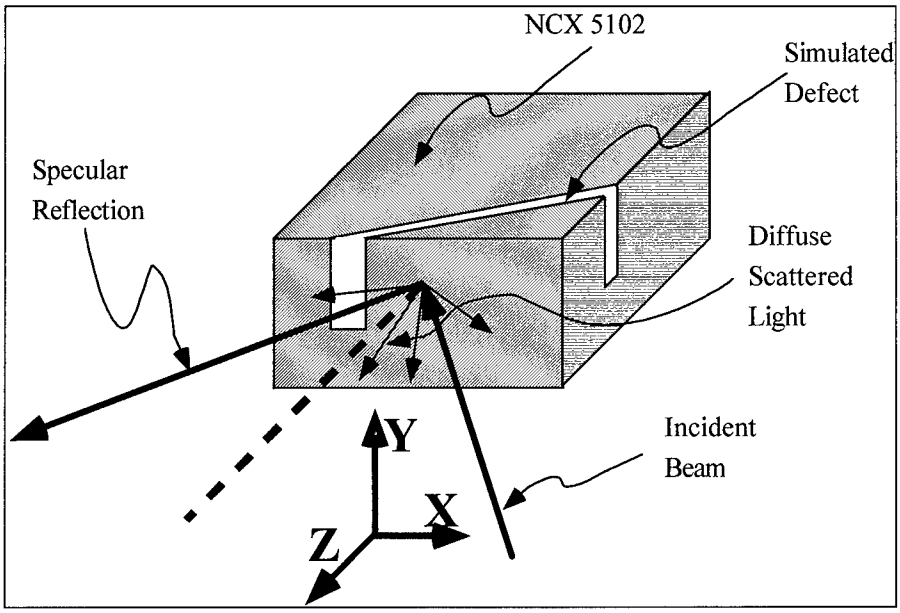


Figure 2

A schematic diagram shows the geometry used for subsurface defect detection in a ceramic material. The probe beam is incident at an angle with respect to the front surface of the ceramic so that the specular reflection is rejected by the imaging system. A saw cut was made in the ceramic at a slight angle to the front surface in order to simulate a void-like defect at varying depths.

NCX 5102 was developed for high temperature applications such as in components of gas turbine engines. The choice and concentration of second phase material has a significant effect on the optical properties of a ceramic[6].

The dimensions of the ceramic sample are 3x4x10 mm. The 3x4 mm surface is positioned at an angle to the incident signal beam so that the specular reflection from the surface is not collected by the imaging system, as shown in Fig. 2. A subsurface defect is simulated in the sample by cutting a grove approximately 200μm wide and 2 mm deep. The saw cut is at a small angle to the surface so that as the ceramic is translated horizontally in the X-direction the void, as seen by the probe beam, moves deeper into the ceramic.

Fig. 3 shows an X-Z image through the sample reconstructed from the individual range-gated scans in Z. As the probe beam propagates through the ceramic it is both scattered and absorbed.

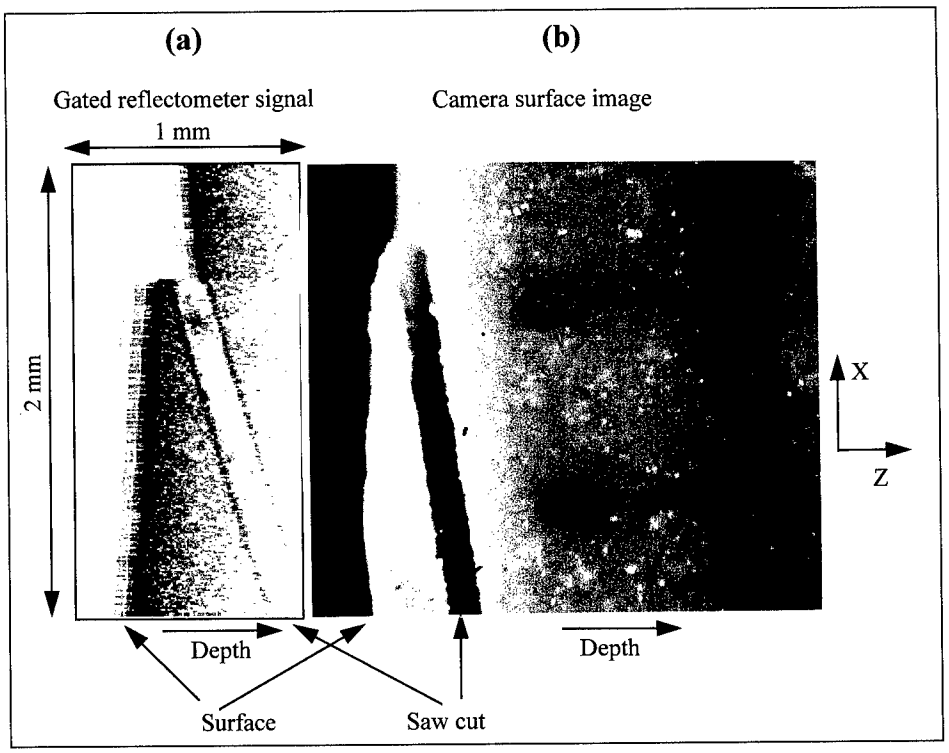


Figure 3

(a) An X-Z image through the middle of the NCX 5102 sample reconstructed from the individual depth scans in Z. (b) A camera image of the top surface of the ceramic, showing the saw cut.

The scattering results from grain-boundary reflections which are enhanced by the refractive index mismatch, Δn, of the second phase material. In the absence of voids or defects, the return from the scattered light decreases with depth into the ceramic. We see the reduction in signal level below the surface as expected. We also clearly see the increased scattering from the front and back surfaces of the saw cut. Current experimental conditions define the depth at which the void can be detected to about 500 μm below the surface.

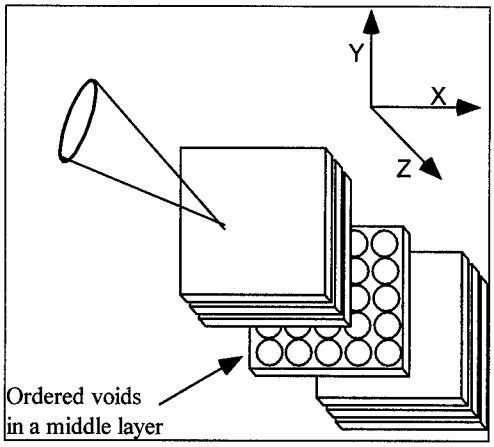


Figure 4

A schematic diagram showing internal structure of the ceramic composite material with ordered voids. Individual ceramic layers are 25 μm thick and the circularly shaped voids are present in some layers below the surface.

The second object we studied was a composite structure made from a ceramic known as PZT 5A. It is advantageous under some circumstances to use multi-layered components rather than single piece devices. This structure, as shown in Fig. 4, was made at NRL from multiple ceramic layers and was provided with an ordered array of holes in some layers.

Fig. 5 (a) shows an X-Z image through the above structure reconstructed from the individual range-gated scans in Z. We can clearly see the individual layers below the surface. The dark regions in the fourth layer can be attributed to the voids. These regions are followed by increased signal in the fifth layer which we attribute to the void to ceramic interface. To see the voids below the surface directly, we performed a scan in the X-Y plane with the light focused on the fifth layer in Z. The results are shown in Fig. 5 (b). The circular voids are well defined in the range-gated image.

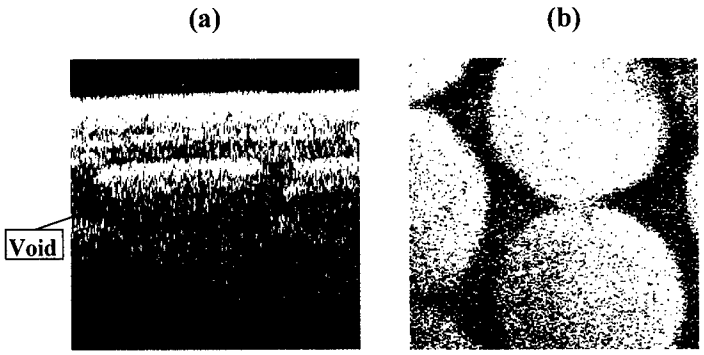


Figure 5

(a) An X-Z image through the ceramic composite structure, reconstructed from the individual depth scans in Z. The individual layers are clearly seen. (b) An X-Y image of the fifth layer below the surface of the same structure.

SUMMARY: In conclusion, we have demonstrated the use of a range-gated reflectometer using scattered light for the detection of subsurface defects in ceramic materials. In particular, we have shown that subsurface voids as deep as 500 μm can be detected with a depth resolution of approximately 10 μm. The range-gating technique used in this work is based on a low-coherence fiber interferometer. An advantage of optical-gating techniques over non-gated light scattering techniques is that they not only detect the presence of a defect but also determine its depth below the surface. Future work will include characterizing the size and type of defects which can be detected using optical-gating techniques in various ceramic materials. We will explore the relation between fabrication and finishing and crack formation in various ceramic materials. In addition, we will extend the optical wavelength further into the IR, in order to increase the depth at which defects can be detected.

This work was supported by the US Office of Naval Research.
* NRL Code 6382
† NRL Code 6380
‡ NRL Code 5600.2

REFERENCES:

1) H.H. Xu, and S. Jahanmir, "Simple Techniques for Observing Subsurface Damage in Machining of Ceramics," *J. Am. Ceram. Soc.*, **77**(5), 1388-1390 (1994).
2) M.N. Menon *et al.* "Creep and Stress Rupture Behavior of an Advanced Silicon Nitride: Parts I-III," *J. Am. Ceram. Soc.*, **77**(5), 1217-1241 (1994).
3) K.E. Amin, "Radiograph and Ultrasound Detect Defects Early," *American Ceramic Society Bulletin,* **74**(1), 76-81 (1995).
4) R. Komanduri *et al.*, "Advanced Polishing/Finishing and NDE Procedures," *ARPA Ceramic Bearing Technology Annual Review*, pp.10-14(1994), Baltimore MD.

5) J.A. Slotwinski, N.N. HSU, and G.V. Blessing, "Ultrasonic Measurements of Surface and Subsurface Structure in Ceramics," *Machining of Advanced Materials*, NIST special publication **#847**, 117-123 (1993).
6) W. A. Ellingson *et al.,* "Detection of Subsurface Defects in Machined Silicon Nitride Ceramics by Optical Scattering Methods," *Machining of Advanced Materials*, NIST special publication **#847**, 147-157 (1993).
7) M. Bashkansky *et al.,* "Nonlinear Optical Field Cross-Correlation Techniques for Imaging through a Strong Scattering Material," pp. 228-237, *Proc. SPIE Inverse Optics III*, **2241**(1994).
8) J.G. Fujimoto *et al.* "Femtosecond Optical Ranging in Biological Systems," *Opt. Lett* **11**(3), 150-152(1986).
9) E. A. Swanson *et al.* "High-speed optical coherence domain reflectometry," *Opt. Lett* **17**(2), 151-153(1992).
10) M. Bashkansky, C.L. Adler, and J. Reintjes, "Coherently Amplified Raman Polarization Gate for Imaging through Scattering Media," *Opt. Lett.* **19**(5), 350-352(1994).
11) R. C. Youngquist, S. Carr, and D. E. N. Davies, "Optical coherence-domain reflectometry: a new optical evaluation technique", Opt. Lett. **12**, 158 (1987).
12) K. Takada, I. Yokohama, K. Chida, and J. Noda, "New measurement system for fault location in optical waveguide devices based on an interferometric technique", Appl. Opt. **26**, 1063 (1987).

POOR SURFACES AND INTERSECTIONS OF SURFACES STILL CAUSE TROUBLE JUST LIKE THEY USED TO DO

Dr. David Krashes
Massachusetts Materials Research, Inc.
P.O. Box 810
West Boylston, Massachusetts 01583

James J. Scutti
Massachusetts Materials Research, Inc.
P.O. Box 810
West Boylston, Massachusetts 01583

Abstract: For nearly half a century engineering designers, materials engineers, and metallurgists have known that poorly machined surfaces and sharp "inside" corners lead to cracking, fatigue, and disastrous failures. Yet, products with these types of deficiencies continue to be produced, having been designed with what is intended to be an ample design factor to avert a potential problem. So often, these products fail. In this paper, we will discuss the effects of surface finish and corner radii on product durability, and provide photographic examples to illustrate the types of failures that occur due to deficient conditions. These failures are well known to failure analysts and engineers in manufacturing, but are worth reviewing here. We will then challenge engineers to specify and inspect for surface finish and fillet radius requirements for their products.

Key Words: Fatigue; fracture; notch; sharp radius; stress concentration; stress raiser; surface finish.

INTRODUCTION: Deficiencies of surfaces and surface intersections cause so many product failures that it is worthwhile pointing out again what the problems are and then issuing a "call" that product designers and materials engineers enumerate surface finish and fillet radius requirements and that quality control departments inspect for compliance with those requirements. Current engineering practice generally is to specify alloy grade, material composition and mechanical properties very carefully. Only if surface finish and radii of intersections are considered critical are they also specified. Then, during manufacturing, the alloy grade, composition, and mechanical properties are usually checked by quality control; but the surface finish and radii are frequently not checked. There may be a lack of appreciation by product designers of the deleterious effect that inadequate radii at intersections and poor surface finish can have on product life.

The types of situations that cause failures are well known to failure analysts and materials engineers in manufacturing. They are published in many books. Several examples are given

here, not to illustrate the method of failure analysis, but simply to be a refresher. We don't give alloy compositions, heat treatments, and service conditions because we want to lessen attention on the failure analysis and increase attention on the problem.

STRESS RAISERS: A sharp radius at the intersection of surfaces and a rough surface finish are two examples of stress raisers, or "stress concentrations." Any geometrical discontinuity in a stressed solid, such as a notch, a spline, a keyway, or a hole, results in an increase in the stress at the discontinuity above the average stress away from this feature [1].

Detailed data showing the effects of stress raisers on the local stresses at discontinuities are well documented by Peterson [2]. The data are presented in the form of stress concentration factors (K_t) for many different geometrical discontinuities stressed in different ways. The stress concentration factor is the ratio of the maximum stress to the nominal stress at the discontinuity. The stress concentration factor generally increases with *increasing* depth and *decreasing* radius at the discontinuity. A rough surface finish, therefore, can result in a high stress concentration factor since the many discontinuities (i.e. "peaks and valleys") at the surface have small radii. Similarly, widely used features in machinery components such as the geometrical discontinuities described above are also stress raisers on a larger scale.

High stress concentrations can affect the performance of products in different ways. If a component is loaded statically or sustains a single load application at a magnitude that exceeds the capability of the material, a crack initiates and propagates (often rapidly) from the stress concentration site. If the static load is constant and well below the yield strength, while the component is exposed to an aggressive environment, the sustained stress is increased at the discontinuity and a failure mode known as stress corrosion cracking [3] can initiate at the stress raiser.

If fluctuating or cyclic loads are applied to the component, the presence of stress raisers reduce the fatigue life of the component. The degree to which rough surfaces affect the lives of components is accounted for in one of the fatigue strength modifying factors known as the "surface factor" [4]. This factor, a fraction less than one, is applied to the fatigue strength obtained with laboratory specimens which have been carefully machined and polished. The factor depends on the surface finish, produced by such processes as grinding, machining, chemical machining or etching, hot rolling, forging or casting, and on the tensile strength of the material.

Sharp radii at the intersections of surfaces also affect fatigue properties. The reduction in fatigue strength due to the presence of a sharp radius is known as the fatigue-notch factor, K_f. This factor varies with the severity and type of the notch, the tensile strength of the material, the type of loading, and the stress level [5].

The impact of surface discontinuities must be considered for both static *and* cyclic load conditions. Under some circumstances, such as for rotating shafts with keyways, the nominal stress through the cross section of the component at the discontinuity may be low, but the

reduction in fatigue strength may be very high, due to a high K_f, which could result in premature failure.

CASE HISTORIES: First, a trunion shaft approximately 4" diameter that failed in fatigue starting from origins on the inner surface. Figure 1 shows the fracture with two arrows pointing to origins on the rubbed (post fracture) surface. Figure 2 shows the step-like fracture progression so typical of fatigue, and Figure 3, with an arrow pointing to a tiny crack just starting, shows the roughly machined inner surface in which the origins were found. Surfaces machined inadequately for the service that the component must endure are a major source of failures of load-bearing and rotating components. Poor surface preparation can include simple roughness, tears, gouges, and chatter.

Another example of poor surface is in Figures 4 and 5. Thread rolling laps caused nuts to be loosely tightened on bolts; and the looseness ultimately allowed vibration and fatigue failure. The looseness was due to surface finish of threads: slivers (laps) of steel were able to tear free from the rolled threads during tightening and cause galling between nuts and bolts. This led to the nuts appearing to be adequately fastened when actually they were not.

Sharp radii where surfaces intersect in keyways or where diameters change on rotating shafts are so commonly understood as a problem that they should not have to be discussed in modern times by materials engineers and designers. Yet, the problem is still major and causes serious financial losses in downtime and repairs for some manufacturers. Figure 6 shows a crack emanating from a sharp corner in a keyway of a huge printing press roll. When the other end of this roll, made the same way, failed catastrophically, the total loss was in the millions of dollars. Figure 7 shows this fracture opened to reveal the fracture origin (arrow) at this sharp corner. Here, a single overload at the sharp corner caused a rapid brittle fracture.

Poorly ground surfaces may contribute to fractures, as shown in Figures 8, 9, and 10. This shaft, about 6" diameter and 30' long from a drinking water delivery system also failed in a brittle fashion from the sharp inside corner (origin of crack) in a keyway. The inner surface of the keyway showed considerable grinding "chatter" and checking in the vicinity of the origin revealed additional cracking starting along a machining mark, as Figure 11 shows. Thus, the poor surface may have added to the stress concentration at the keyway corner.

Surface finish anomalies can occur due to processes other than machining. Chemical etching was performed on .004" thick strips, to facilitate bonding to a high frequency resonator. Etching produced many tiny transverse grooves on the strips. When the components were installed in service, the cyclic loads sustained during resonating caused fatigue cracks to initiate and propagate from these surface anomalies, as shown in Figure 12.

The important point is that these failures are recent. They occurred in modern products. Despite the ready availability of knowledge, designers and manufacturers are making stress concentrating mistakes. These failures could all have been avoided by proper specification of surface finishes and radii by the designers, and by quality control use of surface finish and radius measuring devices.

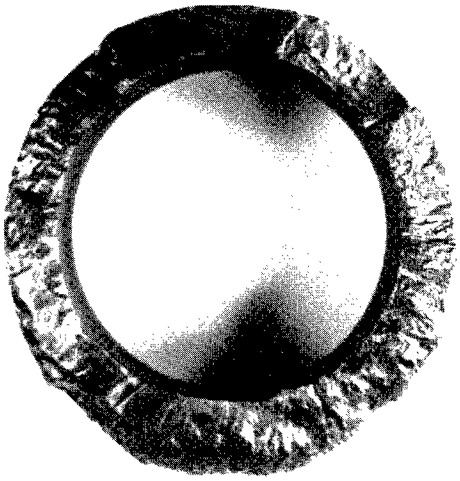


Figure 1. Fracture of 4" trunion. Mag: 0.6x

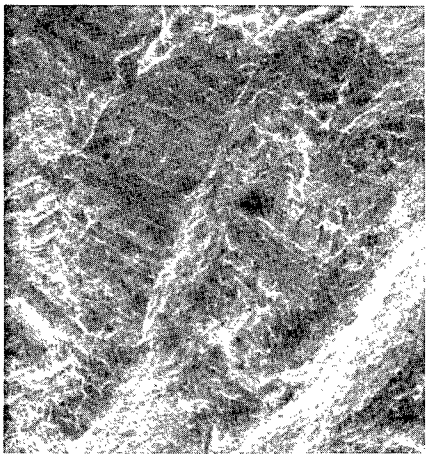


Figure 2. Fatigue striations from one origin. Mag: 42x



Figure 3. Crack (arrow) starting in machining groove. Mag: 85x

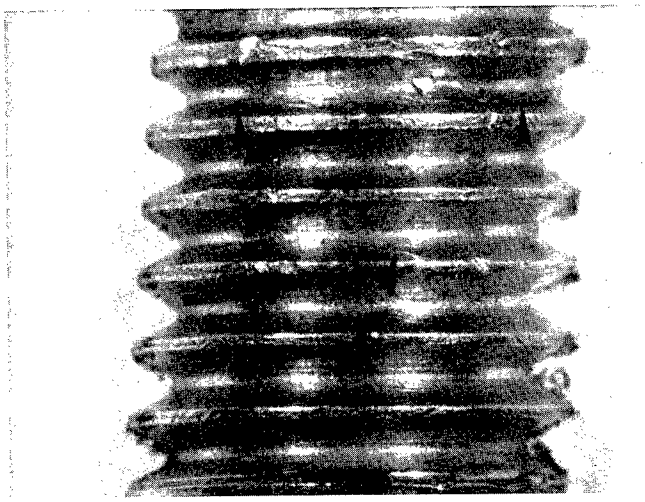


Figure 4. Rolled threads on bolt. Arrows mark defects. Mag: 8x

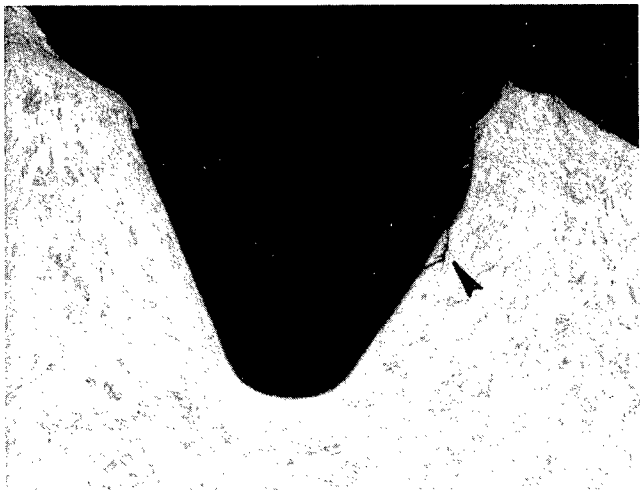


Figure 5. Cross section shows lap re-entering surface. Mag: 64x

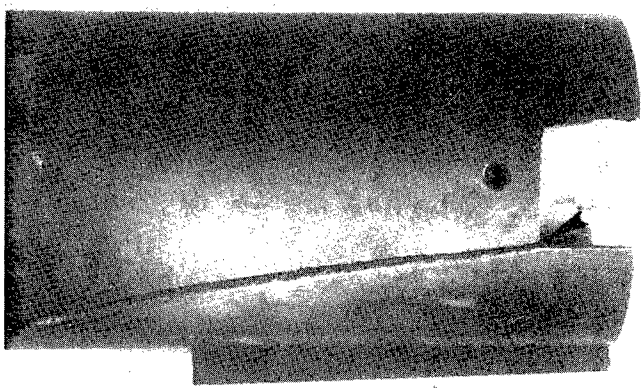


Figure 6. Sharp corner. Mag: 1.2x

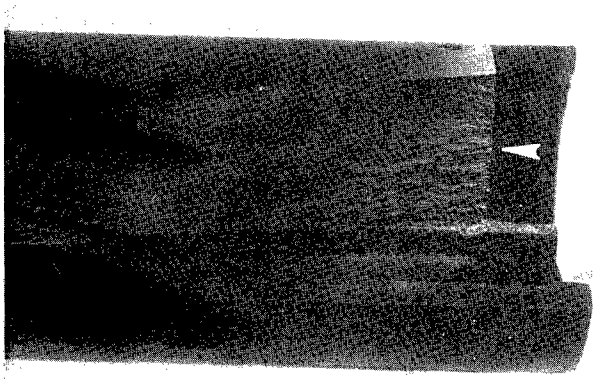


Figure 7. Arrow shows fracture origin at corner of keyway. Mag: 1.2x



Figure 8. Arrow shows fracture origin at corner of keyway. Mag: 0.4x

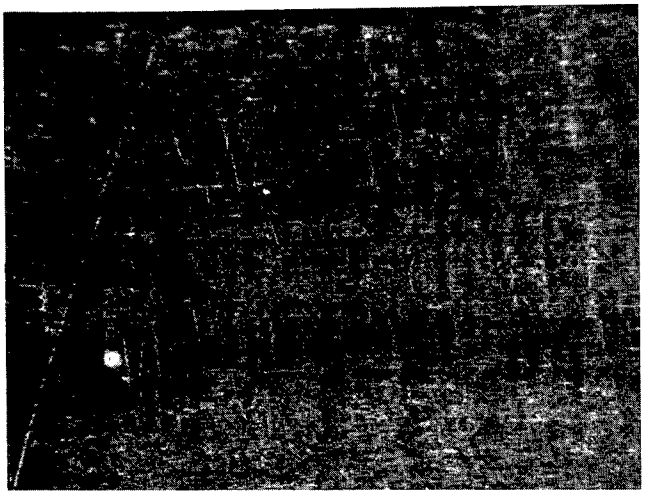


Figure 9. Grinding checks (the vertical pattern) on keyway surface near origin. Mag: 16x

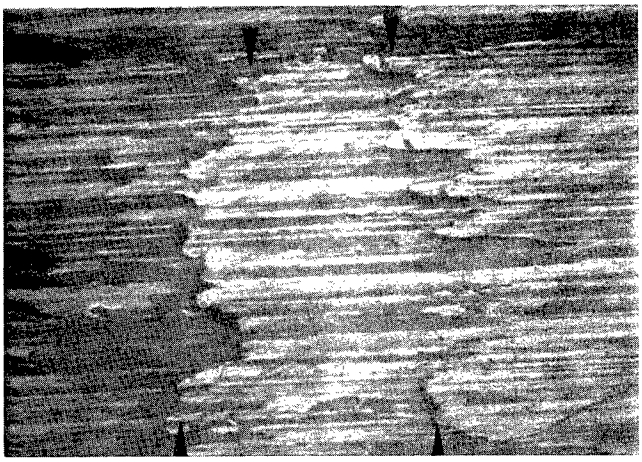


Figure 10. The grinding checks (arrows) in detail. Horizontal pattern is grinding grooves. Mag: 500x

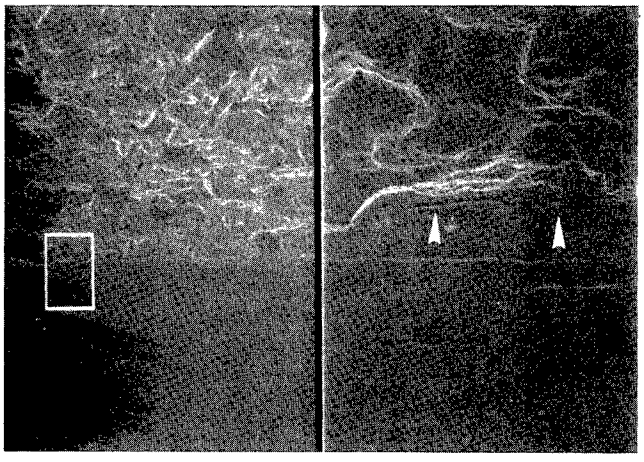


Figure 11. Box at left shows edge of fracture (upper) and keyway surface (lower). Mag: 50x At right: higher magnification view of area in box at edge of fracture, crack (arrow) starts at machining mark . Mag: 400x

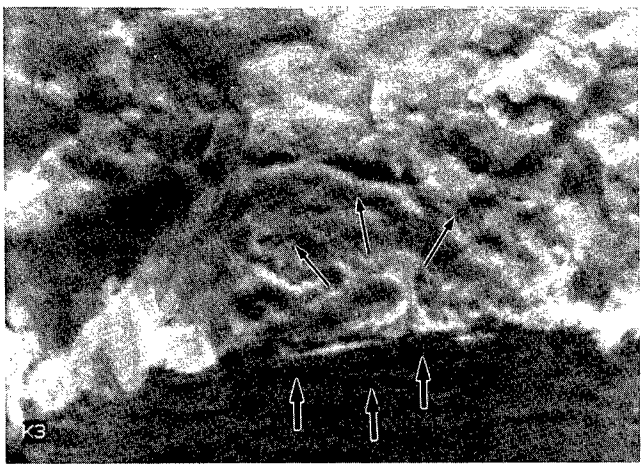


Figure 12. Fatigue crack (small arrows) initiated at groove on surface. Other grooves indicated by large arrows. Mag: 2500x

What should be done? Ultimately, manufacturing standards for most mechanical components, not just the critical ones, must include criteria for surface finish and radii at intersections of planes. Interestingly, ASTM seems to have no standards that relate to the subject and ANSI has only one.

FUTURE EFFORTS: To convince ourselves of the importance of this subject, statistics from failure analysis laboratories all over the country should be gathered, just as the United States Consumer Products Safety Commission gathers statistics on causes of accidents from hospitals. (The Commission uses this data as a basis for issuing product regulations.) To start, requests to review three years' analyses could be issued to two dozen well-qualified independent failure analysis laboratories. They would be asked to determine percentages of failures of mechanical components attributable to stress concentration, surface finish, and radii. This should not be a major undertaking.

The results would indicate if there is a need to survey many more laboratories in a comprehensive study. If the statistics show that poorly fabricated surfaces and intersections are causes of failures, there would be a strong impetus to governmental agencies and manufacturers to find ways to overcome the problem. Specifications would certainly be part of the answer.

CONCLUSIONS: Attention to often overlooked details such as surface finish and radii at the product design stage, as well as during manufacture can significantly improve the durability and performance of products. These features should be specified for critical and noncritical components alike. Development and implementation of specifications defining surface finish and radius requirements for machinery components would generalize the requirements for different configurations, applications, loading conditions, and operating environments. Adherence to these requirements would prevent many failures caused by such discontinuities. There should be a study of failure analysis histories to compile our collective experience with stress concentrations and causes of failures.

REFERENCES

1. G.E. Dieter, Mechanical Metallurgy, McGraw-Hill, Inc., New York, 1976, pp. 66-70.
2. R.E. Peterson, Stress-Concentration Design Factors, John Wiley & Sons, Inc., New York, 1974.
3. R.H. Jones, ed., Stress Corrosion Cracking, ASM International, Materials Park, Ohio, 1992, pp. 1-40.
4. J.E. Shigley, and C.R. Mischke, Mechanical Engineering Design, McGraw-Hill, Inc., New York, 1989, pp. 282-283.
5. G.E. Dieter, Mechanical Metallurgy, McGraw-Hill, Inc., New York, 1976, p. 422.

FAILURE ANALYSIS #1

GEAR CRACK PROPAGATION INVESTIGATIONS

David G. Lewicki
Vehicle Propulsion Directorate
U.S. Army Research Laboratory
NASA Lewis Research Center
Cleveland, Ohio 44135

Roberto Ballarini
Department of Civil Engineering and
Mechanical & Aerospace Engineering
Case Western Reserve University
Cleveland, Ohio 44106

Abstract: Analytical and experimental studies were performed to investigate the effect of gear rim thickness on crack propagation life. The FRANC (FRacture ANalysis Code) computer program was used to simulate crack propagation. The FRANC program used principles of linear elastic fracture mechanics, finite element modeling, and a unique re-meshing scheme to determine crack tip stress distributions, estimate stress intensity factors, and model crack propagation. Various fatigue crack growth models were used to estimate crack propagation life based on the calculated stress intensity factors. Experimental tests were performed in a gear fatigue rig to validate predicted crack propagation results. Test gears were installed with special crack propagation gages in the tooth fillet region to measure bending fatigue crack growth. Good correlation between predicted and measured crack growth was achieved when the fatigue crack closure concept was introduced into the analysis. As the gear rim thickness decreased, the compressive cyclic stress in the gear tooth fillet region increased. This retarded crack growth and increased the number of crack propagation cycles to failure.

Key Words: Crack propagation; finite element modeling; fracture mechanics; gears; rims.

Introduction: A common design goal for gears in helicopter or turboprop power transmissions is reduced weight. To help meet this goal, some gear designs use thin rims. Rims that are too thin, however, may lead to bending fatigue problems and cracks. The most common methods of gear design and analysis are based on standards published by the American Gear Manufacturers Association. Included in the standards are rating formulas for gear tooth bending to prevent crack initiation [1]. These standards can include the effect of rim thickness on tooth bending fatigue [2]. The standards, however, do not give any indication of the crack propagation path or the remaining life once a crack has started. Fracture mechanics has developed into a useful discipline for predicting strength and life of cracked structures.

Ahmad and Loo [3] applied fracture mechanics to gear teeth to illustrate the procedure and estimate crack propagation direction. Honda and Conway [4] also applied fracture mechanics to simulate tooth crack propagation, compute threshold loads, and calculate tooth life. Flasker and Jezernik [5] applied fracture mechanics to gear teeth to estimate stress intensity factors and gear life. Researchers at Tohoku University in Japan performed a series of analyses and experiments to determine the effect of residual stress on crack initiation and propagation [6,7]. Also, Daniewicz, et al. [8] developed a comprehensive, self-contained analysis package to refine the spur gear

bending fatigue theory using fracture mechanics. Lastly, Flasker and Pehan [9] described their method for calculating crack propagation in gear teeth using fracture mechanics. Much of the work of the above references considered only an initial crack and propagation paths were not considered. Many of the references that did consider crack propagation assumed the propagation occurred in a straight path. In addition, experimental validation of the cited analyses was sparse. Finally, no work using fracture mechanics was performed for thin-rim gears.

The objective of this study was to determine the effect of gear rim thickness on crack propagation life. From an extensive study [10], linear elastic fracture mechanics was used to analyze gear tooth bending fatigue in standard and thin-rim gears. Finite element computer programs were used to determine stress distributions, estimate stress intensity factors, and model crack propagation. Various fatigue crack growth models were used to estimate crack propagation life. Experimental tests were performed to validate predicted crack propagation results.

Fatigue Crack Growth: Many machine elements, such as gear teeth, are cyclicly loaded in application. The overall fatigue life of such components may be represented by three distinct phases: 1) crack initiation, 2) crack propagation, and 3) final failure. Once crack initiation has occurred, fracture mechanics may be used to estimate crack propagation fatigue growth rate and time to final failure.

The most universally used method to calculate crack propagation fatigue crack growth was postulated by Paris and Erdogan [11]. Considered were purely mode I loaded specimens subjected to cyclic load. Further considered was unstable crack growth such that the stress intensity factor grew with increasing crack size. Paris postulated that the rate of crack growth with respect to number of stress cycles was a logarithmic relationship with the stress intensity factor range as

$$\frac{da}{dN} = C(\Delta K)^n \tag{1}$$

where da is the change in crack length for dN number of stress cycles, ΔK is the range of the mode I stress intensity factor at a given time, and C and n are material constants. The material constants, C and n, must be determined by some experimental means.

Further research of fatigue crack growth has shown that there exists three important factors not considered in the Paris model. First was the effect of load ratio, R, on crack growth (R = minimum cyclic load / maximum cyclic load). Second was the instability of crack growth observed when the stress intensity factor range approached the material's fracture toughness index, K_{IC}. Third was the presence of a stress intensity threshold factor, ΔK_{th}. The stress intensity threshold factor is the highest stress intensity factor in which no crack growth would occur. The Collipriest crack growth model [12] accounts for these effects where

$$\frac{da}{dN} = C(K_{IC}\Delta K_{th})^{\frac{n}{2}} \cdot \exp\left\{\ln\left(\frac{K_{IC}}{\Delta K_{th}}\right)^{\frac{n}{2}} \cdot \tanh^{-1}\left[\frac{\ln\left(\frac{\Delta K^2}{(1-R)K_{IC}\Delta K_{th}}\right)}{\ln\left(\frac{(1-R)K_{IC}}{\Delta K_{th}}\right)}\right]\right\} \tag{2}$$

In addressing applications to gears, Inoue, et al. [7] describes fatigue crack growth of gear bending fatigue tests. Here, crack growth equations were derived as a function of crack depth through a gear tooth. The expression derived for crack growth rate da/dn, as a function of stress intensity range, ΔK, was

$$\frac{da}{dN} = \begin{cases} \dfrac{\lambda}{(1-\alpha^{\eta})}\left(\Delta K^{\eta} - \Delta K_{th}{}^{\eta}\right) & \text{for } \Delta K_{th} \le \Delta K \le \Delta K_C \\ \dfrac{\lambda}{(1-\alpha^{\eta})}\dfrac{\Delta K^{\eta} K_{IC}{}^{\eta}}{\left(K_{IC}{}^{\eta} - \Delta K^{\eta}\right)} & \text{for } \Delta K_C < \Delta K < \Delta K_{IC} \end{cases} \tag{3}$$

where the parameters K_{IC}, α, ΔK_C, ΔK_{th}, η, and λ were all estimated as a function of tooth hardness [7].

Crack Propagation Simulation: The analysis of the current study used the FRANC (FRacture ANalysis Code) computer program described by Wawrzynek [13]. FRANC is a general purpose finite element code for the static analysis of cracked structures. FRANC is designed for two-dimensional problems and is capable of analyzing plane strain, plane stress, or axi-symmetric problems.

Among the variety of capabilities, a unique feature of FRANC is the ability to model a crack in a structure. FRANC uses a method called "delete and fill" to accomplish this. To illustrate, the user would first define an initial crack by identifying the node of the crack mouth and coordinates of the crack tip. FRANC will then delete the elements in the vicinity of the crack tip. FRANC will next insert a rosette of quarter-point, six-node triangular elements around the crack tip to model the inverse square-root stress singularity [14,15]. Finally, FRANC will fill the remaining area between the rosette and original mesh with conventional six-node triangular elements. The user can then run the finite element equation solver to determine nodal displacements, forces, stresses, and strains.

A further unique feature of FRANC is the automatic crack propagation capability. After an initial crack is inserted in a mesh, FRANC models a propagated crack as a number of straight line segments. For each segment, FRANC models the crack tip using a rosette of quarter-point elements. FRANC then solves the finite element equations, calculates the stress intensity factors, and calculates the crack propagation angle. After the crack propagation angle is determined, FRANC then places the new crack tip at the calculated angle and at a user-defined crack increment length. The model is then re-meshed using the "delete and fill" method described above. The procedure is repeated a specific number of times as specified by the user. In the current study, the stress intensity factors were determined from the calculated nodal displacements using the displacement correlation method [16]. The method of Erdogan and Sih [17] was used in the current study to determine the crack propagation angle.

Once the stress intensity factors are determined for each segment, the predicted number of crack propagation cycles can be estimated using the fatigue crack growth models. Regardless of the model used, the crack growth rates, da/dN, were of the form

$$\frac{da}{dN} = g(\Delta K) \tag{4}$$

where $g(\Delta K)$ is given by Eq. (1) for the Paris relationship, Eq. (2) for the Collipriest relationship, or Eq. (3) for Inoue's method. The predicted number of crack propagation cycles for the i_{th} crack segment, N_i, was estimated by

$$N_i = \frac{a_i - a_{i\text{-}1}}{g(\Delta K_i)} + N_{i\text{-}1} \tag{5}$$

where a_i was the crack length of the i_{th} segment, $a_{i\text{-}1}$ was the crack length of the $(i\text{-}1)_{th}$ segment, $N_{i\text{-}1}$ was the number of cycles of the $(i\text{-}1)_{th}$ segment, and $g(\Delta K_i)$ was the average crack growth rate of the i_{th} and $(i\text{-}1)_{th}$ segments. Note that a_1 was the initial crack length, N_1=0, and i varied from 2 to the total number of segments.

Gear Finite Element Modeling: Basic gear tooth geometry data was input to a tooth coordinate generation computer program. The tooth coordinate generator program used the method of Hefeng, et al. [18] to determine the tooth coordinates. The output was tooth coordinate and rim coordinate data which defined a single-tooth sector of a gear. This output was used by a commercial available pre- and post-processing finite element analysis software package [19]. This package created the finite element mesh of the complete gear. FRANC then used this mesh and performed crack propagation simulations.

Figure 1 shows a sample finite element mesh of an uncracked gear. The tooth geometry used modeled that of the test gears of the NASA Lewis Spur Gear Fatigue Rig (described in the following section). The analysis used 8-node, plane stress, quadrilateral finite elements. The mesh was refined in the region of the loaded tooth for improved accuracy. The model of Fig. 1 had 2353 elements and 7295 nodes. Material properties used were that of

AISI 9310 steel. The tooth load was placed at the highest point of single tooth contact. For boundary conditions, four hub nodes were fixed. In addition, gears with various rim thicknesses were modeled. The parameter describing the rim thickness was the backup ratio, m_B, where

$$m_B = \frac{b}{h} \tag{6}$$

where b was the rim thickness, and h was the tooth whole depth. Gears with various backup ratios were modeled by incorporating slots in the model. All cases used the same finite element mesh for the loaded tooth.

Test Facility: Crack propagation experiments were performed in the NASA Lewis Spur Gear Fatigue Rig (Fig. 2). The test stand operated on a torque-regenerative principle in which torque was circulated in a loop of test gears and slave gears. Oil pressure was supplied to load vanes in one slave gear which displaced the gear with respect to its shaft. This produced a torque on the test gears, slave gears, and connecting shafts proportional to the amount of applied oil pressure. A 19 kW (25-hp), variable-speed motor provided speed to the drive shaft using a belt and pulley. The lubricant used for the gears, bearings, and loading system was a synthetic paraffinic oil. The test gear lubricant was filtered through a 5-micron fiberglass filter.

Test Gears: The test gears were 28-tooth, 8-pitch, 20° pressure angle external spur gears with a face width of 6.35 mm (0.25 in.). The teeth had involute profiles with linear tip relief starting at the highest point of single tooth contact and ending at the tooth tip at an amount of 0.013 mm (0.0005 in.). All test gears used in the experiments were fabricated and machined from a single batch of material. The test gear material was consumable-electrode vacuum-melted AISI 9310 steel. The gears were case-carburized and ground. The teeth were hardened to a case hardness of R_c 61 and a core hardness of R_c 38. The effective case depth (depth at a hardness of R_c 50) was 0.81 mm (0.032 in.). Two different test gear designs were considered. The first was a thick-rimmed gear with a backup ratio of m_B=3.3 (Fig. 3(a)). The second was a thin-rimmed gear which incorporated slots (Fig. 3(b)). The backup ratio of the thin-rimmed gear was m_B=0.3.

It was believed that tooth bending fatigue cracks would be difficult to initiate based on the load capacity of the test rig. Due to this, notches were fabricated in the fillet region (loaded side) on one tooth of each of the test gears to promote crack initiation. The notches were fabricated using electrodischarge machining (EDM) with a 0.10-mm (0.004-in.) diameter wire electrode. The nominal notch dimensions were 0.20 mm (0.008 in.) in length and 0.13 mm (0.005 in.) in width along the full face width of the tooth. The notches were located at the same location for both test gears. This location was at a radius of 40.49 mm (1.594 in.) on the fillet, which was the position of the greatest tensile stress for the solid gear (m_B=3.3). The notches produced a stress concentration factor of approximately three as determined using a finite element analysis.

Instrumentation: The standard test rig instrumentation monitored test gear speed, oil load pressure, test gear and slave gear oil pressure, and oil temperatures. Also, overall test stand vibration was monitored using an accelerometer mounted on the top housing. In addition to the standard facility vibration sensor, an advanced vibration processing diagnostic system was installed in the test stand to help assist in crack detection. Crack propagation gages were used in the experiments to determine fatigue crack growth. Special gages were fabricated for installation in the tooth fillet region of the test gears. The gages had ten circular strands with an inner radius of 1.52 mm (0.060 in.) and an outer radius of 3.05 mm (0.120 in.) (Fig. 4). The strands were designed to break as the crack propagated though them, which in turn, increased the electrical resistance of the gage (Fig. 4(a)). Figure 4(b) shows the installation of a gage in the fillet region of a notched tooth. A gage was installed on each side of the tooth flank for each gear instrumented with crack gages. The electrical resistance of the crack gages were monitored along with the load cycle count to estimate cycles as a function of crack length. The information from the rotating crack gages was transferred through brush-type slip rings. Also, an infrared tach sensor was used to measure number of load cycles.

Measured Gear Fatigue Crack Growth: The thin-rimmed gear was used in test 1. The test was run at 89 N·m (786 in.·lb) torque and 10,000 rpm speed for 6.5 hr, at which time rim fracture occurred. Figure 5 plots the

number of load cycles as a function of the measured crack length. The crack gage results indicated the crack growth was non-uniform throughout the tooth face width. A crack started on the rear flank of the tooth at the tip of the notch and reached an initial size of 0.46 mm (0.018 in.) at 1,060,000 cycles. The crack continued to propagate through the rear flank but did not reach the front flank until approximately 2,680,000 cycles. At 2,910,000 cycles, the crack reached a size of 0.64 mm (0.025 in.) on the front flank, but completed propagated through the rear gage by this time. Even though the crack initiation time was not uniform throughout the tooth face width, the crack propagation rate was uniform. This was indicated by the similarity in slopes of the curves in Fig. 5 for gages 1 and 2.

The thick-rimmed gear was used in test 2. This gear was run at 136 N·m (1200 in.·lb) torque and 10,000 rpm speed for 15 minutes, at which time tooth fracture occurred. Figure 6 gives the processed crack propagation results for test 2. Note that the crack initiation and crack propagation was fairly uniform throughout the tooth face width for this test.

Comparison of Predicted and Measured Crack Growth: The FRANC computer program was used to simulate crack propagation and calculate mode I stress intensity factors as a function of crack length. The predicted stress intensity factors were then used with three different fatigue crack growth models (Paris, Collipriest, and Inoue) to estimate crack propagation life.

A comparison of predicted crack propagation cycles using the Paris, Collipriest, and Inoue methods is shown in Fig. 7. For this, the thin-rimmed model (m_B=0.3) was used to simulate the test gear of Fig. 3(b). An initial crack of 0.64 mm (0.025 in.) was placed in the tooth fillet at the location of the maximum tensile stress. Crack propagation was then simulated and the mode I stress intensity factor as a function of crack length is given in Fig. 7(a). From this, six different fatigue growth cases were considered. The first four cases used the Paris equation and material constants of AISI 9310 specimens from experiments of Au and Ke [20]. The fifth case used the Collipriest equation and AISI 9310 material constants from Forman and Hu [21]. The load ratio used was R=-2.6 as determined from the finite element analysis. The sixth case used Inoue's method and the material constants of the SCM415 material (SCM415 is a high-strength Japanese steel, similar in properties to AISI 9310). The predicted number of cycles per crack length varied significantly among the cases studied (Fig. 7(b)). Note that the cycles were defined as the number of crack propagation cycles after an initial crack of 0.64 mm (0.025 in.).

Predicted crack growth for the m_B=0.3 and 3.3 gears were compared to the measured crack growth from the experiments. Again, the six different prediction schemes as mentioned above were used. The predicted number of crack propagation cycles using the sixth schemes were, on the most part, extremely low compared to the measured number of cycles from the experiments. To account for this, the concept of fatigue crack closure was investigated. Elber [22] performed crack experiments on aluminum alloys and deduced that residual compressive stresses existed near the crack tip region due to plastic deformation. These residual stresses reduced the effective stress intensity factor range (and thus, increased crack propagation life) and provided a better fit to experimental data than other empirical expressions. Elber proposed an effective stress intensity range ratio, U, such that

$$\Delta K_{eff} = U(\Delta K) \tag{7}$$

where ΔK_{eff} was the effective stress intensity factor range. Elber then used the effective stress intensity factor range in the Paris fatigue crack growth model. In addition, Elber defined U through experimental studies as a linear function of the load ratio, R.

The concept of fatigue crack closure was applied to the current gear crack experiments and predictions. A study was then conducted to estimate the effective stress intensity factor range ratio for the gear crack experiments. The predicted number of crack propagation cycles using the previously mentions six schemes were plotted versus crack length at a variety of arbitrarily chosen U ratios. For the Paris equation and material constants n=2.954 and

C=8.433 x 10-9 mm/cyc/(MPa$\sqrt{m}$)^n, good correlation between predicted crack cycles and the experiments occurred when: 1) U=0.4 for R=-2.6, and 2) U=0.8 for R=-0.1. Assuming a linear relation between U and R produced

$$U = 0.82 + 0.16(R) \tag{8}$$

Figure 8 shows a sample comparison of predicted and measured crack growth when the fatigue crack closure concept was used. The cycles were defined as the number of crack propagation cycles after an initial crack of 0.64 mm (0.025 in). It should be noted that good correlation was also achieved when the Collipriest equation was used with certain U values. This produced a relationship similar to Eq. 8 but with different coefficients [10].

Figure 9 displays the effect of rim thickness on predicted mode I stress intensity factors and predicted crack propagation cycles. The stress intensity factors were determined from FRANC using the appropriate finite element models. The Paris equation was used along with the effective stress intensity range ratios of Eq. 8. The initial cracks of the various models were placed at the location of the maximum tensile stress in the tooth fillet. The stress intensity factors were lowest for the m_B=0.5 case. This gave the highest predicted number of cycles for the cases studies. The cycles all were defined as the number of crack propagation cycles after an initial crack of 0.28 mm (0.011 in). The stress intensity factors were highest for the m_B=0.3 case. However, the predicted life for this was somewhere between the case of m_B=0.5 and 1.0 due to the fatigue crack closure effect. The cases of m_B=3.3 and 1.0 gave nearly the same predicted life.

Conclusions: Analytical and experimental studies were performed to investigate the effect of gear rim thickness on crack propagation life. The following conclusions were made: 1) Good correlation between predicted and measured gear crack growth was achieved when the predictions used the Paris crack growth equation and the concept of fatigue crack closure. 2) For thin rims, a decrease in rim thickness caused an increase in both the stress intensity factor and the compressive cyclic stress in the gear tooth fillet region. The increase in stress intensity factor promoted crack growth while the increase in cyclic compressive stress tended to retard crack growth and increase the number of propagation cycles to failure.

Acknowledgment: The authors wish to thank Dr. Paul A. Wawrzynek of Fracture Analysis Consultants, Inc., for fruitful discussions and for providing the FRANC program.

References:

1. "Fundamental Rating Factors and Calculation Methods for Involute Spur and Helical Gear Teeth", ANSI/ AGMA 2001-B88, American Gear Manufacturers Association, Alexandria, VA, 1990.
2. Drago, R.J., and Lutthans, R.V., "Combined Effects of Rim Thickness and Pitch Diameter on Spur Gear Tooth Stresses", Journal of the American Helicopter Society, Vol. 28, Jul., pp 13-19, 1983.
3. Ahmad, J., and Loo, F.T., "On the Use of Strain Energy Density Fracture Criterion in the Design of Gears Using Finite Element Method", ASME Paper No. 77-DET-158, presented at the Design Technical Conference, Chicago, IL, Jun., 1977.
4. Honda, H., and Conway, J.C., "An Analysis by Finite Element Techniques of the Effects of a Crack in the Gear Tooth Fillet and its Applicability to Evaluating Strength of the Flawed Gears", Bulletin of the JSME, Vol. 22, No. 174, Dec., pp. 1848-1855, 1979.
5. Flasker, J., and Jezernik, A., "The Comparative Analysis of Crack Propagation in the Gear Tooth", Proceedings of the International Conference of Application of Fracture Mechanics to Materials and Structures, Freiburg, West Germany, Jun., pp. 971-982, 1983.
6. Kato, M., Inoue, K., Deng, G., and Jeong, B.S., "Strength Evaluation of Carburized Gear Teeth Based on Fracture Mechanics", Proceedings of the KSME/JSME Joint Conference "Fracture and Strength '90", Seoul, Korea, pp. 248-253, 1990.

7. Inoue, K., Kato, M., Deng, G., and Takatsu, N., "Fracture Mechanics Based Evaluation of Strength of Carburized Gear Teeth", Proceedings of the JSME International Conference on Motion and Power Transmissions, Hiroshima, Japan, Nov., pp. 801-806, 1991.
8. Daniewicz, S.R., Collins, J.A., and Houser, D.R., "The Stress Intensity Factor and Stiffness for a Cracked Spur Gear Tooth", Journal of Mechanical Design, Vol. 116, No. 3, Sep., 1994.
9. Flasker, J., and Pehan, S., "Crack Propagation in Tooth Root With Variable Loading", Communications in Numerical Methods in Engineering, Vol. 9, No. 2, Feb., pp. 103-110, 1993.
10. Lewicki, D.G., "Crack Propagation Studies to Determine Benign or Catastrophic Failure Modes for Aerospace Thin-Rim Gears", Ph.D. Dissertation, Case Western Reserve University, May 1995.
11. Paris, P.C., and Erdogen, F., "A Critical Analysis of Crack Propagation Laws", Journal of Basic Engineering, Vol. 85, pp. 528-534, 1963.
12. Collipriest, J.E., "An Experimentalist's View of the Surface Flaw Problem", The Surface Crack: Physical Problems and Computational Solutions, American Society of Mechanical Engineers, pp. 43-61, 1972.
13. Wawrzynek, P.A., "Discrete Modeling of Crack Propagation: Theoretical Aspects and Implementation Issues in Two and Three Dimensions", Ph.D. Dissertation, Cornell University, 1991.
14. Henshell, R.D., and Shaw, K.G., "Crack Tip Finite Elements Are Unnecessary", International Journal for Numerical Methods in Engineering, Vol. 9, pp. 495-507, 1975.
15. Barsoum, R.S., "On the Use of Isoparametric Finite Elements in Linear Fracture Mechanics", International Journal for Numerical Methods in Engineering, Vol. 10, No. 1, pp. 25-37, 1976.
16. Tracey, D.M., "Discussion of 'On the Use of Isoparametric Finite Elements in Linear Fracture Mechanics' by R.S. Barsoum", International Journal for Numerical Methods in Engineering, Vol. 11, pp. 401-402, 1977.
17. Erdogan, F., and Sih, G.C., 1963, "On the Crack Extension in Plates Under Plane Loading and Transverse Shear", Journal of Basic Engineering, Vol. 85, pp. 519-527, 1963.
18. Hefeng, B., Savage, M., and Knorr, R.J., "Computer Modeling of Rack-Generated Spur Gears", Mechanism and Machine Theory, Vol. 20, No. 4, pp. 351-360, 1985.
19. P3/PATRAN User Manual, PDA Engineering, Costa Mesa, CA, 1993.
20. Au, J.J., and Ke, J.S., "Correlation Between Fatigue Crack Growth Rate and Fatigue Striation Spacing in AISI 9310 Steel", Fractography and Materials Science, ASTM STP 733, pp. 202-221, 1981.
21. Forman, R.G., and Hu, T., 1984, "Application of Fracture Mechanics on the Space Shuttle", Damage Tolerance of Metallic Structures, ASTM STP 842, pp. 108-133, 1984.
22. Elber, W., "The Significance of Fatigue Crack Closure", Damage Tolerance in Aircraft Structures, ASTM STP 486, pp. 230-242, 1971.

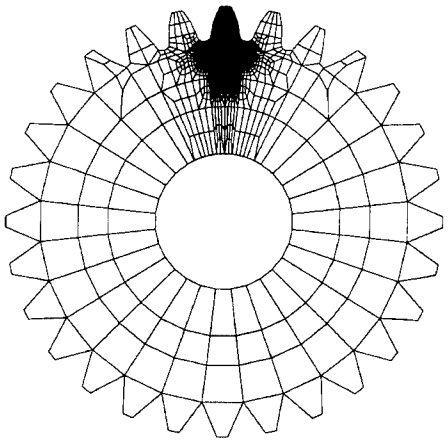


Figure 1.—Finite element model of gears used in crack propagation studies, solid model, m_B = 3.3.

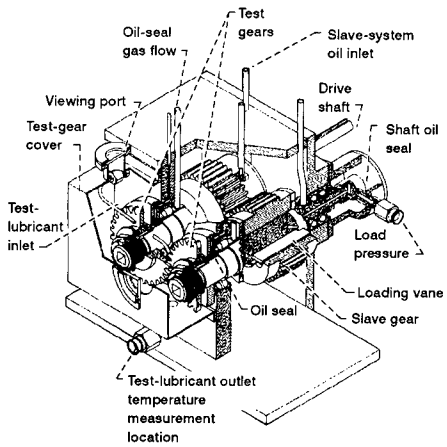
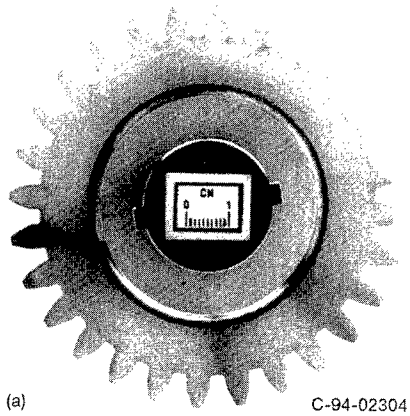


Figure 2.—NASA Lewis Spur Gear Fatigue Rig.



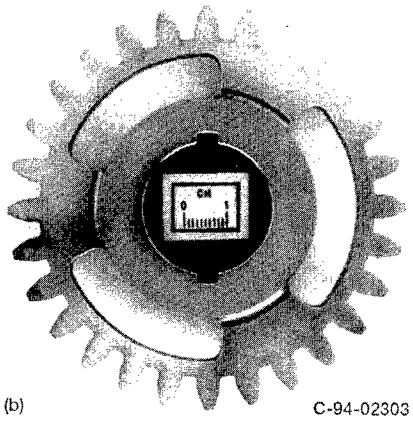
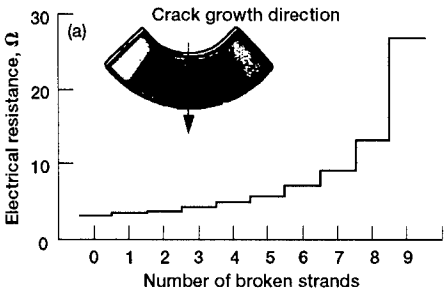


Figure 3.—Test gears used to determine effect of rim thickness on crack propagation. (a) m_B = 3.3. (b) m_B = 0.3.



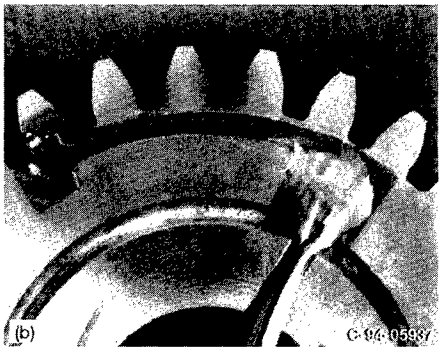


Figure 4.—Specialized crack propagation gages for gear tooth crack growth measurements. (a) Increase in gage electrical resistance as the number of broken strands increase. (b) Installation of crack propagation gage on test gear.

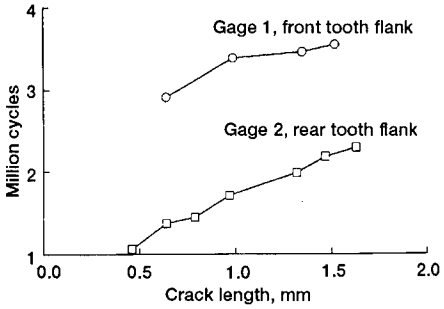


Figure 5.—Crack propagation fatigue growth for test 1, $m_B = 0.3$.

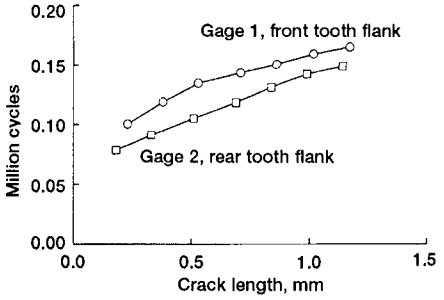


Figure 6.—Crack propagation fatigue growth for test 2, $m_B = 3.3$.

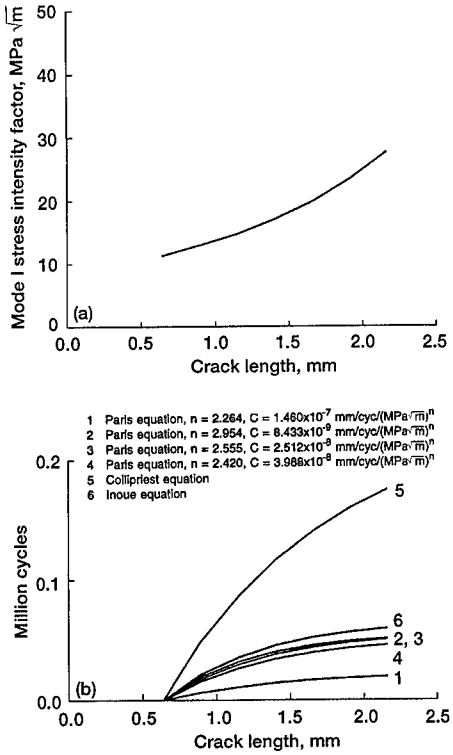


Figure 7.—Comparison of predicted crack propagation cycles using Paris equation, Collipriest equation, and Inoue equation. Model for m_B = 0.3. (a) Mode I stress intensity factor. (b) Life comparison.

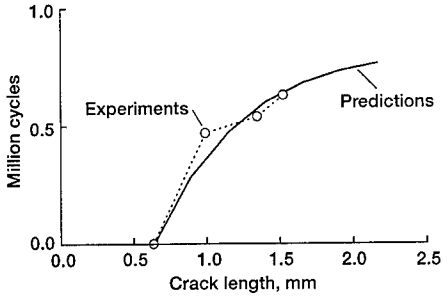


Figure 8.—Comparison of predicted crack propagation cycles to experiments. Paris fatigue crack growth model, n = 2.954, C = 8.433 x 10^{-9} mm/cyc/(MPa$\sqrt{m}$)^n, R = –2.6, U = 0.4, used for predictions. Test 1, gage 1, front flank, for experiments.

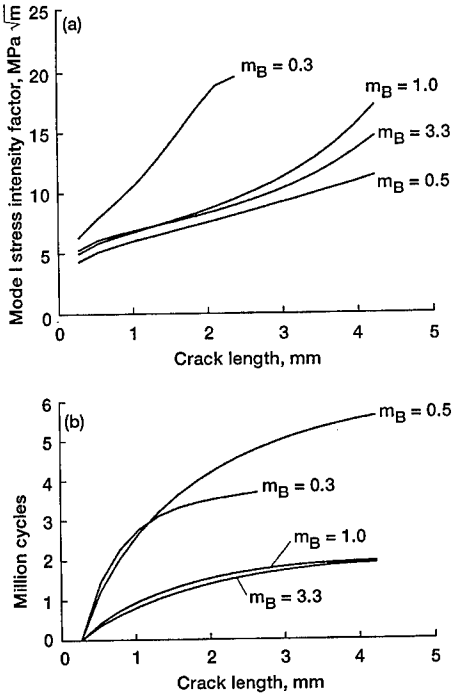


Figure 9.—Effect of backup ratio on stress intensity factors and crack propagation cycles. (a) Mode I stress intensity factors. (b) Crack propagation cycles, Paris fatigue crack growth model, n = 2.954, C = 8.433 x 10^{-9} mm/cyc/(MPa$\sqrt{m}$)^n, U = 0.82 + 0.16 R.

DETECTING GEAR TOOTH FATIGUE CRACKS IN ADVANCE OF COMPLETE FRACTURE

James J. Zakrajsek
National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135

and

David G. Lewicki
U.S. Army Research Laboratory
Lewis Research Center
Cleveland, Ohio 44135

Abstract: Results of using vibration-based methods to detect gear tooth fatigue cracks are presented. An experimental test rig was used to fail a number of spur gear specimens through bending fatigue. The gear tooth fatigue crack in each test was initiated through a small notch in the fillet area of a tooth on the gear. The primary purpose of these tests was to verify analytical predictions of fatigue crack propagation direction and rate as a function of gear rim thickness. The vibration signal from a total of three tests was monitored and recorded for gear fault detection research. The damage consisted of complete rim fracture on the two thin rim gears and single tooth fracture on the standard full rim test gear. Vibration-based fault detection methods were applied to the vibration signal both on-line and after the tests were completed. The objectives of this effort were to identify methods capable of detecting the fatigue crack, and determine how far in advance of total failure positive detection was given. Results showed that the fault detection methods failed to respond to the fatigue crack prior to complete rim fracture in the thin rim gear tests. In the standard full rim gear test all of the methods responded to the fatigue crack in advance of tooth fracture; however, only three of the methods responded to the fatigue crack in the early stages of crack propagation.

Key Words: Gear; Fatigue; Diagnostics; Failure prediction

Introduction: Drive train diagnostics is one of the most significant areas of research in rotorcraft propulsion. The need for a reliable health and usage monitoring system for the propulsion system can be seen by reviewing rotorcraft accident statistics. An investigation of serious rotorcraft accidents that were a result of fatigue failures showed that 32 percent were due to engine and transmission components [1]. In addition, governmental aviation authorities are demanding that in the near future the safety record of civil helicopters must match that of conventional fixed-wing jet aircraft.

This would require a thirtyfold increase in helicopter safety. Practically, this can only be accomplished with the aid of a highly reliable, on-line Health and Usage Monitoring (HUM) system. A key performance element of a HUM system is to determine if a fault exists, as early and reliably as possible. Research is thus needed to develop and prove various fault detection concepts and methodologies.

For rotorcraft transmissions, a critical element of a reliable HUM system is the accurate detection of gear tooth damage. A gear tooth fatigue crack, if undetected, can propagate into complete tooth fracture or gear rim fracture, both of which can result in a potentially catastrophic failure of the transmission system. This paper addresses this mode of gear failure by applying gear fault detection techniques to vibration data from several spur gear crack propagation fatigue tests, in which crack length was monitored using special crack gages.

A number of methods have been proposed to provide early detection of gear tooth damage. Method FM4 was developed by Stewart [2] to detect isolated damage on gear teeth. Methods NA4 and NB4 were developed at NASA Lewis [3,4] to detect general damage on gear teeth. McFadden [5] proposed a method to detect gear tooth cracks using the instantaneous phase of the demodulated time signal. A variation of the energy ratio reported by Swansson [6] was constructed to detect damage using the energy in the frequency spectrum. Ma [7] proposed using the energy operator to track the non-linear energy associated with a damaged gear. Verification of these detection methods with experimental vibration data along with a comparison of their relative performance is an integral step in the overall development of an accurate means to detect gear tooth damage.

In the view of the above, it becomes the objective of this research to determine the relative performance of the detection methods as they are applied to experimental data from a spur gear fatigue rig at NASA Lewis. The vibration signal from three spur gear crack propagation tests were monitored and recorded. The primary purpose of these tests was to verify analytical predictions of fatigue crack propagation direction and rate as a function of gear rim thickness [8]. Two of the tests used thin rim gears, which resulted in complete rim fracture. One test used a standard full rim test gear, which resulted in complete tooth fracture. Results of each gear fault detection method are compared, and overall conclusions are made regarding the performance of the methods in detecting the fatigue cracks in advance of total fracture.

Theory of Fault Detection Methods: All of the methods in this investigation utilized vibration data that was processed as it was collected. The vibration data was converted to digital form and time synchronously averaged to eliminate noise and vibration incoherent with the period of revolution of the spur test gear. The averaged data was then resampled by linear interpolation to provide exactly 1024 samples over two complete revolutions of the test gear. This was done to optimize the frequency resolution when converting the data to the frequency domain. This averaged and resampled data was used as the input to the diagnostic methods discussed below.

FM4 was developed to detect changes in the vibration pattern resulting from damage on a limited number of teeth [2]. A difference signal is first constructed by removing the regular meshing components (shaft frequency and harmonics, primary meshing frequency and harmonics along with their first order sidebands) from the time-averaged signal. The fourth normalized statistical moment (normalized kurtosis) is then applied to this difference signal. For a gear in good condition the

difference signal would be primarily Gaussian noise, resulting in a normalized kurtosis value of 3 (nondimensional). When one or two teeth develop a defect (such as a crack, or pitting), a peak or series of peaks appear in the difference signal, causing the normalized kurtosis value to increase beyond the nominal value of 3.

NA4 is a method developed at NASA Lewis Research Center to detect the onset of damage, and also to continue to react to the damage as it increases [3]. A residual signal is first constructed by removing the regular meshing components (shaft frequency and harmonics, primary meshing frequency and harmonics) from the time-averaged signal. The fourth statistical moment of the residual signal is then divided by the current run time averaged variance of the residual signal, raised to the second power. This operation normalizes the kurtosis in NA4; however, it is normalized using the variance of the residual signal averaged over the run up to the current time record, where NA4 is being calculated. With this method, the changes in the residual signal are constantly being compared to a weighted baseline for the specific system in "good" condition. This allows NA4 to grow with the severity of the fault until the average of the variance itself changes. NA4 is dimensionless, with a value of 3, under nominal conditions.

NB4 is another parameter recently developed at NASA Lewis [4]. NB4 is similar to NA4 in that it uses the same operation to normalize the kurtosis. The major difference is that instead of using a residual signal, NB4 uses the envelope of a bandpassed segment of the signal. NB4 is a demodulation technique, in which the signal is first band-pass filtered about the dominant meshing frequency. A bandwidth of ± 17 sidebands was used for the spur gear tests. Using the Hilbert transform, a complex time signal is then created in which the real part is the band-pass signal, and the imaginary part is the Hilbert transform of the signal. The envelope is the magnitude of this complex time signal and represents an estimate of the amplitude modulation present in the signal due to the sidebands. Amplitude modulation in a signal is most often due to periodically reoccurring transient variations in the loading. The theory behind this method is that a few damaged teeth will cause transient load fluctuations which can be observed in the envelope of the signal. NB4 is also dimensionless, with a value of 3 under nominal conditions.

Another demodulation technique was developed to detect local gear defects, such as gear tooth fatigue cracks [5]. The basic theory behind this technique is that a gear tooth defect will produce sidebands that modulate the dominant meshing frequency. In this method the signal is band-passed filtered about a dominant meshing frequency, including as many sidebands as possible. The Hilbert transform is then used to convert the real bandpassed signal into a complex time signal or analytic signal. Using the real and imaginary parts of the analytic signal, the instantaneous phase (I.P) can be estimated from the filtered sidebands. Teeth with a fatigue crack will cause a lead or lag in tooth contact during meshing, resulting in transient changes in the gear rotation. These transient changes in rotation will dominate the I.P. function. The peak-to-peak level of the instantaneous phase over one complete revolution of the gear is then determined in order to quantify the relative variance of the I.P. at each point in the run.

One method investigated is a variation of the energy ratio technique [6]. The energy ratio was originally constructed to be a robust indicator of heavy uniform wear. The energy ratio is found by dividing the standard deviation of the difference signal by the standard deviation of the regular meshing components, as defined previously. In this study, the energy ratio is found by dividing the

sum of the non-mesh related amplitudes in the frequency spectrum by the sum of the mesh related amplitudes (meshing frequencies plus first 6 sidebands). This parameter was used to quantify the trend seen in the frequency spectrum of increasing amplitudes at non-mesh related frequencies as the damage increased.

A technique using the "energy operator" was also investigated in this study. First developed for speech production modeling, the energy operator was recently proposed for gear defect detection recently by Ma [7]. The energy operator is a non-linear function that reacts to impulsive energy in a vibration signal. An impulse in the time-averaged vibration signal caused by a damaged gear tooth is reinforced by the energy operator, thus allowing the impulse to be more easily detected. In this study, the time averaged signal is first band-passed filtered about the primary meshing frequency using ±20 sidebands. The discrete form of the energy operator, as given in Ref. 7 is then applied to the band-passed signal. In order to detect any reinforced impulsive events in the energy operator, the normalized kurtosis is then applied to the energy operator results. The normalized kurtosis, as described earlier, reduces the results to one nondimensional number with the value of 3 under nominal conditions.

Apparatus and Gear Damage Review: The spur gears shown in Figs. 1 and 2 were part of a series of crack propagation tests conducted on a spur gear fatigue test rig at NASA Lewis Research Center. To initiate cracks, a notch was placed in the fillet region (loaded side) on one tooth of each of the test gears. The notches were located at a radius of 40.49 mm (1.594 in.) on the fillet, which is the position of the maximum tensile stress. The spur gear test rig was then used to fail the spur gear specimen through bending fatigue. The primary purpose of these tests was to verify analytical predictions of fatigue crack propagation direction and rate as a function of gear rim thickness [8]. To verify the crack growth rate, crack detection gages were installed on each side of the tooth with the notch. As the crack propagated it broke through each of the nine crack wires that circumferentially surrounded the fillet area on each side. The nine crack wires cover the first 1.62 mm (0.064 in.) of crack length, thus allowing the crack magnitude and rate to be recorded as a function of cycles, or run time. During the tests, vibration data from an accelerometer mounted on the gear shaft bearing housing was captured using an on-line program running on a personal computer with an analog to digital conversion board and anti-aliasing filter. The 28-tooth instrumented test spur gear meshes with a 28-tooth standard rim spur gear at a nominal speed of 10 000 rpm in the fatigue test rig. The spur gears have a face width of 6.35 mm (0.25 in.) and a pitch diameter of 88.9 mm (3.50 in.). Each test started with a one hour break-in period at a low load level, after which the load was increased and the test continued until complete fracture occurred. A test load of 89 N-m (786 in-lbs) was used for runs 1 and 2. For run 3, a test load of 136 N-m (1200 in-lbs) was used.

Runs 1 and 2 used the thin rim gear geometry illustrated in Fig. 1. Crack length, as recorded by the crack gages, is plotted as a function of run time for run 1 in Fig. 3. The first hour was the break-in period, thus the crack started on the rear side only at approximately 1.75 hours after full load was applied. As seen in this figure, the crack gages cover approximately the first 50% of the total crack length to failure. For runs 1 and 2, failure is defined as the complete fracture of the gear rim, as seen in Fig. 1. The lines connecting the last crack gage points to failure is for reference purposes only. Due to instrumentation problems the crack gages did not function correctly for run 2, and no crack length plot is available.

Run 3 used the full rim gear geometry illustrated in figure 2. Crack length is plotted as a function of run time for run 3 in figure 4. The break-in period is not included in this plot, thus the crack started shortly after the full load was applied. In this run, the crack gages cover only the first 18% of the total crack length to failure. For run 3 failure is defined as complete fracture of the tooth, as seen in figure 2. Although the crack gages only cover 18% of the total crack length, they cover 75% of the time from the start of the crack to failure. Two regions can be seen in figure 4, with the first representing a moderate crack growth rate (first 75% of propagation time). The last 25% of the crack propagation time represents a region of accelerated crack growth rate, where the final 82% of total crack length is achieved.

Discussion of Results: Results of applying the diagnostic methods to crack propagation runs 1, 2, and 3 are illustrated in Figs. 5, 6, and 7, respectively. Figure 8 gives the frequency spectrum of the time-averaged signal for each of the runs at the start of each test, following the one hour break-in period. Figure 9 illustrates the time-averaged vibration plots for the last data point (end of test point) from each of the runs. The vibration plots in this figure are over one complete revolution of the test gear.

Based on the results, the gear fault detection methods were unable to give early indication of the fatigue crack in the thin rim gear tests. As seen in Fig. 5, the diagnostic parameters were applied to the vibration data from run 1 after the initial one hour break-in period. Each of the parameters are plotted as a function of run time and, at the top of the plots, as a function of crack length also. As seen in Fig. 5, with the exception of the kurtosis of the energy operator, only parameters NA4 (Fig. 5(b)) and NB4 (Fig. 5(c)) reacted to the damage. However, NA4 and NB4 were unable to detect the crack until complete rim fracture resulted, thus giving no advanced warning of the impending failure. The kurtosis of the energy operator appears to have reacted to the crack early in the run, as seen in Fig. 5(f). This very large reaction at the start of crack propagation is not repeated in runs 2 and 3. It is unclear at this time as to why this ideal type of reaction occurred only on run 1. Figure 6 plots the results of the parameters for run 2, also a thin rim gear test. Although no crack propagation plot is available for this run, the results are similar to those seen in run 1. The fault detection methods give positive indication of the damage only at or near completion of the test.

All of the gear fault detection methods reacted to the fatigue crack prior to complete tooth fracture in the full rim spur gear test. The plots in Fig. 7 do not include the initial one hour break-in period for run 3. Each of the parameters are plotted as a function of run time, and as a function of crack length (top of plot). As seen in Fig. 7, all of the parameters react substantially to the crack in the region of accelerated crack growth. As stated earlier this is the region where the last 82% of total crack length is achieved in only 25% of the total crack growth time. Of the normalized, nondimensional parameters, NB4 peaks at a value of 107, the kurtosis of the energy operator peaks at 32, NA4 peaks at 17, and FM4 peaks at a value of 5. All of these parameters are referenced to the value of 3 under nominal conditions.

Although all of the methods were able to detect the crack during the last stage of crack propagation in run 3, more advanced warning of the crack is needed for this potentially catastrophic event. As seen in Fig. 7, NA4 (Fig. 7(b)), NB4 (Fig. 7(c)), and the kurtosis of the energy operator (Fig. 7(f)) are the only parameters that show an increasing trend above the nominal value during the first phase of crack propagation. A method of establishing an exceedance level was used to determine

relative warning times for the three parameters. This method uses a value of five standard deviations over the mean value for the exceedance level, where the mean level and standard deviation are based on the data prior to crack growth [9]. As seen in Fig. 7(f), based on this exceedance level, the kurtosis of the energy operator gives the earliest warning of damage. It reaches the exceedance level at 0.083 hours, or at 46% of the total crack propagation time, at which the crack has only advanced to approximately 7% of its total length. As seen in Figs. 7(b) and (c), NA4 and NB4 reach the exceedance level at 0.111 hours, or at 62% of total crack propagation time, at which the crack has only advanced to approximately 11% of its total length. The warning times based on these three parameters are nearly twice as long as that given by the other fault detection methods. In addition, NA4, NB4, and kurtosis of the energy operator were able to detect the crack during the more moderate crack growth rate region.

The additional compliance associated with the thin rim gears along with the geometry of these gears may have inhibited the parameters from detecting the fatigue crack in runs 1 and 2. Figure 8 plots the frequency spectrum of the first data point after the break-in period for each run. The spectrums from the thin rim gear tests, Figs. 8(a) and (b), show a more complex frequency distribution as compared to the spectrum from the full rim gear test (Fig. 8(c)). As seen in Fig. 8(c), the frequency spectrum is primarily composed of the primary gear mesh frequency (f1 = 4.7 kHz), harmonics of the mesh frequencies, and the corresponding sidebands about these frequencies. The non-uniform geometry of the thin rim gear coupled with the added compliance associated with the thin rim will tend to produce a more complex meshing pattern than a standard rim gear. This is illustrated in the plots shown in Fig. 9. All of the plots in this figure are of the last data point for each run, thus complete rim fracture has occurred for runs 1 and 2 (Figs. 9(a) and (b)), and complete tooth fracture has occurred for run 33 (Fig. 9(c)). The tooth fracture induced impulse seen in Fig. 9(c) is clearly visible and highly localized. The rim fracture induced impulse seen in Figs. 9(a) and (b) are nonlocalized and much more difficult to distinguish in the signal. The additional flexibility of the thin rim gears may not allow the cracked tooth to create the impulse in the vibration signal, and thus compensates for the fault until complete fracture. In addition, the nonuniform geometry of the thin rim gear complicates the vibration signal, possibly masking the dynamics of the tooth crack.

Conclusions: Based on the results of applying a variety of gear fault detection techniques to experimental data from spur gear crack propagation tests, the following conclusions can be made.

(1) In the thin-rim spur gear tests, the gear fault detection techniques investigated were unable to give early indication of the fatigue crack prior to rim fracture. The additional compliance associated with the thin-rim gears may have compensated for the crack, thus minimizing the dynamic influence of the crack on the vibration signal until total rim fracture.

(2) In the full rim spur gear test, all of the techniques investigated reacted to the fatigue crack, in varying degrees, prior to complete tooth fracture. The more robust reactions to the fatigue crack occurred in the last stages of crack propagation, as the crack accelerated to complete fracture.

(3) The methods NA4, NB4, and the energy operator gave the earliest reactions to the fatigue crack in the standard spur gear test. These techniques reacted to the fatigue crack during the early stages, as the crack was growing at a moderate rate.

References:

1. Astridge, D.G.: Helicopter Transmissions - Design for Safety and Reliability. Inst. Mech. Eng. Proc., Pt. G-J Aerosp. Eng. vol. 203, no. G2, 1989, pp. 123–138.
2. Stewart, R.M.: Some Useful Data Analysis Techniques for Gearbox Diagnostics. Report MHM/R/10/77, Machine Health Monitoring Group, Institute of Sound and Vibration Research, University of Southampton, July 1977.
3. Zakrajsek, J.J.; Townsend, D.P.; Decker, H.J.; An Analysis of Gear Fault Detection Methods as Applied to Pitting Fatigue Failure Data. Proceedings of the 47th Meeting of the Mechanical Failures Prevention Group. Office of Naval Research, Arlington, VA., 1993, pp. 199–208.
4. Zakrajsek, J.J.; Handschuh, R.F.; and Decker, H.J.: Application of Fault Detection Techniques to Spiral Bevel Gear Fatigue Data. Proceedings of the 48th Meeting of the Mechanical Failures Prevention Group. Office of Naval Research, Arlington, VA., 1994, pp. 93–104.
5. McFadden, P.D.: Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration. J. Vib. Acoust. Stress Reliab. Design, vol. 108, no. 2, April, 1986, pp.165–170.
6. Swansson, N.S.: Application of Vibration Signal Analysis Techniques to Signal Monitoring. Conference on Friction and Wear in Engineering 1980, Institute of Engineers, Australia. Barton, Australia, 1980, pp. 262–267.
7. Ma, J.: Energy Operator and Other Demodulation Approaches to Gear Defect Detection. Proceedings of the 49th Meeting of the Society for Machinery Failure Prevention Technology. Vibration Institute, Willobrook, Illinois, 1995, pp. 127–140.
8. Lewicki, D.G.: Crack propagation Studies to Determine Benign or Catastrophic Failure Modes for Aerospace Thin-Rim Gears. Ph.D. Dissertation, Case Western Reserve University, May 1995.
9. Rose, H.J.: Vibration Signature and Fatigue Crack Growth Analysis of a Gear Tooth Bending Fatigue Failure. Proceedings of the 44th Meeting of the Mechanical Failure Prevention Group. Office of Naval Research, Arlington, Virginia, 1990.

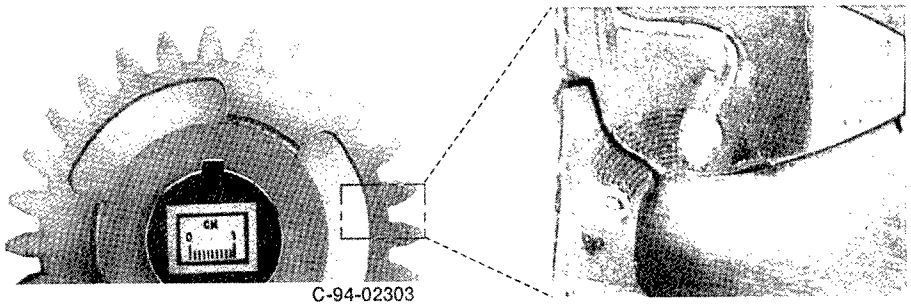


Figure 1.—Gear geometry configuration for runs 1 and 2.

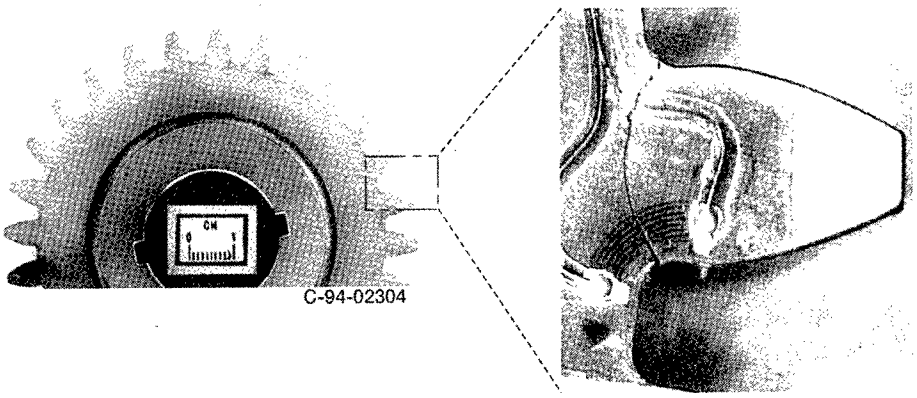


Figure 2.—Gear geometry configuration for run 3.

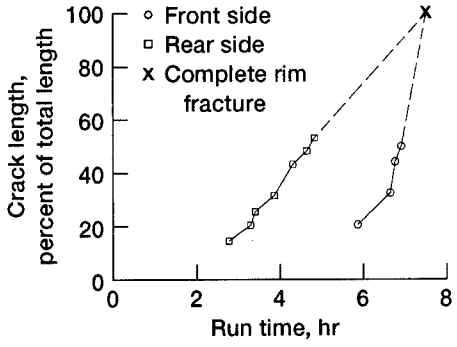


Figure 3.—Crack length as a function of run time for run 1.

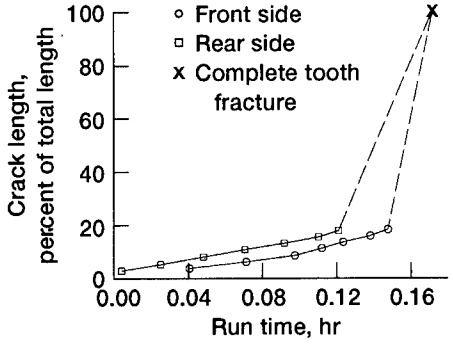


Figure 4.—Crack length as a function of run time for run 3.

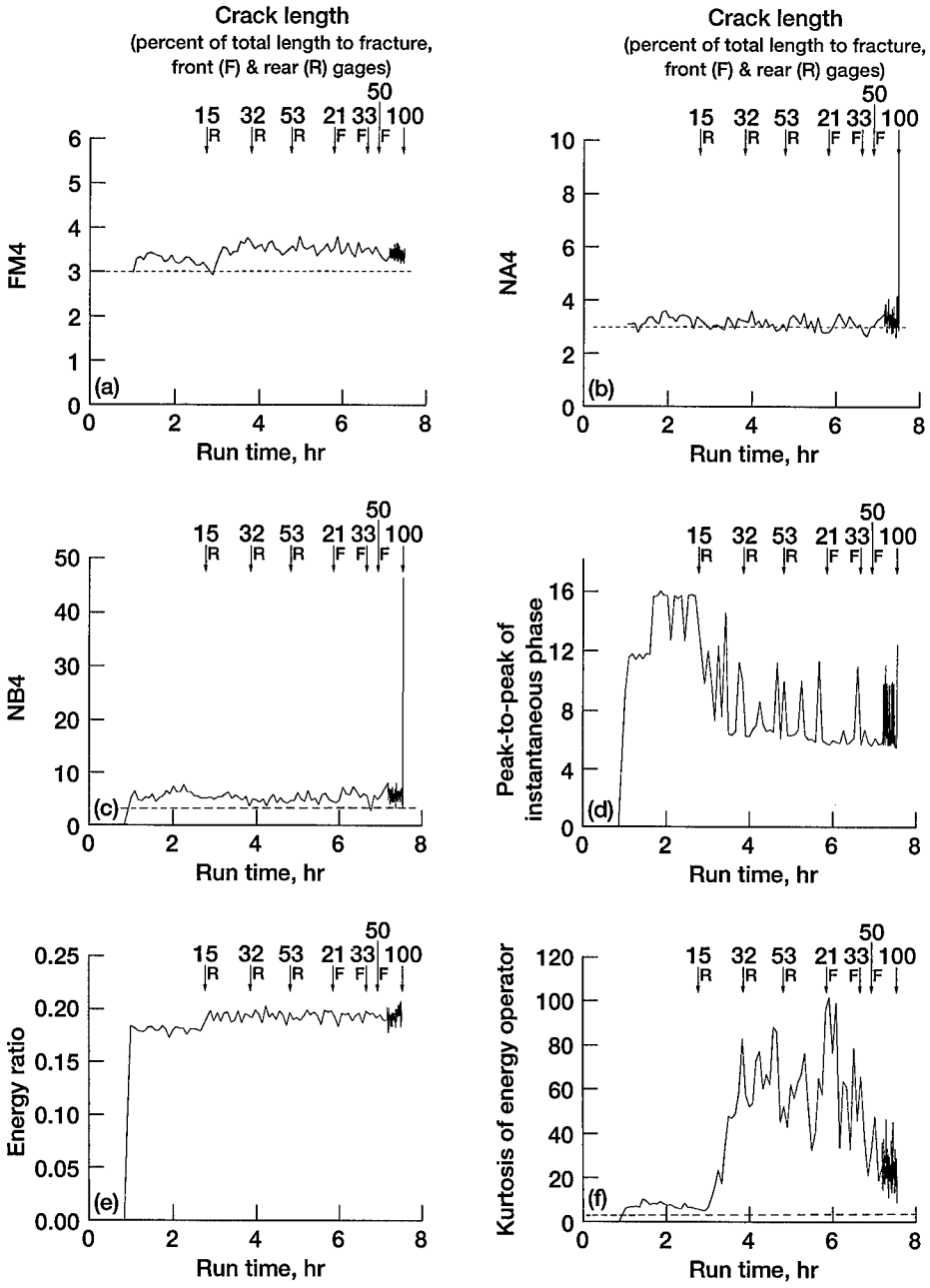


Figure 5.—Response of gear fault detection parameters for run 1.

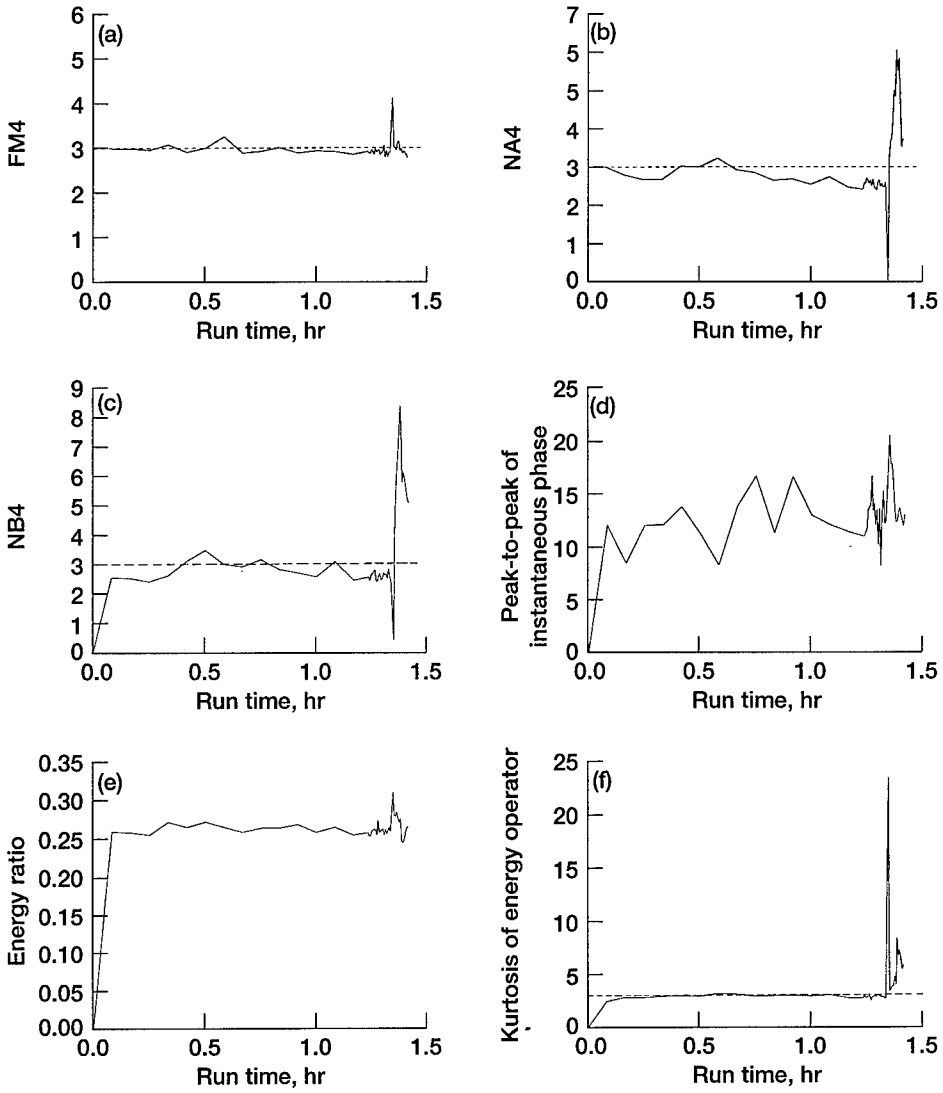


Figure 6.—Response of gear fault detection parameters for run 2.

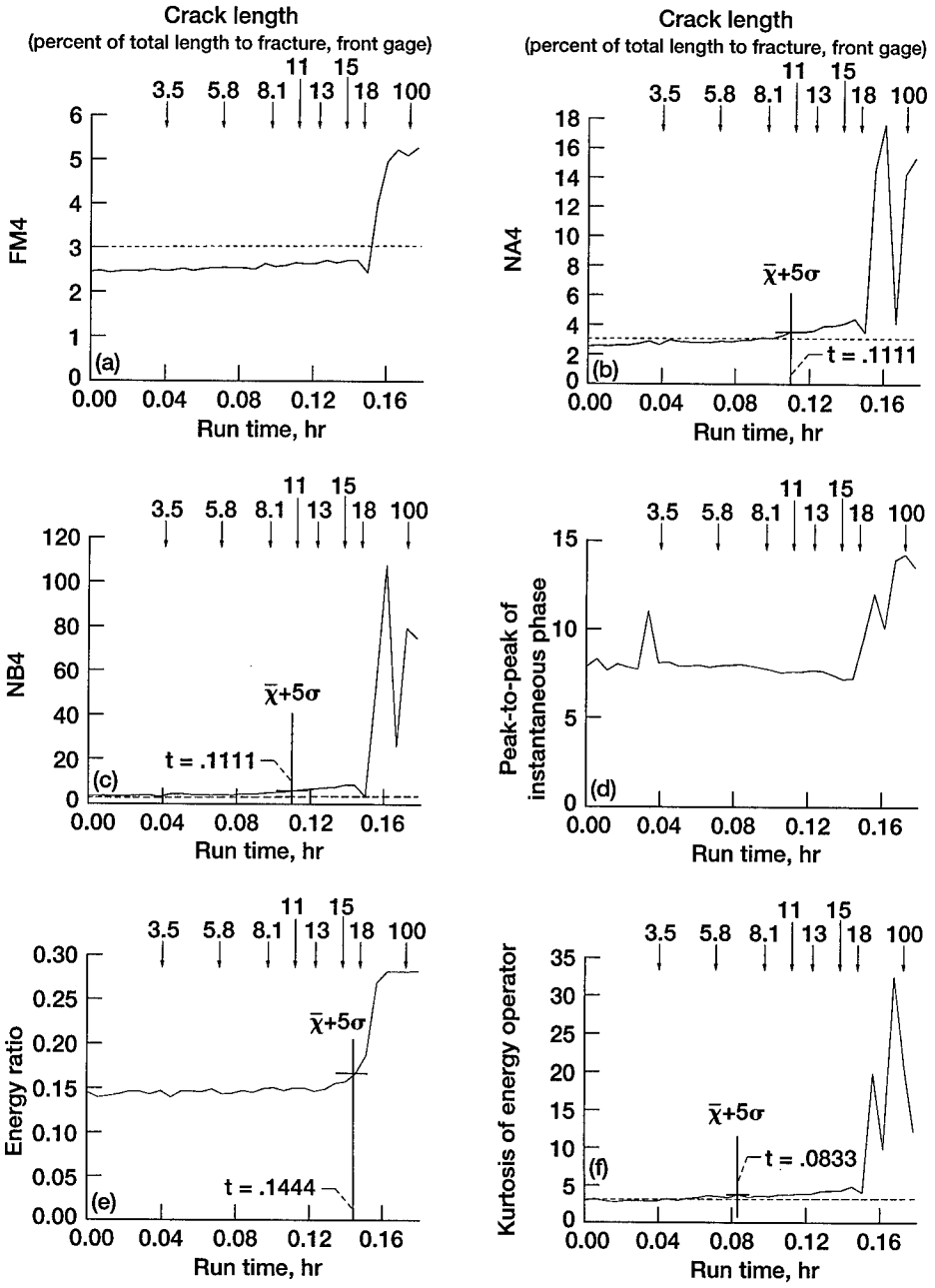
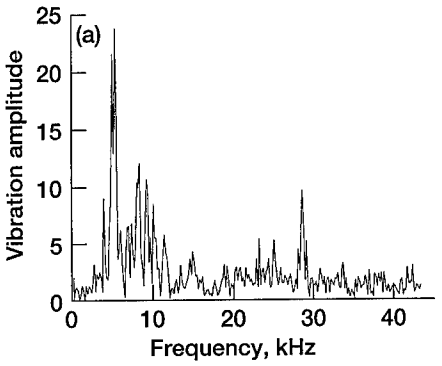
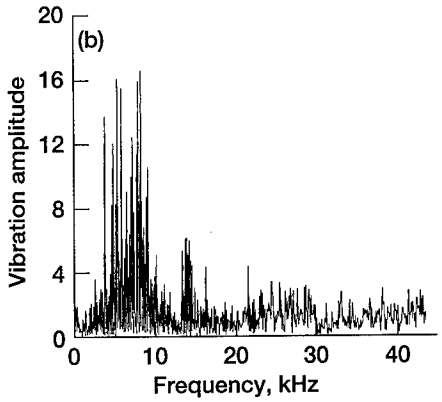


Figure 7.—Response of gear fault detection parameters for run 3.





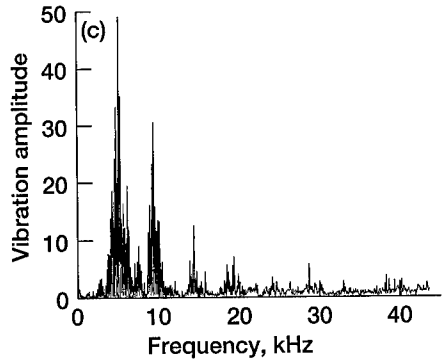
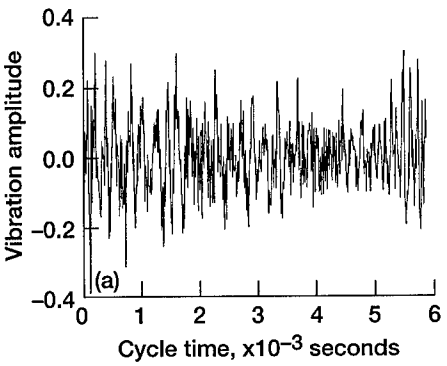
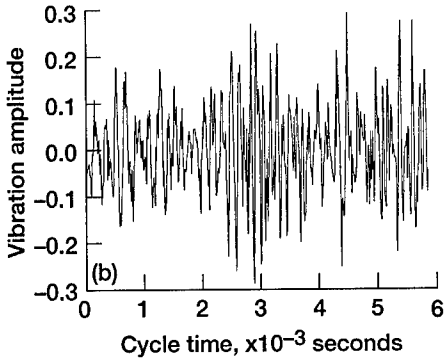


Figure 8.—Frequency spectrum at start of: (a) run 1, (b) run 2, (c) run 3.





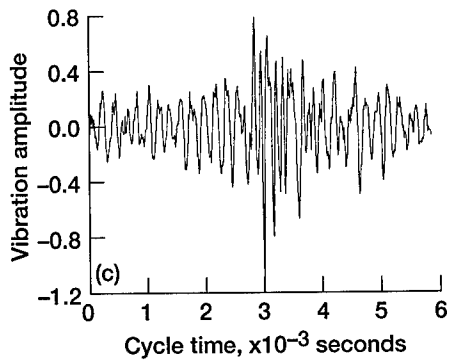


Figure 9.—Vibration plot at end of: (a) run 1, (b) run 2, (c) run 3.

METALLURGICAL EXAMINATION OF A FAILED SPIRAL BEVEL GEAR

Marc Pepi
U.S. Army Research Laboratory
Materials Directorate
Aberdeen Proving Ground, Maryland 21005

Abstract

A failed spiral bevel transmission gear from an Army cargo helicopter was first sent to the primary contractor for analysis and then the Army Research Laboratory. The failure occurred during a training flight at Ft. Meade, Maryland, where a reported "loud bang and shutter" forced an immediate landing. Subsequent inspection of the Number 2 engine revealed a 15-20 degree section of the gear had fractured and penetrated the transmission housing. The gear was fabricated from X-2M steel, a high hot hardness alloy. Light optical microscopy of the failed gear section revealed characteristics consistent with a fatigue failure, including a relatively flat fracture surface and beach marks. The fracture origin was located at a defect characterized by a darkened half-moon shaped region. The darkened topography was considered evidence that a pre-existing crack was present during the manufacturing process of the gear. Energy dispersive spectroscopy of the blackened surface revealed the presence of sodium, indicating the crack was open to the black oxide finish process. The pre-existing crack was oriented perpendicular to the direction of grinding, indicating the possibility of a grinding crack. Metallography confirmed that a pre-existing crack was a grinding burn, since evidence of rehardening and retempering were observed. Also a possible contributory factor to crack propagation was the presence of a carbide network in the carburized case of the X-2M steel part. Metallographic examination of the damping ring groove area (fracture origin location) showed a deeper than acceptable carburized case. It was later learned that this region was not masked, and was subject to a double carburization by mistake. Further visual examination of the failed component revealed small grinding cracks at the corner of the tapered surface. Although the morphology of the origin was featureless, fractographic examination of a secondary fracture (most likely an additional grinding crack) did reveal an intergranular "rock-candy" morphology. This was evidence that the crack propagated along the carbide network. It was concluded that the double carburization sequence of the damping ring groove section led to the formation of carbide networks within the carburized case. This scenario created a very hard surface, rendering it sensitive to grinding. Improvements have since been added to the manufacturing sequence including an additional magnetic particle inspection to determine the existence of cracks and nital etch inspection to locate possible grinding burns. In addition, the grinding process for burr removal in the damping ring groove region had been replaced with a less detrimental turning process. Also, the manufacturer has agreed to consider an addition to the material specification which would include micrographs of unnacceptable case microstructures.

Key Words: Carbide networks; Carburization; Failure analysis; Metallurgical examination; Spiral bevel gear; X-2M steel.

Background

A spiral bevel transmission gear (Figure 1) from an Army cargo helicopter failed during a training flight leading to an immediate landing. Inspection of the number 2 engine transmission revealed that an 8-tooth segment (out of a 35-tooth gear) had fractured and penetrated the transmission housing. The failed component was fabricated from X-2M steel (a modified H11 tool steel), and

carburized to a required case hardness of 59-64 HRC. The broken parts were subject to visual examination/ light optical microscopy, magnetic particle inspection, chemical analysis, metallography, hardness testing/case depth measurement, X-ray diffraction, scanning electron microscopy (SEM) and energy dispersive spectroscopy (EDS). Transmission electron microscopy/striation counts of the failed part were performed by the contractor.

Visual Examination/Light Optical Microscopy

Visual examination and light optical microscopy of the fracture surfaces revealed characteristics consistent with a fatigue failure, including smoothness of fracture and beach marks (Figure 2). The beach marks and radial lines on the fracture surface revealed that the origin was located in the damping ring groove portion of the gear at a darkened half-moon shaped defect (Figure 3). The pre-existing crack was oriented perpendicular to the direction of grinding, suggesting the possibility of a grinding crack. The darkened region had a featureless topography resembling that of a steel surface which has been exposed to a heat treatment atmosphere. Energy dispersive spectroscopy (EDS), in conjunction with a scanning electron microscope (SEM), within the darkened area revealed the presence of sodium. This finding was considered evidence that a pre-existing crack was present during the manufacturing process, since sodium nitrate and sodium dichromate are widely used to black oxide finish steel components. EDS spectra outside this darkened region failed to detect the presence of sodium. Further visual examination and magnetic particle inspection of the remaining component revealed small grinding cracks located in the same area as the fracture origin.

Magnetic Particle Inspection

Magnetic particle inspection was performed by the contractor on the failed part before sectioning. Several small grinding cracks were noted emanating from the damping ring groove region. The surface length of the cracks ranged from 0.015-0.025 inch.

Chemical Analysis

The elemental composition of the X-2M steel was analyzed to determine conformance to the governing specification. The carbon content was determined by infrared detection combustion, while the sulfur was automatic titration combustion. The contents of the remaining elements were determined through direct current plasma emission spectroscopy. The molybdenum content was slightly lower than specified, however, the remaining elements conformed to the governing specification. Table I lists the results of this analysis.

Table I
Chemical Analysis
Weight Percent

	C	Si	Mn	S	P	W	Cr	V	Mo	Fe
Gear	0.15	0.86	0.28	<0.001	<0.004	1.33	5.17	0.42	1.25	Rem
Specification	0.13-0.16	0.80-1.10	0.20-0.40	0.010 max.	0.015 max.	1.20-1.50	4.75-5.25	0.40-0.50	1.30-1.50	Rem

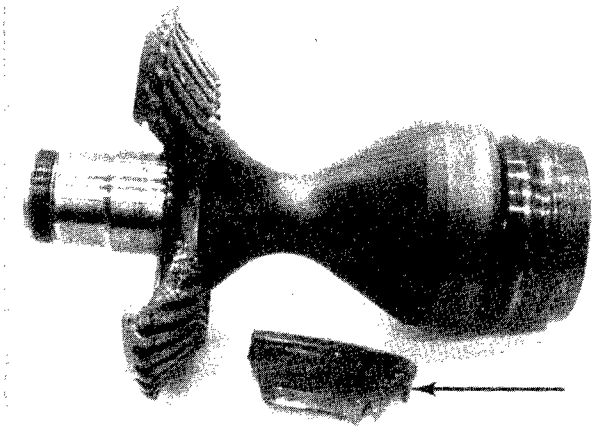


Figure 1 Optical Macrograph of the failed spiral bevel gear and the fractured eight-tooth segment (denoted by the arrow). Reduced 70%.

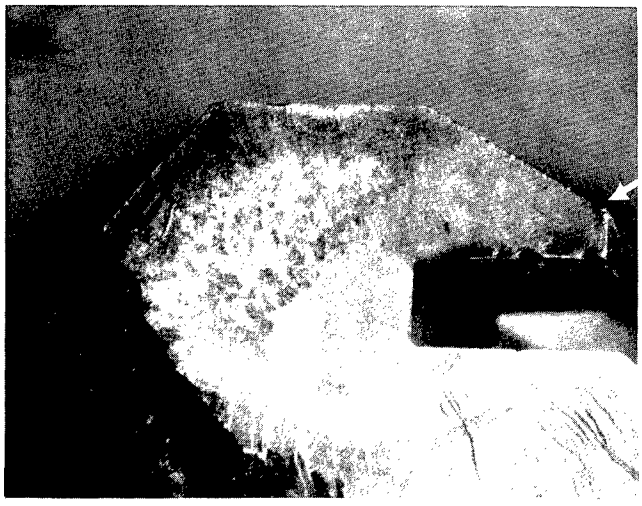


Figure 2 Optical micrograph of the primary fracture surface showing evidence of fatigue (beach marks). Arrow denoted fracture origin. Mag. 7.5x.

Metallography

A sample was sectioned through the origin and metallographically prepared to confirm that the pre-existing defect was a grinding crack. The sample showed evidence of rehardening (white region) and retempering (dark region) as shown in Figure 4. This pre-existing crack was oriented perpendicular to the direction of grinding, consistent with a grinding crack. Other grinding burns and grinding cracks were also noted adjacent to the fracture origin. In addition, metallography revealed what appeared to be continuous carbide networks (CCN's) in the region of the gear which contained the origin (Figure 5). These networks were discontinuous in high stress areas, such as the gear root and flank. Continuous carbide networks are generally caused by an excessive carbon potential during the carburization process, and can act to embrittle the case and reduce the fatigue limit of the component under bending fatigue conditions. Also, these networks render a surface sensitive to grinding, and if cracking occurs, it usually follows the path of the networks. Metallography combined with microhardness testing of the damping ring groove area (fracture origin location) revealed a greater case depth than specified (see Hardness/Case Depth Measurement section for results). The core structure was consistent with the prior treatment (duplex structure of free ferrite within tempered martensite).

Hardness/Case Depth Measurement

The case hardness was measured directly from a section of the component. The hardness was measured utilizing the superficial Rockwell 15-N scale, since the lighter major load (15kg) posed less of a risk of penetrating the case depth compared to the major load of the Rockwell "C" scale (150kg). The depth of the HR15-N hardness indentations was calculated to ensure the 15kg load did not penetrate the case and measure a composite of both the case and the core. The depth of penetration of the HR15-N diamond indenter was calculated from the following formula [1]:

$$(100\text{-HR15-N}) \times 0.001\text{mm} = \text{depth of penetration}$$

Applying this formula, a hardness of 88.9 HR15-N (the lowest reading measured) had a depth of penetration of 0.00044 inch which was safely within the required case depth of 0.030-0.050 inch. The readings were subsequently converted to the Rockwell "C" scale using standard conversion charts. Table II lists the results of case hardness testing. The average converted HRC measurement (61 HRC) conformed to the governing requirement of 59-64 HRC.

Table II
Case Hardness Measurements
HR15-N Scale
Major Load 15kg

Reading	HR15-N	Eq. HRC	
1	90.0	60	
2	88.9	57	
3	90.1	60	
4	91.0	62	
5	90.1	60	
6	90.6	61	
7	91.1	62	
8	92.4	66	
9	91.5	63	
10	90.3	60	
Average	90.6	61	Required 59-64

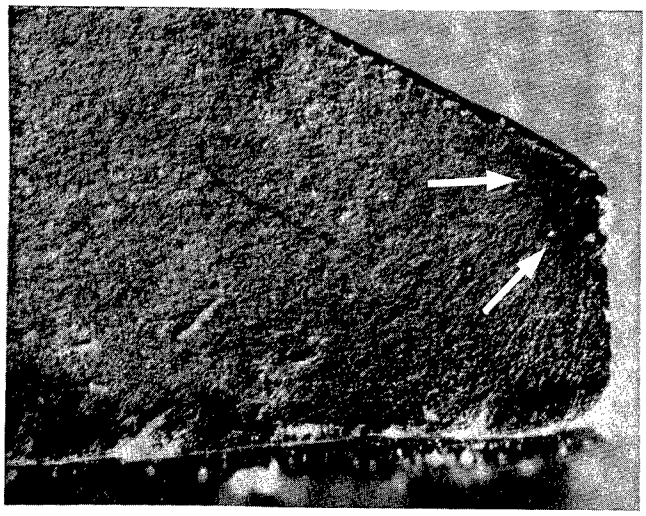


Figure 3 Magnified view if the processing defect at the origin. The grinding crack is denoted by arrows. Mag. 40x.

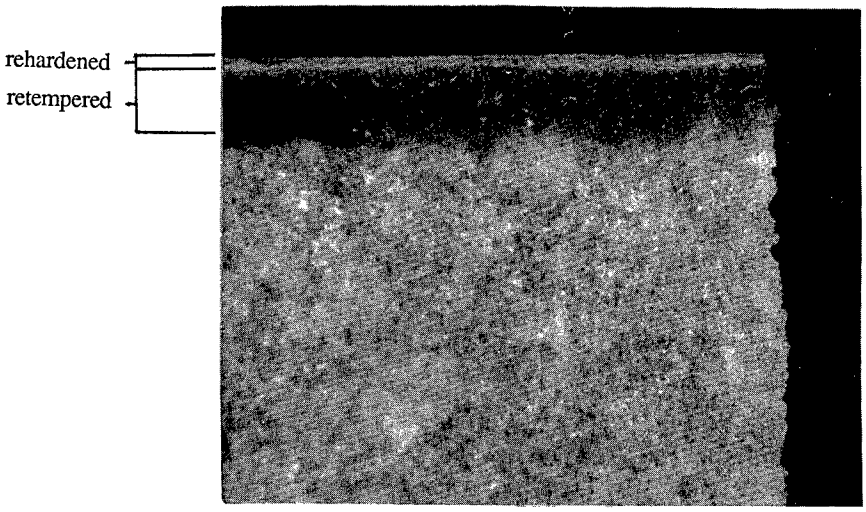


Figure 4 Metallographic section through the origin. Note the rehardened white layer, and the retempered darkened layer caused by the grinding burn. Mag. 200x.

The core hardness was measured utilizing the Rockwell "C" scale. A total of four sections were analyzed. The results are listed in Table III. Each reading conformed to the governing requirement of 36-44 HRC.

Table III
Core Hardness Measurements
HRC Scale
Major Load 150kg

Reading	Section 1	Section 2	Section 3	Section 4	
1	38.7	38.7	40.4	39.6	
2	38.7	39.2	40.5	40.4	
3	38.8	40.6	40.2	40.5	
4	38.6	40.2	40.7	40.2	
5	38.3	39.8	41.0	40.5	
6	39.7	40.3	40.7	40.7	
Average	38.8	39.8	40.6	40.3	Required 36-44

The effective case depth was measured from four metallographically prepared samples. Three samples represented gear tooth roots, and one sample represented the damping ring groove region. The case depth was defined as the perpendicular distance from the surface to a point where the microhardness of the part was 513 VHN (approximately 50 HRC). The case depths were measured directly from photomicrographs taken at 50x magnification. Table IV contains the results of case depth measurements. Each root measurement satisfied the governing requirement. However, a case depth of 0.075 inch was measured from the damping ring groove sample, which exceeded the maximum allowable case depth. It was later learned that the masking was inadvertently omitted before the second carburization cycle by the subcontractor contributing to this excessive case depth.

Table IV
Case Depth Measurements
Vickers Microhardness Scale
500 gram load

Sample	Region	Case Depth (Inch)
1	Root	0.053
2	Root	0.049
3	Root	0.044
4	Damping Ring	0.075
	Requirement	0.030-0.050

X-Ray Diffraction

X-Ray diffraction was conducted by the primary contractor at two locations adjacent to the fracture origin. The results were consistent with typical residual stress profiles for shot-peened X-2M steel. X-Ray diffraction was utilized by ARL-MD to determine the amount of retained austenite within the structure of the failed part. The specification states that the maximum allowable amount of retained austenite shall be 20%. Two samples were analyzed, and results of 8.81 and 8.94% retained austenite were obtained. These values conformed to the governing specification.

Scanning Electron Microscopy/Energy Dispersive Spectroscopy

The extent of fatigue propagation observed on the fractured part was 1.65 inch followed by a ductile overload morphology. The fracture origin was located at a defect within the damping ring groove region. The defect was "half-moon" shaped and had a depth of approximatley 0.005 inch and a surface length of approximatley 0.015 inch. The defect had a darkened topogrpahy, and was considered evidence of a pre-existing crack opened to the surface during the manufacturing sequence of the gear.

Energy dispersive spectroscopy of the defect revealed the presence of sodium. This was evidence that the crack was open to the surface during the black oxide finish process. Sodium is typically found in the black oxide salts used by the subcontractor. Spectra obtained outside the darkened area contained no evidence of sodium.

Fractographic examination of a secondary through-fracture revealed an intergranular morphology within the case of the damping ring groove area (see Figure 6). The remaining morphology consisted of ductile dimples, indicative of overload fracture. No evidence of fatigue was noted in this section. This finding was evidence that a grinding crack had propagated along the continuous carbide networks.

Transmission Electron Microscopy/Striation Count

The contractor obtained two stage chrome/carbon replicas from the primary fracture surface of the failed gear. These replicas were examined utilizing the transmission electron microscope (TEM) to determine an approximate fatigue striation spacing. The fatigue striations were difficult to locate, as is the case with most high-strength alloys, and were only located in isolated regions on the fracture surface. Two regions were analyzed which exhibited measureable striations. These regions were approximately 0.50 inch and 1.05 inch from the fatigue origin, respectively. The average spacings are listed in Table V. The points were used to fit a standard crack growth curve, which assumed that fatigue cracks propagate at a much slower rate near the origin. Assuming one striation represented one load cycle, it was estimated the fatigue crack propagated from the origin to overload region in approximately 20,344,000 cycles. It should be noted that the striation spacing near the origin was unobtainable due to the magnitude of the spacings. The spacings, typically in the 10^{-9} inch range were beyond the resolution of both the replicating tape and the TEM.

Table V
Average Fatigue Striation Spacing
TEM

Sample Location	Average Spacing (Inch)
0.50 inch from origin	8.3×10^{-7}
1.05 inch from origin	3.5×10^{-6}

Conclusions

The primary fracture origin was located at a pre-existing processing defect, which was a half-moon shaped darkened region. This defect was determined to be a grinding crack which emanated from a grinding burn, and was considered a contributing factor to the failure of this component. The grinding crack/burn was characterized by an orientation perpendicular to the direction of grinding, a half-moon shape, a darkened, featureless topography, the presence of sodium in the darkened area, and a rehardened and retempered structure adjacent to the fracture origin. Although the morphology of the primary fracture origin (darkened area) was featureless, it was possible that the continuous carbide networks within the carburized case facilitated grinding crack propagation.

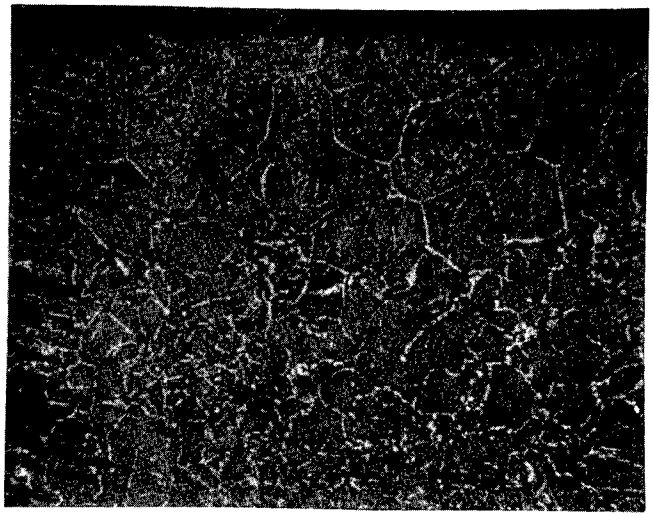


Figure 5 Metallographic cross-section through the damping ring groove region which was double carburized inadvertently. Note the carbide networks in the carburized case formed by this excess carbon. Mag. 300x.

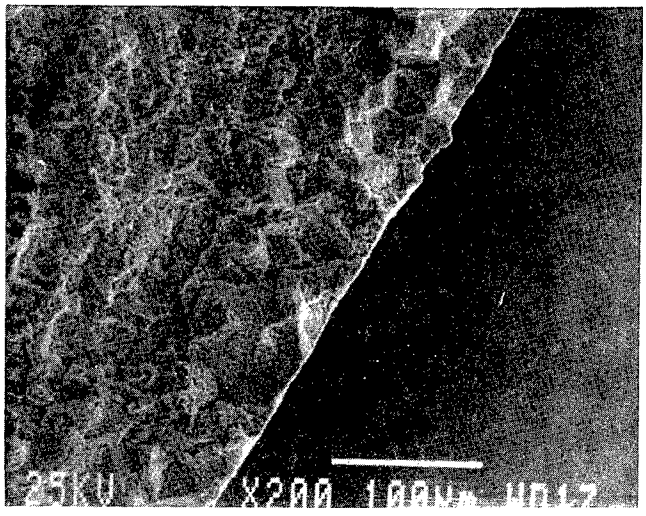


Figure 6 Secondary fracture sample from the damping ring groove area found to contain an intergranular morphology. A grinding crack most likely propagated along the carbide networks. Mag. 200x.

Corrective Measures

This failure was the second spiral bevel transmission gear to fail in three years. The first failure occurred when a 9-tooth segment of the gear fractured. Although major fire damage made that investigation difficult, it was determined that a fatigue crack had grown before final fracture. The origin of that failure was also determined to be the damping groove region. Steps were taken by the contractor to ensure this type of failure would not recur, including an additional magnetic particle inspection in the manufacturing process. In addition, grinding in the damping ring groove has been eliminated and replaced with a turning procedure which should keep heat generation to a minimum and ensure that defects related to grinding will not occur. In addition, a nital etch with 10x magnification step has been added to the inspection process, which should also help to detect future defects. The contractor has also agreed to consider an addition to the material specification which would consist of micrographs of unacceptable case microstructures. This would aid the subcontractor in determining rejectable parts before they are placed into service.

References

[1] American Society for Metals (ASM) Handbook, Volume 8, Mechanical Testing, "Rockwell Hardness Testing", p. 75, 1992.

THE DEVELOPMENT OF A COST BENEFIT ANALYSIS METHOD FOR MONITORING THE CONDITION OF BATCH PROCESS PLANT MACHINERY

B.S. Rajan
GlaxoWellcome Operations, Ulverston
Cumbria, U.K.

Brian J. Roylance
Department of Mechanical Engineering
University of Wales, Swansea, U.K.

Abstract: Whereas it is self-evident that large potential cost benefits exist for high capital cost - high consequential loss plant, the scope for achieving realistic cost benefits for lower value, batch process equipment is less obvious and is also difficult to quantify. The results presented in this paper represents an attempt to determine a suitable basis for establishing cost benefits for plant machinery typified by that utilised for the manufacture of pharmaceutical products. In this preliminary investigation, a mathematical model has been devised for pump operation throughout the whole plant. It is based on actual data obtained over a five year period in which a maintenance cost prediction is established and the advisability of utilising condition - based maintenance strategy is decided.

Key Words: Cost Benefit Analysis; Maintenance; Mathematical Model; Pumps;

Introduction: Maintenance of running machinery has been a need as long as machinery has existed. From the onset of the Industrial Revolution until the end of the Second World War machinery was repaired as it broke down. There were no philosophical considerations of any other type of maintenance practice. It was implicit in the running and maintenance of Marine Machinery [1] published in 1965 that this was the philosophy to be followed. However, since then there have been a numerous studies of maintenance practice [2] and a recognition that costs could be as much as 80% of machine capital costs per annum. Hence it is worth analysing whether other philosophies are most cost effective. The current maintenance costs for British industry is estimated to be £20 billion and therefore the scope of prospective saving can be spectacular.

Initially the move away from breakdown maintenance was driven by a recognition that breakdowns had considerable consequential costs if a continuous manufacturing process was interrupted without warning. It was also realised that manning for the peaks in maintenance activity was a very inefficient use of manpower and hence the next philosophy to be advocated was that of preventive maintenance. In this philosophy, machinery was serviced at regular

intervals and all worn parts replaced. (This could be considered as one form of condition monitoring where by a full intrusive examination was carried out). The time intervals used for this maintenance activity were decided empirically as detailed records were not available in most areas. In many cases they were done to suit either the resources available or at convenient time intervals (e.g. 26 weeks, 1 year etc.).

In the late 60's and early 70's, in the defence industry, especially in operational units, where machine *availability* was the overriding concern, a philosophy of using the reliability data that was available to establish preventive maintenance intervals was suggested. As the data on machine availability and repair times were available in most combat and continuous running situations for key machinery, MTBF (Mean Time Between Failure) and MTTR (Mean Time To Repair) times were determinable. Thus, statistical analyses could be carried out and confidence limits for machine availability theoretically established. Various elegant mathematical models are available for establishing ideal preventive maintenance intervals [3,4]. However, data to establish this is still not available for most mainstream industry. (The nuclear industry is a particular exception since they have to *prove* the reliability of plant; the National Centre for Systems Reliability was established to collate this data).

As more industries collate real data relating to *their* plant and *their* running conditions, data becomes available which can allow mathematical models to be constructed for general plant. The main problem with this approach is deciding the confidence limit at which normal industry can sustain preventive maintenance costs. Further, when one deals with very general large populations of machines, the levels of maintenance to guarantee machine availability to a high confidence limit becomes very expensive. It is also recognised that if there are components which fail in a truly random manner (e.g. rolling element bearings) the system has only limited usefulness. Nevertheless, Reliability Centred Maintenance (RCM) is now an accepted strategy and there are a number of plants actively employing this approach [5].

From the early 70's, instruments have been used for field testing and measurement of various parameters which give an insight into machinery condition. The first (and still the most widely used) parameter is vibration. This has been used for many years in machinery diagnostics as a trouble shooting method. The technical expertise needed to carry out this function was initially very high and so it was the preserve of a few specialists. However, it was soon realised that overall vibration levels could be easily measured and used as a trending tool to make judgements on machine condition. On a wider front, Michael Neale and Associates were commissioned by the Department of Industry in 1970 to produce a Guide to Condition Monitoring which for the first time also considered the economics of condition monitoring and its place in the maintenance regime [2].

It is quite evident that cost benefits are very high in the case of high capital value plant (see Appendix 1 - A case study). It is not so evident in the case of lower value plant at what level condition monitoring regimes should be pitched. The present study is an attempt to investigate this situation and establish a cost benefit analysis method which could be used in such cases.

It is now accepted that vibration analysis is not the only technique available. For particular classes of machinery (e.g. rail locomotives, earth moving equipment etc.) a more suitable method would be to monitor the oil condition. Oil condition monitoring, and maintenance systems based on it are now common in many industries. Other methods include acoustic emission, stress wave level, temperature measurement; all of which give insight into machinery condition. Thus, the present generation of strategies includes Condition Based Maintenance (CBM). A subset of this philosophy is where CBM is supplemented by low level intrusive inspection (using modern optical equipment or other devices) so that reliability can be **guaranteed** for a particular length of time. This is specially useful in applications such as Aircraft or Nuclear Plant where the consequences of failure are particularly high.

In summary the main approaches to maintenance now available are:

1. Breakdown Maintenance
2. Time-based, planned preventive maintenance
3. Reliability-centred maintenance
4. Condition-based maintenance

Cost Effectiveness of Maintenance Regimes: The comparative life cycle costs of the four maintenance philosophies has not been fully investigated. A start was made by Jardine [3] who looked at different types of machine characteristics in order to arrive at replacement, inspection and overhaul decisions. Implicit in his methodology was an assumption that maintenance was to be part of operating costs but these would vary in a simple manner with time. Hence, as operating costs varied, a replacement decision would become the correct option and this could be mathematically calculated. A number of different possibilities were considered where operating costs decreased with age, increased with age, had to function within a finite (bounded) time horizon, and where the unit was a stand-by machine with all the above possibilities. Using these types of model, he also attempted to develop equations for inspection decisions (intrusive and hence, essentially a "planned maintenance" approach). These models are theoretically defensible but the type of data required to support the equations is generally unavailable. Further, the problem of components with random failure modes is not addressed.

Moubray [5] disputes what a failure is and concludes that this could be perceived very differently by people with differing viewpoints. An example quoted is that of a hydraulic system where a small leak is seen as failure by the Safety Officer, a much larger leak as failure by the engineer and only the complete stoppage of the system as failure by the production staff. Therefore, costs become dependent on the "failure" being investigated.

The authors consider that failure should be judged against performance standards and costs similarly calculated against these standards. However, it is important to take into account consequential costs, which are usually the largest item in a breakdown maintenance regime.

Carter [4] employed a similar approach to calculate whole life cycle costs using a statistical approach in which he defines a failure in terms of a machine failing to perform (the "production" definition).

The current investigation is aimed at establishing a cost benefit analysis method based on evidence obtained from actual maintenance data.

Background to Data Collection: The data used to evaluate the system is acquired from the pharmaceutical industry and is characterised by some unique features:

1. The product is small in volume terms but is of very high value

2. A large number of the production stages are batch process and hence there is machinery which is idle time between batches.

3. As the quality requirements of the product are extremely tight, all machinery used has to work within very stringent operational parameters as this is one way of ensuring consistent quality of the product.

Methodology: The system proposed for the cost benefit analysis is based on a standard spreadsheet package so that it can be easily transferable. It is also based on actual cost data for both direct and consequential costs.

It was recognised at the outset that there would be differences between various classes of machines and hence all machines on-site were divided into groups by machine class. The single group with the largest number of machine trains was "pumps".

Other classifications include fans, compressors, gearboxes, generators and dryers.

Classification: The pump population was divided up according to type and power rating as shown in Table I.

Of the c.6500 machine trains on-site c.3000 are "pumps". Approximately 400 machine trains of all types are subjected to high level vibration monitoring (once a month surveillance with full Fourier analysis of the results and complete diagnostics being generated by an artificial intelligence system) using portable data collection systems. Four hundred pumps are subject to low level vibration monitoring (using only overall vibration levels to establish trends). In addition, forty machines are subjected to monthly oil condition monitoring and analysis, and 40 machines to checks using ultrasonic stethoscope and semi-intrusive examinations (using optical equipment such as introscopes and endoscopes) conducted, as requested, by maintenance engineers.

1. Costs for these services were assessed using three main criteria to ascertain the total costs

2. Capital cost of equipment used using discounted cash flow with a 10 year amortisation period

3. Labour costs. These were assessed at two separate rates. The low level vibration monitoring programme was undertaken by craftsmen who had served a recognised apprenticeship, whilst

all other services were provided by technicians who, additionally, had a three year, in-house, training programme in inspection.

4. Software and equipment maintenance costs. For the low level vibration monitoring carried out there were no maintenance costs since the purchase of the equipment. However, a nominal sum has been taken (equivalent to 5% of purchase price per annum) to allow for any future costs.

Solartron [6], in their model, also consider some cost benefits associated with:
1. Operational safety effects on plant and environment
2. Personnel safety risk
3. Operational issues excluding safety, e.g. severity of machine duty
4. Technical uses

The pharmaceutical industry is highly regulated as there are many fire and bio-hazards associated with it. Hence, on analysis, condition monitoring does not change their safety effect on either the plant or the environment. Therefore, this factor was not considered in the current model. The same situation applies in the case of personnel safety risks.

Operational issues, including severity of duty were found to reveal its effects on the consequential costs of failure and this is included in the analysis of consequential costs. Other technical issues that were considered by Solartron related to selection of in-line monitoring vs. hard wired surveillance methods vs. hand held data collection methods. Since the plant being considered here is primarily a batch production plant with very few machines on continuous duty, only hand-held data collection equipment was considered for this case study.

The need for a mathematical model to assess the cost benefits becomes apparent when the fact that the savings could be small when low capital value equipment is considered. But if sufficiently large numbers of such machines are taken into account, then the situation could alter. To determine the point at which each regime of condition monitoring becomes cost-effective requires a mathematical model so that the situation can be analysed.

Mathematical Model: To be able to *predict* costs, the model proposed is in two parts:

1. Costs of damage to the machine itself
2. Consequent costs to the process (or product)

All costs have been converted into current costs by compounding their book costs on the date of acquisition, or of repair, by using the published RPI figures within the spreadsheet.

Costs of damage to the machine: A number of models have been proposed [4] but in every case the cost of repair has been treated as one of the "knowns". To be able to take the condition monitoring "decision" it was felt that an ability to *predict* this cost was essential and hence the proposed model uses the capital cost (which is known in every case) and a factor K_d which is

derived from analysis of the model. A similar index K_q will be used to predict consequential costs where appropriate.

A simple model is proposed:

Cd = Ci*Ip*Ic*Ipr*Kd
where, Cd = Direct costs of machine breakdown
Ci = Initial cost of machine corrected to present day value
Ip = Power Index
Ic = Criticality Index
Ipr = Process Index
and Kd = Direct costs factor

The Power Index is a method by which the relationship of power to costs was brought into the analysis in a non-dimensional form. A number of alternatives were tried but offered no significant advantages over a simple relationship on a scale of 1 to 10. As the powers of the pumps in the study varied from fractional kilowattage to c.250 kW the relationship shown below was used.

Power (kW)	3	8	15	30	50	80	110	150	200
Ip	2	3	4	5	6	7	8	9	10

The Criticality Index is a measure of how critical the machine is to the process of which it is a part. This index is again a number on a scale of 1 to 10 which was determined by studying the process and evaluating the criticality of the machine. If, for instance, a machine was one of a series in which a stand-by machine could be put directly into service without interrupting the process, the criticality index was automatically set at 1. On the other hand, if machine failure had the result of immediately stopping the entire process, then it had a criticality rating of 10 (e.g., the failure of a motor or gearbox on the stirrer drive of a reactor vessel would immediately halt the reaction and would have a criticality index of 10). If the machine failure was such that the process could be maintained but only at reduced throughput then I_c = (Through put available/throughput possible) x 10.

The Process Index was evaluated on the basis of how much of value had been added to the product at that point in the overall process. This index is again a number on a scale of 1 to 10. Hence, all machines at the stage where the raw material is fermented would have a process index of 1, while all machines in the finishing suites would have a process index of 10. Figure 1 shows the general process flow through the plant.

Using these factors and costs, the direct costs factor was evaluated and plotted against power. This could then be used to *predict* direct costs of breakdown of any machine in that particular class.

The frequency of breakdown was evaluated from pump field data collected since September 1989. A total of 329 pumps in the solvent recovery area (see Appendix 3) were studied and 711

failures took place over a 5 year period giving a machine reliability of 0.57 per machine per year (using the definition as proposed by Carter (6) after Carhart:

R(t) = Cumulative probability function of occurrence of survival

Planned maintenance costs and inspection costs were calculated using the company standard hourly rates and published prices for spares. Capital costs were taken from the company records and adjusted for inflation using the published RPI figures. These figures were then compared with new quotations for the same machines from the manufacturers so that confidence could be gained in this method. The comparison showed that if prices were extrapolated from pre. 1990 data the deviations became very large for smaller pumps, though (taking a 10% deviation as the maximum acceptable) larger pumps could be regressed over a longer period.

To determine k_d values, the raw data acquired from the 329 pumps was analysed using curve fit routines using several methods. It was evident from this that the best fit was obtained from using the expression $y = m\hat{}x+b$ (Figure 2) and hence this is used in the subsequent predictive work.

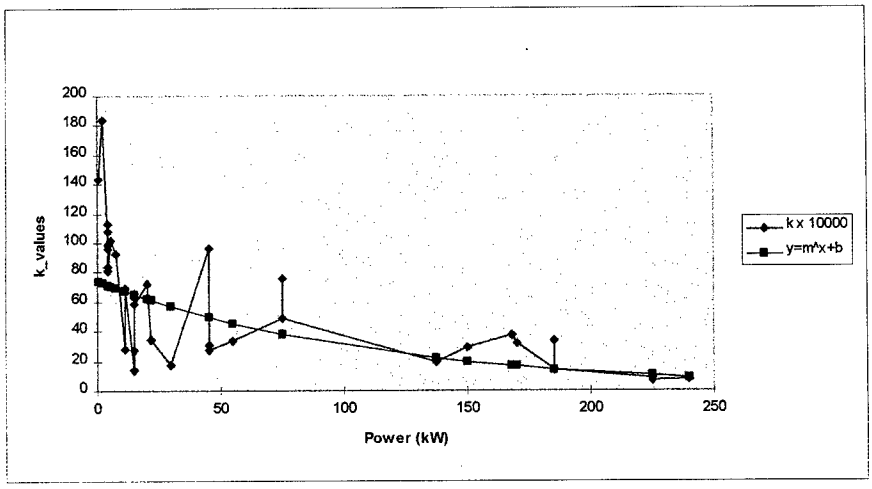


Figure 2 Curve Fit for K_d Values

Analysing the data for the period from 1990 (when low level vibration monitoring was commenced for pumps with less than 15 kW motors) to the present, the average saving by using vibration measurements to trigger maintenance against a time-based system, was £1124 over the five year period, i.e., £224.80 per annum per pump. The monitoring costs were a one off purchase of a vibration meter in 1990 at £1750 (which has now been completely written off) and labour costs at 1.5 hrs per pump per annum, which, at the company's present rates, equates to £21 and hence, there is a saving of £203.8 per pump per annum.

From the above figures, it is evident that if payback of two years is used as the criterion, and the current cost of a vibration monitor is £4500, then a minimum of 11 pumps must be within the group for the system to be more cost effective than a time-based planned maintenance system.

Breakdown maintenance costs were difficult to acquire but using a small sample of pumps that had broken down it was estimated that breakdown costs were 1.8 times of planned maintenance costs per pump because of greater damage sustained within the pump. Using the data to estimate pump reliability gave a figure of 0.57 per pump per year. The planned maintenance system generated 0.86 pump overhauls per pump per year. With properly targeted planned maintenance, there were almost zero breakdowns and hence, with the population of 329 pumps studied, breakdown maintenance costs were only £7.7 per pump per year greater than if planned maintenance was used.

If a high level condition monitoring system was to be used, the capital cost of the instrumentation and software would be £26000. The cost of data acquisition and analysis is £81.9 per machine train per annum. If the capital costs are amortised over 5 years, the saving would be £127.1 per pump per annum. Hence, for this system to be cost effective with a payback period of two years, a minimum of 205 pumps need to be monitored.

These calculations are now coded into a spreadsheet which makes an assessment of whether condition monitoring will be cost effective over whatever payback period is selected. The "Frontsheet" of this MS Excel spreadsheet is shown in Appendix 2. Two separate examples are shown, one of a 22 kW pump and another of a 240 kW pump. The decision arrived at in each case was based on the predicted costs of repair which were as shown. The actual costs of repair of these pumps were £510 and £1050, respectively which gives errors of 20.9% and 9.5% respectively.

The cost saving per pump per year by using condition monitoring is given by:

C_p= 0.8 *Predicted repair costs - (Capital cost of condition monitoring equipment/ No. of machines in the group) - (Running costs of condition monitoring)

Conclusions: The conclusions of the study show that:

1) A simple model can be used to predict maintenance costs for the population of pumps. These costs are related to the capital cost of the pump. The agreement between predicted costs and actual costs become increasingly better as pump powers increase.

2) Capital costs of purchase can only be meaningfully extrapolated using the original costs and the RPI figures for a comparatively short time span (< five years).

3) The costs of maintenance using the different regimes in decreasing order of costs are:

   a) Breakdown maintenance
   b) Planned (time based) maintenance
   c) High level vibration monitoring based maintenance
   d) Low level vibration monitoring based maintenance

4) The above conclusions should be read with the following caveats:

   a) Breakdown maintenance is only marginally more expensive than planned maintenance. A slight increase in pump reliability will bring down breakdown maintenance costs below that of planned maintenance.

   b) The extra information acquired as part of a high level vibration monitoring based maintenance system does not lead to any greater reliability of the pumps than if a low level vibration monitoring based system were used.

   c) The break even point rises dramatically between using the low level system and the high level system and hence, in the introductory stages, a low level system will deliver benefits much more positively than a high level system.

   d) The above study was composed of pumps, all of which are fitted with stand-by machines and hence failure has no consequential costs. When the study is extended into Fans, Compressors, Generators, Dryers, etc. many of which have no stand-by machines, the consequential costs will play a dominant role.

References

1. The Institute of Marine Engineers, The Running and Maintenance of Marine Machinery, 5th Edition, 1965.

2. Michael Neale and Associates (for the Department of Industry), A guide to the Condition Monitoring of Machinery - HMSO 1970.

3. A.K.S. Jardine and S. Kennedy; Maintenance Replacement and Reliability, Pitman Pub. Ltd, 1973.

4. A.D.S.Carter, Mechanical Reliability, 2nd Edition, Macmillan 1986.

5. Solartron Instruments, Cost benefit analysis methods for Condition Monitoring - Technical note, No. 27, 1994.

Acknowledgements

The authors wish to acknowledge with grateful thanks the kind permission of GlaxoWellcome Operations UK Ltd to present and publish this paper.

Pump Type	Power Rating - kW				
	0-3	3-8	8-15	15-30	30-50
Centrifugal	√	√	√	√	√
Positive Pressure	√	√	√	√	√
Single Stage	√	√	√	√	√
Centrigugal	√	√	√	√	√
Positive Pressure	√	√	√	√	√
Multi-Stage	√	√	√	√	√
Centrifugal	√	√	√	√	√
Vacuum	√	√	√	√	√
Single Stage	√	√	√	√	√
Centrifugal	√	√	√	√	√
Vacuum	√	√	√	√	√
Multi-Stage	√	√	√	√	√
Positive Displacement	√	√	√	√	√
Positive Pressure	√	√	√	√	√
Single Stage	√	√	√	√	√
	50-80	80-110	110-150	150-200	>200
Centrifugal	√	√	√	√	√
Positive Pressure	√	√	√	√	√
Single Stage	√	√	√	√	√
Centrigugal	√	√	√	√	√
Positive Pressure	√	√	√	√	√
Multi-Stage	√	√	√	√	√

Table 1 - Classification of Pump Type and Power Rating

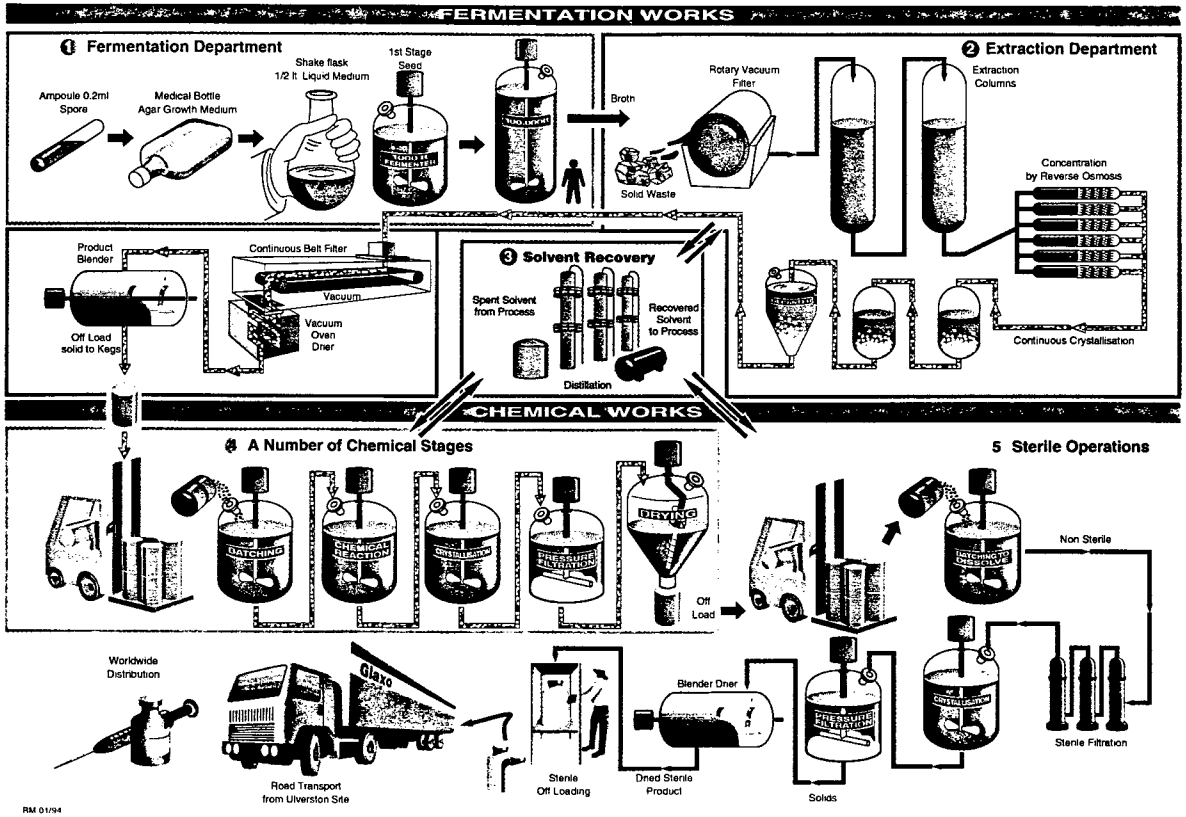


Figure 1 Schematic Representation of Pharmaceutical Products Process

Appendix 1

A two stage centrifugal fan was supplied as part of a new installation. This fan has two bearings, both, Cooper Split type, with the first stage overhung beyond the second bearing. The fan is powered by a 110 kW motor. The capital cost of the fan (normalised to 1995 prices) was £96,890.

Post installation vibration checks carried out on the fan gave cause for concern. Alignment was considered partially responsible and the drive motor was realigned. Subsequent vibration readings showed a steady deterioration and analysis of the spectra showed presence of bearing outer race problems at *both* fan bearings. The main installation contractor and the fan manufacturer were invited to comment as the process on which it was being used was now in full production.

The comments received were that the vibration levels were quite acceptable for normal operation and that the machine was covered by a maker's guarantee.

After nearly 8 months operation the fan completely self-destructed. During the 8 months of operation the vibration levels had shown steady deterioration. It cost the manufacturer £47,130 to repair the fan (normalised to 1995 costs). It was estimated that modifications costing less than £5,000 could have avoided the incident.

Appendix 2 - Examples of Master Sheets Used for Assessment of Cost Effectiveness of Pumps

Give power of pump in kW	22	240
How many pumps are there in the group	10	10
What was the capital cost of the pump at time of purchase	£4,644	£17,084
What year was it purchased	1991	1991
Predicted capital cost of purchase on current date	£5,270	£19,388
Is there a stand-by pump in the circuit (Y or N)	y	y
If no stand-by pump enter criticality index	2	2
Pump process index	5	7
Enter hourly rate for labour	£14.00	£14.00
Predicted pump repair cost	£645	£1,168
Predicted pump condition monitoring cost (p.a.) - (l.l)	£63.50	£63.50
Predicted pump condition monitoring cost (p.a.) - (h.l)	£249.98	£249.98
Payback period for assessment	2	2
Is use of low level condition monitoring justified	Yes	Yes
Is use of high level condition monitoring justified	No	No
Capital cost of low level Condition Monitoring Equipment	£4,250	£4,250
Capital cost of high level Condition Monitoring Equipment	£20,000	£20,000
Year of Purchase	1995	1995
Corrected cost of low level equipment	£4,250	£4,250
Corrected cost of high level equipment	£20,000	£20,000
Cost saving with use of l.l. equipment	£452	£871
Cost saving with use of h.l. equipment	£266	£685
Years for payback on low level equipment	9.4	4.0
Years for payback on high level equipment	75.2	29.2

FAILURE ANALYSIS OF A PITCH LINK SELF-LOCKING NUT MS 17825-10

Scott Grendahl
U.S. Army Materials Research Laboratory
Aberdeen Proving Ground, MD 21005-5059

Abstract: The Aviation and Troop Command (ATCOM) shipped a broken self-locking nut to the Army Research Laboratory-Materials Directorate (ARL-MD) and requested a metallurgical investigation to determine the probable cause of failure. The component was received on 31 October 95. The hexagon shaped, castellated self-locking nut, MS 17825-10, was attached to the lower end of the pitch link located on an army attack helicopter. The outside of the nut contained the manufacturer's designation which consisted of the letter "G" stamped 5 times along a longitudinal section of the part. The nut was fabricated from non-corrosion resistant steel and is to be used at temperatures up to 250°F. During assembly while applying torque, the nut cracked and split open. Further inspection revealed that the nut was cracked at one of the castellations.

Light Optical Microscopy: The self-locking nut received was split open approximately 1/8 of an inch along its longitudinal axis as shown in Figures 1. The cadmium plating was relatively intact with no evidence of general corrosion or pitting observed. The fracture surface contained slight discoloration from post fracture corrosion. This discoloration from corrosion can be seen in Figures 2 and 3. Upon sectioning, it was determined that the crack originated from a slightly subsurface area of the nut near the O.D. as seen in Figures 4 and 5 (arrows denote origin and inclusion band). The radial lines and chevron pattern converge to an area adjacent to the O.D. of the nut opposite the castellated end. A long narrow band was noted on the fracture surface traveling parallel to the O.D. edge of the nut opposite the threads as shown in Figures 2 - 5 (arrows denote band). This region displayed a distinctly different appearance from the remaining fracture. It was deduced that this artifact was a result of an elongated inclusion from primary processing. The inclusion encompassed the entire length of the nut. The fracture likely progressed under uniform loading conditions as evidenced by the lack of distinct crack arrest marks in the origin area. The fracture surface morphology was generally ductile throughout with a considerable amount of inclusions aligned along the longitudinal axis, remnant of the extruding direction of the part. There was a shear lip on the thread tips and on the castellated end of the nut which indicated final fracture.

Metallography: According to the governing specification, MIL-N-25027G, *Nut, Self-Locking, 250 Deg. F, 450 Deg. F, and 800 Deg. F*, the maximum allowance for a defect depth for nuts made from bar or wire with a thread size of 0.625 inches is 0.017 inches as listed in Table VII of that specification entitled Limits of Depths on Laps, Seams, and Inclusions. Also contained in that specification, is "Figure 2" entitled Acceptable and rejectable defects of self-locking nuts as revealed by magnetic particle or fluorescent penetrant inspection. This figure shows that laps or seams that intersect an edge in line with beam slots (castellations) passing through the center of the hex flats are rejectable.

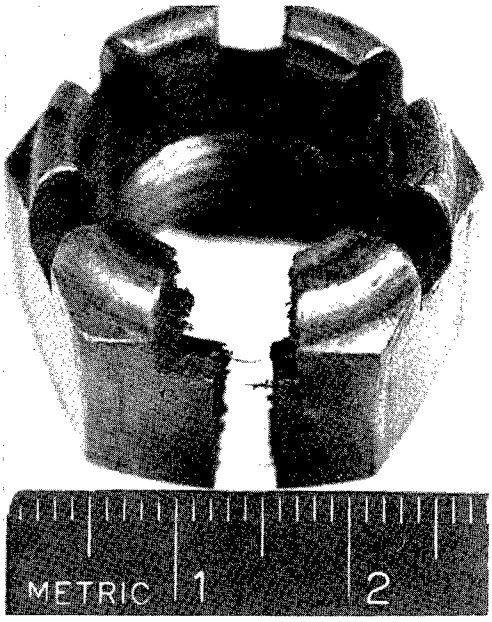


Figure 1. As received condition of failed component, scale in centimeters.

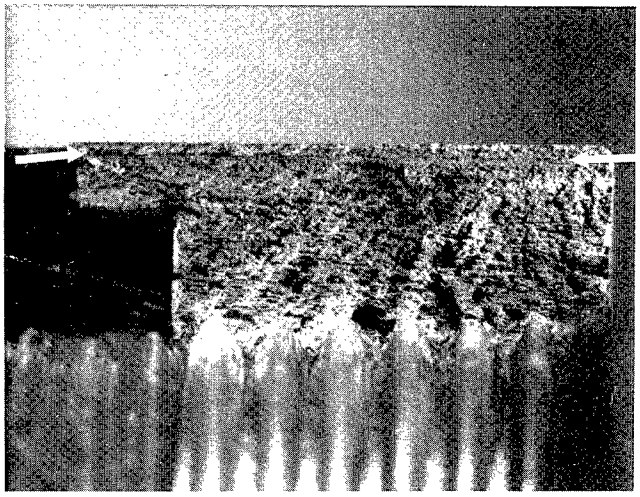


Figure 2. Micrograph of fracture half "A" showing corrosion. Mag. 10x

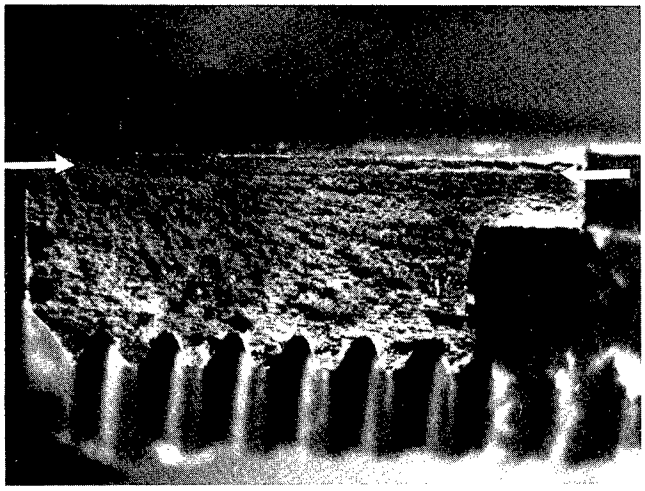


Figure 3. Micrograph of fracture half "B" showing corrosion. Mag. 10x

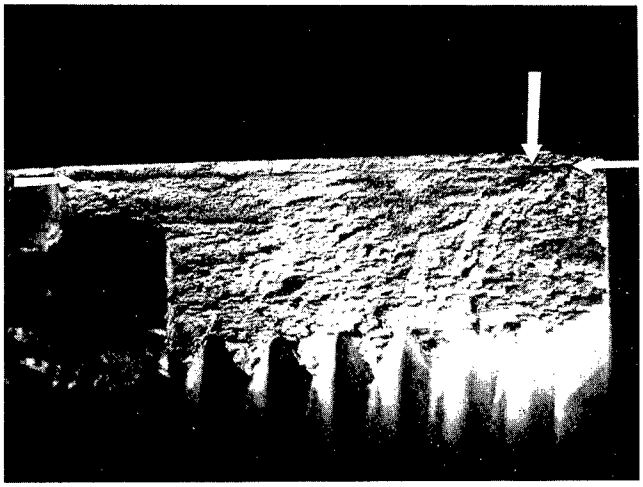


Figure 4. Micrograph of fracture half "A", arrow denotes origin. Mag. 10x

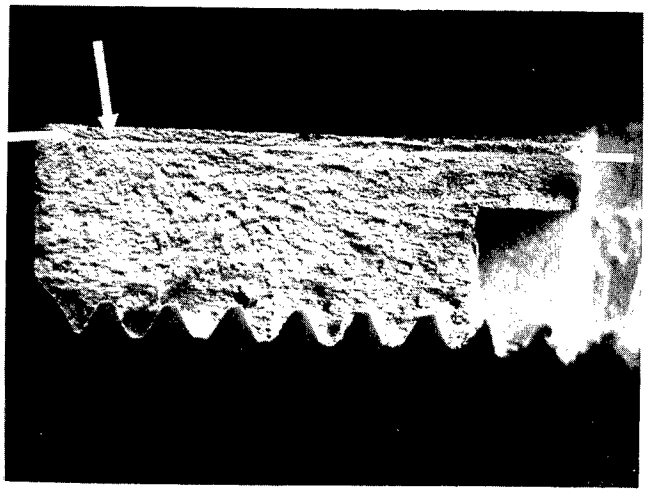


Figure 5. Micrograph of fracture half "B", arrow denotes origin. Mag. 10x

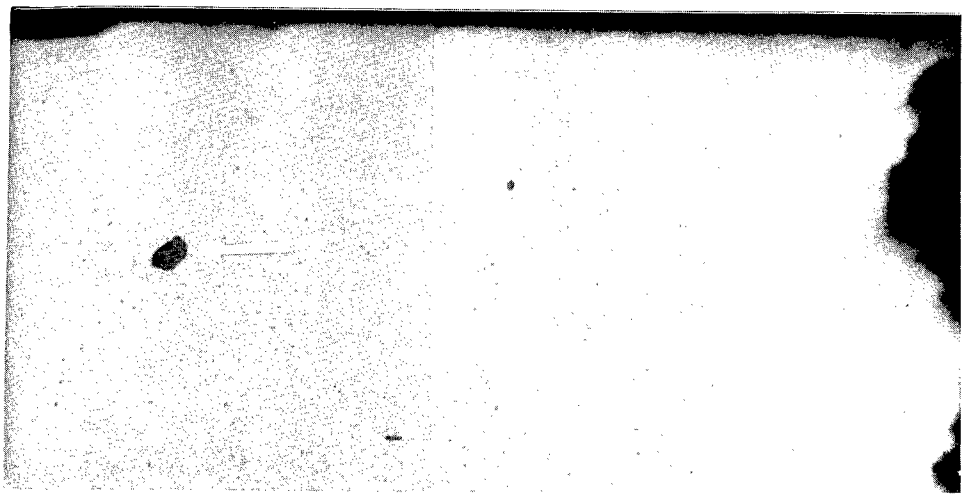


Figure 6. Micrograph of failure inclusion size and depth. Mag. 200x

The failed part has a thread size of 0.625 inches and therefore falls under the above criteria. A cross-section of the inclusion from the failed part can be seen in Figure 6. This section was taken from an area 0.025 inches from the fracture surface at the end opposite the origin. The inclusion depth at this location is 0.010 inches as measured from the photomicrograph per ASTM-B-487. This is within the limit set forth in MIL-N-25027G. It is important to note that the governing specification is probably not referring to inclusions of the type that caused this failure in setting the limits on depth. The specification is most likely referring to abnormalities or discontinuities open to the surface that can not be labeled as laps or seams. Inclusions of this severity are indeed rare since proper primary processing techniques usually eliminate them. The maximum depth of this inclusion was contained on the fracture surface and can be seen in the Scanning Electron Microscopy section of this report. Mn-S stringer inclusions are very common in steels of this type, however the inclusion that caused this failure was approximately 100 times larger in cross sectional area than normal inclusions of the same type.

The microstructure of the nut was further investigated. There was no cleanliness rating or inclusion content limit found within the governing specification. However, the other inclusions within the nut matrix were large and numerous. Although not required, the measure of cleanliness for this material was a type A (sulfide type) level 5, the worst, according to ASTM-E-45 entitled Chart for Determining the Inclusion Content of Steel. A tempered martensitic lathed structure was noted which was consistent with the prior heat treatment for this material.

Chemical Analysis: Chemical analysis was performed to determine which one of the four prospective alloys (AISI 1137, 11L37, 4130, 8740) was utilized in the fabrication of this component. A section of the nut was subjected to atomic absorption and inductively coupled argon plasma emission spectroscopy. The carbon and sulfur contents were analyzed by the LECO combustion method. Results of the analysis showed the nut compared most favorably with alloy 11L37, as shown in Table I. The carbon content of the material was out of specification as it was more than 0.1% higher than the requirement of 0.32 - 0.39%. Having a higher carbon content in steels can lead to deleterious effects in the mechanical properties of the material.

Table I
Chemical Composition of the Pitch Link Nut
Weight Percent

Element	Component	1137	11L37	4130	8740
Carbon	0.50	0.32-0.39	0.32-0.39	0.28-0.33	0.38-0.43
Manganese	1.42	1.35-1.65	1.35-1.65	0.40-0.60	0.75-1.00
Phosphorus	0.012	0.04 max.	0.04 max.	0.025 max.	0.025 max.
Sulfur	0.108	0.08-0.13	0.08-0.13	0.025 max.	0.025 max.
Silicon	0.05	*	*	0.15-0.35	0.15-0.35
Chromium	0.05	*	*	0.80-1.10	0.40-0.60
Nickel	0.02	*	*	0.25 max.	0.40-0.70
Molybdenum	<0.01	*	*	0.15-0.25	0.20-0.30
Lead	0.2	*	0.15-0.35	*	*
Iron	balance	balance	balance	balance	balance

* - Not listed with specification

Mechanical Properties:

Surface Roughness Measurements

The surface roughness of the component was verified by a Mitutoyo Surftest 401 analyzer. Readings were measured on the exterior surface of the nut in both the longitudinal and transverse directions. Specification MS 17825 requires the nut have a surface finish of 125 µin. or smoother. Each of the ten readings met this requirement, as shown in Table II.

Table II
Surface Roughness Measurements, µin

Transverse	Longitudinal
69	50
74	46
78	54
69	40
54	61
MS 17825	125 µin

Hardness Testing

Hardness testing was performed on a mounted and polished cross-section of the nut. Testing was conducted even though specification MS 17825 does not require hardness nor ultimate tensile strength. The Rockwell "A" scale was utilized, due to the light major load (60 kg) which allowed for more readings on the sectioned surface. A total of ten readings were taken on a longitudinal and transverse section of the nut. The results are listed in Table III.

Table III
Hardness Testing Rockwell "A" Scale
60 kg Major Load

Transverse	Longitudinal
67.4	66.0
67.5	64.5
67.7	64.7
67.8	65.8
67.3	65.2
Average	66.4

Scanning Electron Microscopy and Energy Dispersive Spectroscopy: The fracture surface of the failed nut was characterized utilizing Scanning Electron Microscopy (SEM) and Energy Dispersive Spectroscopy (EDS). The surface, with the exception of origin area and the inclusion that caused the failure, was composed entirely of ductile dimples (D) with elongated stringers characteristic of an overload failure of an extruded part. The fracture surface is shown in Figure 7, note the inclusion band that caused failure denoted by arrows. Figure 8 shows the origin area (arrow denotes origin). It can be observed at higher magnification in Figure 9 and 10. The Mn-S stringer

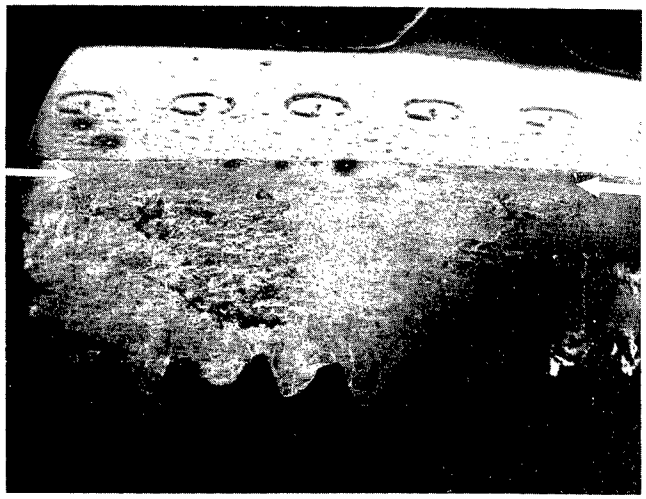


Figure 7. SEM fractograph of the nut, arrows denote line of inclusions. Mag. 10x

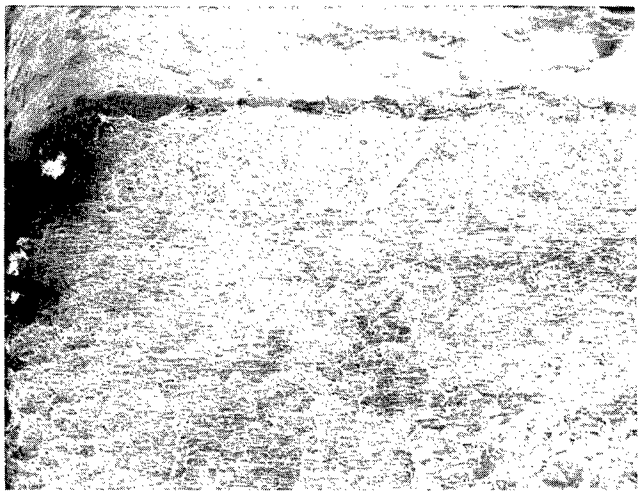


Figure 8. SEM micrograph of the origin area, arrow denotes origin. Mag. 75x

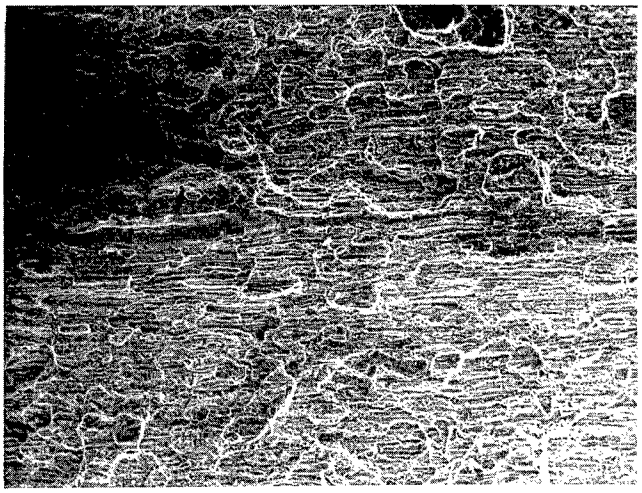


Figure 9. SEM micrograph of the inclusion that caused failure. Mag. 200x

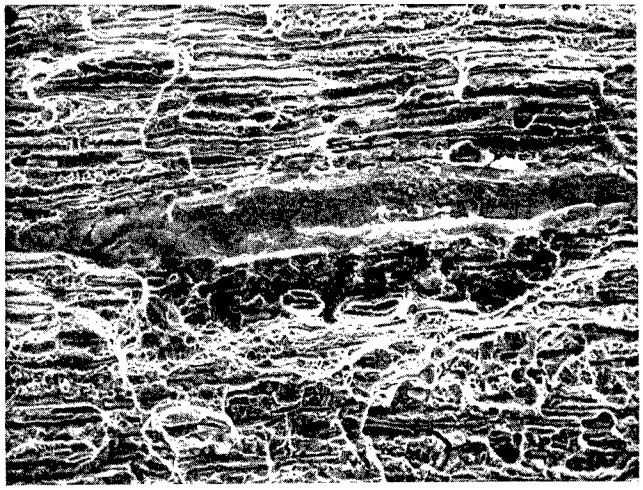


Figure 10. SEM micrograph showing Figure 9 at high magnification. Mag. 500x

that caused this cavity was not present at the origin but the inclusion must have measured at least 0.0015 inches in diameter as is approximately 0.013 inches in depth from the surface. Fracture morphology near the origin was mixed-mode, both transgranular (T) and ductile (D) modes of fracture are shown in Figure 11. Cadmium rich areas were found along the Mn-S stringer that caused failure indicative of a path open to the outside surface may have existed while the nut was in the plating bath. These cadmium deposits are shown in Figure 12 denoted by arrows. Figure 13 depicts the fracture surface morphology of the majority of the nut and the ductile dimples aligned along the extrusion direction. Shear lips (SL) were also noted on the tips of the threads and the castellated end indicating that these areas were the last to fail. The fracture morphology of the shear lips is shown in Figure 14. The inclusions on the fracture surface were identified to be manganese-sulfide through Energy Dispersive Spectroscopy (EDS) utilized in conjunction with the SEM. Inclusions of this severity can be deleterious to the mechanical properties of the material. Inclusions of the magnitude of the one causing this failure may obviously be catastrophic to the component. Figure 15 maps out the entire fracture surface morphology indicating the respective fracture modes. EDS aided in determining the fact that the inclusion band contained cadmium. This provided evidence that the defect may have been open to the surface during the cadmium plating process and occurred as a result of primary processing operations as opposed to a service or installation related anomaly. The use of EDS also helped identify that the nut contained lead as verified by chemical analysis.

Failure Scenario: The MS 17825-10 self-locking nut that secures the lower end of the pitch link failed by overload. This failure was the result of a large inclusion which was located slightly subsurface between beam slots or castellations on the nut. The crack progressed by one or by very few high loading cycles until final fracture occurred. The failure was also assisted by the low degree of "cleanliness" due to the other numerous elongated inclusions within the matrix of the material. The nut failed while applying torque and installation loads due to the existence of the defect. Evidence substantiating this scenario is the fact that the crack origin was determined to be at the base of this defect, the cadmium found within the defect, the mixed mode of fracture at the origin, the overall ductile mode of fracture, and the lack of crack arrest marks emanating outward from the inclusion that caused failure.

Conclusion: The failure of the MS 17825-10 Pitch Link Self-Locking Nut that secures the lower end of pitch link on the helicopter was the result of an inherent material defect. A large (0.0015 inch diameter) subsurface Mn-S inclusion caused the component to fracture during installation. Cracking originated along the base of the subsurface flaw approximately 0.015 inches below the surface. The inclusion may have been exposed to the outside surface of the nut due to the presence of cadmium within the flaw, although a direct path could not be found. The crack progressed by overload along the longitudinal axis of the nut. Crack propagation was aided due to the size and length of this inclusion and the significant amount of other Mn-S stringers within the material.

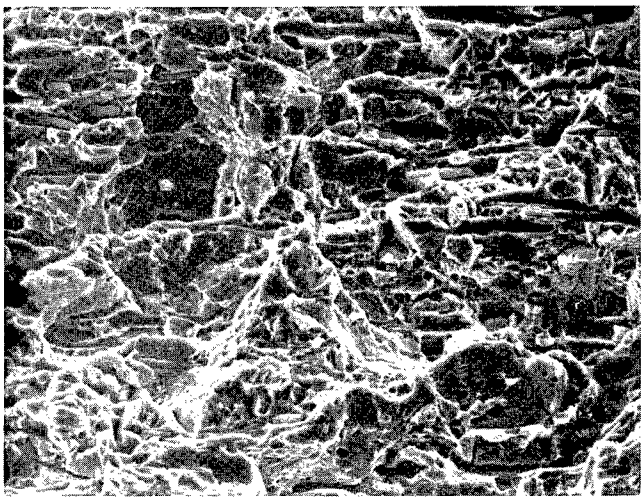


Figure 11. SEM micrograph of transgranular-ductile morphology at origin. Mag. 1000x

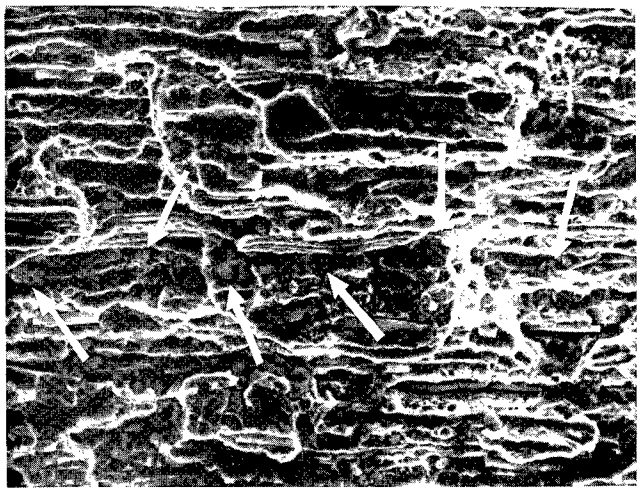


Figure 12. SEM micrograph of cadmium line along inclusion line. Mag. 1000x
Arrows denote existence of cadmium.

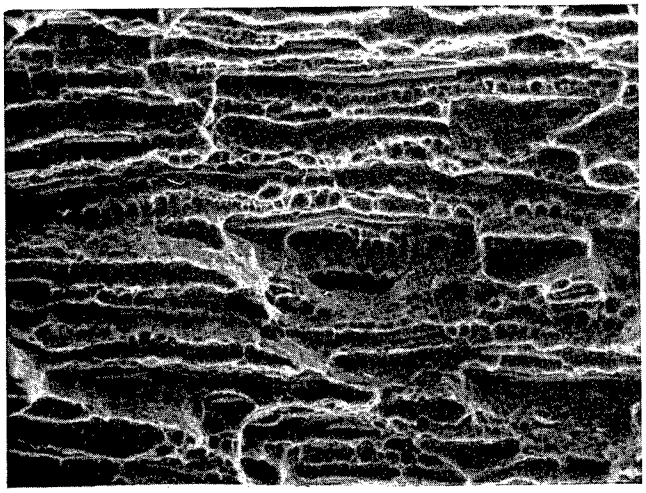


Figure 13. SEM micrograph of the elongated ductile morphology. Mag. 1000x

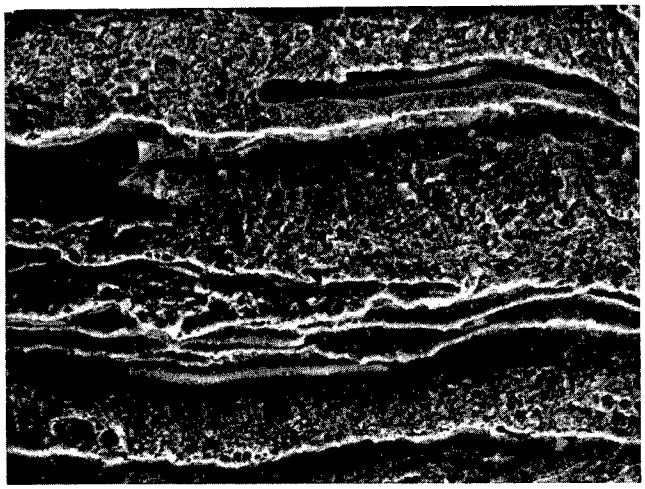


Figure 14. SEM micrograph of the morphology of the shear lip region. Mag. 1000x

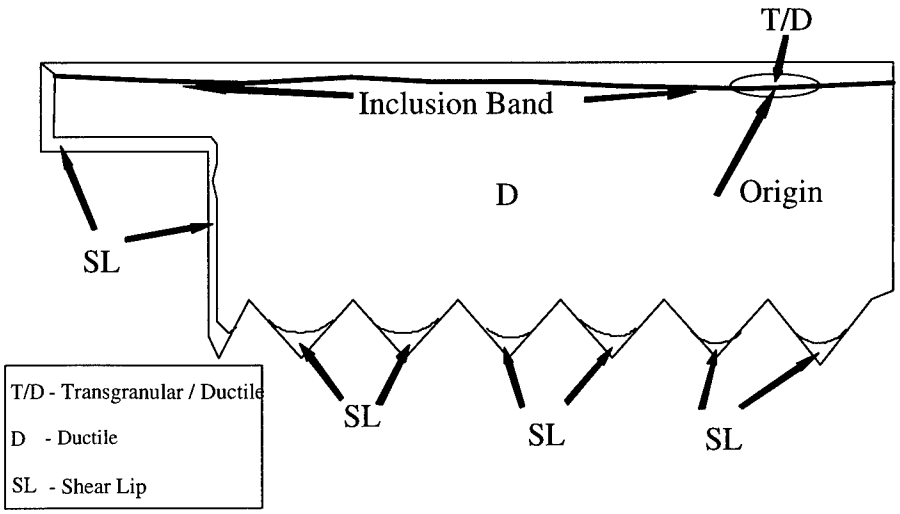


Figure 15. Map of fracature surface morphology for the failed component.

FAILURE ANALYSIS #2

FAILURE ANALYSIS OF MODULAR LIFT LINK, LUG MOUNTING, SELF-LOCKING NUTS

Scott Grendahl
U.S. Army Research Laboratory
Aberdeen Proving Ground, Aberdeen, MD

Abstract: The U.S. Army Research Laboratory-Materials Directorate (USARL-MD) conducted a failure analysis of two self-locking nuts from an army attack helicopter, at the request of the U.S. Army Aviation and Troop Command (ATCOM). Light optical and scanning electron microscopy were utilized to characterize the fracture halves of the broken nuts. The self-locking nuts are governed by Specification MS21042, *Nut, Self-Locking, 450°F, Reduced Hexagon, Reduced Height, Ring Base, Non-Corrosion Resistant Steel.* During phase maintenance on the lift beam in the pylon area, maintenance crews noticed the bolts securing the lift link support lug were not properly seated. Further inspection showed that the left side self-locking nuts were cracked and/or missing. The right side nuts were also cracked.

Key Words: Failure Analysis; Stress Corrosion Cracking (SCC); Hydrogen Assisted Cracking (HAC)

Visual Examination and Light Optical Microscopy: The two failed components were further designated as failures "A" and "B". Nut "A" was sectioned such that the fracture surfaces could be examined. The fracture origins were determined to be at the root of the first thread for both Nuts "A" and "B", and are shown in Figures 1 - 4 (arrows denote origins). Nut "B" contained two separate fracture surfaces, and subsequently both were examined. These two fracture surfaces of Nut "B" can be seen in Figures 3 and 4. They were further designated as fracture "B1" and "B2" for convenience. Note the radial lines which emanate from the fracture origins. The cadmium plating adjacent to the fracture surfaces was worn and damaged but no evidence of significant corrosion was observed. There was also no corrosion visible on the remaining fracture surfaces. A shear lip region was observed extending around one end of Nut "A", however no such failure mode was observed on Nut "B". In addition, beyond the third or fourth threads of both nuts the fracture morphologies revealed the direction of extrusion. This was evidenced by the sets of parallel lines (which represent grain flow) running perpendicular to the threads. The threads of both nuts were slightly damaged in that the tips of the threads were worn away, however this was determined to be post-fracture induced.

Metallographic Examination: The nuts were sectioned such that longitudinal (parallel to the thread direction) and transverse (perpendicular to the thread direction) specimens

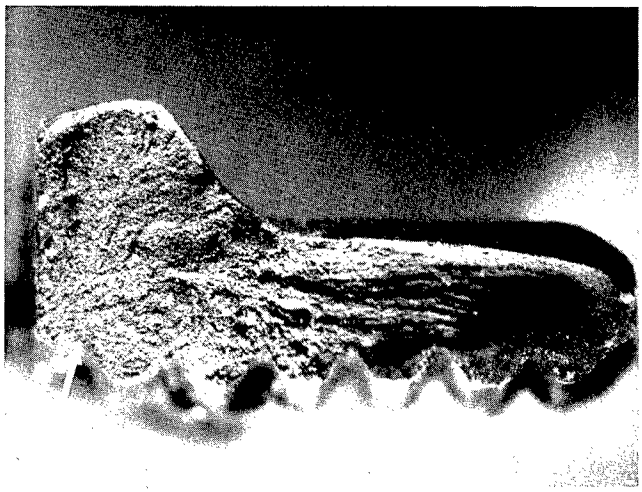


Figure 1. Micrograph of the fracture surface of Nut "A". Mag. 20x

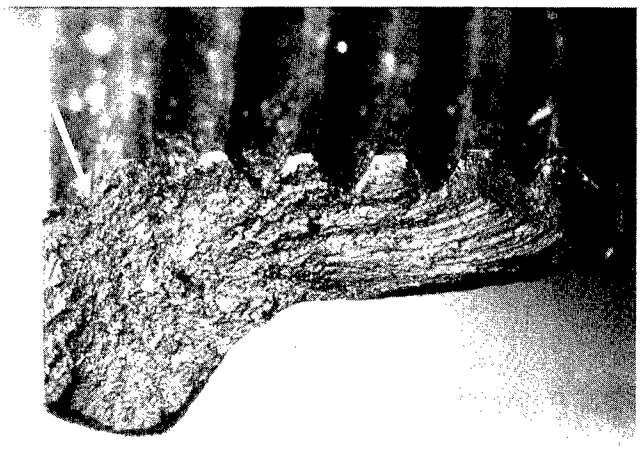


Figure 2. Micrograph of the mating fracture surface of Figure 7. Mag. 20x

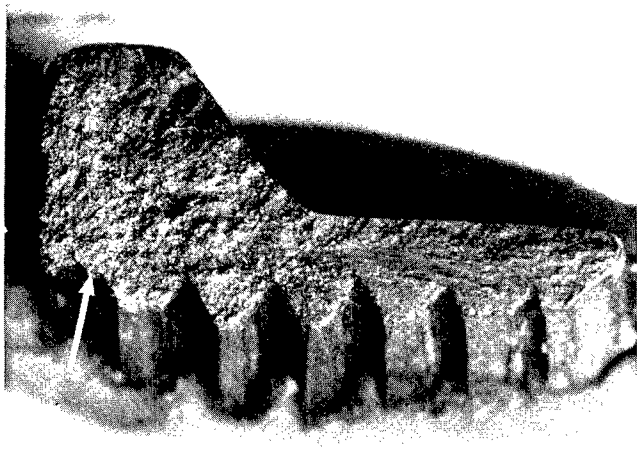


Figure 3. Micrograph of the fracture surface of Nut "B1". Mag. 20x

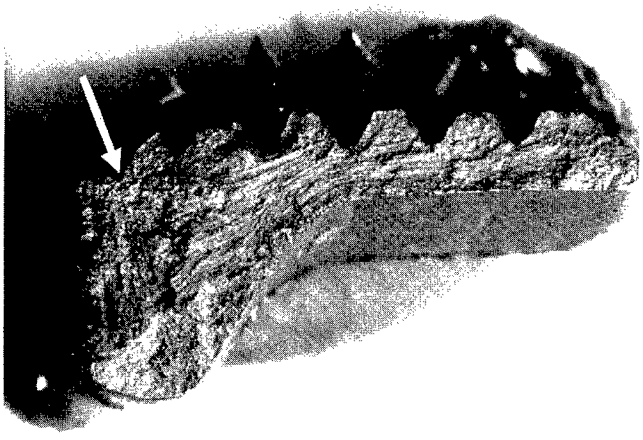


Figure 4. Micrograph of Nut "B" fracture surface "B2". Mag. 20x

were prepared. Metallographic preparation included mounting the samples in phenolic powder, rough polishing utilizing silicon carbide papers ranging from 240 to 600 grit, and fine polishing utilizing 3 micron alumina and 0.025 micron colloidal silica. No internal cracks were noted within the transverse or longitudinal specimens. Internal cracking is sometimes seen in HAC failures when it occurs within the plating bath. However, if the nuts failed primarily as a result of stresses induced during torquing, these internal cracks may be absent. The structures showed very fine martensitic lathe characteristics and were consistent with the prior heat treatment of the nuts.

Hardness Testing: The mounted samples were subjected to Knoop microhardness testing to determine conformance to the guideline of 49 HRC maximum (as required in Specification MS21042). Rockwell "C" scale readings could not be measured directly on the part due to geometric constraints, or on the mounted samples because of their small size. Therefore, Knoop readings were performed and converted to the Rockwell "C" scale. A total of five readings were taken on each sample. A load of 500 gmf was utilized. The results are listed in Table I and Table II. The converted hardness values conformed to the governing specification.

Table I. Nut "A"
HRc Hardness Values (Converted)
Knoop 500 gmf

Longitudinal	Transverse
48.8	49.0
47.8	48.3
48.6	48.7
48.5	48.1
48.2	48.3

Table II. Nut "B"
HRc Hardness Values (Converted)
Knoop 500 gmf

Longitudinal	Transverse
48.9	48.9
47.8	49.0
48.6	49.0
48.7	48.6
48.6	48.7

Chemical Analysis: Chemical analysis was performed to verify the elemental composition of the alloy. Specification MS21042 designates seven different alloys which may be used. The alloys listed were AISI C1035, C1042, 1050, 4027, 4037, 8630 and

8740. The carbon content was determined through combustion-infrared detection, the sulfur by combustion-automatic titration and all other elements through direct current plasma emission spectroscopy. The results compared most favorably with alloy AISI 8740, as shown in Table III. However, the carbon content of Nut "B" and the sulfur content of Nut "A" each slightly exceeded the maximum limits set forth in AISI 8740. This nonconformity suggested a possible lack of melt control. The minute additional amounts of carbon and sulfur would most likely have no adverse affects on mechanical properties, and were considered negligible.

Table III.
Chemical Analysis
Weight Percent

Elem.	Nut A	Nut B	C1035	C1042	1050	4027	4037	8630	8740
C	0.439	0.463	0.31-0.38	0.40-0.47	0.47-0.55	0.25-0.30	0.35-0.40	0.28-0.33	0.38-0.44
Mn	0.84	0.88	0.60-0.90	0.60-0.90	0.60-0.90	0.70-0.90	0.70-0.90	0.70-0.90	0.75-1.00
Si	0.30	0.031	0.15-0.35	0.3 max	0.3 max	0.15-0.35	0.15-0.35	0.15-0.35	0.15-0.35
P	0.018	0.020	0.04 max	0.04 max	0.04 max	0.035max	0.035max	0.025max	0.025max
S	0.026	0.022	0.05 max	0.05 max	0.05 max	0.04 max	0.04 max	0.025max	0.025max
Cr	0.55	0.57	*	*	*	*	*	0.40-0.60	0.40-0.60
Ni	0.46	0.50	*	*	*	*	*	0.40-0.70	0.40-0.70
Mo	0.21	0.21	*	*	*	0.20-0.30	0.20-0.30	0.15-0.25	0.20-0.30
Cu	0.079	0.085	*	*	*	*	*	0.35 max	0.35 max
Fe	rem.	rem.	rem.	rem.	rem.	rem.	rem.	rem.	rem.

Scanning Electron Microscopy and EDS: The fractures of Nut "A" and Nut "B" were classic cases of HAC with service stress related failure. SEM fractographs of Nut "A" and "B" failures are shown in Figures 5, 6 and 7, 8 respectively. The fracture morphologies at the crack origins of Nuts "A" and "B" were intergranular (I) as shown in Figures 9 and 10. These SEM micrographs show the traditional "rock candy" features, as well as some secondary cracking. X-ray mapping (utilizing energy dispersive spectroscopy) of the entire fracture surfaces did not reveal a layer of cadmium on or near the fracture origins. The absence of a cadmium film suggests the cracking most likely occurred in service rather than in the plating bath. The intergranular regions transitioned to large mixed-mode (D/I) regions consisting of intergranular decohesion and ductile dimples covering nearly half of the fracture surfaces. These mixed mode morphologies are shown in Figures 11 and 12. The mixed-mode morphologies transitioned abruptly to ductile dimples (D) as shown in Figures 13 and 14. This abrupt change in morphology represents the areas of fast fracture beyond the third or fourth threads. Figures 13 and 14 also show the ductile dimples aligned in the extrusion direction of the components. Shear lips (SL) were also noted on Nut "A", indicative of the last region to fail. Light optical and scanning electron microscopy did not reveal any gross anomalies at the fracture origin. The schematics shown in Figure 15 and 16 outline the different



Figure 5. SEM micrograph of the fracture surface of Nut "A". Mag. 15x

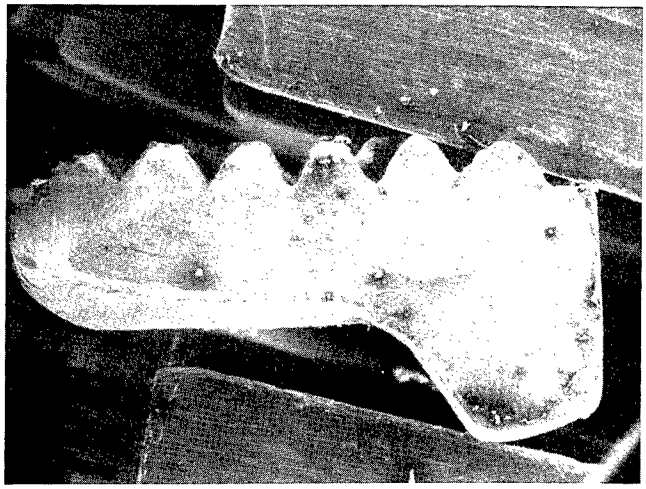


Figure 6. Mating fracture surface of Nut "A" Figure 19. Mag. 15x



Figure 7. SEM micrograph of the fracture surface of specimen "B1". Mag. 15x

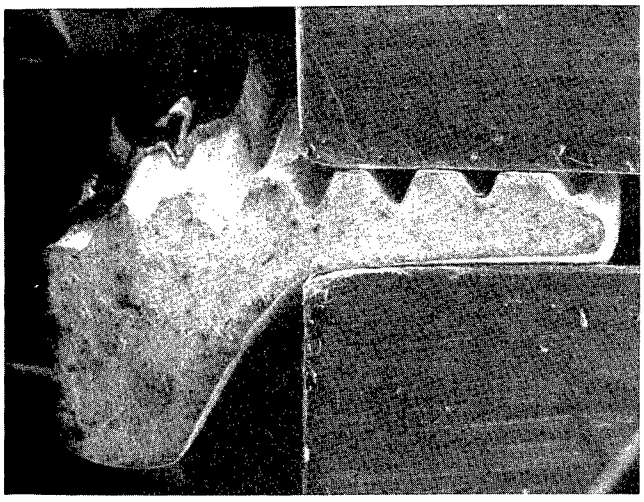


Figure 8. SEM micrograph of the fracture surface of specimen "B2". Mag. 15x

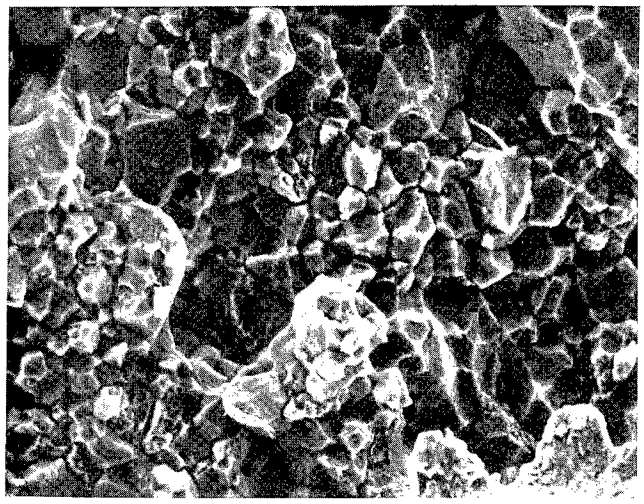


Figure 9. SEM fractograph of morphology at the origin of Nut "A". Mag. 1000x

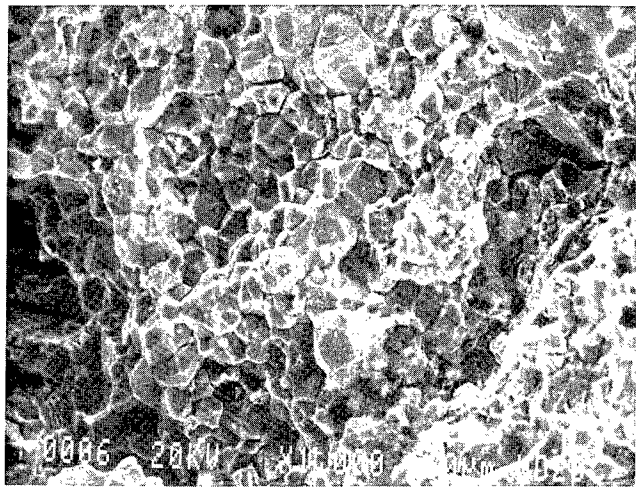


Figure 10. SEM fractograph of morphology at the origin of Nut "B". Mag. 1000x

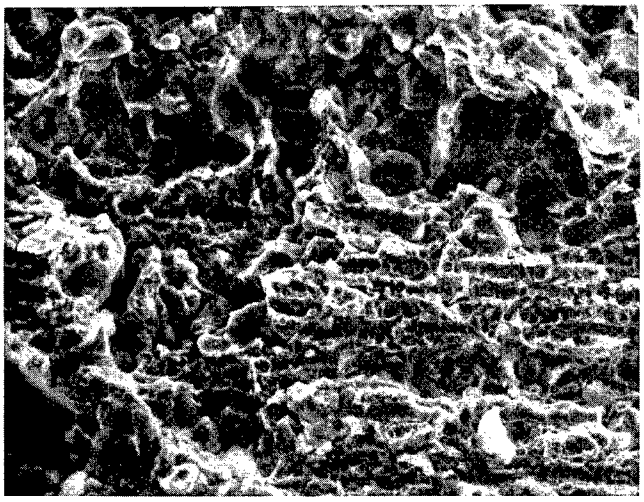


Figure 11. SEM fractograph of mixed-mode morphology of Nut "A". Mag 1000x

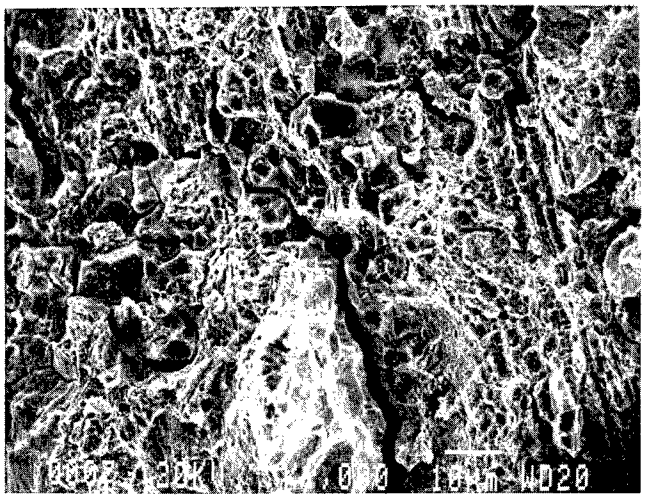


Figure 12. SEM fractograph of mixed-mode morphology of Nut "B". Mag 1000x

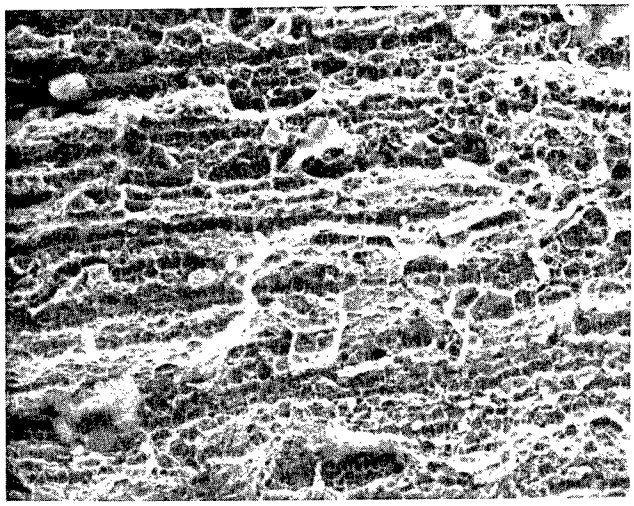


Figure 13. SEM fractograph of ductile dimple morphology of Nut "A". Mag 1000x

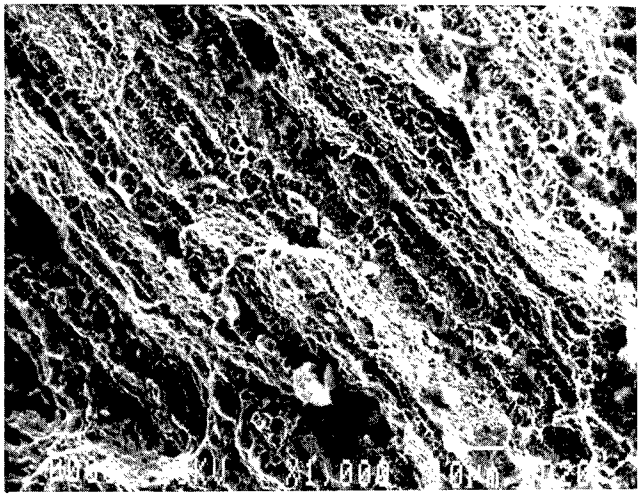


Figure 14. SEM fractograph of ductile dimple morphology of Nut "B". Mag 1000x

morphologies noted on the mating fracture surfaces of Nut "A" and the two different fracture surfaces of Nut "B", respectively.

Discussion: A part subjected to stress corrosion cracking (SCC) exhibits a fracture morphology similar to that noted in this investigation. However, the lack of corrosion and corrosion products on or near the fracture surfaces and the lack of significant crack branching ruled out SCC as a failure mode. Therefore, the failure was attributed to hydrogen assisted cracking. The most likely source of the hydrogen in the fabrication of this component was the cadmium plating process. Specification QQ-P-416 is an electro-deposition plating method, which creates nascent hydrogen that diffuses into the parts during plating. A post-bake is recommended for high strength steels to prevent hydrogen embrittlement. Baking subsequent to plating is required for moderate to high strength steels, usually at 375°F within four hours of plating to evenly disperse the hydrogen which had diffused into the component. Baking time is usually dependent on the strength of the component and the geometry of the part. Problems of embrittlement can arise if the parts are a) plated in an improper bath, b) not post-baked, c) post-baked after a prolonged period of time after the plating process, or d) post-baked at an incorrect temperature or bake time. It is well documented that the absorption of hydrogen into a steel results in a loss of ductility over a sustained stress. However, this could not be substantiated by ARL-MD since tensile specimens could not be fabricated due to the small size of the failed components.

Conclusion: The self-locking nuts under investigation failed as a result of hydrogen assisted cracking (HAC). The primary source of the hydrogen would most likely have been related to a manufacturing operation (electro-deposited cadmium plating) as opposed to the service environment. Evidence substantiating this conclusion was the intergranular mode of fracture and the absence of corrosion on the fracture surfaces. The service life of these components was not known. However, it is possible that the nuts may have been broken for some time before the failures were discovered. ARL-MD has experienced HAC related failures associated with the manufacturer of these nuts and its cadmium plating procedures in the recent past. The nuts were cadmium plated (electro-deposited) in accordance with QQ-P-416, Type II, Class 2 per specification MS21042. These nuts also had a hardness between 48-49 HRc, which converts to a UTS between 238-246 KSI. It is important to note that QQ-P-416 Rev. F does not recommend electro-plating parts with a UTS above 200 KSI, because of the susceptibility of high strength steels to embrittle from nascent hydrogen.

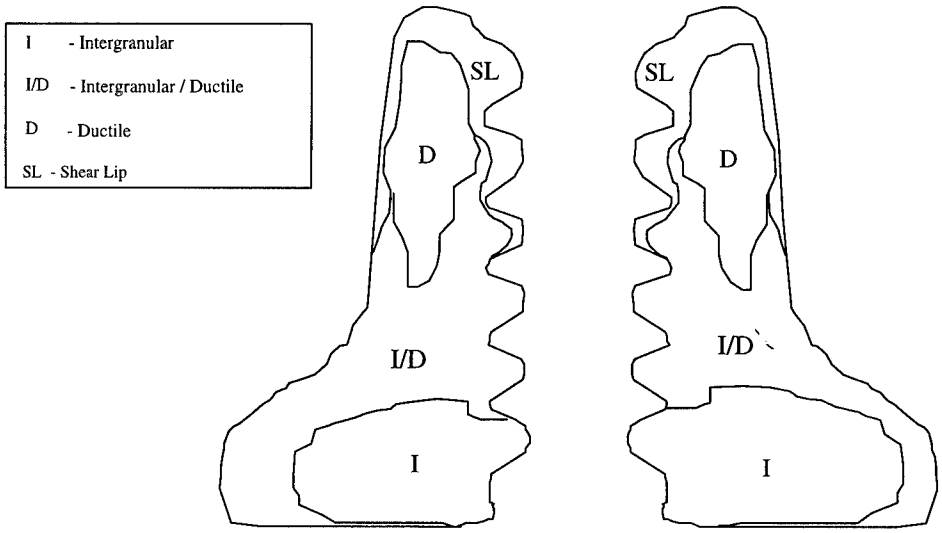


Figure 15. Map of failure modes for mating fracture halves of Nut "A"

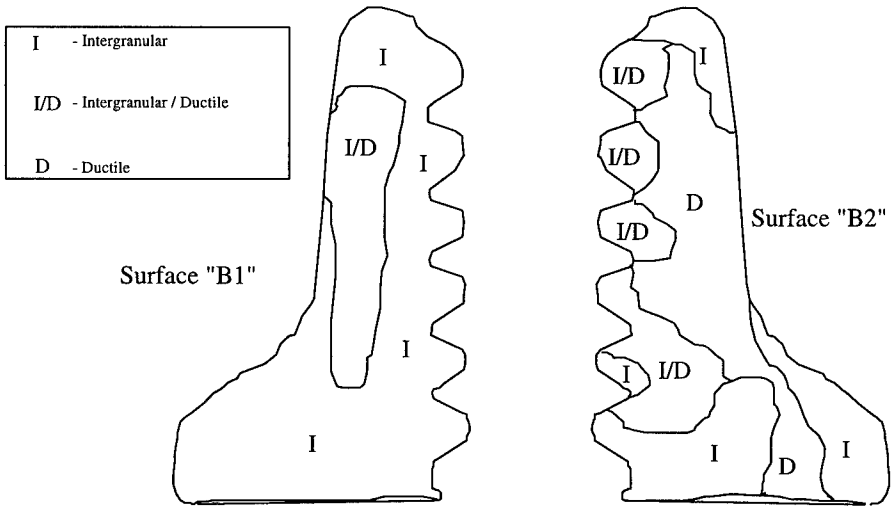


Figure 16. Map of failure modes of fracture surfaces of Nut "B"

MULTISENSOR MONITORING OF GEAR TOOTH FATIGUE FOR PREDICTIVE DIAGNOSTICS

Grant A. Gordon, Amulya K. Garga, and Clark Moose
The Pennsylvania State University
Applied Research Laboratory
PO Box 30
State College, PA 16804 USA
gag100@psu.edu

Abstract: Successful machine diagnostics is critically dependent on the collection and processing of prognostic features that relate back to failure precursors. Since the significance of gear tooth failure is recognized as a critical component to the overall condition of drive train mechanical systems, single gear tooth failure has been examined. By employing a special jig to orient and constrain the gear samples, Hertzian loading was applied along a single contact line on the gear tooth to simulate the conditions seen during operation. Optical, ultrasonic and mechanical sensors measured a variety of observables including load, deflection, and acoustic emission. After monitoring the fatigue test with these three noncommensurate sensors, features of the data could consistently be related to crack growth phenomena. Data collection, analysis, and interpretation are discussed for spur gear samples that show both the absence and presence of cracks and support the validity of the extracted features as failure precursors. Cyclostationary analysis, an advanced signal processing techniques, was used to promote earlier indication and sharper resolution of these measures. The results demonstrate the potential for using nontraditional sensors and techniques, which are more amenable to commercial use, for an in situ monitoring system.

Key Words: Multisensor monitoring; condition based maintenance; cyclostationary; acoustic emission; gear tooth fatigue

INTRODUCTION: The ability to predict the remaining lifetime of a mechanical system is an area that, although is not presently realized, commands considerable attention from both commercial and military organizations. Effective machinery prognostic systems hold the promise of being able to reduce maintenance costs, improve safety margins, increase mission readiness and reduce waste through implementation of retirement for cause maintenance programs. Critical to this objective is the ability to estimate the time required for fault initiation and to track the progression of these failure precursors. Since even modern manufacturing methods can not create truly flawless materials, the notion of flaw initiation is one of semantics. Flaws are incipient within in the material, so effectively our objectives should be to detect, identify and track the progression of flaws with increased accuracy. Therefore, it is generally accepted [e.g. 1] that improvements in the prediction of remaining service life of a component will depend on this ability.

Traditionally the prediction of remaining lifetime has be done on a statistical basis. Large population databases are used to establish relationships between operating parameters and mechanical endurance. Because of their statistical nature however, the results of these studies are not directly applicable to an individual part or component. This fact is generally overlooked or mitigated by ensuring that significant safety 'margins' are applied before the data is employed in a maintenance regiment. This approach results in potentially large amounts of waste, since parts are replaced based on a conservative statistical estimate of lifetime rather than a more direct measure. There is also the potential that a failure will occur before the statistically predicted lifetime, which motivates the need for more accurate condition monitoring from the perspective of safety. In this paper we investigate techniques that can detect and track failure precursors in individual components rather than populations of parts. These features could potentially be used to develop a robust prognostic model that predicts the remaining useful life of the system by employing data fusion and intelligent reasoning techniques [2].

EXPERIMENTAL: As faults grow from micromechanical initiation sites and gain dominance over other potential failure sources, they manifest themselves in a number of ways through out the system. This redundancy of information about the physical failure process suggests that different measures or perspectives can be combined to establish a more robust description of the underlying system. Furthermore, we expect that certain measures, or our ability to probe them, will be more sensitive to early fault detection while other will display attributes such as improved source identification.

In this paper we employed three sensing methods: acoustic emission, mechanical deflection and load measurements, and optical deflection measurements, to monitor the fatigue crack growth in a set of typical aircraft drive train gears. The gears were 32-tooth spur gears made from commercially accepted hardened steel. After the gears were modified by cutting alternating

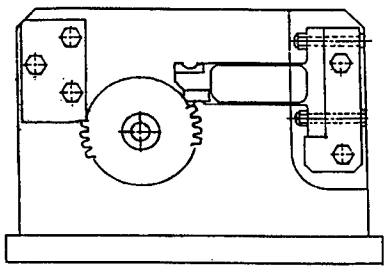


Figure 1. Gear fatigue test fixture

groups of four teeth from the periphery of the gear, they were placed in a test fixture (Figure 1) based on a Boeing gear fatigue apparatus [3]. This test fixture constrains the tooth loading along a line of contact at the highest point of single tooth contact. A Materials Testing System (MTS) cyclically drove the load frame ram onto the compliant armature of the fixture in a stress controlled manner. After an initial ramp up to the maximum load condition, the ram cycled at a rate of 20 Hz, Figure 2, between a minimum compressive load of 900 lb. and a maximum compressive load of 10,000 lb. Load and position data were collected from the MTS sensor head at rate of 100 samples per sec until the axial position of the loading head exceeded a preset deflection limit. This limit, which established the end of the testing sequence, was set at an 0.002 in. increase beyond the initial maximum deflection established at the onset of loading. Commensurate with this data collection, acoustic emission and optical data were collected with sensors that were triggered and synchronized with the MTS loading frame.

OPTICAL MEASUREMENTS: The optical setup used to measure deflection, was based on a knife edge detection arrangement, Figure 3. To accommodate optical beam access into the crowded environment we mounted the mirror as close to the loaded gear tooth as possible, on the loading jigs fixture arm. Any movement experienced by the loaded tooth results in deflections of the mirror since the load arm insert and the gear tooth was in intimate contact throughout the test. The deflection of the laser beam by the reflecting surface of the mirror varied the amount of light that passed by the knife edge and was focused on the detector by collecting optics. As the mirror tilts, the optical beam path deflects by d at the knife edge according to the expression

$$d = l \tan(2\varphi) \tag{1}$$

l is the distance to the knife edge from the vibrating mirror, and ϕ is the rotation of the sensing mirror. By positioning the knife edge to intercept a portion of the beam, the amount of light

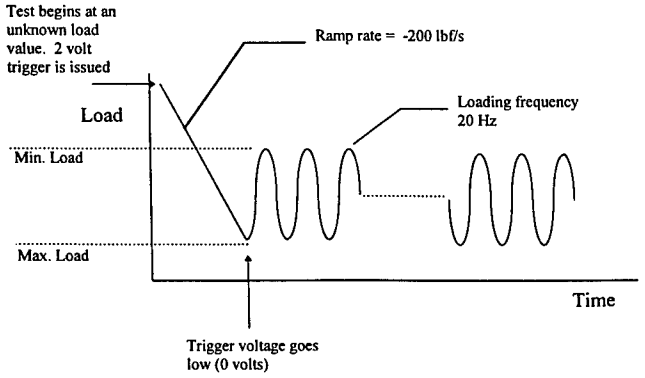


Figure 2. MTS loading profile

focused onto the detector will be modulated by the motion of the sensing mirror. For an collimated laser beam, operating in the TEM_{00} mode, the spatial variation of intensity has a radial Gaussian distribution given by

$$I(r)=\frac{2P_t}{\pi\omega^2}\exp(-\frac{2r^2}{\omega^2}) \tag{2}$$

where r is the distance from the beam center, w is the halfwidth of the beam (the points where the intensity drops to e^{-2} times that of the center) and P_t is the total power in the beam [3]. The voltage output of the photo diode will be proportional to the light power imaged onto the detector, defined by the unmasked portion of the beam that falls on the detector

$$P=\frac{2P_t}{\pi\omega^2}\iint\exp(-\frac{2r^2}{\omega^2})rdrd\theta \tag{3}$$

with the limits of integration given as function of d.

Voltage data was collected at a rate of 500 samples per second and stored on a personal computer using PC based software and a 12 bit A/D card. Since the deflection due to crack growth/opening were seen to be much smaller than the elastic deflection of the gear tooth caused by cyclical loading, it was imperative that the entire dynamic range of A/D be utilized. Alignment of the optical beam was performed so that the laser beam was captured by the collecting optics and focused on to the active area of the photodiode throughout the entire displacement cycle. Next, the intensity of light was adjusted using a variable attenuator so that

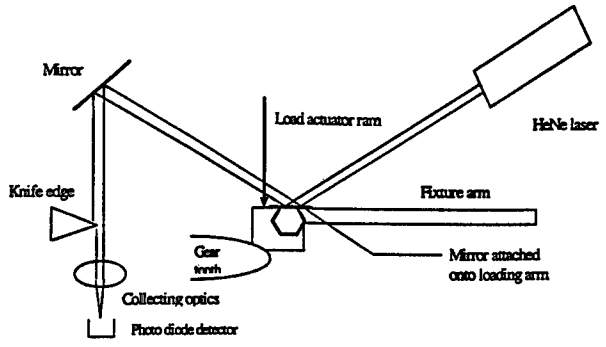


Figure 3. Knife edge detection set up used for optical detection of gear tooth deflection

the maximum light intensity did not saturate the photodiode device. After the fatigue loading ended, a static calibration curve was generated to relate deflection to voltage output by driving the ram into a known deflection position and recording the output voltage. Unfortunately the dynamic loading conditions produced deflections that differed substantially from the statically loaded deflections.

ACOUSTIC EMISSION MEASUREMENTS: Acoustic Emission (AE) is an elastic stress wave generated by the rapid release of energy within a material. Plastic deformation, crack initiation, and crack growth from fatigue or corrosion produce acoustic emission signals. The energy that is converted to AE signals comes from the material itself. As a result, this nondestructive technique is sensitive to initiation and propagation of defects. Several parameters are necessary to define an acoustic emission event or hit. Figure 4 illustrates the primary terms used to describe the signal from one AE event. The AE signal is monitored as the voltage response as a function of time, V(t), from the transducer. The noise level refers to the background noise. Successful AE testing relies on the ability to detect the signals of interest above the background noise. Background noise problems range from negligible to severe, depending on the test conditions. For fatigue testing, mechanical noise sources include: test machine noise, flow noise from pumps and valves, and all types of frictional processes (e.g., guide fixtures, contact points, and pivot points). One of primary ways of excluding background noise is the use of a threshold. When the amplitude of an AE signal exceeds this voltage

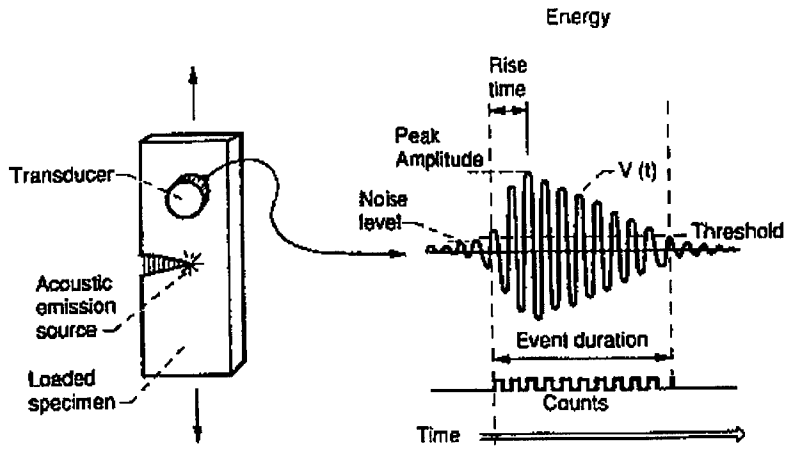


Figure 4. Definition of acoustic emission event parameters

threshold, an event or 'hit' is recognized and the event parameters are recorded. These specific signal parameters include: the peak (maximum) amplitude, the rise time, the signal duration, the signal counts, and the signal energy. The peak amplitude represents a measure of the signal strength and therefore governs the detectability of the event. The rise time is the time from the first threshold crossing to the signal peak. This parameter is useful for source discrimination and signal filtering. Duration is defined as the time from the first threshold crossing to the last and may be used for source discrimination. Counts (threshold crossing counts) are the number of times the AE signal crosses the threshold and gives a simple measure of the signal size. The signal energy is the calculated area under the rectified signal envelope and provides a combined view of peak amplitude and signal duration.

Acoustic emission data was collected during each fatigue test using a LOCAN-AT acoustic emission unit made by Physical Acoustics, Inc. Each gear was instrumented with a broadband (100-1000 kHz) acoustic emission transducer. The transducer was cemented on a flat region of the gear just above the tooth being loaded. The transducer position was selected for protection of the transducer, location near the tooth being tested, and coupling of the transducer to a flat region of the gear. Calibration of the transducer response in this orientation was performed using the standard mechanical pencil lead break test on the test tooth. This calibration test produced an 80-84 dB amplitude response indicating strong coupling of the transducer to the gear and high sensitivity of any acoustic emission events from the tooth region. The acoustic emission (AE) system was synchronized with the start of the cyclic loading via a voltage trigger from the MTS control panel. Acoustic emission data was collected on a hit-driven basis, meaning that data was only collected when a signal threshold value was reached.

DATA ANALYSIS AND RESULTS: Figure 5 shows macrographs of two gears visually inspected after the cyclically fatigue testing was complete. Both gears were driven in the same manner and reached the same level of deflection before the loading cycle was halted. Notice that the gear on the right, made from VASCO-X2, does not show any significant cracking. In contrast, the gear on the left, made from AMS-6265, which shows a severe, 0.7 cm long, crack starting from the tooth root. The data sets from these two gear teeth were used in the ensuing data analysis.

In order to provide features that might be useful to a prognostic model, our approach was to examine changes within the data with respect to a dependent variable that relates back to the fatigue loading condition. A convenient and effective observable which provides a measure of the load experienced by the tooth was found to be given by

$$Accumulated\ load = \int_{0}^{cycles} PdN \tag{4}$$

For the sinusoidal driving function used in our experiment this parameter simplifies to the product of mean load multiplied by the number of cycles, $P_{mean}N$. The deflection data collected from the MTS sensor was processed by using a moving average filter to remove the 20 Hz cyclical variation that arose due to the elastic deflection response of the gear tooth to the cyclical

loading. This processed result was then subtracted from a base line average value established from the initial portion of the data record. This result provided a differential moving average deflection value as a function of accumulated load. The most revealing feature obtained from the acoustic emission data was the relationship between accumulated acoustic emission energy and the accumulated load parameter. This relationship is physically founded since we expected the strain energy released during crack tip advances to generate significant acoustic emission (AE). Unfortunately it was difficult to discriminate these AE events from the high background noise. The optical data was initially processes in a manner similar to the MTS deflection data. A high pass filter was applied to remove the 20 Hz response of the dynamic loading and the resulting signal was squared to provide a mean squared voltage value plotted against accumulated load. The results of these analysies for ther two different set of gear data are compared in Figures 6 - 8.

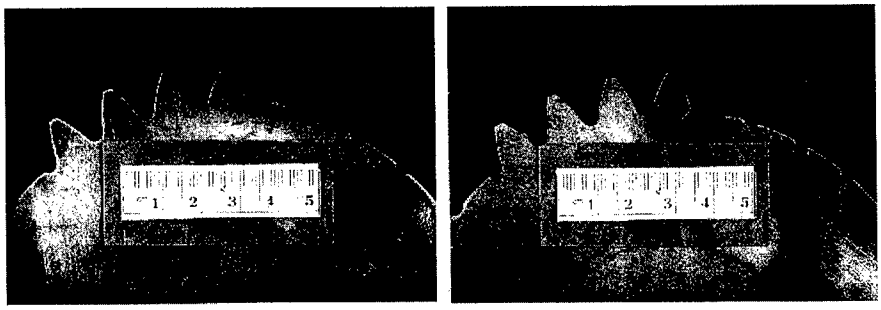


Figure 5. Macrographs of gear teeth after fatigue loading. The right image shows a gear made from AMS-6265, while the left image shows a gear made from VASCO-X2

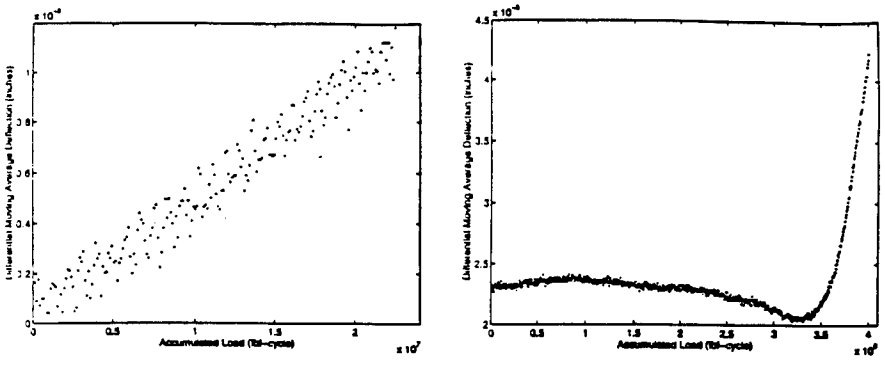


Figure 6. Comparison of differential moving average deflection calculated from AMS-6265 (right side) and VASCO-X2 (left side) gear data

In each of the respective figures, the analyzed data from the cracked, AMS-6265, is shown on the right hand side, while the left hand side shows the computed results from the VASCO-X2 gear data. Notice that the results for the AMS-6265 data all consistently indicate a dramatic deviation, from an initial trend, at an accumulated load value of ~ 3.4 lbf-cycle. The uncracked gear, VASCO-X2 does not indicate this behaviour. In the case of the AMS-6265 gear we believe that these departures from initial trend values occured when the micromechanical mechanism of gear fatigue transitioned from crack initiation to crack growth/propagation. Closer inspection of the results for the AMS-6265 gear showed that the high rate of AE energy release predated the

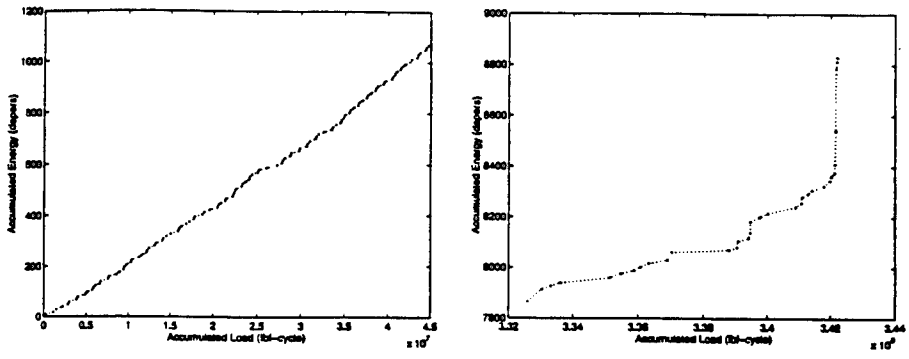


Figure 7. Comparison of accumulated acoustic emission energy calculated from AMS-6265 (right side) and VASCO-X2 (left side) gear data

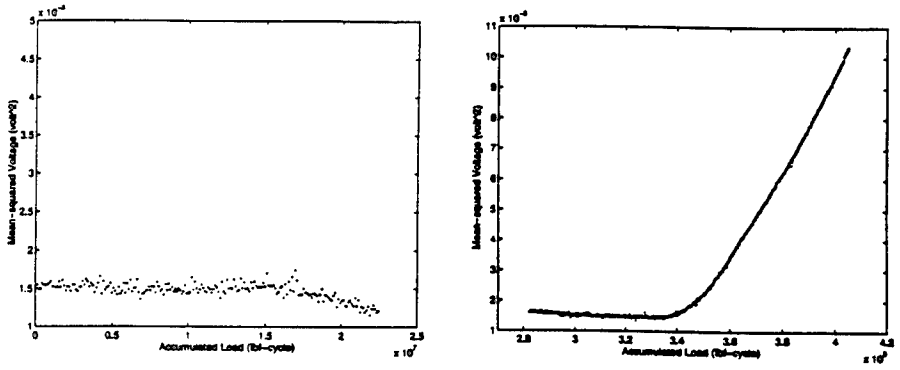


Figure 8. Comparison of mean squared voltage values calculated from photodiode response of AMS-6265 (right side) and VASCO-X2 (left side) gear data

moving average deflection feature by ~ 2500 cycles, or 3 % of the total number of cycles. The results from the optical data analysis were also consisitent which the moving average deflection feature. However the high sampling rate of the optical data provided an opportunity to investigate more sophisticated analysis proceedures.

SIGNAL PROCESSING BY CYCLCOSTATIONARITY: The periodic nature of the loading suggested application of cyclostationary signal processing techniques for studying the frequency spectrum of the observed deflection signal. These techniques have been successfully applied to various problems in communication systems. We selected the optical photodiode voltage measurements for our analysis due to the higher sampling rate used, which allows higher harmonics to be studied. Cyclostationary processing should provide a better estimate of the harmonic content of the observed signal which gives an indication of abnormal mechanical activity. In this section we present some *preliminary results* of cyclostationary processing of the observed signal.
We use the definition in [5] of the cyclic spectral-correlation density (SCD) of a signal $x(t)$ is defined as

$$S_x^{\alpha}(f) = \lim_{B\to 0}\frac{1}{B}\left\langle [h_B^f(t) * x(t)\exp(j\pi\alpha t)][h_B^f(t) * x(t)\exp(-j\pi\alpha t)]^* \right\rangle \qquad (5)$$

where $h_B^f(t)$ is a one-sided narrowband filter with center frequency f and bandwidth B, * denotes

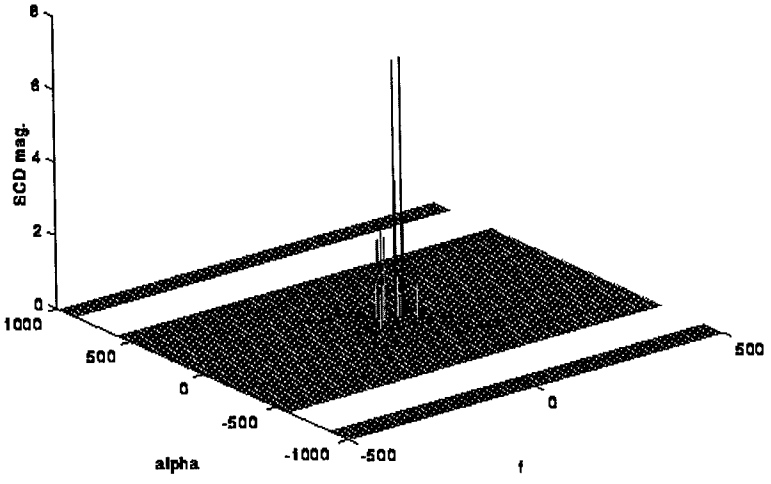


Figure 9. SCD magnitude of the photodiode voltage signal prior to crack propagation (70000-75000 cycles)

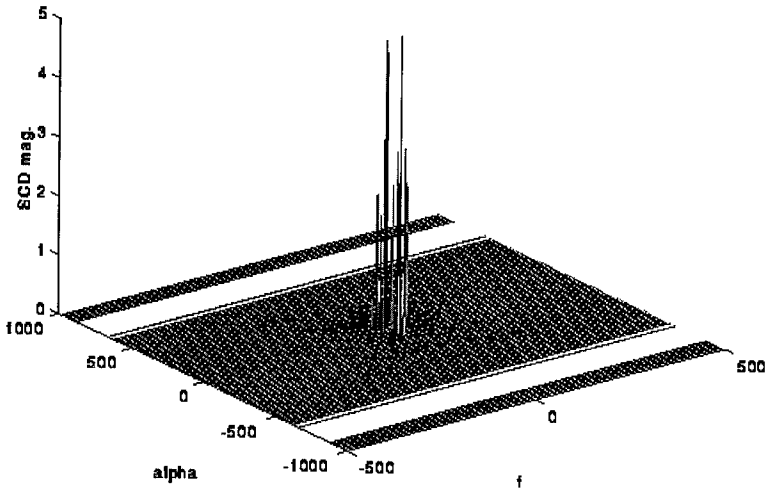


Figure 10. SCD magnitude of the photodiode voltage signal during crack propagation (75000-84000 cycles)

the convolution operator, the superscript * indicates the conjugation operation, and $\langle\rangle$ indicates the time-average operation. The above operation is performed for various values of α to obtain the entire SCD. We have shown in Figures 9 and 10 the magnitude of the SCD of the laser photodiode voltage signal during the last part of the run which resulted in a cracked tooth. The two line spectra clearly indicate a change in the nature of the signal. In particular, there are more spectral lines in Figure 10 than in Figure 9 but magnitude of the lines in Figure 10 is smaller than the lines Figure 9. This indicates that the fatigue in the tooth is observable in the spectral content of the deflection signal. This fact is corroborated by the presence of a crack during the visual inspection of the gear after the completion of the test and by other signal processing analysis described above.

CONCLUSIONS: We have described an experiment where multisensor data was collected during gear tooth fatigue studies. Although the number of data sets was not large, our analysis extracted several features which could be useful in a failure prediction model. These features were shown to be in agreement with the observed fatigue induced crack damage. We have also employed new techniques specifically designed for cyclostationary systems, such as rotating machinery. Having shown the viability of this approach, we believe that further investigation into to determine whether this approach can be utilized for predicting remaining useful life of machinery is warranted.

ACKNOWLEDGMENTS: The authors would like to thank Guillermo Astete for assisting in data collection and Dr. Steve Schell of Penn State's Electrical Engineering Department for providing the Matlab routines for the cyclic spectral analysis discussed in this paper.

REFERENCES:

[1] Barkov, A. and Barkova, N. "Condition assessment and life prediction of rolling element bearings - Part1," *Sound and Vibration*, June 1995.

[2] Hansen, R.J., Hall, D.L., and Kurtz, S.K. "A new approach to the challenge of machinery prognostics," *Transactions of the ASME*, **117**, pp. 320-325, April 1995.

[3] Lemanski, A.J., and Rose, H.J. "Evaluation of advanced gear materials for gear boxes and transmission," Boeing Technical Report, No. D210-10345-1, 1971.

[4] Collier, R.J., Burckhardt, and Lin, L.H. Optical Holography, Academic Press, Inc., NY, NY, 1971.

[5] Scott, I.G. Basic Acoustic Emission, Gordon and Breach Science Publishers, NY, NY, 1991.

[6] Gardner, Exploitation of spectral redundancy in cyclostationary signals, *IEEE Signal Processing Magazine*, pp. 14-36, April 1991.

PREVENTION OF CRANKCASE EXPLOSIONS IN RECIPROCATING COMPRESSORS

V. A. Cox

Shell Research Ltd
Thornton Research Centre
P.O. Box 1
Chester CH1 3SH
United Kingdom

Abstract: Three ignitions occurred in quick succession in the crankcase of a large, reciprocating, hydrocarbon gas compressor. Protective systems limited the effects and led to safe system shutdown. An investigation team found the ignition source to be an electrical discharge and resultant arcing, between the piston rod oil scraper rings and their housing. The insulating properties of the pad fitted between the non-drive end bearing and earth were found to deteriorate with time. A simple monitoring system was fitted to indicate the duration of the pad's effective life.

Experimental work found that the double tangential rings used for low-pressure sealing were ineffective at pressures less than 0.5 bar, although they sealed well at higher pressures. These rings could be energised by the application of nitrogen buffer gas. In evaluating a proposal to replace these rings with side-loaded pressure rings to API 618, it was found that certain manufacturing quality and housing design features could reduce their sealing effectiveness to less than that of the original type. A manufacturing quality plan agreed with the manufacturer has overcome these problems.

Modifications incorporating these findings have been applied to the compressor, and to three others, without any recurrence of the problem.

Key Words: API SLPR; arcing; hydrocarbon gas; ignition; LP seals; nitrogen buffer gas; surface finish.

1. INTRODUCTION:

Four identical reciprocating gas compressors are utilised in a similar, natural gas duty. The machines have six, two-stage cylinders and are electrically-driven, consuming almost 10 MW. Three ignitions occurred in quick succession in the crankcase of one of these machines. The immediate extent of damage was similar in all three events; crankcase pressure relief valves had burst, emission of smoke and a minor emission of noise. Protective systems, including fire and gas detection and module UV detectors, limited the effects to the immediate surroundings of the machine and led to its safe shutdown. No personnel were injured and no equipment was damaged.

An investigation team studied the machine component parts, ancillary equipment, and three sister machines, to discover both the source of the ignition and the reasons for the presence of an explosive mixture in the crankcase. Ultimately, a research project that studied the performance of a range of low-pressure axial sealing rings was instituted. The total duration of the investigation was more than two years.

The original layout of sealing and venting components of the four compressors is shown in Figure 1.

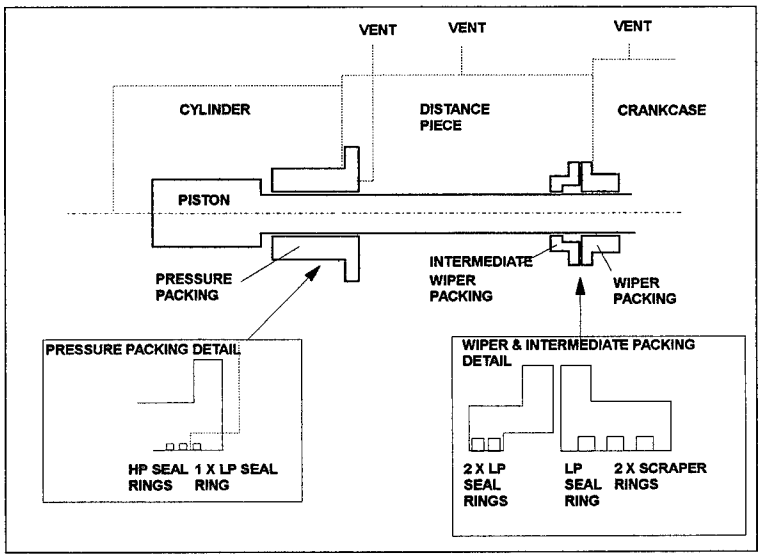


Figure 1. Arrangement of sealing components

2. MACHINE COMPONENT STUDIES:

In the initial, post incident study, the compressor was systematically dismantled in order to ascertain both the causes of gas ingress to the crankcase and the ignition source. No abnormal markings were found on any main or connecting-rod big-end bearings, crosshead pin, slipper bearings or piston rods. No defect was found in the drive motor or other electrical systems. All pressure packings were dismantled and examined in detail. Their appearance was normal. The piston rings and rider bands from all cylinders were also of normal appearance.

Examination of the elements from the piston rod wiper packings found unusual markings on those taken from one cylinder but not on the others. Damage observed on the mating faces of the oil scraper rings is shown in Figure 2. No damage was seen on the faces adjacent to the containers, nor on the containers themselves. Metallographic examination of a cold-mounted section of a typical damaged wiper element showed shallow areas of microstructural transformation typical of the localised pitting that results from electrical discharge. Such transformation indicated that the local temperature had been raised to at least 240°C. A

micrograph of this area is shown in Figure 3. Confirmation that this microstructure was typical of electrical arcing was obtained by comparison with laboratory-induced arcing on an undamaged ring.

Identical damage was later identified on wiper packing elements of two sister reciprocating compressors.

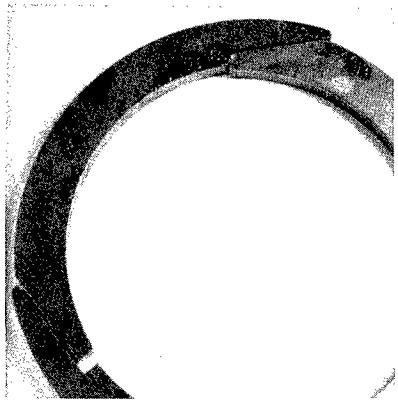


Figure 2. Appearance of damaged rings

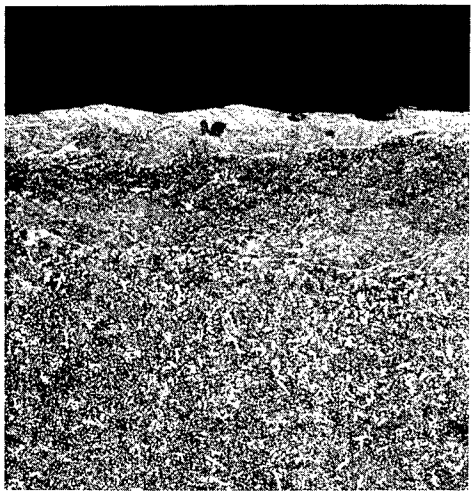


Figure 3. Microstructure of a damaged area.

3. POSSIBLE IGNITION SOURCES:

Five possible sources of ignition were considered during the investigation. No other sources are considered to be possible. No evidence was found for: a hot-spot within the rotating or reciprocating components of the compressor; discharge of static electricity generated by the passage of oil through the filters; discharge of static electricity generated by oil shear between piston rods and wiper packings; and pyrophoric ignition. They are mentioned here solely for completeness.

3.1. An electrical defect within the machine.

Examination at the time of the incident found that the insulating pad beneath the non-drive end bearing of the drive motor may have been bridged, due to the attachment of protective guards and bearing temperature measurement probes. All potential conductors were removed immediately from all four machines. Further monitoring after this time showed that arcing damage continued to occur. It was determined to examine the insulating properties of the pad beneath the non-drive end bearing.

A simple technique was evolved using a volt meter to indicate the potential difference between the bearing pedestal and frame of the machine. From an initial value of about 300 mV for a new pad with fully reinstated and cleaned metal components, a steady deterioration was found to occur until values as low as 10 mV were recorded. A strong relationship was found between the severity of wiper packing arcing damage and low measured voltages. Investigation showed the two main contributory causes of this deterioration to be shorting of the insulation by gradual leakage of carbon and dirt-impregnated lubricating oil, and filling of the insulated bolt/dowel holes by salt from seawater deluges or washdown.

Changing the insulation type was found to be ineffective in overcoming the problem. The solution was provided by a combination of factors:

- varnishing and proper installation of insulation
- sealing of dowel/bolt holes with silicone mastic during installation
- attention to housekeeping in the pedestal vicinity, for example oil spills, etc.
- cleaning of insulating pad when measured voltage fell below 150 mV.

3.2. Mechanism of arcing

The following theory is advanced in explanation of all of these observations. During normal operation the crankshaft is supported by a hydrodynamic film of oil at the main bearings. This film may be considered to be fully insulating. The drive motor has an insulated non-drive-end bearing, the drive-end radial and axial loads being carried within the compressor. Thus the motor rotor and crankshaft are isolated from the frame of the machine in all respects other than the connecting rods.

Under ideal conditions contact between the big-end bearing and its journal is also inhibited by a hydrodynamic oil film. However, previous work carried out these machines suggested that oil film thickness might be marginal at the TDC and BDC positions. It is known that a high voltage can pass across a bearing oil film when very limited asperity contact is achieved.

Stresses at the cross-head bearing normally exceed those at the big-end bearing, loads being similar but the projected area being considerably less. Thus a low oil-film thickness at the big-end is likely to be duplicated at the cross-head bearing. Electrical contact would then be possible between the crankshaft and piston rod at bottom-dead-centre and to a lesser extent at top-dead-centre also. The cross-head slippers are relatively lightly-loaded and flooded with lubricant at crankcase temperature. A good oil film may be assumed between them and their guides at all times.

At the two dead-centres the piston rod travel reverses, rings in the wiper and pressure packing shuttling from one side of their containers to the other. Work carried out in compressor research at Thornton has shown a very strong surface tension or 'stiction' effect to exist between lubricated rings and container walls. Shuttling of wiper rings may therefore not be instantaneous, but will occur after a short delay determined, for each ring, by the relative proportions of surface tension in opposition to frictional drag from the piston rod. The face of the ring that is nearest to the crankshaft (inboard) in each set is smooth, the others in the set being grooved for oil scraping and having a smaller contact area and, therefore, surface tension force. Thus there is a tendency for the further outboard ring in each set to shuttle a

short time before the inboard one. At this instant the inboard ring is in close contact with the container but the rod, flooded with oil, is moving under it. However the outboard ring is in close contact with the rod under the influence of the garter spring. Electrical potentials between the rings change from both at earth to one at earth, one at rod, then revert to both at earth for the remainder of the stroke. Any current flow from the rod to earth undergoes a rapid 'break and make' at the ring interfaces, causing sparking, high temperatures and metal transfer.

4. EXPLOSIVITY OF CRANKCASE ATMOSPHERE

The investigation considered three possible causes of an explosion in a crankcase, given an ignition source. No evidence was found for either oil cracking, due to high temperature or shear rates, or oil mist ignition, but they are included here for completeness.

4.1. Gas ingress

The presence of gas in the crankcase was known from routine flash-point measurements, which had shown gradual deterioration over time since commissioning. This was considered to be the most likely cause of the ignitions.

The proportions of liquid and vapour phases in the crankcase were calculated at various levels of gas contamination. It was found that for mixtures of the hydrocarbon gas in service and the crankcase lubricant, the vapour phase and input gas had a straight line relationship. Very small amounts by weight of the gas entering the crankcase dissolved in the oil, the remainder occupying the void or leaving by the vent. At the minimum allowable flash point, 0.020%w of methane, equivalent to 336 litres of gas at NTP, would be contained in the full sump capacity of 1200 litres. In view of the possible flow rate of gas into the crankcase, it must be assumed that the space above the lubricant contained a rich mixture of gas and air. Severe reductions in the flash temperature of the crankcase oil are an indication of high levels of gas. However, this work showed that measurement of oil flash point gives only minimal guidance to the explosivity of the vapour in the crankcase.

These results were used to establish safe operating standards on input of explosimeter test results of the crankcase vapour, which was commenced immediately.

4.3.1. Gas transport mechanisms

The source of gas within the distance piece and crankcase was considered to be inefficiency of the low-pressure (LP) axial seals. All four LP seals fitted in the locations shown in Figure 1 were of the type shown in Figure 4(a).

The double tangential seal is designed to operate with a small axial clearance, which allows the ring to reciprocate slightly with rod motion. The volume displaced by a ring on each stroke equals the product of the ring's cross sectional area and the axial clearance, 1244 mm^3. Provided that a differential pressure exists across the seal, gas will pass it solely due to the effects of reciprocating motion. Leakage of gas to the crankcase by this mechanism could total 134 litres/hour when all six cylinders are taken into consideration.

Small pressures, typically between .15 and .4 bar.g., had been recorded in both the pressure packing and distance piece vent lines, due to backpressure resulting from long piping runs. The resulting pressure gradient between pressure packing and crankcase enabled a significant flow of gas into the crankcase, where it mixed with air already present. This is the explanation for the presence of an explosive mixture in the crankcase.

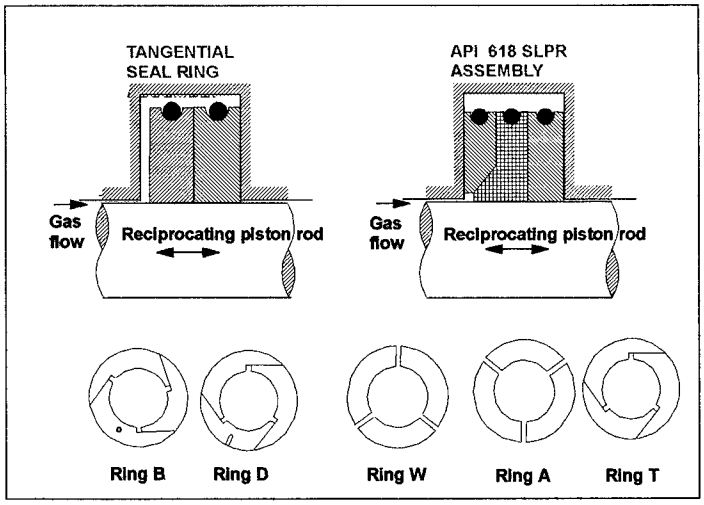


Figure 4. Low-pressure axial seal types, (a) double tangential, (b) SLPR

5. MEASURES TAKEN TO PREVENT FURTHER IGNITIONS.

5.1. Sealing ring modifications

A test rig was constructed for the purpose of measuring typical leakage of gases. The rig, shown in Figure 5, comprised two seal containers mounted back-to-back, with a gas entry port between them. Leakage past the seals was directed to passages equipped with Rotameter flow measurement meters. Elastomeric seals fitted outboard of the containers provided leak-free sealing and were replaced after each test. The piston rod used for this work was of 50 mm diameter and its stroke rate was 300 per minute, identical with that of the compressors.

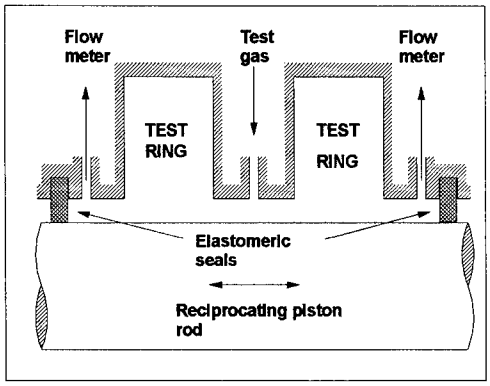


Figure 5. Low pressure seal test rig arrangement

Initial testing was concentrated on the evaluation of commercial quality, LP seals. The early findings were that sealing performance was very dependent upon the quality of manufacture. As shown in Figure 6, whereas some rings sealed with reasonable efficiency others, of apparently equivalent manufactured quality, performed poorly. Immediate improvements to the surface finish of all rings were agreed with the manufacturer.

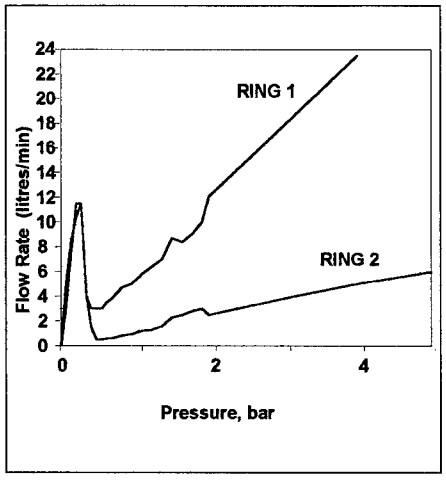


Fig. 6 Leakage of leaded bronze tangential sealing rings to vent

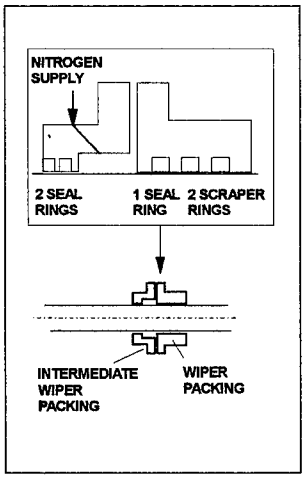


Fig. 7 Nitrogen buffer arrangement

Also shown in Figure 6 is an important finding of this phase of the work. The performance of all seals was poor at the range 0 to 0.5 bar, i.e. the very range within which it would be called upon to operate in practice. This is believed to be a function of friction between the seal and

rod, which allows the seal to shuttle from one side of the container to the other, effectively "pumping" gas downstream. As an interim measure it was proposed to inject nitrogen into the space between the intermediate packing and the gas seal of the wiper packing at a pressure of 1 bar. A sketch of the arrangement is shown in Figure 7. No provision was made for nitrogen gas to exit the void, unlike a typical purge. The primary intention of this modification was to energise the seals and to overcome the frictional drag effects of the rod, a buffer of increased pressure being a secondary advantage. A nitrogen blanket, or any high capacity inert system, was not possible in the application for logistic reasons.

Operating experience with this modification has been extremely good. The application of nitrogen gas has resulted in significant savings in used-oil discards, due to reductions in the rate of flash point depression, and explosimeter readings have become either very low or zero.

The test rig was next used to evaluate API 618 SLPR type seals, with a view to replacing the LP seals. The first rings tested were made to the finish standards of those in service in other applications, but these were found to have leakage rates considerably exceeding those of the later LP seals. Examination of these rings after testing showed the main cause to be coarseness of the turned finish of the 45 degree face. The least grooving that remained after machining led to friction between rings, inhibiting the even and constant conversion of radial spring tension to axial thrust. Discussions were held with the manufacturer to define the best repeatable commercial finish that could be provided on sealing surfaces. Agreement was reached that the chamfer faces would be fine-turned to a very high standard and that all radial seal faces would be lapped to 6-12 micro-inch CLA. A second set made to this standard was evaluated, giving the improved results shown in Figure 8.

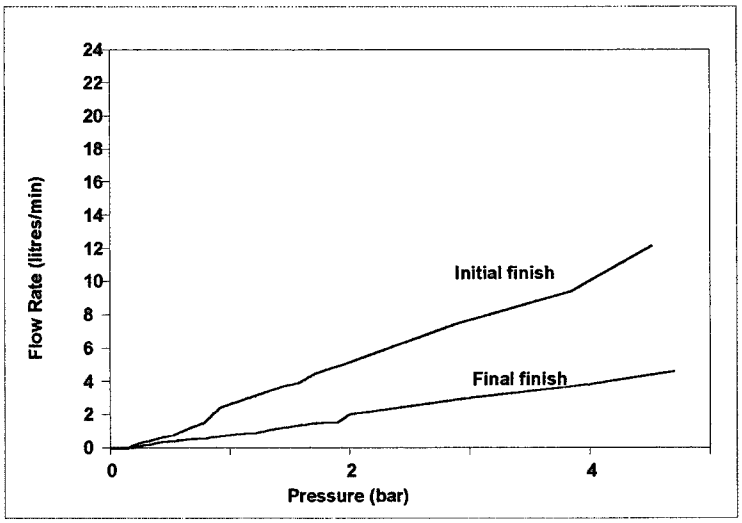


Figure 8. Leakage of leaded bronze SLPR rings to vent.

CONCLUSIONS

Investigation into the crankcase ignitions found that an explosive mixture of gas and air was ignited by electrical arcing. The presence of gas in the crankcase resulted from a deficiency in the design of the tangentially-cut piston rod seals, which were incapable of sealing efficiently. The effectiveness of this type of seal was improved by energising the rings with a nitrogen buffer of 1 bar pressure. Improved seals, of the type specified in API 618, were found to be ineffective until their surface finish had been improved to the highest practicable standards.

Ignition of the explosive mixture occurred when electrical arcing occurred at the rod wiper packings. The driving current was provided by recirculating eddy currents which were caused by a gradual reduction in the properties of the insulating pad at the non-drive end of the electric drive motor. Voltages involved were of the order of 0.3 mV.

Acknowledgement

The author wishes to thank Peter Collins for the design and construction of the rig used in these studies, and Chris Clements for his work in carrying out the seal leakage research described in this paper

EXPERIMENTAL AND NUMERICAL STUDY OF VIBRATION MONITORING APPLIED ON GEAR TRANSMISSION SYSTEMS

J.Mahfoudh, C.Bard and D.Play
Laboratoire CASM
Institut National des Siences Appliquées de LYON
20,av. A.EINSTEIN-69621 VILLEURBANNE Cedex-FRANCE.

Abstract: Predictive maintenance asks for efficient indicators and methods to detect incipient failures. Numerous studies have been made in this domain, but some difficulties stay to isolate the defected elements, to foresee through signal analysis, causes that produce faults, especially when different technogical elements are involved simultaneously (gears, bearings, ...). In the first part of this paper, experimental study have been performed on gear faults: manufacturing faults (profile error, deviation, ...) and operating faults (wear, pitting, ...) in order to establish a kind of data base for fault identification. Then, operating faults were examined and a procedure of identification of faults nature have been developped. Statistical treatement of both time and time-frequency (wavelet analysis) domains give arise to robust indicators. The second part deals with the definition of thresholds for machinery stops. Calculations of the dynamic behaviour of the geared system show that how a gear fault modify the gear profile and the kinematic transission error, while stiffness modification have almost no significant effect on the system response.

Key Words: Diagnostics; condition monitoring; maintenance; vibration analysis, signal processing; numerical simulation; gears

INTRODUCTION: This study takes place in a research program dealing with the condition monitoring of gearboxes, and particularly the analysis of the vibration signatures of spur and helical gear faults [1]. Machine monitoring using vibration measurements is rapidly established itself as a worthwhile techniques in modern industry. The analysis permit to characterize results of faults and gear damages [2, 3, 4]. The research program includes several steps: the study of design and manufacturing parameters as the tooth profile modification, misalignement, pitch module, center axe distance, ... This part enable users to optimize gear design parameters and the definition of the reference state of the studied system. The second step is the study of effects produced by local and distributed faults on spur and helical gears. The aim is to isolate significant parameters from the system response which indicate the presence and the nature of faults [5] and then establish warning level for fault progressing. In order to accomplish this task, experimental and numerical simulation have been undertaken in the same time. In this paper, only faults on gears will be considered. Vibration from bearing housing seems to be the suitable system response to be used. The working method is as follow:
- making tests for gears without fault,

- create fault on one gear, then reproducing the same set of tests,
- processing the acceleration signal from bearing housing in different domains (time, frequency and time-frequency),
- extracting the influence of signal parameters through a statistical design procedures.

Measurement were carried out for several operating conditions (speeds and loads), descriptors of bearing acceleration levels were compared with and without faults. Finally, and in order to define warning levels, the studied system is modelled using the Finite Elements Method [6]. The aim is to correlate the calculated dynamic load evolution with respect to the gear fault dimension, and indicator evolution issued from experimental measurements.

EXPERIMENTAL PROCEDURE

TEST RIG: The test stand (Fig. 1) is constituted of two shafts of 60 mm in diameter, mounted on two rolling bearings and coupled with gears. Test gears were clamped with nuts at the operating end of each shaft and centred by involute splines so as to limit variations of eccentricity. After gear clamping, a special nose was fixed at each shaft end, in order to measure their angular positions during motion with optical encoders (C1-C2), used for transmission error measurement [7]. The output shaft (1) and its rolling bearings were fixed on a special mounting, made of thick intermediate plates. It permits to impose small misalignments. The input shaft (2) was driven by a 120 kW DC motor. The output shaft is braked by a DC motor. Rotating speed was varied between 0 and 6000 r/min and was feedback controlled. The input torque was varied independently from 0 to 150 Nm. The active part of the apparatus was fixed on a 7 tons rigid frame (3). Vertical and horizontal bearing accelerations have been measured with piezoelectric accelerometers A1 and A2 respectively. Other parameters (speed, torque, oil temperature) were also recorded for each test.

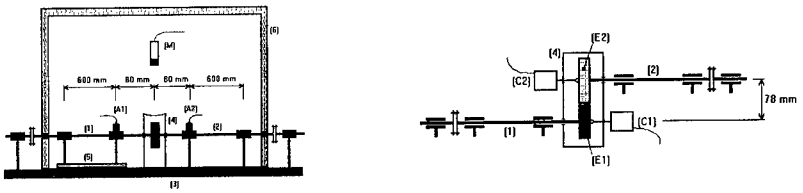


Fig. 1: Schematic arrangement of the test gear apparatus

SIGNAL ANALYSIS: Stored bearing acceleration signals have been processed using the asynchronous FFT. The spectrum is obtained from an average of 16 successive FFT signal acquisitions of 1024 points. Note that for this analysis, the amplitude of the first five harmonics of gear meshing frequency (H1, H2,..H5), the sum of the first five harmonics and the root mean square value (RMS) of the signal are considered. Beside time domain waveform study and spectral analysis, Fast Wavelet Transform analysis (FWT) has been used. Details and development of this technique are defined elsewhere [8]. This is a projection technique similar to the FFT. However, rather that simply decomposing the signal into sinusoids of varying frequencies, the data is represented as a projection onto a special function. This means that data set can be represented as time translations of a mother wavelet (a basis function) and as time dilatations (i.e. expanding the time scale of observation). Then, amplitude of the FWT coefficients can be visualized versus time (Fig. 2). The x-axis represents the time (ms), while the

y-axis represents octave bands of frequency (0-48 Hz, 48-97 Hz, etc ...). The FWT technique can be interpreted as octave band filters.

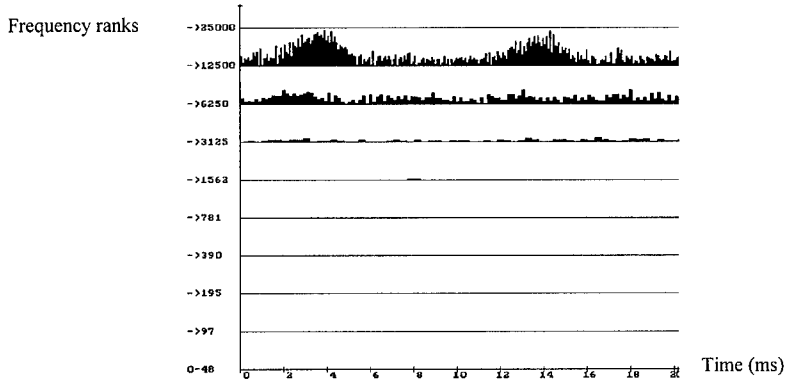


Fig. 2 : Wavelets representation. Spur gear.Distributed fault (12 daNm, 6000 rpm)

The statistical analysis has been also considered in order to quantify the effect of gear faults. Thus, for time domain and for wavelet coefficient analysis, the parameters are: Peak value, Peak to peak value, Standard deviation (s), Skewness, Kurtosis, Root Mean Square value (RMS) and Crest factor (peak/ RMS).

EXPERIMENTAL RESULTS: In a first time, the influence of design and manufacturing parameters on the dynamic behaviour of spur and helical gears (AGMA 12 quality and 51/52 teeth) was studied. The considered parameters were: deviation of 4', inclination of 4', tooth profile modification, center axis distance (78.02, 77.89 mm), module (2.1, 1.52 mm), contact ratio (1.3, 2, 3) and overlap ratio (0, 1, 2). Measurements have been done for several speeds (1000, 2000, 3500, 4750 and 6000 rpm) and several loads (0, 3, 6, 9 and 12 daNm). The conclusions of this part are summarized in Table 1.

Studied parameters	**Effects on acceleration**
Torque	For the harmonics H1 & H2 & for the sum of the first five harmonics, the influence is important, it is not so for the RMS.
Speed	The RMS is very sensitive for the different harmonics.
Contact ratio	The level of the different descriptors decreases with an increasing of contact ratio.
Deviation	Give arise to an increase of different descriptor levels.
Inclination	Does not bring any change in the level of the different descriptors.
Tooth correction	Give arise to a decrease of different descriptor levels.
Center axis distance	Important increase of the different descriptors has been remarked.
Pitch module	An increase of the module gives arise to a decrease of the level of different descriptors.

Table 1: Effects of design parametres on bearing acceleration

Local fault: The operating conditions are five speeds and five loads as above. For each test, the statistical parameters were calculated and the well-known Taguchi experimental design procedure [9] have been applied. The purpose of this analysis consists of first, defining the test to be realized, and second applying variance analysis procedure which enables us to extract the influence of parameters from the whole set of parameters being studied. The characteristics of the test gears are presented in Table 2.

characteristics	spur gear	helical gear
number of teeth	36/38	36/38
contact ratio ε_α	1.3	1.3
overlap ratio ε_β	0.0	1.0
helix angle	0.0	21°
face width of the tooth (b)	17.5 mm	17.5 mm
tooth thickness (e)	3.31 mm	3.31 mm
total depth of the tooth (h)	4.96 mm	9.30 mm
pitch module	2.0 mm	2.0 mm

Table 2: Characteristics of test gears

A local fault has been created along the pitch line of one tooth . It had a width of 0.863 mm and a 0.169 mm depth (Fig. 3). Tests have been made and signals obtained for spur gear with and without local fault have been processed. Beside the increase in amplitude of time domain waveform and of spectrum, due to fault, an increase of the FWT coefficients is also noted. This increase exists always in the last two octaves considered in the visualization.

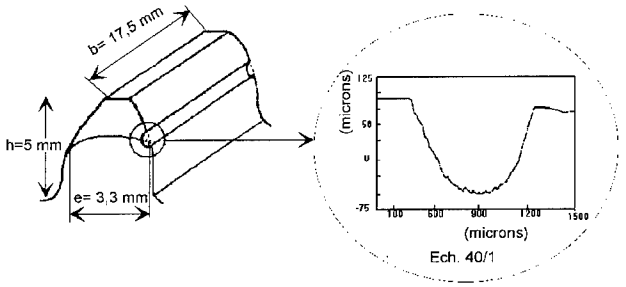


Fig. 3: Visualization of local fault

Statistical parameters of the three treatments have been calculated. Noting here that, in the case of the FWT analysis, only the last two octaves will be considered. A Taguchi experimental design procedure has permitted to extract the parameters that have been influenced by this type of fault. The final results of the Taguchi procedure is expressed in the Table 3. In the first column, the studied factors (fault, torque and speed) are noted. The other columns contain the statistical parameters with their influences in percents. Fault has an influence of more than 80 % on four of the seven studied parameters.

studied factors	parameters						
	Mean	Peak	RMS	s	skew	kurt.	crest factor
Fault	87 %	83 %	85 %	86 %	17 %	14 %	19 %
Speed	9 %	6 %	6 %	8 %	16 %	16 %	7 %
Torque	1 %	4 %	4 %	2 %	3 %	0 %	2 %
Residual	3 %	7 %	5 %	4 %	64 %	70 %	62 %

Table 3 : Influence of different statistical parameters

This procedure is applied for the two other signals. In summary, the parameters which have the largest influence are for :

- Time domain: peak value, peak to peak value, standard deviation and RMS,.
- Signal spectrum : H1, H2, $\sum_{i=1}^{5} H_i$ and RMS,
- Wavelet coefficients: peak value, mean, standard deviation and RMS.

Concerning the helical gear, the shape of the created local fault was similar to that of spur gear. This fault was of 0.859 mm width and its depth was of 0.033 mm. Tests have been performed and the statistical parameters have been calculated for the three treatments. The Taguchi statistical design procedure have been also applied. The same significant parameters as those considered in the case of spur gear tests have been found.

Uniformly distributed fault : In the case of local faults in spur and helical gears, good results have been obtained concerning the detection of incipient failure by statistical parameters, and particularly those of FWT coefficients. Now, the sensibility of these parameters in detecting a uniformly distributed fault will be considered. This fault has been induced on a helical gear of the same characteristics as the previous local fault gear. Test was recorded after every one hour. The speed of rotation was 3500 rpm with a 90 Nm torque. Lubrication was removed approximately 40 hours after the start of the test and at this time a fault uniformly distributed has been induced by scuffing. The results issued from the treatment of the recorded signal show that during the early period of the tests, no change have been pointed out for all the three signal presentations and also in all of their statistical parameters. While after inducing of the fault, amplitudes of time domain waveform, spectrum and wavelet coefficients were largely modified. This fault gives a modulation in time domain waveform that appears clearly with Fast Wavelet Transform analysis. This phenomenon can be attributed to an increase of the eccentricity effect on gear noise and vibrations. In fact, definition of uniformly distributed fault gives a single image of the one tooth fault. However, it can be reasonably considered that scuffing differs from one gear tooth to an another tooth. Then, new circular pitch and rotations occur. Tables 4 and 5 show the statistical parameters of fault. Results obtained from spectrum and FWT analysis are practically the same. On these tables, one can find the statistical parameters values with their rate of change, in percents, with respect to a baseline. The test after one hour is considered as a baseline against which test parameter are compared. It can be observed that a minimum of 10 % difference appears between the two considered gear situations. It can be notted too, that with an uniformly distributed fault, gears out of service with about 10% of level increase, while local fault gives more than 50% of level increase and gears still almost available. That is why, we have developped a methodology to identify the nature of fault, based on the use of statistical tests,

applied on the last two frequency rangs of the FWT analysis: KHI2 test identifies distributed fault and extreme values test, local fault [5].

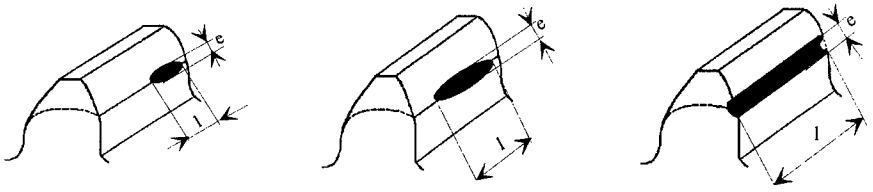
	statistical parameters			
	Peak (dB)	peak to peak (dB)	s (dB)	RMS (dB)
after one hour	66.1/ 0 %	72.1/ 0 %	57.2/ 0%	57.2/ 0 %
after 4 hours	69.3/ 5 %	75.1/ 4 %	61.4/ 7 %	61.4/ 7 %
after 8 hours	68.6/ 4 %	74.2/ 3 %	60.8/ 6 %	60.9/ 7 %
after 16 hours	69.8/ 6 %	75.3/ 4 %	61.7/ 8 %	62.0/ 8 %
after 32 hours	66.8/ 1 %	72.4/ 1 %	58.9/ 3 %	58.9/ 3 %
with fault	77.5/17%	82.9/15%	69.6/22%	69.6/22%

Table 4: Time domain parameters

	statistical parameters			
	peak (dB)	mean (dB)	s (dB)	RMS (dB)
after one hour	106.9/ 0 %	95.2/ 0 %	91.9/ 0 %	96.3 /0 %
after 4 hours	111.1/ 4 %	97.2/ 2 %	94.5/ 3 %	99.0/ 3 %
after 8 hours	107.6/ 1 %	96.0/1 %	94.6/ 3 %	98.4/ 2 %
after 16 hours	112.4/ 5%	97.8/3 %	96.3/ 5 %	100.2/ 4 %
after 32 hours	108.5/ 2%	96.4/1 %	93.6/ 2 %	98.2/ 2 %
with fault	118.5/11%	104.2/10%	104.5/14%	107.3/11%

Table 5: FWT coefficients parameters .

THRESHOLDS: In order to define thresholds, the maximum dynamic load at meshing have to be compared to a limit value defined by the nature of used materials and the geometry of teeth. The direct measurement of this dynamic load seems to be a delicate operation, thus effect from the dynamic loading can be emphasis either by experimental or numerical means. This step has been limited to the effect of local fault on spur gear. The tested gears are AGMA 12 quality with 36/38 teeth, a contact ratio of 1.3 and a pitch module of 2 mm. The chosen local fault simulates pitch line pitting on a single tooth. The single line of pitting is modelled as a strip of metal removed at the pitch line, a grinding wheel has been used to create this fault. Three levels have been created, their positions were determined using a marking compound (Fig 4).



level N° 1 level N° 2 level N° 3

Fig. 4: Visualisation of different levels of local fault

The geometrical characteristics of local faults for tested gear are presented in Table 6. Unfortunately, it has not been possible to obtain the same thickness and depth of the fault, but the same order of magnitude is obtained.

	Width (l)	Depth (p)	thickness (e)
Level N° 1	4.4 mm	0.043 mm	0.4 mm
Level N° 2	8.7 mm	0.072 mm	0.75 mm
Level N° 3	17.5 mm	0.23 mm	0.97 mm

Table 6: Geometrical characteristics of local fault

The operating conditions were, as usally, five speeds and five applied loads. Only results of RMS values versus speeds for those three levels in time and spectral domains (Fig. 5) will be presented. The acceleration r.m.s. level obtained with and without faults are compared.

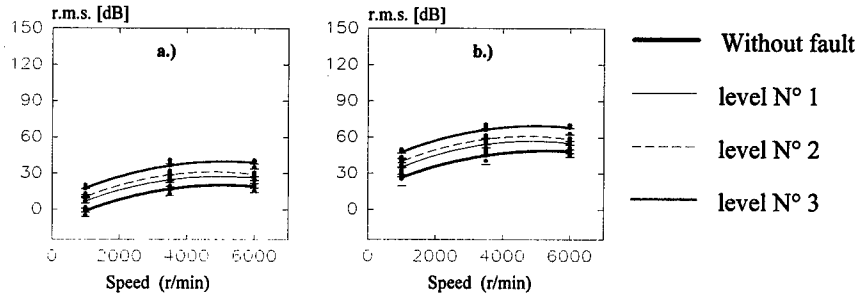


Fig. 5: Effect of local faults on r.m.s. : (a) Time domain, (b) Spectral analysis

An increase in the geometric area of the fault gives an increase of the signal output, in both time and frequency domains. The same trends are observed with respect to torque, but the acceleration r.m.s. increase versus torque is linear. General trends observed for this type of fault have been also mentioned previously in the literature.

Fitting between experimental and numerical dynamic results is now well known as the global mechanical architecture behaviour is taken into account through realistic bearing stiffness and damping [10]. A mathematical model has been developed to simulate tooth line pitting effect using general equations of motion for torsional response of a geared system and an elastic strain energy calculation of the stiffness function of a gear pair. The Finite Element Method is used to predict the dynamic behaviour of the test rig equiped with faulted gears. The excitations are introduced by the fluctuations of meshing rigidity and static transmission error. These parameters were calculated by taking into account all the design parameters of the gears, using a software based on the Finite Prism Method [11]. This technique allowed a 3D evaluation of cylindrical gear elastic deformations with a fine definition of the load sharing and pressure distribution. For each meshing position, potential areas of gear contact are located, and the unloaded transmission error and relative distances between teeth are calculated and stored. Elastic deformations are then calculated. Local and global deformations are evaluated separately, and surface deformations are estimated by the well known Boussinescq theory. Final results give pressure distributions along

the tooth face width, load sharing between different tooth pairs, stress, meshing stiffness and loaded transmission error. Thus, introducing precise values of meshing stiffness and static transmission errors, finite element modelling of the test rig has been performed. This modelling has been used for the prediction of dynamic effects of gear faults. Shafts were described with single rotor elements with two nodes and six degrees of freedom per node and gears are modelled by rigid disk connected by meshing stiffness. The description of meshing was associated to a 6x6 stiffness (and mass) matrix which connects the 6 degrees of freedom of gear centres. The general form of equations of motions is :

$$M.\ddot{X} + C.\dot{X} + K(\theta^*).X = K(\theta^*).\left\{\xi_o(\theta^*) + \varepsilon_o(\theta^*)\right\} \qquad (1)$$

where θ^* is a space parameter that describes the nominal progression of meshing. It could be expressed by $\theta^* = \Omega^*.t$, where Ω^* is the nominal angular velocity [12]. $\mathbf{K}(\theta^*)$ is the general form of the stiffness matrix (it depends on θ^* due to the meshing stiffness variations) and $\xi_o(\theta^*) + \varepsilon_o(\theta^*)$ is a specific vector which simulates excitation associated with the static transmission error variations and the unloaded transmission error[13]. M and C are mass and damping matrix. This system of equations is solved using the step by step Newmark techniques. A preliminary analysis begins by a geometrical description of the gear teeth, including profile modifications, manufacturing errors, gear bodies position including eccentricity, centre distance faults and misalignments. Then, gear meshing progress is analysed during a large angular period. This rigidity was then applied on the dynamic model and the tests rig was modelled by 41 elements with 251 degrees of freedom. The chosen modal damping was 0.02 for all modes of the stucture and 0.06 for mesh.

RESULTS: Local gear faults, described in the previous experimental part, have been introduced in this analysis by their geometrical description on a single tooth flank. The effects on meshing stiffness are presented in the figure 6. The results show that fault levels N°1 and N°2 have practically no significant effect on meshing stiffness variation, but level N°3 gives a variation of 50%. The geometrical variation due to gear fault can lead to a modification of the kinematic motion, a pseudo-polygonal effect appears due to flat bearing segment along the tooth profile caused by the fault geometry.

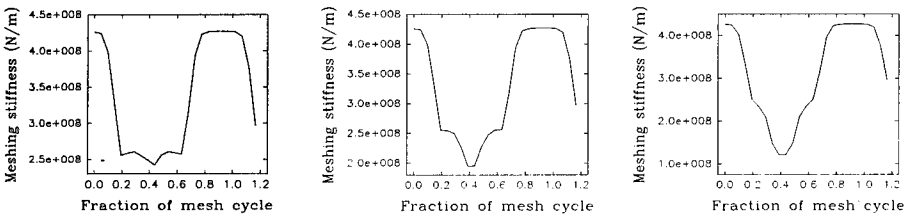


Fig. 6: Effects of faults on meshing stiffness evolution

The static transmission error has been determined using a kinematic simulation of mesh for fault level 3 only (Fig. 7) as for the others levels, this excitation is considered negligible.

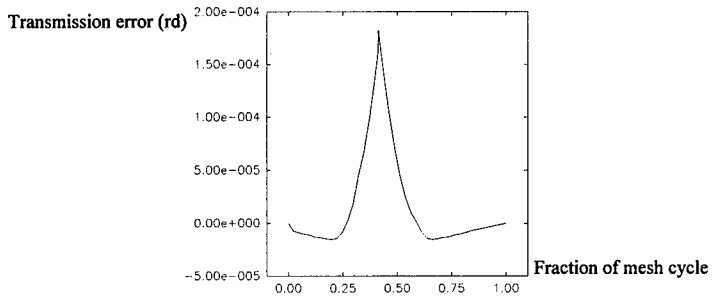


Fig. 7: Transmission error-fault level N°3

Calculations have been performed for different values of torque: the acceleration r.m.s. values of the spectral analysis have a linear variation with respect to torque, so, the r.m.s. values are calculated for several speeds and the maximum value of torque (120 Nm), in a steady state regime at each step of speed (50 r/min) (Fig.8).

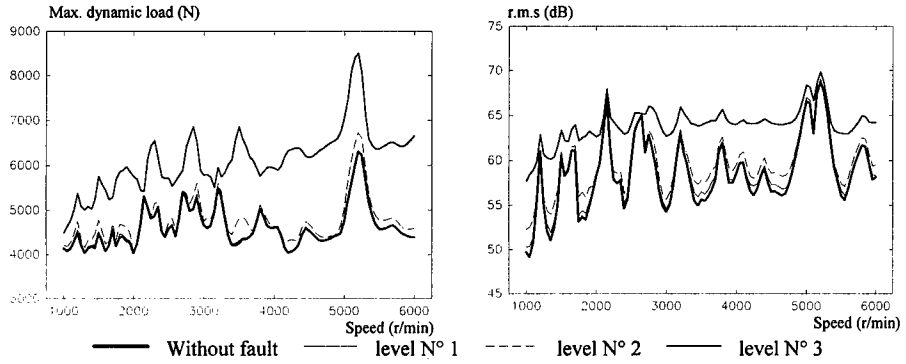


Fig. 8 : RMS acceleration values & Max. dynamic load

The maximum increase of the acceleration RMS is about 10 dB, while it was about 22 dB from measurements. Nevertheless, general trends are almost the same. We can observe that the increase of RMS values is almost negligible around pics, this is due to the fact that the effects of tooth deflection at those speeds are more important than the modification of tooth mesh rigidity or static transmission error due to fault.

CONCLUSION: Three topics have been considered in this work, the influence of design parameters on the dynamic behaviour of gearbox, which enable users to optimize gear's design. Then the influence of local and uniformely distributed faults have been studied, this part permit to establish a data base on the dynamic behaviour of gears with faults. Discriptors have been defined in order to indicate early the presence of failure. The last part deals with the elaboration of numerical model in order to establish thresholds. The same general trends have been observed

in both experiments and calculations. The next step is to determine the maximum stress due to this maximum dynamic load and to make comparaison with respect to the limit value for rupture. The calculations were not performed yet, What we are looking for now is to adjust the numerical model and particularly the simulation of excitation forces and the adjusting of different parameters as damping and static transmission error. Experimental work have been done to study the response of gearbox with faults on bearing, in the same time we are working to elaborate a numerical model to take into account the dynamic behaviour of bearing as a rolling bodies, the aim is to study gearbox dynamic behaviour with faults on both gears and bearing.

REFERENCES

[1]-ALATTASS M., Maintenance des machines tournantes: signature de defauts d'engrenage droits et helicoidaux, Thèse de doctorat, I.N.S.A de Lyon, 1994, 198 p.

[2]-RANDALL R.B., A new method of modelling gear faults, Journal of mechanical design, ASME, 1982, 104, pp 259-267.

[3]-ESHLEMAN R.L., The role of sum and difference frequencies in rotating machinery fault diagnosis, 2nd international conference on vibration in rotating machines, Cambridge, 1980, pp 145-149.

[4]-TAYLOR J.L., Fault diagnosis of gears using spectrum analysis, 2nd international conference on vibration in rotating machines, Cambridge, 1980, pp 163-168.

[5]-ALATTASS M., MAHFOUDH J. & PLAY D., Morphological study of vibratory signals, Conference on the acoustical and vibratory surveillance methods and diagnostic techniques, Paris, 1995, pp 456-474, CETIM.

[6]-MAHFOUDH J., BARD C., ALATTASS M. & PLAY D., Simulation of gearbox dynamic behaviour with gear faults, Second International Conference on Gearbox Noise, Vibration &Diagnostics, LONDON, 1995, pp.91-100, IMECH E.

[7]-BARD C., REMOND D. & PLAY D., New transmission error measurement for heavy load gears, Proc. of the 1994 International Gearing Conference, 1994, University of Newcastle upon Tyne, UK, pp.393-399.

[8]-Remond, D. , Bard, C., and Play, D., A Systematic Approach of the effect of Design Parameters on Spectral and Wavelet Gear Noise Analysis, JSME International Conference on motion and power transmission- Hiroshima, Japan- pp 128-134 (Nov. 1991)

[9]-Fauchon, J., "Plans d'Experiences" pp 99, INSA- LYON (1991). cours polycopié du département de génie mécanique développement.

[10]-BARD C., Modélisation du comportement dynamique des transmission par engrenages, Thèse de Doctorat, INSA de LYON, 1995, 292 p.

[11]-OLAKOREDE A., PLAY D., Load sharing load distribution and stress analysis of cylindrical gears by Finite Prism Method in a CAD environment, Proc. Design Productivity Int. Conf., Feb. 1991, Honolulu, Ragsdell & T.Holt, pp 921-927.

[12]-BLANKENSHIP G. W. & SINGH R., A comparative study of selected gear mesh interface dynamic models, International Power Transmission and Gearing Conference, A.S.M.E, 1992, Vol. 1 pp 137-146.

[13]-ÖZGUVEN N. & HOUSER D.R., Mathematical models used in gear dynamics - a review, Journal of Sound and Vibration, 1988, N° 121(3) pp 383-411.

EQUIPMENT MONITORING & CONDITION DIAGNOSIS

Low-Speed Mini-Motor Failure Recognition Using Fuzzy Theory

Khor Chin Siong and Kikuo Nezu.
Department of Mechanical Engineering. Gunma University
1 - 5 - 1 Tenzin-cho, Kiryu-shi, Gunma-ken, Japan 376

Abstract

This paper proposal a method of fault diagnosis for low-speed mini-motor with acoustics signals under the motor is not disassembling condition. Due to there are two uncertain events between fault of motor and acoustics signals of motor. These events will be described by fuzzy logic. A new probability concept of fuzzy events will apply in the method. The condition of mini-motor running will be described by membership grade set of feature value of acoustics signals. By using the inference method a new membership set will result. The new membership set not only can judges the motor quality, but also can recognize the kind of motor failure. The high frequency spectrum energy of acoustics signals and low frequency spectrum energy of acoustics signal is applied in the method as optimal feature value. Actual experiment results and numerical simulation's results all show the method of the combination fuzzy theory is very effectively method for fault diagnosis of low-speed mini-motor.

Key Words : diagnosis : low-speed mini-motor : acoustics signal : frequency spectrum : fuzzy theory.

1. Introduction

The manufacturing industry is surrounded with grave situation increasing year by year. Automation on manufacturing line is propulsion with factory automation. However, automatic examining process of production quality is still dropout, and present situation of examination has to rely on skilled worker. On the examining process of mini-motor for the automobile, the examination is performed using electric current wave-form, sound, vibration, etc. [1] [2] [3], which are measured by human vision, touch, sense of hearing, under noload condition. The standard judgement cannot describe quantitatively and clearly. It seems to be influenced by skilled worker's experience perception or condition. Therefore, it is desirable to offer the standard examining process and auto diagnostic system, by which the production of inferior works is suppressed and judgement method of reliability and with high-grade positive answer is supported [7] [8]. On this research, we have used the manufacturing company's mini-motor, analyzed the rotating sound into computer digital data and developed the standard examining process and verified the algorithm of judgement. We also applied noncommittal criterion with fuzzy system, and verified the possibility of automatic diagnosis.

2. Mini-motor Acoustic Signatures

In mini-motor, the components investigate in this paper. Fig.1 shows out ward and main dimension. The mini-motor was construction assembly by stator, armature and end bracket.

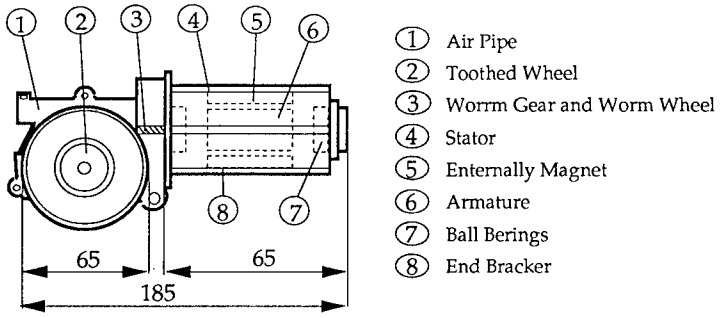


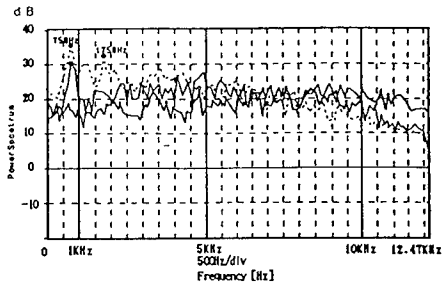
Fig. 1 Outward and main dimension of mini-motor

The data use in this project were collected with microphone in bench tests conducted by Electricity de France. The test voltage of mini-motor is 13.5V. The load of torque is 2.6kgf.m. The time dependent data were transformed using FFT techniques into low-frequency (10 ~ 1000 Hz) and high-frequency (10 ~ 12 KHz) spectra. Data are collected using an microphone placed in the radial direction to the loading of the mini-motor. The microphone is establish near by the End Bracket about 300mm in turn is stuck to the rolling element. From these measurement, spectra are generated using FFT techniques. Spectra are averaged over 8 samples with a HANNING window [4]. The spectra have 1,024 points. The acoustic voltage is -10V ~ +10V and sampling intervals of 0.04ms.

We got the judgement of 92 motors from manufacturing company. They were rotated in forward direction and backwards direction, and we picked up the total of 184 acoustic signals. Acoustic description are also calculated that describe overall acoustic levels in the various components. The technique described here is based on the information contained in these descriptor and the spectra.

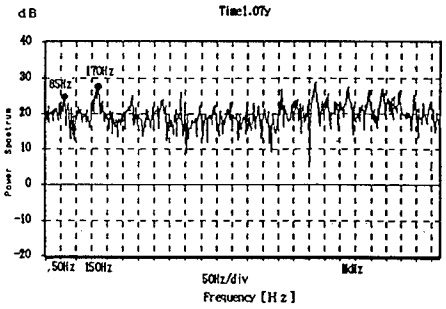
Generally, each machine defects produces a set of acoustic components that allows for recognition of different faults. For mini-motor, there are formulas for calculating the frequency regions in which features may appear in the spectra. For example, in identifies the frequency at which defects of the inner race would appear and there are similar formulas for faults of the outer race and rolling element.
Noise frequency = NZ / 60
N : Integral number Z : Number of gear

The specification of mini-motor acoustic signal is shown in Fig.2. Low speed mini-motor of effective in the range of 0 to 1KHz is extremely high, cause as electromagnetism niose.



Fine line : Failure motor A frequency model
Dotted line : Failure motor B frequency model
Thick line : Good motor frequency model

Fig. 2



Low-frequency spectra

Fig. 3

15 times at the basic of electricity wave as 750Hz also occurred electromagnetism noise. Range of 1 KHz to 5 KHz frequency spectra is high because of resonance noise. Power spectra of defect motor is higher rather than good motor. Range of 5KHz to 10 KHz frequency spectra as stability area also range of 10KHz to 12.47 KHz is good and failure characteristic area and high power spectrum is assume as failure motor. Fig.3 is enlarge from low-frequency spectra. It show the mini-motor bearing occurred the signal by 50 times as 85Hz and 100 times as 170 KHz peak level rate.

3. Parameter Estimation

In order to process data using fuzzy logic, the input data must be scaled to the operating range of algorithms. For this project, we adoption 4 spectra values of discred time signals and 3 spectrum of frequency band as characteristic parameter.

1) Average x of total absolute value $\sum x = \sum_{i=1}^{N} |x_i|$

$$\overline{x} = \frac{\sum x}{N}$$

2) Skewness α of standard deviation $\delta = \frac{\sqrt{\sum_{i=1}^{N} (|x_i| - x)^2}}{N-1}$

$$\alpha = \frac{1}{\delta^3} \cdot \frac{\sqrt{\sum_{i=1}^{N} (|x_i| - x)^3}}{N-1}$$

3) Kurtosis

$$\beta = \frac{1}{\delta^4} \cdot \frac{\sqrt{\sum_{i=1}^{N} (|x_i| - x)^4}}{N-1}$$

4) Crest Factor

$$C f = X_{max} / X_{RMS}$$

5) Total range of low-frequency spectra

LP = $\sum_{i=1}^{N}$ (0 ~ 1 Khz)

6) Total range of high-frequency spectra

HP = $\sum_{i=1}^{N}$ (10K ~ 12.47 KHz)

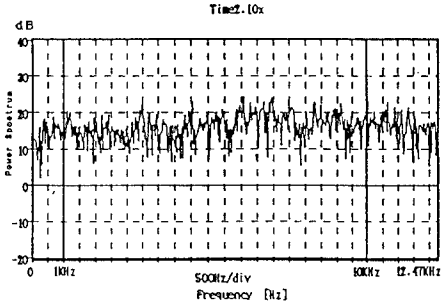
7) 15 times frequency spectra based on electricity sources wave.
For fuzzy logic, the various values for each pattern are calculated and spectra scaled to the appropriate range.

4. Fuzzy Methodology

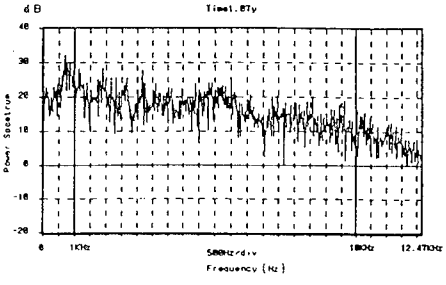
In this paper, several problem concerning with quality will take as probability set, indicate as condition vector. And this vector detention function μ_Y (Yi) as
Y = {y1, y2, y3, ...yn}
n dimension as defect clarify.
As several cause of defect occurred deal as each characteristic set. The characteristic vector belonging to detraction function of μ_X(Xi) as
X = {x1, x2,x3,.....xm}
m dimension as a variety of characteristic.
Generally, the judgement will decide by 2 rate of region [0,1] as detention function Xi, as good and failure. However, all kinds of mini-motor's characteristic Xi and quantity are not relation clearly in ratio. Must of existence quantity will not indicate clearly. Example stator resonance part processing not order occurred metal noise. The situation of μ_A difficult to express destination in two rate logic.
Then based on fuzzy logic concept, the existence ambiqlious height can accuracy and exactly detonate in variety value logic. The height of a fuzzy net A evaluates the possibility of finding X at least one element which fits the predicate A exactly. A is normalized when such an element can be exhibited. Owing to cause interpretation of the equation A ○ R = B, where R is fuzzy relation. A is fuzzy set of symptoms as shown as above. B is a fuzzy disease. R represents the membership function and is estimated from a set of patients on whom symptoms are observed and who are diagnosed by reliable physicians. For good effective to estimate the defect we provide a new probability concept of fuzzy to analysis the failure [5]. The fuzzy functioning projection is to complementary the problem.

5. Results

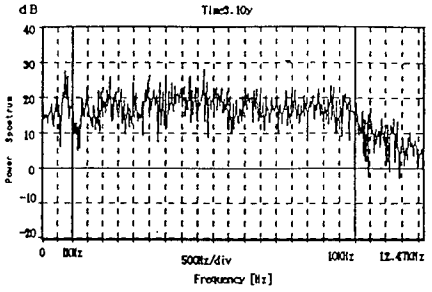
The preprocessing fuzzy logic was used to identify possibility faulty system. The input to



Good motor frequency model
Fig. 4

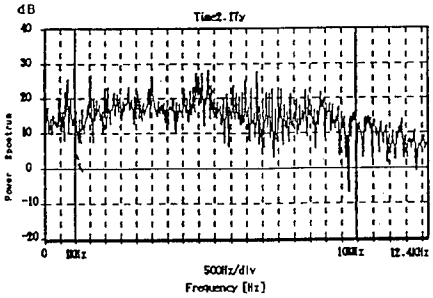


Failure motor frequency model
Fig. 5



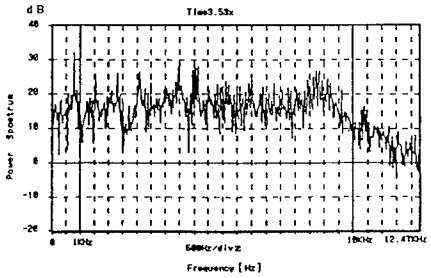
Usual system detection error frequency model

Fig. 6



Usual system detection error for metal noise frequency mode

Fig. 7



Metal noise frequency mode

Fig. 8

the system were the normalized values of the high-frequency descriptor, low-frequency descriptor, time signal descriptor and corresponding differences of these descriptor between parameter. The differences provide information about sudden increases in the descriptors (rate) and are essential in the detection of certain faults. The output of fuzzy system indicate absence of a fault and one indicate a possibly faulty signature. For total of patterns of failure the system can estimate and diagnostic the defect accuracy.

Number	A	B	C	D	E	Defect {0,1}
210 x	0.35	0.35	0.35	0.35	0.35	0.35
217 y	0.50	0.65	0.45	0.75	0.60	0.75
310 y	0.64	0.65	0.64	0.74	0.60	0.74
353 x	0.86	0.71	0.65	0.75	0.65	0.86
107 y	0.55	0.90	0.85	0.90	0.85	0.96

A : Main body defect, cause of impact element
B : Part processing not order occurred metal noise
C : Armature abnormal defect signal noise
D : Cycling occurred electromagnetism noise
E : Resonance noise

6. Conclusions

We are working on the implementation of meteorology for interpreting acoustic measurements based on fuzzy theory. The anticipated advantage of developing such system is the possibility of automating and diagnostic process for mini-motor and low speed mini-motor diagnostic systems that complement traditional signal analysis by dealing with the nonlinear characteristic of the signal. The flexibility of using fuzzy theory for spectra data has been establish. Under this study, we have achieved 92% accuracy while maintaining the diagnostic range 82% of the original system.This system also can point out the possibility of defects.

Using the low-frequency and high-frequency spectra adds to the robustness of our system because we able to detect incipient faults a well as severe defects. In addition, the information provided by expert may be used to fuzzy theory and possibly faulty signatures. We feel that fuzzy theory can provide a meteorology for improving the analysis of spectra for acoustic signal analysis and may provide a system to estimate and diagnostic of mini-motor.

[1] Toshio Toyota, Peng Chen : A study of the Automatic diagnosis of rolling element bearing (Part 1) Japan Mechanical Forum 1993.
[2] Takuzo Iwatsubo, Shozo Kawamura, A. Kanie : Application of fuzzy algorism of the diagnosis of rotating machinery, The 71st Japan Mechanical Association Conference Vol. E, 2509.
[3] C.S. Khor, K. Nezu : Application of fuzzy modeling to the diagnostics of the motors, 10th Fuzzy System Symposium, SB 1-1, 717-718 (1994).
[4] CK Mechefske, J.Mathew : Fault detection and diagnosis in low speed rolling element bearing part I, Mechanical System and Signal Processing (1992) 6(4), 297-307.
[5] C.S. Khor, K.Nezu and others : The study of inference for calculation rate using fuzzy theory, 5th Non-Engineering Fuzzy Symposium (1995). 5-27.
[6] C.S. Khor,K.Nezu,Y.Kotani : A study of fault of mini-motor using vibration diagnosis method. Japan AEM. 1995. 12. 897-908.
[7] H. Yamakuchi and others : Diagnosis system based on vibration measurement for rotating machines. Mitsubishi Industry Technical Report. Vol. 24 No. 5 (1987-9).
[8] T. Takise : A diagnosis method by ratational noise of the small size assembly using ball bearings. Japan Acoustic Journal. Vol. 44 No. 6 (1988)

AUTOMATIC MACHINERY FAULT DETECTION AND DIAGNOSIS USING FUZZY LOGIC

C.K. Mechefske, J. Del Mar and D. Prendergast
Department of Mechanical Engineering
Victoria University of Technology
P.O. Box 14428
Melbourne, Victoria, Australia, 8001

Abstract: Machine condition monitoring incorporates a number of machinery fault diagnosis techniques. Many of these machinery fault diagnostic techniques involve automatic signal classification. In this paper Fuzzy logic techniques have been applied to classify frequency spectra representing various bearing faults. The frequency spectra have been processed by four common Fuzzy set shapes: linear, triangular, S-curve and Pi curve. The application of basic Fuzzy logic techniques has allowed Fuzzy numbers to be generated which represent the similarity between two frequency spectra. Correct classification of six different bearing fault spectra was observed when the frequency spectra were represented by Pi curves. The degree of membership of each individual spectrum with respect to the other spectra, however, indicated a certain degree of overlapping. Further investigations must be conducted in order to optimise the ability to classify spectra with a certain degree of overlapping or masking.

Key Words: Fault detection, fault diagnosis, frequency spectra, Fuzzy logic, machinery faults, rolling element bearings.

INTRODUCTION: Unexpected failures of machinery and automated systems can be reduced by incorporating planned maintenance and condition monitoring strategies. Condition monitoring strategies involve detection, identification, prediction and correction of faults during the operation of a system [1]. However, identification of machinery faults can be difficult in systems with a high degree of complexity, hence introducing uncertainties in the area of fault diagnosis.

Recently there has been a significant amount of research effort put into developing and implementing useful automated machinery fault detection and diagnostic tools. Most of these tools have been based on various pattern recognition schemes, knowledge based systems (expert systems) or artificial neural network systems. The main thrust of the work has been towards developing systems that are not only objective in their treatment of data and presentation of results but also flexible, thereby being applicable in a wide range of situations.

Fuzzy logic has gained a wide acceptance in other fields as a useful tool for blending objectivity with flexibility, particularly in the area of process control. The work reported on in this paper involves the use of Fuzzy logic principles to categorise vibration signals and thereby provide a means of objectively detecting and diagnosing machinery faults.

Fuzzy logic is also proving itself to be a powerful tool in knowledge modeling [2]. Its ability to handle ambiguities makes it a useful tool, especially in the area of diagnostics. Its application as

a diagnostic technique for machinery faults may prove to be advantageous to companies who operate complex automated machines. The realisation of Fuzzy logic as a useful diagnostic technique will require further research and development.

FUZZY LOGIC: Fuzzy logic provides a method of reducing and explaining system complexity [2]. It deals with system uncertainties and ambiguities. It mimics human reasoning and allows variables such as time, acceleration, force, distance, etc., to be represented with a certain degree of uncertainty. Fuzzy logic allows the membership of a variable within a group to be esstimated with a particular degree of uncertainty. Application of Fuzzy logic to machinery fault diagnosis should allow the membership of an unknown frequency spectra to be determined with respect to the known frequency spectra of particular faults.

Fuzzy logic represents system parameters as normalised values between zero and one. The uncertainties and ambiguities associated with a system parameter can then be quantified in terms more easily interpreted by humans. For example; Is the temperature of 75°C high or low? If we know that 100°C is definitely high and 60°C is definitely low, 75°C may be considered somewhat more low than high but still not low. Fuzzy logic allows us to quantify the grey area between high and low rather than simply considering every temperature below 80°C (the mid point) to be low and every temperature above 80°C as high. So called crisp boundaries are made Fuzzy but in a quantified manner. The actual degree of membership of a system parameter (temperature) in a particular group (low) is indicated by the values between zero and one inclusive [2]. A membership of zero means that the value does not belong to the set under consideration. A membership of one would mean full representation of the set under consideration. A membership somewhere between these two limits indicates the degree of membership. The manner in which values are assigned membership is not fixed and may be established according to the preference of the person conducting the investigation.

Fuzzy sets can be represented by various shapes. They are commonly represented by S-curves, Pi curves, triangular curves and linear curves. The choice of the shape of the Fuzzy set depends on the best way to represent the data. The degree of membership is indicated on the vertical axis. In general, the membership starts at zero (no membership) and continues to one (complete membership). The domain of a set is indicated along the horizontal axis. The fuzzy set shape defines the relationship between the domain and the membership values of a set.

The S-Curve moves from no membership at its extreme left-hand side to complete membership at its extreme right-hand side. The inflection point of the S-Curve is at the 0.5 membership point. The S-curve can also represent declining membership. It is defined by three parameters; its zero membership value (α), its complete membership value (γ) and its inflection point (β). (See Figure 1.) The domain values of an S-Curve can be determined from the following relationships [2].

$$S(x;\alpha,\beta,\gamma) = \begin{cases} 0 & \rightarrow x <= \alpha \\ 2((x-\alpha)/(\gamma-\alpha))^2 & \rightarrow \alpha <= x <= \beta \\ 1-2((x-\gamma)/(\gamma-\alpha))^2 & \rightarrow \beta <= x <= \gamma \\ 1 & \rightarrow x >= \gamma \end{cases}$$

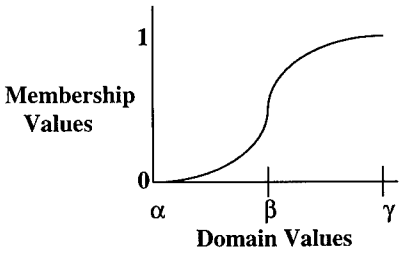


Figure 1. S-Curve

Pi curves are the preferred and default method of presenting Fuzzy numbers where membership values have lower and upper bounds [2]. The Pi curve represents full membership at its central value. A smooth descending gradient is then observed on either side of the central value where the membership approaches zero along the domain. The parameters of the Pi curve are; the central value (γ) and the width of half of the Pi curve (β). (See Figure 2.)

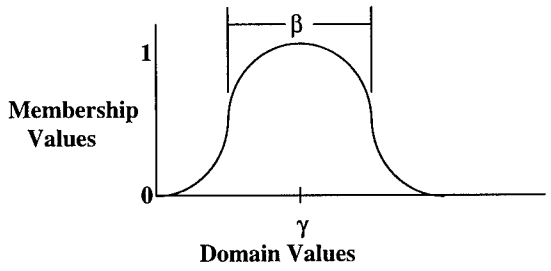


Figure 2. Pi Curve

The domain values of the Pi curve can be determined from the following relationships [2] where the Pi Curve is made up of ascending and descending S-Curves.

$$\mathbf{Pi}\ (x;\beta,\gamma) = \begin{cases} S(x;\gamma-\beta,\gamma-\beta/2,\gamma) & \rightarrow x <= \gamma \\ 1\text{-}S(x;\gamma,\gamma+\beta/2,\gamma+\beta) & \rightarrow x > \gamma \end{cases}$$

The linear Fuzzy set is the simplest Fuzzy set shape being basically a straight line. For an increasing linear Fuzzy set, no membership begins at the extreme left hand side. The line then increases linearly to the position where it represents complete membership on the right hand side. (See Figure 3.)

Fuzzy sets represented by triangular curves are similar to the Fuzzy sets represented by the Pi curve. The apex of the triangle is the central value and represents complete membership. The

left and right edges of the triangle represent membership that tends toward non membership in the set. (See Figure 4.)

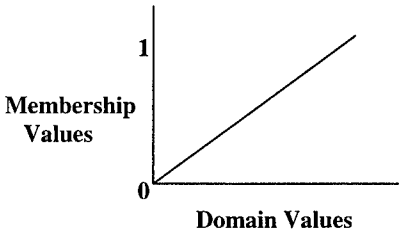


Figure 3. Linear Curve

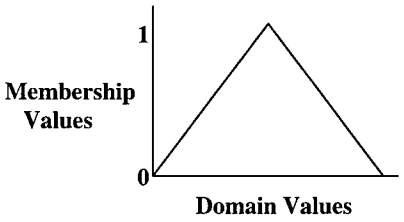


Figure 4. Triangular Curve

FUZZY SET OPERATORS: The main difference between Boolean logic and Fuzzy logic is that Boolean logic only has two states, one and zero. Fuzzy logic deals with uncertainties and hence, considers all values between zero and one. However, like conventional sets there are specifically defined operation for combining and modifying Fuzzy sets. These basic operations provide fundamental tools for Fuzzy logic. The basic Fuzzy logic operations are shown below.

Intersection: $A \cap B = \min(\mu_A[x], \mu_B[y])$
Union: $A \cup B = \max(\mu_A[x], \mu_B[y])$
Complement: $A = 1 - \mu_A[x]$

Above, **A** and **B** are sets of Fuzzy numbers and μ_A and μ_B are the individual Fuzzy membership values. The intersection of two Fuzzy sets involves taking the smallest membership value of the two Fuzzy sets. The union of two Fuzzy sets involves taking the largest membership value of the two Fuzzy sets. The complement of a Fuzzy set is determined by inverting the Fuzzy set shape at each point. For example, the compliment of an increasing linear Fuzzy set is a decreasing linear Fuzzy set [2]

FUSION FRAME WORK: An individual frequency spectrum can be identified with a class of spectra with similar characteristics. To obtain a final membership value of a sample of a particular class, Fuzzy logic operators can be used. First the union operator is used to combine all the available information about the different spectral features. For example, if frequency measurements were made for a particular fault at different time intervals, all the magnitude values taken at 50 Hz at the different time intervals would be combined. This procedure has also been found to be useful when the aim is to maximise all the information of the same feature coming from different sensors [3]. To finally establish the overall membership value of the spectra requires a more selective process. To narrow the selection of membership values obtained from the union operation, the intersection operation is used to obtain the final membership value of a particular class of spectra. The intersection operation makes establishing the final membership value of a particular class a more selective process, if there are a number of existing classes [3].

METHODOLOGY: The aim of this work was to investigate the use of basic Fuzzy logic concepts for possible application as a machinery fault diagnostic tool. This diagnostic technique will be capable of automatic and objective machinery fault classification.

The data used in this research were frequency spectra obtained from a test rig used to conduct low speed rolling element bearing tests [4][5]. The frequency spectra obtained were representative of different bearing conditions. These bearing conditions were; Fault 1 - Inner race fault, Fault 2 - No fault, Fault 3 - Outer race fault, Fault 4 - Combination of outer race and rolling element faults, Fault 5 - Combination of outer, inner and rolling element faults, Fault 6 - Rolling element fault. The Pi curve, linear curve, S-curve and the triangular curve were used to represent the frequency spectra of the various bearing conditions. The main purpose of the investigation was to determine which Fuzzy set shape best represents the frequency spectra of the various bearing faults and how well the technique classified faults.

Programs coded in C were written to allow the various bearing fault conditions to be classified with respect to each other. The C programs implemented basic Fuzzy logic data fusion techniques with the corresponding Fuzzy shape to classify the various bearing conditions.

The way this investigation was approached was initially to investigate each of the frequency spectra of the bearing conditions. The purpose of investigating each of the frequency responses was to determine the characteristic acceleration ranges of each of the bearing conditions. This would establish the ranges of each bearing condition when represented in a Fuzzy set of a particular shape. It was observed that when the frequency spectra were super imposed onto each other, there was a certain degree of overlapping. For this reason, the Fuzzy set shapes also displayed a certain degree of overlapping.

RESULTS: The programs written read in all the frequency spectral responses of a particular bearing condition to act as a data bank. One of the known frequency spectrum corresponding to a particular bearing condition was then input into the program. This spectrum was then considered as an unknown spectrum. If the spectrum corresponds to a particular bearing condition, its membership value will be high, indicating membership to that particular fault. If

the unknown signal does not correspond to a particular fault condition it will have a lower membership value, indicating only partial or non-membership.

The tables given below are square matrices containing the various bearing conditions investigated. Only results from trials using the triangular and Pi curves are reported here. Linear and S-curves are not appropriate for the type of data representation used for the frequency spectra under consideration. Table 1 shows the results of the triangular curve classification trial and Table 2 shows the results of the Pi curve classification trial.

Table 1: Triangular Results

FAULTS:	1.	2.	3.	4.	5.	6.
1.	**1.00**	0.85	0.78	1.00	1.00	0.65
2.	1.00	**0.99**	1.00	1.00	1.00	1.00
3.	0.82	0.77	**0.98**	0.00	0.00	0.57
4.	0.54	0.00	0.42	**0.66**	0.00	0.35
5.	0.00	0.00	0.00	0.00	**1.00**	0.00
6.	0.67	0.82	0.79	0.00	1.00	**0.84**

Table 2 : Pi Curve Results

FAULTS:	1.	2.	3.	4.	5.	6.
1.	**0.99**	0.96	0.98	0.85	0.00	0.96
2.	0.95	**0.98**	0.97	0.00	0.00	0.95
3.	0.97	0.95	**0.99**	0.94	0.00	0.98
4.	0.89	0.00	0.92	**0.98**	0.59	0.63
5.	0.00	0.00	0.00	0.03	**1.00**	0.00
6.	0.97	0.98	0.98	0.85	0.00	**0.99**

DISCUSSION: Linear curves, triangular curves, S-curves and Pi curves are the common set shapes used to represent data in Fuzzy logic. It was possible to code the various curves in the programming language of C to obtain membership values. A Fuzzy fusion framework based on basic Fuzzy logic operations was then used to obtain the final membership value of the unknown frequency spectral response data with respect to the known frequency response values stored in the program's data bank.

Tables 1 and 2 represent square fault matrices with the six known faults being compared to six known fault classes in a blind test. The six known fault classes are in the vertical columns and are actually groups of 15 sample spectra representing the different faults placed in the program's data bank. Six representative faults across the top row were considered to be unknown and were inputs to the programs of the various Fuzzy shapes. The order of the faults was the same for both the vertical column and the top horizontal row. For any particular Fuzzy set shape the results should show highest membership values in the diagonal of the table. This would indicate

that the frequency spectra data input into the program displayed a high degree of membership to the corresponding known fault class.

The Fuzzy set shape which produced the most desirable results was the Pi curve. The results in Table 2 show membership values that were highest along the diagonal. This indicates that the corresponding faults had the highest degree of membership or can be considered most similar. In many cases there are significant differences between the various trial classifications. In some cases the value in the diagonal is much higher than the other numbers, in other cases the value is only slightly higher. This is to be expected as some of the frequency spectra are in fact more similar than are others.

Use of the triangular curve for Fuzzy classification trials (results shown in Table 1) resulted in promising classification results but also some errors. These errors occurred where there is an obvious similarity between the frequency spectra being compared. Some modifications to the data fusion method employed may provide grester numerical distinction between similar spectra. Perhaps some combination of Fuzzy logic set operations and classical statistical analysis would improve the results.

CONCLUDING COMMENTS: This work has investigated the use of basic Fuzzy logic techniques as a machinery fault diagnostic technique. The work conducted has displayed the potential of basic Fuzzy logic to classify frequency spectra according to the likely fault condition which they represent. However, further investigation must be conducted to optimise its ability to classify spectra when overlapping or masking exists. Its ability to classify and identify machinery faults shows considerable potential. With further research and development as a diagnostic tool, Fuzzy logic may have an important role to play in machine condition monitoring.

REFERENCES

1. Clark, R., Fault Diagnosis in Dynamic Systems, Prentice Hall International Ltd., London, 1989.

2. Cox, E., The Fuzzy Systems Handbook - A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems, AP Professional. Publishing, New York, 1994.

3. Russo, F., Fuzzy Methods for Multi-sensor Data Fusion, IEEE Transactions on Instrumentation and Measurement, Vol. 43. No. 2., 1994

4. Mechefske, C.K., and J. Mathew, Fault Detection and Diagnose in Low Speed Rolling Element Bearings II: The use of Nearest Neighbour Classification. Mechanical Systems and Signal Processing, Vol.6, p309, 1992

5. Mechefske, C.K., and J. Mathew, Fault Detection and Diagnosis in Low Speed Rolling Element Bearings, Part I: The use of Parametric Spectra, Mechanical Systems and Signal Processing, Vol.6, No.4, p297-307. 1992.

TECHNOLOGY OVERVIEW: SHOCK PULSE METHOD

Louis E. Morando
Vice President and National Sales Manager
SPM Instrument Inc.
351 N. Main St.
Marlborough, CT 06447

Abstract: The name, SPM, is derived from the technology that SPM Instrument developed and patented in the early 70's in Sweden. The Shock Pulse Method is the monitoring and analyzation of high frequency compression (shock) waves generated by a bearing while rotating. From this research, empirical data was developed and patented to measure the theoretical film thickness of the lubricant in the rotating bearing along with an analysis of the overall condition of the bearing surfaces.

The way these signals are separated is really what makes this technology unique. Unlike vibration analysis that monitors a broad vibration band and then tries to isolate unique frequencies; SPM has developed a means to only "look" at the high frequency signals of antifriction bearings. Having ensured that the signal quality truly reflects a bearing signal, the development of a defined data base by SPM became practical. The ability to analyze lubrication changes versus surface damage becomes more practical and repeatable.

Through years of testing, this data base has been developed and perfected so as to represent the "True" operating condition of the bearing being monitored. Regardless of whether the bearing is 5 days old or 5 years old the reading taken represents the operating condition at that time.

Key Words: Code; condition; decibel; kilohertz; lub; pulse; resonant; shock

Condition monitoring means different things to different people. When you look at the most common problems of rotating equipment, the bearing condition is of most concern. With over 20 years of worldwide experience, SPM certainly finds this to be true. **(Fig.1)**

When one looks at what is conventionally used for machine analysis we see a vibration time signal. Ninety nine per cent of that signal is affected by rotational forces, which is helpful for problems that exhibit repetitive signals and therefore show up at discrete frequencies. The difficulty in monitoring antifriction bearing signals along with the other rotational forces is that the vast majority of bearing damage is not a repetitive signal and hence not always found at a discrete frequency. **(Fig.2)**

Shock pulses consist of a string of pulses with varying magnitudes. The strength of the individual pulses, and the ratio between stronger and weaker in the overall pattern, provide the raw data for

bearing condition analysis. The magnitudes of these pulses are dependent on the bearing surface condition and the peripheral velocity of the bearing (rpm and diameter). In undamaged bearings, the shock level varies with the thickness of the lubricant film between the rolling elements and raceway. The relationship between stronger and weaker pulses, however, is hardly effected. Surface damage causes an increase of up to 1000 times in shock pulse strength, combined with a marked change in the ratio between stronger and weaker pulses. Shock pulse values are translated into measurements of oil film thickness or surface damage, whichever applies. **(Fig.3)**

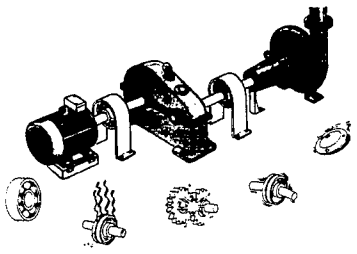
Only a small percentage of bearings fail because the natural fatigue limit of the bearing steel has been reached. For the large majority of bearings, metal fatigue starts early because the rolling elements and raceways are not properly separated by a protective lubricant film. The cause of this is often insufficient or incorrect lubrication. Improper mounting, dirt in the bearing, electrical currents, and machine vibration also cause reduced bearing life.

In a given bearing application, a number of factors that influence the lubricant film are constant. These factors include static and dynamic load; the geometry of the bearing housing, shaft and bearing; rolling velocity and necessary pre-load. Other factors can be influenced by maintenance personnel, making it possible to have a significant impact on bearing condition and life. These factors include pre-load (due to incorrect mounting), shaft alignment, total load (by correcting alignment and pre-loading), lubricant supply to the bearing, lubricant type and lubricant quality. Bearing temperature depends on constant factors such as velocity and environment, as well as lubrication. **(Fig.4)**

The L_{10} value is the life expectancy stated by the bearing manufacturers, representing the time it takes for 10% of the bearings to fail. Bearing life is also a function of the bearing's lambda value, which is the ratio of the oil film thickness to the surface roughness. A value of approximately 1.2 is generally needed to reach L_{10} life. In ball bearings that have less surface roughness than roller bearings, a LUB No. of 3 corresponds to Lambda = 1.2 and thus to 100% of the projected bearing L_{10} life. In roller bearings, which have rougher surfaces, the corresponding LUB No. is 6. The SPM curves show the relationship between LUB No. and the L_{10} life for roller bearings (upper) and ball bearings (lower). The SPM LUB No. is a measurement of a relative oil film thickness. When the lubricant film in the measured bearing is thin, the shock pulse meter will indicate a low LUB No. Checking and, if possible, improving the lubricant supply is often the easiest remedy and the first remedy that should be tried to improve operating condition. Operating with an insufficient LUB No., a bearing will in time develop irreversible surface damage. The reality is, while bearing manufacturers have defined bearing L_{10} life, few applications can achieve the lubrication level necessary to meet that life. The Shock Pulse Method allows the user to maximize the LUB No. to whatever the bearing can sustain under its operating conditions (RPM, Load, Temperature). "True" Condition Monitoring will allow the end user to obtain actual lubrication condition the first time a reading is taken and not rely on the user to develop the data base. **(Fig.5)**

Extensive research was undertaken to establish the relationship between the shock pulse signal from a bearing and its lubrication condition. An examination was made of a large number of bearings of different sizes and types. Variables examined include rpm, load, as well as lubricant

Condition monitoring



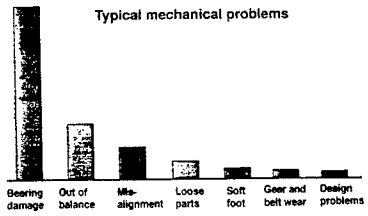
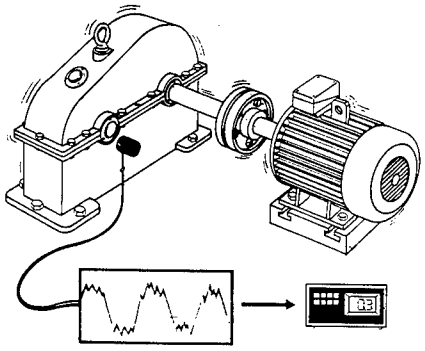


Fig 1



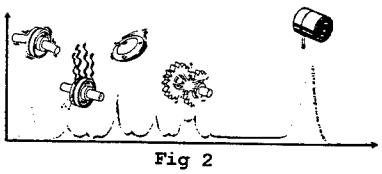
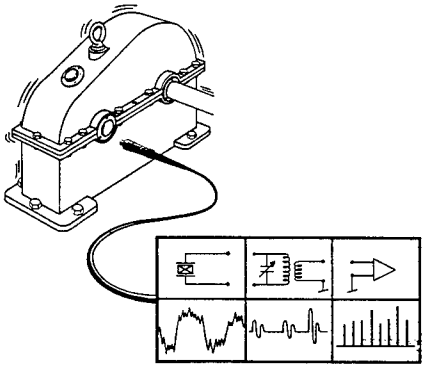


Fig 2



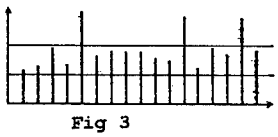


Fig 3

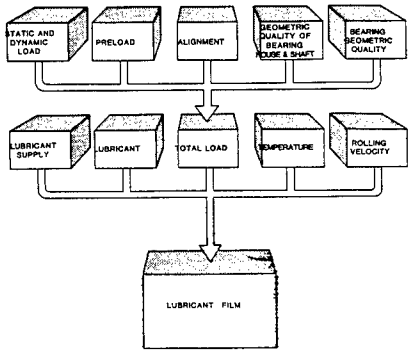


Fig 4

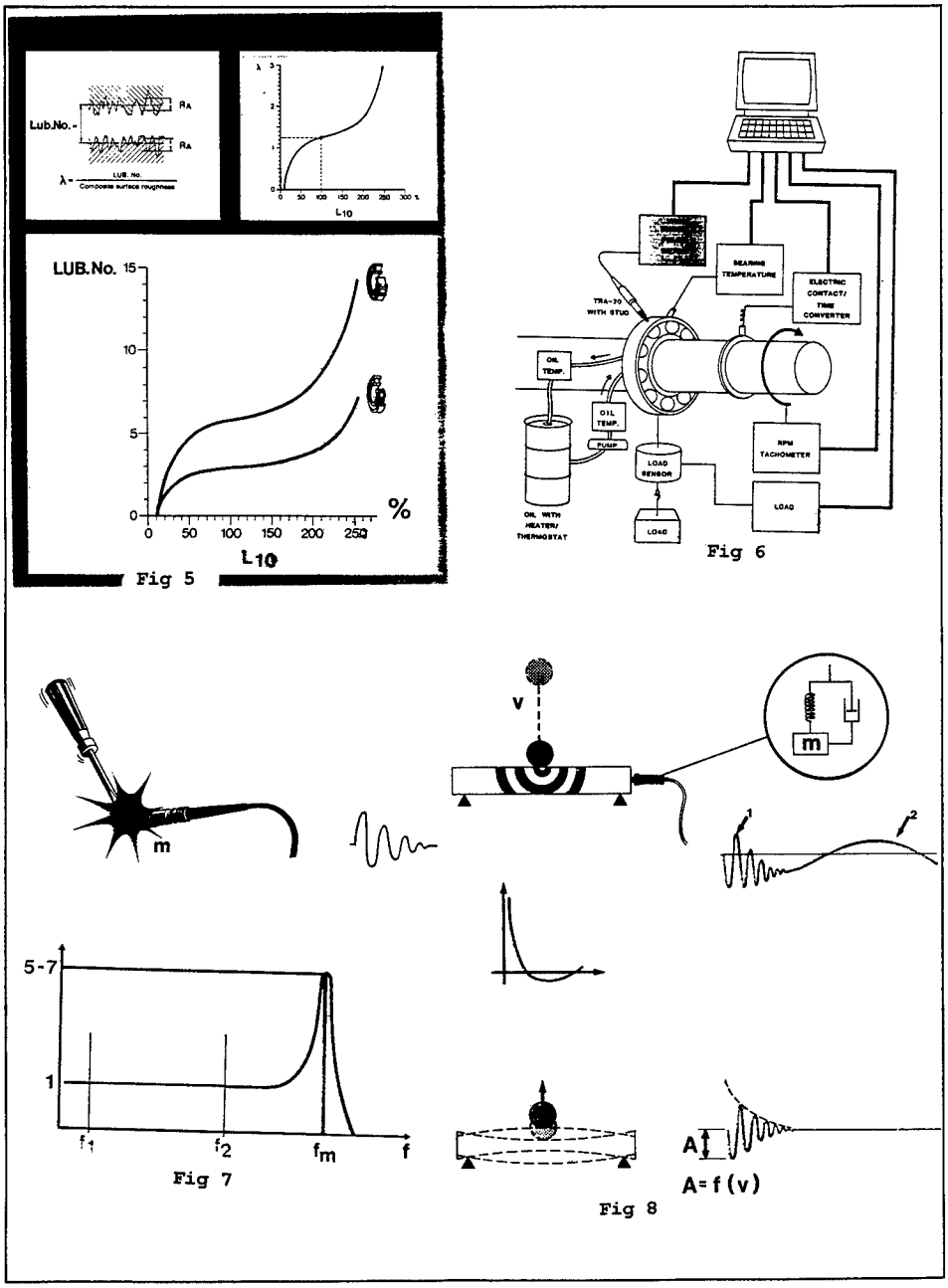
temperature, quality and supply. Measurements taken included shock pulse intensity(decibel) and pattern, bearing temperature, and the amount of electrical contact through the bearing. Shock pulse measurements were initially collected from all types and sizes of ball and roller bearings lubricated with kerosene and then loaded. No interruption occurred in the electrical contact through the bearings. These readings, representing the bearing's shock levels under dry running conditions. When subtracted from their shock measurements obtained from lubricated bearings, the difference in shock value, delta dB, was related to the time of no current flow. The dB difference values were mathematically converted into units for measuring oil film thickness (LUB No.). The scale was selected so that one unit of LUB No. change is approximately one micro inch change in oil film thickness. **(Fig.6)**

A shock pulse transducer contains a reference mass (m) and responds with a dampened oscillation when hit by a shock wave, here produced by hitting the transducer with a screw driver. Attached to the reference mass is a piezoelectric crystal that produces a voltage when compressed by the movement of the reference mass. This voltage is proportional to the amplitude of the oscillation and thus to the energy of the shock wave. The principle is somewhat similar to the one used in the accelerometers for vibration measurement. Shock pulse transducers, however, are mechanically and electrically tuned to operate at their resonance frequency of 32 KHz (f_m) where the resulting bearing signal is strongest. This provides a very sensitive transducer with a very favorable signal-to-noise ratio. **(Fig.7)**

At the moment of impact between two colliding bodies, a pressure wave spreads through the material of both bodies. When the wave front hits the shock pulse transducer, it will cause a dampened oscillation of its reference mass (m). The peak amplitude (A) is a function of the impact velocity (v) at the moment of impact. So for Shock Pulse Technology the mass of the objects is not a factor in the accuracy. During the next phase of the collision, both bodies will start to vibrate. The frequency of this vibration is a function of the mass of the colliding bodies. The initial pressure waves transient quickly dampens out. It causes the reference mass of the shock pulse transducer to vibrate at its own resonance frequency of 32 KHz. The resulting voltage is a function of the vibration amplitude and thus proportional to the impact voltage.**(Fig.8)**

Shock pulse analyzers measure signal strength in dB_{SV} (decibel shock value) at two different levels. **HR** = High Rate of signal occurrence, the level at which 1000 shocks per second can be counted; and **LR** = Low Rate of signal occurrence, the level at which approximately 50 shocks per second can be counted. **HR**, **LR** and the difference between them (+dBsv), are used to evaluate the signal and determine the bearing's operating condition.

A microprocessor evaluates the signal. It needs input data defining bearing type and rolling speed. Rolling speed is calculated from rpm and mean diameter (D_m), and input as an SPM NORM number. Bearing type (ball, roller, single and double roll) is input as the SPM TYPE, number 1 through 8. Bearing Condition is described by a letter code (A, B, C and D) and by a Green-Yellow-Red scale. CODE A(Green) means a good bearing, B(Yellow) means a good bearing with severe LUB problems, C(Yellow) is a bearing with early damage and D(Red) is more severe damage. The LUB (lubrication) number is the relative oil film thickness so whenever it is displayed a major goal would be to maximize it and then maintain that LUB number over


Lub.No.-
Ra
Ra
LUB. No.
Composite surface roughness
λ
L10
LUB.No.
15
10
5
0
0 50 100 150 200 250
%
L10
Fig 5
BEARING TEMPERATURE
ELECTRIC CONTACT/ TIME CONVERTER
OIL TEMP.
OIL TEMP.
PUMP
LOAD SENSOR
RPM TACHOMETER
LOAD
LOAD
OIL WITH HEATER/ THERMOSTAT
Fig 6
m
v
m
1
2
5-7
1
f1
f2
fm
f
Fig 7
A
A=f(v)
Fig 8

operating time. The COND (Condition) number indicates the degree of surface damage and once it appears it must be tracked so as to pick up increasing damage severity. As a result, "True Condition Monitoring" is available for bearings in both portable and continuous monitoring systems. **(Fig.9)**

Use machine specifications to identify the location of bearings and their construction. Measurement points should be selected and prepared with care to be sure that future readings are repeatable. Shock pulses are generated in the interface between the rolling element and the raceway, in the loaded part of the bearing. From the outer raceway, shock pulses are transferred to the bearing housing. To select a measurement point, the signal path must contain only one mechanical interface, that between the bearing and the bearing housing. The signal path between the bearing and the measurement point must be straight, solid, should be in the loaded region of the bearing housing. With the many bearing applications it is not always possible to meet these requirements and signal dampening occurs. In 1992 newly developed Shock Pulse Methodology eliminated this problem and now provides even greater accuracy for analysis. This dampening phenomenon actually has many causes. Low RPM, low load, housing material porosity are but a few causes. One quickly realizes that weakened signals are prevalent on most good bearings. Through advancements from SPM R & D, a bearing location can now be "initialized" and the analyzer will automatically calibrate the signal loss for a good bearing. This is accomplished by SPM patented empirical data calculating a compensation number corresponding to the degree of signal loss. When dampening is detected then the compensation number will provide a more accurate evaluation. This new technology is built into the all portable and continuous monitoring systems and in SPM LUBMASTER and CONDMASTER software. **(Fig.10)**

Automatic Bearing Evaluation is the basis of the Shock Pulse Method. With "True" SPM Condition Monitoring, the user need not collect multiple spectrums or have to evaluate them. The portable shock pulse analyzer or the continuous monitoring systems provide direct condition monitoring (e.g., "bearing severely damaged"); Green-Yellow-Red Condition Analysis. **(Fig.9)**

The major benefit of the Shock Pulse Method is providing a direct indication of bearing condition on a Green-Yellow-Red scale. A single reading can provide reliable values for oil film thickness and surface condition. This is very important when monitoring installed bearings on machines for which there are no trends or comparable readings. Performing Condition Monitoring includes making a decision on the type of maintenance required. With bearings, there are three common possibilities: 1) Satisfactory oil film, no surface damage, no special maintenance required; 2) Thin oil film and reduced life expectancy, the questions that must then be asked are: Can the oil film be improved? How can that be done? Is it worth the effort? and 3) Bearing damage, the bearing has to be replaced, the question is when. **(Fig.11)**

With the Shock Pulse Method a CODE Letter D normally indicates bearing damage. Surface damage is irreversible and will increase with time. Whenever verified readings show bearing damage, shorten the measuring interval and trend the Condition No. While it is not possible to say exactly when a damaged bearing will fail, as a guide: Condition No.'s > 45 infer a high risk for failure. Rapidly increasing Condition No.'s also mean a high failure risk. Both factors together should lead to an immediate shutdown and bearing replacement. **(Fig.12)**

A lubrication test can be performed on bearings suspected of bearing damaged. To test grease lubricated bearings for possible causes of high shock values, relubricate with clean grease. Measure the LR before lubricating, then shortly after lubricating, and again a few hours later. If the shock pulse value drops, then rises again, the most likely cause is bearing damage. If the shock pulse value drops and remains low, the contaminated lubricant has been replaced by clean lubricant, with no bearing damage. Lubricate again to see if the LR drops farther. If it does not, then, record the LUB number. If it increases after relubrication then the bearing has been overlubed and the previous measured LUB number is the maximum the bearing can maintain. If the shock pulse values are not affected after lubrication, suspect the signal is mechanical in origin and the hydraulics of the lubricant can not overcome the mechanical forces. Look for misalignment in the drivetrain or in the bearing installation. We know the relative thickness of the lubricant film has a direct influence on bearing service life. In order to achieve better lambda ratios and hence longer service life, the relative film thickness must be maintained and increased when practical. Relative film thickness (as measured by the SPM LUB Number) is influenced by many factors. **(Fig.13)**

The CODE Letter A relates the measured film thickness to a calculated value that the bearing should have according to EHD lubrication theory. CODE Letter B shows that film thickness is less than it should be, and that improvements are possible. Similarly, decreasing LUB No.'s indicate a deterioration of the oil film. In reality, few bearings operate in an environment that allows it to reach a LUB number or a film thickness necessary for full rated bearing life. RPM, load, temperatures are some of the factors that could prevent this. With the SPM Shock Pulse Method you can at least see the Lubrication Condition immediately so as to try to maximize the LUB number to whatever value the application allows. With a maximized LUB number you now have a standard for a lubrication schedule. **(Fig.14)**

To improve lubricant film, it is necessary to understand the factors which influence film thickness. Obviously, a certain amount of lubricant has to reach the area of rolling contact in order to supply the film. The thinner the film, the higher the HR level signal (dBsv), measured in the bearing. **(Fig.15)**

In many cases a supply of fresh grease to the bearing will improve its LUB number or restore it to the previous level. This can have a dramatic effect on bearing life. As the graph shows, increasing the LUB number of a roller bearing from 2 to 8 can extend bearing life by a factor of 10, provided the surfaces were not already affected during the period of near dry running. It is therefore important to lubricate regularly and at optimum intervals, so that the oil film never becomes too thin. **(Fig.16)**

The Shock Pulse Method allows oil condition monitoring over the whole speed range, from low to high rpm. The development of film thickness with speed in this roller bearing goes through three distinct stages: (1) At LUB No.'s below 5 there is metallic contact between the surfaces in this bearing for NORM numbers below 30. The speed at which surface contact ceases differs with bearing type, but lies mostly between NORM numbers 30 and 40. In the low speed range, oil film thickness cannot become greater than the composite surface roughness. Hence the only chance to

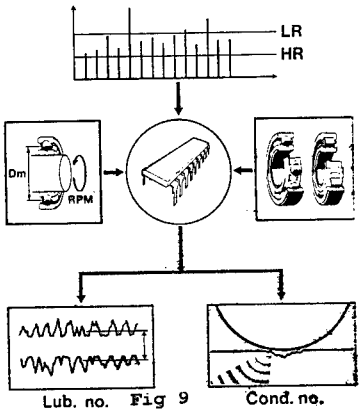


Fig 9

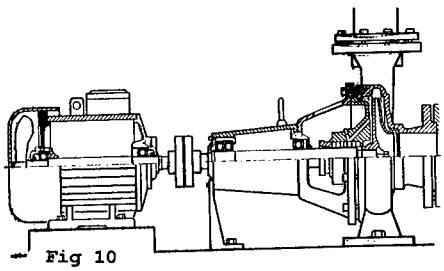


Fig 10

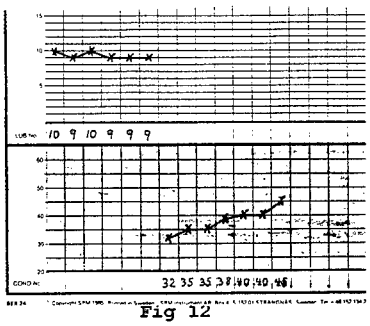


Fig 12

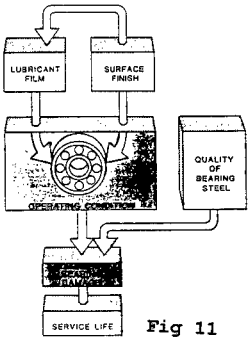


Fig 11



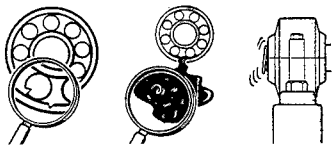


Fig 13

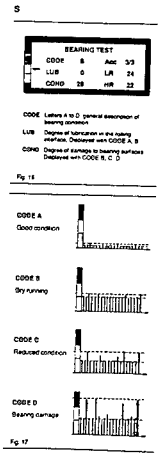


Fig 14

improve film thickness is by using a lubricant with higher viscosity. (2) The bearing surfaces are completely separated by an EHD film, which rapidly increases with speed. For bearings with NORM numbers between 30 and 50 it is normally easy to achieve good lubrication. (3) If speed is further increased, lubricant starvation could set in and oil film thickness tends to drop. The Shock Pulse Method can provide this lubrication condition without the user having to develop his own knowledge base. **(Fig.17)**

Case Study #1- Paper Mill in Oklahoma. Readings taken on 12 consecutive dryer can bearings. Bearings are spherical roller, P.N. 22138 at a speed of 75 RPM. On this particular demonstration a shock pulse analyzer was programmed with the physical information of the bearings and a 30 second reading was taken on each housing at the same location on each housing. The first readings taken_show marginal LUB numbers on locations 1 through 4 and 9 through 12. Locations 5 through 8 showed good lubrication condition. In all cases, the Shock Pulse Method reported that all the bearings were in good condition and without damage, then it displays where there is room for improvement with lubrication. At first it was thought that the lubrication lines to these bearings might be blocked. This was quickly eliminated as a possibility when the maintenance manager expressed agreement with the meaning of the shock pulse readings. It seemed he had for years felt that the lubrication rate was inadequate for the rpm. A few years earlier, the machine speed had been increased with a corresponding increase in oil temperature and bearing failures. Increasing the oil flow rate was not possible because undersized oil return lines caused such an excessive back pressure that oil start leaking through the seals and the flow rate had to be cut back. He requested that the oil return lines be modified and a test area was chosen to include the lines for the middle bearings, 5 through 8. He would have to prove the value of this repiping by monitoring the bearing failure rate between the different locations. He had made the piping changes and then increased the oil flow but he had no method to determine if there was any improvement in the bearing condition. Now the SPM Lubrication Number provided that proof. He then wanted to see if there was further improvement possible so he increased the drop rate slightly on location #5 and the SPM Lubrication Number indicated further improvement. Still more increase in drip rate and the SPM readings reversed themselves-identifying over lubrication. By using the SPM Method the LUB Number or film thickness is optimized on each of the bearings and thereby the bearing life is improved. **(Fig.18)**

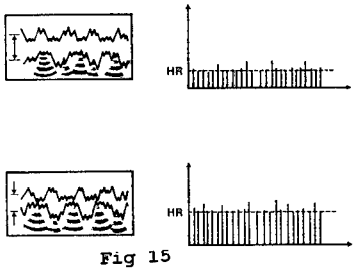
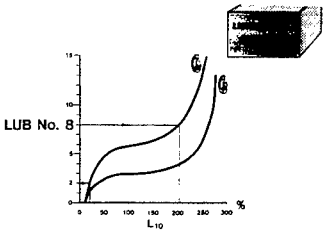


Fig 15



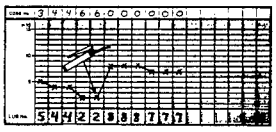


Fig 16

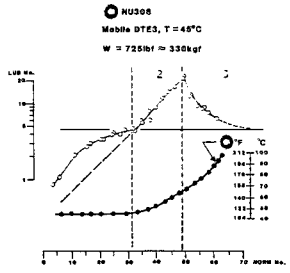


Fig 17

CASE HISTORY #1

Location: Paper Mill in Oklahoma Monitoring 12 Dryer can bearings. Bearing Part number 22138 running at 75 RPM

LOC #	1	2	3	4	5	6	7	8	9	10	11	12
LR	11	10	10	10	2	1	2	1	12	11	10	11
HR	7	7	7	6	-3	-3	-2	-3	7	7	6	7
CODE	A	A	A	A	A	A	A	A	A	A	A	A
LUB	1	1	1	1	4	4	4	4	1	1	1	1

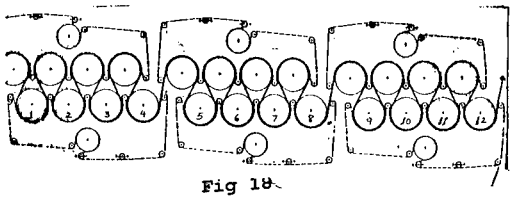


Fig 18

Reliability-Based Maintenance as a Breakthrough Strategy in Maintenance Improvement

by
Grahame Fogel and Dave Petersen

Introduction

Industrial plant maintenance is gaining attention as the next great opportunity for manufacturing productivity improvement. As companies invest in more high-tech and expensive equipment, they become more reliant on the need to reduce equipment redundancy without sacrificing reliability and availability, accomplishing this within an ever decreasing availability of operating capital. E I Dupont has said that maintenance was once its "single largest controllable cost opportunity, representing $100-$300 million per year corporate wide". It is estimated that U.S. Industry needlessly squanders in excess of $200 billion each year on inadequate or unnecessary maintenance procedures. Within the last ten years a wide range of advanced maintenance technologies have been developed which can help manufacturers reduce their maintenance costs while simultaneously increasing plant reliability.

Reliability-Based Maintenance (RBM) has emerged as perhaps the preferred advanced maintenance philosophy in North America. RBM was initially conceived as a solution advocating the logical balance between the four technical strategies of traditional maintenance: reactive, preventive, predictive, and proactive maintenance. Since the Reliability-Based Maintenance "recipe" has evolved to additionally include the appropriate technical strengths of Reliability Centered Maintenance (the RCM Process) and the people/work concept of Japanese-based Total Productive Maintenance (TPM). This broadened formula for Reliability-Based Maintenance has been driven by users and implementers in an effort to incorporate the tangible benefits of all advanced maintenance strategies and philosophies into a single deliverable solution. A historical perspective of the development of the various maintenance strategies follows.

Historical Discussion of Maintenance

In surveying the last sixty years there has been an enormous evolution in the sophistication of machinery used in production processes, mainly driven by the demand for increasing productivity as a competitive issue. This has led the equipment evolution from purely mechanical systems to precision electro mechanical systems with sophisticated computerized controls.

Pre-1930 machinery was robust, overdesigned, and long lasting. The major failure modes were wear or metallurgical. The maintenance plan was simple, machinery was rebuilt after failure by skilled craftsmen. In the 1950's productivity was becomming more of an issue. The prevailing maintenance philosophy was the belief that "machinery failure" was an accepted and unavoidable part of manufacturing life. This led to designing processes which had significant standby capacity and large spares inventory, with a strategy of ever increasing scheduled intervention (addictive maintenance). It was also a time of evolving relations between the workforce and management where Unions played a defining role, creating strict job definitions within the maintenance organization.

Until the early 1970's, most plants worldwide performed maintenance in a reactive, or breakdown, mode. Reactive maintenance is expensive because of extensive unplanned downtime and damage to machinery. With the availability of mainframe computers in the 70's, many companies implemented periodic preventive maintenance strategies to encourage planned maintenance inspection and repair in preference to reactive maintenance. This still dominant maintenance approach typically utilizes maintenance scheduling software to track and schedule calendar-based maintenance activities, and to automatically trigger required work orders. As the adoption of preventive maintenance grew, original equipment manufacturers habitually began to oversubscribe PM recommendations in an attempt to reduce their warranty exposure, thereby increasing overall maintenance costs with needless open-and-inspects.

As maintenance costs ballooned, a maintenance optimization procedure called Reliability-Centered Maintenance (RCM) was developed in the late 70's to help reduce the ever increasing volumes of work orders resulting from the implementation of computerized scheduling. The early RCM procedures were heavily influenced by safety issues because RCM has its origins in the airline industry. About the same time a maintenance philosophy called Total Productive Maintenance (TPM) was gaining momentum, particularly among Japanese manufacturers. TPM advocates a partnership between maintenance and operations departments such that basic maintenance activities (cleaning and inspections) are performed by operators. TPM has been adopted successfully in Asia and some parts of Europe, but it has suffered in North America due to union opposition.

In the mid 1980's, advances in instrument technology coupled with widening adoption of the personal computer provided the capability of "predicting" machinery problems by measuring machinery condition, using vibration, thermal, and ultrasonic sensors. This technology is commonly referred to as Predictive Maintenance (PDM), or condition monitoring. Another more advanced maintenance strategy called Proactive Maintenance (PAM) assists in further extending the failure cycles of plant machinery through the systematic removal of failure sources. Finally, in 1992, Reliability-Based Maintenance was introduced which effectively combines the strengths of all of the aforementioned strategies and philosophies into a single deliverable maintenance solution.

As we review the history of maintenance it is interesting to observe that prior to the early 1970's, the maintenance function was little changed since the beginning the industrial age. There were no improvement strategies developed, no re-engineering attempts, little investment and attention. The perceived purpose was to, first, repair things when they failed, and second, paint the parking lot and mow the grounds for visitors. Until recently, maintenance has always been perceived as a "necessary evil", beyond optimization and improvement.

Today, machinery is a complex hybrid of semiconductor controlled electro mechanical devices designed to operate with a much more demanding duty cycle. The maintenance manager in every manufacturing environment must now ask himself where he and his team stand in terms of whether they are sufficiently equipped, trained and organized to be effectuve abd competitive. Modern maintenance has to be in step with the demands of a much more sophisticated manufac-

turing environment. In order to succeed the basic philosophy of maintenance must continue to evolve in step with the changing demands of manufacturing and competition. An owner, in order to be competitive, requires maximum uptime from the machine operating at near its design capacity. And of increasing concern is the environmental aspects of the effect of plant operation.

Progressive companies are perceiving that maintenance is a worthy investment area, and as such the investment has to be carefully managed and measured for it's returns. From this, the ideas and practices of our breakthrough process have evolved. One difficulty has been the establishment of well accepted metrics of maintenance performance in order to apply tangible criteria against alternative competing investments.

Financial Impacts of Maintenance

Let us explore some of the more significant issues which the maintenance function impacts on a day-to-day basis. First, and most obvious, is production availability. Without 100% process and machinery uptime, we have less than 100% production availability, resulting in lower than planned sales. But can't lost production be made up on weekends? It certainly can if eroding margins are acceptable. Of course, one might consider shipping out of built up inventory so long as this doesn't conflict with company Just-In-Time manufacturing plans.

A second impact of maintenance is product quality. It stands to reason that well-balanced and well-aligned machinery and processes will produce a consistent, higher-quality product. But can't off-quality product be re-worked? It certainly can if eroding margins are of no concern. Of course, one might consider shipping the off-quality product anyway, so long as Quality isn't an issue with the company's customers.

Insurance premiums are another consideration. Many manufacturing facilities purchase "downtime insurance" in case of catastrophic failure. Some major insurance companies provide advanced high-tech maintenance services, the cost of which if utilized by the customer, is offset with lower insurance premiums. NRC regulations require "efforts to predict and prevent machinery failure" in broad terms, punishable by hefty fines. Energy consumption is still another consideration. Manufacturing facilities typically waste significant excess energy in operating poorly aligned and lubricated power transmission systems. How about safety and loss-time injury? Over 50% of loss-time injury accidents occur within maintenance, the majority of which result from the panic pressures of getting equipment back on-line after failure.

An important additional requirement for consideration today is minimal environmental impact from associated production processes. Within this new assignment, maintenance has enormous responsibilities. The consequences for failure cannot only be safety critical but can have enormous negative reputation implications for the corporation. One only has to recall Union Carbide's Bhopal, India disaster as a reminder.

The significant point made is that the implications of an advanced maintenance strategy (or the lack of one) are far-reaching within the corporation. The measured budget line item costs of maintenance typically range between 5-and 15% of total costs depending upon the process, but at what place in the company's income statement are the implications of reactive-type mainte-

nance on production availability, injury avoidance, power consumption, environmental regulations, and insurance premiums represented? Typically not in maintenance. As we begin to recognize the far reaching economic and fiscal impacts of maintenance, we must also recognize that we will need to define new ways to measure maintenance performance.

A Business Perspective of Maintenance

Businesslike demands are often made of the maintenance function, but seldom is the performance of maintenance measured from a business viewpoint. Most operations managers suggest that the maintenance function should be measured by the **uptime** parameter. Maintenance technicians, however, voice that they have little real control over uptime, and that "downtime" is more a result of excessive machinery abuse related to production demands rather than improper maintenance procedures. In fact, maintenance is more commonly measured by the "speed in which machines are back-up-and-running after catastrophic failure".

Another popular measure of maintenance performance is labor overtime. Logic should tell us that labor overtime is not a valid measure of performance. It is more a measure of nonperformance. Put simply, if we demand reduced overtime, we are not seeking improved performance, but instead are simply seeking a reduction in nonperformance.

Maintenance as it is typically measured is a "zero-sum-gain", and in a zero-sum world, all one can do is hope to break-even. Even if maintenance aggressively manages their expenses and comes in under budget, the question has to be asked "What was sacrificed?"..... Availability,...Quality,...Capacity? To reiterate, if these parameters are not measured in relation with each other, how can one truly quantify positive gains in a zero sum world ?

World-class manufacturers monitor their performance by the parametric measures of quality (ergo Deming's influence), cost ("value" is the politically correct term), and delivery (just-in-time). Maintenance should be measured similarly. Maintenance manufactures capacity, so world-class maintenance organizations should be measured against their capacity quality, capacity costs, and capacity delivery.

No company can be a world class competitor if it's factories are not up to the task. It is therefore an issue of the highest strategic importance to create a maintenance function which provides maximum capacity and availability at optimum costs. The concepts of Reliability Based Maintenance provide a framework to achieve "**Breakthrough**" within the maintenance function, measured by a set of metrics consistent with other critical business performance measures within the organization. This marks the beginning of the recognition that maintenance provides a significant opportunity for manufacturing productivity improvement.

Reliability-Based Maintenance

An effective Reliability-Based Maintenance operation is not just a well run refined predictive maintenance effort, but a new philosophy which forces fundamental shifts in the way maintenance is managed and measured. A first step towards breakthrough is to understand the function of maintenance. The function of maintenance in a world-class operating environment is

not to simply maintain, but to provide reliable production capacity and to extend the life of plant assets at optimum cost. The consequences of unreliable capacity are interrupted production schedules, lesser quality, and, most importantly, diminished profits. With reliable production capacity in mind, the most progressive manufacturers are restructuring their maintenance departments from specializing in reacting to breakdowns to organizing for the systematic elimination of machine failure, thereby increasing availability while minimizing maintenance costs.

Reliability-Based Maintenance is an advanced maintenance philosophy which prioritizes plant systems in terms of their impacts on capacity and availability, and forces the appropriate balance of reactive, preventive, predictive, and proactive maintenance strategies to insure maximum capacity and availability while minimizing costs. The seven primary breakthrough concepts advocated by Reliability-Based Maintenance include:

(1) Prioritization of plant systems and failure modes in terms of their impact on capacity and availability,
(2) A business decision of where to invest maintenance resources,
(3) An infusion of available maintenance technologies including preventive, predictive, and proactive technologies,
(4) An increase in the core competency of the maintenance function,
(5) A redefinition of the maintenance function whose mission is to pursue productivity and capacity improvement solutions, through "*Breakthrough*" practices,
(6) An increased awareness throughout the plant of the implications of maintenance decisions, and,
(7) An establishment and acceptance of suitable business performance metrics for maintenance.

Predictive Maintenance technologies combined with a state-of-the-art computerized maintenance management system are the catalyst of change in enabling a breakthrough change in maintenance practices. The implementation of Reliability-Based Maintenance requires a balance of reactive, preventive, predictive and proactive maintenance strategies. These reliability improvement strategies are not independent; they frequently draw upon each other's strengths in achieving reliable plant capacity.

Restructuring of the Plant Reliability Department

The adoption of RBM concepts over traditional maintenance concepts requires a significant redefinition of employee roles and responsibilities, combined with a restructuring of accountability with and between other plant departments.

First, RBM encourages a slow, but deliberate, migration to the work methods promoted by Total Productive Maintenance. This encourages the elimination of traditional organization lines between production and maintenance, lines which many times prevent operators and mechanics from taking immediate action to correct a simple problem. Operators should begin to assume responsibility for the performance of basic housekeeping activities including cleaning, routine inspection, and certain other tasks suggested by maintenance. This strategy further employs a

team approach to continuous improvement wherein production and maintenance systematically maintain and improve equipment effectiveness.

The five "pillars" of TPM as described by its founding father, Seiichi Nakajima, include:

1. Maximizing Equipment Effectiveness
2. Involving Operators in Daily Maintenance
3. Improving Maintenance Efficiency
4. Training to Improve Skill Levels
5. Emphasis on Maintenance Prevention.

A look at how RBM enhances these five pillars follows.

Maximizing Equipment Effectiveness

Improved equipment effectiveness can be achieved through "elimination of the five big losses", the largest contributor being **machine failure**. RBM strategies include predictive technologies which not only predict machine failure, but also proactive techniques which eliminate machine failure and extend life cycles. Additionally, RBM promotes extending machine operating conditions beyond design levels.

Involving Operators in Maintenance

One of the most valuable "predictive" indicators is derived from visual observations. Operations personnel are the closest to equipment, and the persons most likely to be aware of changes in machine or process condition. The RBM strategy emphasizes a close working relationship between maintenance and operations. Operator feedback to the RBM department provides vital information for the root-cause-failure-analysis cycle.

Improving Maintenance Efficiency

The RBM philosophy is based on the use of advanced maintenance technologies to eliminate unneeded preventive activities, and to refine the PM cycle. Predictive technologies dramatically improve efficiency. Additionally, RBM philosophies promote the concept of benchmarking and continuous improvement planning.

Training to Improve Skill Levels

RBM philosophies stress the elevation of skill levels. RBM forces an understanding of preventive strategies, predictive technologies, and proactive work methods. Additionally, predictive technologies often uncover the need for training in other areas such as downtime and change-out practices.

Emphasis on Maintenance Prevention

A strong benefit of RBM is that it forces the accumulation of machinery information and equipment histories. Chronic problems are identified and eliminated. Design deficiencies are identified, and the resulting information is used in future purchases of equipment. Proactive practices embody many maintenance prevention activities including improved purchase specifica-

tions, installation commissioning, and precision alignment, balancing, and lubrication methods.

TPM work methods are essential to Reliability-Based Maintenance. As operations assumes more and more housekeeping activities, valuable free time is created for the maintenance function so as to persurve reliability improvement activities. RBM promotes the development of a more highly-skilled, flexible workforce, and in doing so, optimizes the size and cost of the workforce.

Implementing Reliability-Based Maintenance

PHASE ONE - DISCOVERY

During the initial phase of discovery, the maintenance function must perform some benchmarking activities allowing assessment of current practices, performance measures, objective, and prevailing attitudes. Many have assumed that they can skip this phase and move directly to stabilization, but success rates drop dramatically. The significant point is that one cannot develop a plan of where to go without the knowledge of where one is. There are perhaps two dozen major indices which need to be analyzed during the discovery phase, the results of which produce a viable and workable continuous improvement plan for the difficult period ahead.

Philosophical Elements

- Organization Structure
- Prevailing Attitudes
- Work Methods
- Quality Measures
- Mission Effectiveness
- Performance Objectives and Measures

Procedural Elements

- Work Control Procedures
- Document Control Procedures
- Maintenance Prevention Procedures
- Training Plans
- Major Projects Procedures
- Spares Control Procedures

Technical Elements

- CMMS Program
- Predictive Technologies Program
- Proactive Maintenance Program
- Reactive Maintenance Program
- Preventive Maintenance Program
- Technical Training
- Balance of Maintenance Strategies

A new mission is defined for maintenance with bottom-up and top-down buy-in; coupled with a new set of business-based performance measures.

PHASE TWO - STABILIZATION

The first step of stabilization is to create two key dedicated core functions; 1) Maintenance Planning Group, and 2) Reliability Improvement Group.

These two groups are the hubs for the Reliability-Based Maintenance strategy. The Maintenance Planning Group may exist as a preventive maintenance group already; however its focus needs to be changed from work order generation and tracking to full scope maintenance planning.

Responsibilities of Preventive Planning Group

- Coordination of all preventive maintenance,
- Coordination of minor housekeeping activities performed by operators.
- Coordination of maintenance activities withproduction to obtain the least impact on availabil ity.
- Maintenance work planning, including procedures, tools, parts, inspections, and calibration.
- Work order tracking and cost tracking.
- Machine history file.
- Evaluation of life cycle trends for possible machinery improvement needs.
- Spare parts management.

The Reliability Improvement Group is initially focused on the implementation of predictive technologies in the plant, but also moves toward proactive strategies as the group matures. This group is usually made up of experienced machinery mechanics or technicians, chosen for their machinery maintenance ability as well as their capacity to learn new, advanced maintenance techniques. The typical Reliability Improvement Group at a large plant starts with one or two vibration technicians, one thermography technician, one lubricant engineer and one maintenance engineer. The group should be chosen from existing plant staff and trained to become experts in their area of specialty. At this stage the importance of training cannot be overemphasized.

The Responsibilities of the Reliability Improvement Group

- Operating a self-directed predictive maintenance team integrating all predictive maintenance technologies.
- Providing timely component condition evaluations to Maintenance Planning, with the goal of eliminating all unplanned downtime.
- Performing ongoing SERP analysis.
- Implementing proactive maintenance technologies and methods, including identification of recurring prob lems.
- Tracking performance measures e.g. savings from single event analysis, uptime/downtime, maintenance costs, quality, etc.
- Identification and elimination of unnecessary preventive maintenance activities.
- Identification of equipment design changes to improve reliability.
- Management of the plant's computerized maintenance management system.

More predictive maintenance technologies are added during the second and third year of stabilization. During this stage, machinery condition information is thoroughly integrated with preventive maintenance and proactive technologies, including precision alignment, balancing, and root cause failure analysis. The Reliability Improvement Group will grow to several dozen

people for a large industrial facility, while the preventive planning group may shrink by 30% or more. Also during this stage, the plant attitude toward maintenance begins to change. Management promotes the elimination of breakdowns and employee attitudes change significantly toward predicting and eliminating problem sources.

The performance measurement of the Reliability Improvements Group begins to shift from single event analysis to global improvement indicators.

PHASE THREE - BREAKTHROUGH

This stage is most noteworthy for the attitude change which has occurred within the maintenance department.

Benchmark plants stand out strongly against industrial norms in terms of management and employee emphasis on Reliability-Based Maintenance philosophies. In this stage of aggressive implementation of the strategy, the plant is entirely focused on eliminating breakdowns, permanently eliminating machinery and quality problems, and achieving nameplate (theoretical) productivity or better. The performance measurements for the success of the strategy focuses on shift bottom line measurements consistent with the plant's business - maximizing operating and maintenance costs, and striving for greater profitability and working capital.

Procedures for all aspects of the maintenance business are developed and documented including: document control procedures; major project procedures; and a dynamic flexible master work control procedure.

The work methods within the plant have evolved to a point of full partnership between maintenance and operations in terms of maximizing reliability and controling costs. The Reliabilty Group is now the Reliability-Based Maintenance Department operating with full responsibility

APPENDIX

MFPG/ MFPT PUBLICATIONS

Both printed and microfiche copies of the following MFPG publications whose catalog numbers start with either **AD**, **COM** or **PB** may be obtained from the

National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161

Glossary of Terms AD 721 354

MFPG Proceedings

Meeting Nos. 1-9 (set of five) AD 721 359

Meeting Nos. 1-5
Papers and Discussion on Failure Analysis and Control

Meeting No 6 "Detection, Diagnosis and Prognosis"
December 6, 1968

Meeting No 7 "Failure Mechanisms as Identified with Helicopter Transmissions"
March 27, 1969

Meeting No 8 "Critical Failure Problem Areas in the Aircraft Gas Turbine Engine"
June 25-26, 1969

Meeting No 9 "Potential for Reduction of Mechanical FailureThrough Design Methodology"
November 5-6, 1969

Proceedings

Meeting No 10 "Vibration Analysis Systems" AD 721 912
January 21-22, 1970

Meeting No 11 "Failure Mechanisms: Fatigue" AD 724 475
April 7-8, I970

Meeting No 12 "Identification and Prevention of Mechanical Failures
in Internal Combustion Engines" AD 721 913
July 8-9, 1970

Meeting No 13 "Standards as a Design Tool in Surface Specification
for Mechanical Components and Structures" AD 724 637
October 19-20, 1970

Meeting No 14 "Advances in Decision-Making Processes in Detection,
Diagnosis and Prognosis" AD 721 355
January 25-26, 1971

Meeting No 15 "Failure Mechanisms: Corrosion" AD 725 200
April 14-15, 1971

Meeting No 16 "Mechanical Failure Prevention Through Lubricating Oil Analysis" AD 738 855
November 2-4, 1971

Meeting No 17 "Effects of Environment Upon Mechanical Failures, Mechanisms and Detection" AD 750 411
April 25-27, 1972

Meeting No 18 "Detection, Diagnosis and Prognosis" AD 772 082
November 8-10, 1972

Meeting No 19 "The Role of Cavitation in Mechanical Failures" (NBS SP 394)
October 31-November 2, 1973 COM-74-50523

Note: If PB numbers are not listed, NTIS can identify proceedings by NBS SP number.

Proceedings

Meeting No 20 "Mechanical Failure - Definition of the Problem" (NBS SP 423)
May 8-10, 1974

Meeting No 21 "Success by Design: Progress Through Failure Analysis" (NBS SP 433)
November 7-8, 1974 PB 256 770

Meeting No 22 "Detection, Diagnosis and Prognosis" (NBS SP 436) PB 248 254
April 23-25, 1975

Meeting No 23 "The Role of Coatings in the Prevention of Mechanical Failure" (NBS SP 452)
October 29-31, 1975 PB 257 422

Meeting No 24 "Prevention of Failures in Coal Conversion Systems" (NBS SP 468)
April 21-24, 1976 PB 265 552

Meeting No 25 "Engineering Design" (NBS SP 487) PB 271 743
November 3-5, 1976

Meeting No 26 "Detection, Diagnosis and Prognosis" (NBS SP 494) PB 272 848
May 17-19, 1977

Meeting No 27 "Product Durability and Life" (NBS SP 514) PB 281 878
November 1-3, 1977

Meeting No 28 "Detection, Diagnosis and Prognosis" (NBS SP 547) PB 297 399
November 28-30, 1978

Meeting No 29 "Advanced Composites" (NBS SP 563)
May 23-25, 1979

Meeting No 30 "Joint Conference on Measurements and Standards for Recycled Oil/SystemsPerformance and Durability" (NBS SP 584)
October 23-26, 1979

Meeting No 31 "Failure Prevention in Ground Transportation Systems" (NBS SP 621)
April 22-24, 1980

Meeting No 32 "Detection, Diagnosis and Prognosis: Contribution to the Energy Challenge" (NBS SP 622)
October 7-9, 1980

Meeting No 33 "Innovation for Maintenance Technology Improvements" (NBS SP 640)
April 21-23, 1981 (not available from NTIS)

Meeting No 34 "Damage Prevention in the Transportation Environment" (NBS SP 652)
October 21-23, 1981

Printed copies of the following MFPG publications are available from

Cambridge University Press
110 Midland Avenue
Port Chester, NY 10573

Proceedings

Meeting No 35 "Time Dependent Failure Mechanisms and Assessment Methodologies"
April 20-22,1982

Meeting No 36 "Technology Advances in Engineering and Their Impact on Detection, Diagnosis and Prognosis Methods"
December 6-10, 1982

Meeting No 37 "Mechanical Properties, Performance and Failure Modes of Coatings"
May 10-12, 1983

Meeting No 38 The Proceedings was not published because of the format of the meeting.

Meeting No 39 "Failure Mechanisms in High Performance Materials"
May 1-3, 1984

Meeting No 40 "Use of New Technology to Improve Mechanical Readiness, Reliability and Maintainability"
April 16-18, 1985

Meeting No 41 "Detection, Diagnosis and Prognosis of Rotating Machinery to Improve Reliability, Maintainability, and Readiness Through the Application of New and Innovative Techniques"
October 28-30, 1986

Meeting No 42 The Proceedings was not published. Inquire at the Vibration Institute regarding the purchase of copies of individual papers.

Meeting No. 43 "Advanced Technology in Failure Prevention"
October 3-6, 1988

Printed copies of the following MFPG and MFPT publications may be ordered from the

Vibration Institute
6262 S. Kingery Hwy
Suite 212
Willowbrook, IL 60514

Proceedings

Meeting No 44 "Current Practices and Trends in Mechanical Failure Prevention"
April 3-5, 1990

Meeting No. 45 "Focus on Mechanical Failures: Mechanisms and Detection"
April 9-11, 1991

Meeting No. 46 "Economic Implications of Mechanical Failure Prevention"
April 7-9, 1992 **(No longer available.)**

Meeting No. 47 "The Systems Engineering Approach to Mechanical Failure Prevention"
April 13-15, 1993

Meeting No. 48 "Advanced Materials and Process Technology for Mechanical Failure Prevention"
April 19-21, 1994

MFPT Proceedings

Meeting No. 49 "Life Extension of Aging Machinery and Structures"
April 18-20, 1995

Meeting No. 50 "Technology Showcase - Integrated Monitoring, Diagnostics & Failure Prevention" (A Joint Conference)
April 22-26, 1996