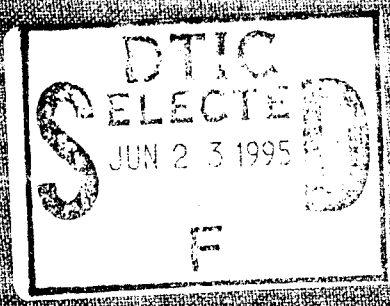
LIFE EXTENSION OF AGING MACHINERY AND STRUCTURES



This document has been approved for public release and sale; its distribution is unlimited.

LIFE EXTENSION OF AGING MACHINERY AND STRUCTURES

Proceedings of the 49th Meeting of the Society for Machinery Failure Prevention Technology

Virginia Beach, Virginia
April 18-20, 1995

Compiled by
Henry C. Pusey
and
Sallie C. Pusey

Accesion For
NTIS CRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification .....
By .....
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1

DTIC QUALITY INSPECTED 3

A Publication of the

Vibration Institute
A NOT-FOR-PROFIT CORPORATION

19950619 112

Copyright © 1995 by
Vibration Institute
6262 S. Kingery Highway
Willowbrook, Illinois 60514
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.

TABLE OF CONTENTS

PREFACE

The 49th Meeting of the Society for Machinery Failure Prevention Technology (MFPT) was sponsored by the Office of Naval Research (ONR), Arlington, VA; Naval Surface Warfare Center (NSWC), Annapolis, MD; U.S. Army Research Laboratory (ARL), Watertown, MA and the Vibration Institute. The conference was held April 18-20, 1995 at the Cavalier Hotel in Virginia Beach, Virginia. Meeting management, program coordination, and proceedings compilation were by the Vibration Institute. MFPT Council Chairman **G. William Nickerson** chaired the Opening Session. The CoChairmen of the Technical Sessions are identified on the title pages for each section in these proceedings. The MFPT Council and the MFPT Program Committee Members are listed separately.

The reader will note that our professional organization is no longer called the Mechanical Failures Prevention Group. Our new name more accurately reflects our structure, objectives and activities. The Technical Committees have not changed significantly, except that they have become more active. The MFPT Society is dedicated to *Achieving Competitive Advantage by Linking Failure Mechanisms and Diagnostics.*

The 1995 THEME was first conceived in response to the concern about aging aircraft and the potential for catastrophic events caused by failure of aircraft structure or machinery. In order to maintain a competitive edge, the airlines must ensure the longest possible service life for their expensive aircraft without compromising passenger safety. There is a logical parallel to this life extension philosophy in the wide variety of machinery used in numerous industrial and manufacturing plants, the electric power industry, transportation systems, maritime systems and many more. The objective of this conference was to provide a forum for the exchange of technical information on ways to improve the competitive posture of American industry by the effective application of available failure technology.

The Keynote Speaker was **Dr. Bernard S. Loeb**, Director of the Office of Research and Engineering, National Transportation Safety Board; his topic was *Prediction of Mechanical Failures: Lessons Learned from Transportation Accidents.* Dr. Loeb was followed in the Opening Session by four distinguished invited speakers. **Mr. Richard H. Menke** from Eli Lilly and Company presented *A Manufacturing Perspective of Reliability.* **Dr. Michael L. Basehore**, Federal Aviation Administration Technical Center, described a program on *Providing for the Continuing Airworthiness of Commuter Airplanes.* **Dr. Steven B. Chase** of the Federal Highway Administration spoke on *Developing New Tools and Techniques for Bridge Inspection of the Future.* **Mr. Richard F. Healing**, Director of Safety and Survivability, Office of the Secretary of the Navy, discussed the Navy's need for *Condition Based Maintenance Equipment with Potential Application to Fleet Safety and Survivability.* There were three Plenary Lectures: **Professor MarjorieAnn E. Natishan** of the University of Maryland on *Learning From Failure*, **Mr. Helmut Thielsch** of Thielsch Engineering on *Life Extension Considerations in Failure, Failure Prevention and Repair* and **Professor Marvin L. Roush** of the University of Maryland on *The Evolution of Reliability Engineering and Its Relationship to Life Extension* . The papers by Professors Natishan and Roush appear in the **Featured Papers** section of these proceedings. Regrettably, none of the other invited papers were available for publication.

Two half-day Mini Courses were presented. One of these provided a basic *Overview of Mechanical Diagnostics Technologies.* Featured lecturers included **Dr. Ronald L. Eshleman**, Vibration Institute, on *Vibration Analysis*; **Mr. Larry A. Toms**, Technical Services, on *Oil Analysis*; **Mr. Craig K. Kelch**, Infraspection Institute, on *Infrared Analysis*; **Mr. Steve F. Smith**, Oak Ridge National Laboratory, on *Motor Current Signature Analysis* and **Professor C. James Li**, Rensselaer Polytechnic Institute, on *Advanced Diagnostic Techniques.* The second course was

presented by **Dr. Rudolph J. Scavuzzo** of the University of Akron. He lectured on *Failure Mechanisms in Polymers - Part II*, a continuation of his Mini Course at the MFPG 48 meeting.

There were four presentations at the conference that are not included in the proceedings because manuscripts were not available in time for publication. The papers are:

Structure-Based Connectionist Network for Diagnosis of Helicopter Gearboxes: K. DANAI and V. JAMMU, University of Massachusetts, Amherst, MA and D. G. LEWICKI, NASA Lewis Research Center, Cleveland, OH

Life Extension by Design: M. W. POE, Mid-Atlantic Associates, Ltd., Lake Ridge, VA

Coin Tossing as a Paradigm for Failure Prediction: J. P. CUSUMANO, Pennsylvania State University, State College, PA

An In-Situ Sensor to Measure Contact Area and Surface Stresses Between Real Tribological Bodies: M. D. BRYANT and J. W. LIN, University of Texas, Austin, TX

New Methods in Time-Frequency Analysis: L. COHEN, Hunter College and Graduate Center, New York, NY

The Cohen paper was a ten minute introduction to the session on **Time Frequency Analysis**. Those interested in the above papers should contact the appropriate author(s).

The Mechanical Failures Prevention Group was organized in 1967 under the sponsorship of the Office of Naval Research for the purpose of stimulating and promoting voluntary cooperation among segments of the scientific and engineering communities in order to gain a better understanding of the processes of mechanical failures. The goals were to reduce the incidence of mechanical failures by improving design methodology, to devise methods of accurately predicting mechanical failures and to apply the increased knowledge of the field to the problems of our present technology. Through the activities of its Technical Committees it is anticipated that the MFPT Society will continue to act as a focal point for any technological developments that contribute to mechanical failure reduction or prevention. MFPG 48 was the first occasion for scheduled open meetings of the four Technical Committees. The attendance was good and the discussions were lively. Committee meetings were again programmed for MFPT 49; the registrants were provided with meeting agendas. As a result of their deliberations specific goals and objectives for each committee will be established and committee activities will be planned. Those interested in working on any of the Technical Committees should contact the appropriate committee chairman. They are identified on the next page of these Proceedings.

On behalf of Dr. Eshleman and the Vibration Institute, I want to thank our sponsors and the MFPT Council for their cooperation in organizing and conducting the 49th MFPT Meeting. We are exploring some exciting possibilities for the future and fully expect that our conferences will continue to provide an effective forum for those who have mechanical failure problems and those who are concerned with machinery failure prevention technology.

Henry C. Pusey
Executive Secretary

MFPT COUNCIL

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

Vice Chairman
MarjorieAnn E. Natishan
University of Maryland
Mechanical Engineering Dept.
College Park, MD 20742

Past Chairman
Henry R. Hegner
ECO, Inc.
1356 Cape St. Claire Center
Annapolis, MD 21401

Executive Secretary
Henry C. Pusey
Consultant-Vibration Institute
4193 Sudley Road
Haymarket, VA 22069

TECHNICAL COMMITTEE CHAIRMEN

Diagnostics and Prognostics
Howard A. Gaberson
NAVFAC Engrg Svc Center
Code 60
Port Hueneme, CA 93043

Failure Analysis
Victor K. Champagne
U.S. Army Research Laboratory
AMSRL-MA-CB-292
Watertown, MA 02172-0001

Life Extension and Durability
Neville F. Rieger
Stress Technology, Inc.
1800 Brighton-Henrietta Town Line Road
Rochester, NY 14623

Sensors Technology
Henry R. Hegner
ECO, Inc.
1356 Cape St. Claire Center
Annapolis, MD 21401

TECHNICAL PROGRAM COMMITTEE

Chairman

MarjorieAnn E. Natishan

Members

Robert A. Bayles
Naval Research Laboratory
Code 6312
Washington, DC 20375-5000

Victor K. Champagne

Howard A. Gaberson

Henry R. Hegner

Paul L. Howard
Paul Howard Enterprises
1212 Clearbrook Road
West Chester, PA 19380

Thomas Moran
Zimmerman Associates
2361 Jefferson Davis Highway
Arlington, VA 22202

Kam Ng
Office of Naval Research
Code 122R4, 800 N. Quincy Street
Arlington, VA 22217

Marc Pepi
U.S. Army Research Laboratory
AMSRL-MA-CB-292
Watertown, MA 02172-0001

John Reintjes
Naval Research Laboratory
Code 5642
Washington, DC 20375-5000

S. Nils Straatveit
Purvis Systems, Inc.
Two Shaws Cove
New London, CT 06320

FEATURED PAPERS

LEARNING FROM FAILURE

MarjorieAnn E. Natishan
Mechanical Engineering Department
University of Maryland
College Park, MD 20742

Abstract: *We can, and must, learn from failures of machinery and structures! We must seek to understand how and why things fail because it is the information gained from failure analyses that allows us to redesign, to make more safe, reliable and economical, the technologies that surround us. It is through understanding failures that we can also advance technology. Understanding why things fail allows us to identify the bounds of current engineering designs. It enables us to stretch those boundaries through a more creative approach to the design and redesign process. Understanding failure mechanisms also allows us to stretch expected service life safely and economically. This paper will present some case histories and their subsequent analyses, including what was learned and implemented.*

KEY WORDS: Failure analyses; Service-life extension; Mechanical design; Condition-based maintenance

Introduction:

For as long as there have been engineered structures, systems and components there have been failures of those structures, systems and components. In most cases failure of machinery or structural components is inconvenient. In some cases it creates a safety hazard. In some cases failure has a strong economic impact. But in all cases understanding what caused the failure provides insight to improving the design and how we use or maintain the system. Unfortunately we tend to learn more through failure than through success. When something fails we learn quickly the limitations of the design. Success tells us little of the limitations! Success only makes us see possibilities for stretching that design success further and results in decreased conservatism. It is only in failure that design limitations are defined and it is typically because of failure that the approach to design is altered. We can learn much from probing failures to their root cause and we can and should implement the knowledge gained to improve designs, increase safety and reliability and decrease manufacturing and operating costs of our technologies.

When to Conduct a Failure Analysis:

Failure analysis tools and methodologies have progressed to a point of becoming somewhat routine, if still relatively expensive. Because of this failure analyses are often conducted as a matter of policy instead of on a need basis. I think this is something of which we also need to be aware. Failure analyses should not ALWAYS be conducted. There are situations in which failure of a particular component has no major impact on

our safety, reliability or economic status. Failure of one component simply means the part is removed, thrown away and replaced with a new part. . . planned obsolescence if you will! If replacing a part is quick, cheap and simple then perhaps spending time on an analysis of the cause of failure is best NOT done. It would not be cost-effective! But where failure has a major impact on our operations then understanding the cause of failure is a requirement!

Why a Failure Analysis Should be Conducted:
Before conducting any failure analysis we must first determine WHY an analysis is to be performed. What will be done with the information gained from a failure analysis? Will the component or system be redesigned? Will the failure mode information be used to model the failure such that time to failure can be predicted for other, similar designs allowing implementation of more effective maintenance schedules? Or is a better understanding of how the component behaves in it's service environment the goal to allow for extension of service life? You must know WHY you want to conduct a failure analysis before defining the analytical steps required for your particular analysis.

How to Conduct A Failure Analysis:
Many books and articles have been written and short courses developed to teach the basic methodologies of conducting a failure analysis. The basics are very straight forward. If a failure analysis has been decided upon then proceeding from the least destructive to the most destructive analysis techniques is critical to preserve evidence! It is imperative that evidence be cataloged and all steps and parts fully documented (written and photographed or diagrammed) to preserve a thorough and careful record. You do NOT want to destroy any evidence before all possible information is gained. This is the basic premise of conducting an analysis of failure. With this in place there are many different techniques available for the analysis and someone trained in conducting failure analysis will know which are most suited for a particular situation.

Once a root cause of failure or a failure mechanism has been established the objective for conducting the failure analysis can be met. There are three basic areas that can be impacted by an analysis of failure that I'd like to discuss today;

1. creating new approaches to design that include less conservatism,
2. redesign to improve safety and reliability, and
3. service life extension.

New Design Approaches
Based on historical successes with mechanical designs, engineers tend to try to push the limits of conservatism on new designs in an effort to improve functionality or decrease costs; building components that are lighter, stronger, faster, cheaper, etc. In many cases, especially before the common use of computers for modelling service conditions, it was difficult to gain a clear understanding about how the new design would behave in the real world application. It wasn't until a failure, and subsequent analysis, that the "real" behaviors, or limitations, of a new design were understood. The failure of the Tacoma Narrows Bridge in Washington is a classic example of this - and one I'm sure we've all watched films of in engineering school! In trying to make a more graceful, lighter weight suspension bridge engineers designed a stiffer, stronger, but thinner, bridge structure. Not clearly understood at the time was the effect of bridge thickness-to-width ratio on

aerodynamic behavior due to wind. Analysis after this famous failure showed that this aerodynamic response was indeed a very important factor in bridge design and now aerodynamic testing and analysis of new bridge designs is common place. But this new design criteria would not have been established had it not been for the failure analysis and willingness to incorporate what was learned into design practice. The film of that failure has been used in almost every engineering school I've been associated with as a tool for emphasizing the point of the necessity of learning from failure!

An analysis which changed the approach to mechanical design of aircraft was conducted on several de Havilland Comets that failed in the early 1950's. This was the first passenger aircraft designed to fly at altitudes great enough to require pressurization of the cabin. However, fatigue in general and fatigue cycling of the aircraft structures due to this pressurization in particular was not clearly understood, nor were the effects of square windows on concentration of stresses in a local area. Several of these aircraft failed in mid flight after only flying 900 to 1250 flights - well under the amount thought at the time to be required for enough accumulation of fatigue damage to cause failure. At first severe weather was blamed but after two failures in clear weather other causes of failure were pursued. Detailed analysis of as many pieces of one of the failed aircraft as could be recovered showed fatigue crack initiation at a rivet hole near the corner of one of the squarish windows on top of the aircraft. Cracking proceeded aft from the window to a circumferential stiffener. The crack then followed the stiffener circumferentially until the skin peeled off the aircraft causing catastrophic failure of the structure. Stress analysis of the window/rivet geometry showed that indeed stresses were increased in this location. Fatigue testing to simulate service conditions were then conducted to determine more precisely the effects of low cycle fatigue on the structure. This testing showed that fatigue failure could occur at even low numbers of cycles with this particular design. This information was used to redesign the aircraft to include only round windows and to increase the safety margin at rivet holes. Initiation of better inspection techniques during manufacture and routine maintenance also resulted from this analysis. But the information gained from this failure was also incorporated into designs of new aircrafts and, in fact, pushed the direction of aircraft design and prototype testing to incorporate the effects of low cycle fatigue due to cabin pressurization.

Failure of the Liberty Ships during the second World War also resulted in a changes in the approach to design of ships. Liberty ships were designed to be built quickly and at a low cost by using the new techniques of welding instead of rivets. Large sections of steel plate were welded together to form the hull sections and then these sections were circumferrentially welded together to form the ship hull. Unfortunately, ductile-to-brittle transition (DBT) behavior of the steel plates was known only slightly and the effects of welding on the temperature (DBTT) at which this transition occurred was not known at all. The other factor in these failures was the fact the welding technology and subsequent inspection of weld quality was not as good then as it is now. The fast pace of shipbuilding resulted in poor quality welds that had a DBTT well above that of the parent steel. This resulted in brittle, catastrophic failure of the hulls of these ships; some while sitting in port under no loading other than residual stresses. Analysis of these failures pointed up the shortcomings of welding processes at the time and the need for weld quality inspection techniques. Further research has certainly resulted in a better merging of the two technologies (welding and quality inspection). But another shortcoming

identified through analysis of these failures was in the understanding and ability to predict brittle fracture of steel. Development of the field of fracture mechanics and incorporation of fracture mechanics analysis into the design phase of components and structures came about largely because of the failure of the Liberty Ships. Test methodology to fully characterize brittle fracture characteristics of materials and stress analysis techniques to better predict brittle behavior conditions are now common practice. They were developed in an effort to understand material and structural behavior that lead to catastrophic failure in conditions that were not understood at the time of initial design and fabrication.

Redesigns:

Less famous than the failures mentioned above, perhaps, but useful and illustrative nonthe- less are several failures the occurred in machinery components of Naval vessels more recently. Analyses of these failures didn't result in far reaching changes in the approach to design but the information gained from failure analyses was incorporated into newer, more failure safe redesigns of the components.

Ferralium 255 is a duplex stainless steel that was developed to take advantage of the corrosion resistance of stainless steel, the strength of ferrite and the ductility of austenite. It is comprised of 50% ferrite comprising the matrix and approximately 50% austenite precipitates. It is a macroscopically homogeneous material. This material, when used as a casting, is typically heat treated to obtain it's duplex structure. It is aged in the dual phase region to precipitate austenite. This material was used in the manufacture of very large bow plane components for submarines. During fabrication of these components some fabrication welding, as well as some repair welding, was required. The repair welding was required to "fill out" some regions of large casting defects. During routine quality assurance testing regions adjacent to the welds were found to have unacceptably low ductility. Analysis of the microstructure of these large castings revealed that the localized heating from welding the components resulted in a very brittle, overaged condition in some regions of the castings. Heat treatment of these large components was ruled unfeasible due to their large size. Instead it was decided to redesign the components in an effort to eliminate the problems exhibited by the cast and welded components. The redesign included changing processing from casting to forging to eliminate the need for weld repairs and eliminating fabrication weld through a change in component geometry. This redesign successfully eliminated the danger of brittle failures in these components and has advanced knowledge of the behavior of Ferralium 255 under various heat treat conditions.

Failure of two hydraulic control valve actuator spools raised concerns regarding system reliability and prompted performance of a failure analysis to determine the causes and factors contributing to these failures. Of primary importance was the determination of whether these failures were caused by design or material selection problems (and were thus systemic) or were simply a "bad" lot of material. The analysis involved detailed fractographic and metallographic analysis of the failed spools to identify the failure mode and any metallurgical factors contributing to failure, and a fracture mechanics-based critical flaw size evaluation for this actuator spool material. Results of these analyses identified the cause of failure as a very small region of intergranular cracking which was too small to detect with current nondestructive inspection techniques but exceeded the

critical flaw size for this material for brittle, catastrophic failure. The cause of the small region of intergranular failure was identified as environmentally-assisted cracking (EAC) possibly caused by a contaminant in the hydraulic fluid. The valve actuator was fabricated using 440C stainless steel which is known to be very susceptible to EAC in a variety of environments. The 440C actuators were also heat treated to a very high hardness to eliminate galling and wear on these components. Unfortunately this high hardness also resulted in a very low fracture toughness for this material. The results of this analysis suggested that the problem was one of material selection rather than a random poor lot of material. In this case a redesign of the component to include selection of a different material or using the 440C material at a lower bulk hardness but incorporating a surface hardening treatment to prevent galling was recommended.

This last failure is an unfortunately all too common occurrence where the information on the behavior of this material (the 440C stainless steel) in various environments and heat treat conditions was known but the information never found it's way into the design phase of this component. It points to the requirement of fully researching all materials considered in the design as well as including materials engineers on the design team! It is imperative that failure analysts/metallurgists, design and maintenance engineers work together to come up with more knowledge-based designs, as well as, more accurate models of life cycles.

Service Life Extension:

The last major area that I feel is strongly impacted by failure analysis results is extension of service life of machinery systems. In designing a system the design engineer usually starts out with some goal for the expected service life for his system; the length of time in which the system is expected to function safely and reliably with little maintenance or overhaul. The engineer then specifies geometry and material to satisfy these goals. Making some assumptions as to existing flaw sizes, loading and environmental conditions can then be modelled and predictions concerning the behavior of the system over it's life time can be made. This allows for materials selection and component geometry design that is cost effective while still preserving some conservatism. But, with economics the way they are and with the improvement of computer modeling techniques, better understanding of material behavior and test methods to simulate service conditions, extension of service life beyond those early goals is becoming common practice. It is now imperative that failure mechanisms be clearly understood particularly the effects of prior service history on mechanical behavior and failure modes of components. This is where the importance of basic studies of the effects of long term loading (cyclic or static) and of load and environment variations on failure mechanisms plays strongly. By conducting failure analyses on similar components and systems we increase our knowledge of expected failure modes and we can incorporate this information into our models and databases to enables more accurate predictions of component behavior over time. Improving the accuracy of our models will then enable us to better predict when servicing intervals or inspection intervals should be conducted. Understanding failure modes enables us to design on-line condition monitoring systems that alert us to impending failure in time to schedule maintenance and repairs safely. And with an accurate understanding of the failure modes we can safely stretch the service life of aging systems without an extensive increase in cost.

Examples of service life extension programs include the nuclear power industry, marine vessels, aircraft, both commercial and military and bridges. These service life extension programs depend on accurate analysis of failure modes and mechanisms, understanding the details of the environmental and loading conditions and their effect on the system and on improvements to nondestructive inspection (NDI), diagnostic and modeling capabilities. But it is mainly through understanding failure that a direction can be set in the development of these last three technologies (NDI, diagnostics and modeling).

Conclusion:

I hope that I have illustrated the importance of failure analysis and understanding failure mechanisms to prevention of mechanical failures. Many of you are working on technologies to detect, diagnose and prognose machinery condition either on-line or during routine inspection intervals. These endeavors will ultimately result lower cost of operation, reduced conservatism in design and maintenance and increased operating safety. But the success of these technologies is soundly based in understanding failure! It is also based on the ability of the failure analysis, materials, design and maintenance engineers to work together to improve our understanding of the real overall life cycle of our system. With this expanded knowledge base (including failure modes, maintenance requirements and design goals) we can more effectively design safe, reliable, economic structures and machinery systems.

THE EVOLUTION OF RELIABILITY ENGINEERING AND ITS RELATIONSHIP TO LIFE EXTENSION

Marvin L. Roush, David Weiss, and Xiaozhong Wang
Center for Reliability Engineering
Materials and Nuclear Engineering Department
University of Maryland
College Park, MD 20742

Abstract

The paper starts with a discussion of what is meant by Reliability Engineering, its relationship to life extension and the ways it is changing. A discussion is provided of the required aspects of Designing for Reliability and Maintainability. Designing for reliability requires a full understanding of the root causes of failure and the physics of failure for the predominant failure modes. Reliability-Centered Maintenance is briefly discussed and an overview given of the Reliability Engineering Program at the University of Maryland.

Key Words: Physics of failure, reliability engineering, robust design

Introduction

Reliability Engineering began to take shape as a discipline following World War II, as systems became more complex and it was recognized that special efforts were required to better understand individual failure mechanisms and provide the tools for reliability analysis of these complex systems. One could say that the initial efforts at "reliability engineering" were efforts to extend the useful life of components and systems.

Extending the useful life of a component or system can be thought of in terms of basic concepts of stress and strength. For many mechanical and electrical systems, the system will fail when the applied load exceeds the strength of the system. Given a distribution of stress and strength, the probability of failure can be calculated as the probability that a stress from the stress distribution exceeds a strength from the strength distribution. As systems deteriorate, the strength distribution may shift to lower values. Thus, life extension can be thought of as process of system design or redesign which either restores the strength distribution to its original value or shifts it to higher values of strength. These are also the underlying concepts of reliability engineering.

1. What is Reliability?

Reliability is generally concerned with failures during the life of a product. To evaluate reliability, we need to know when and why systems fail. Because of the engineering uncertainty associated with the applied stress and the system strength, the concept of probability is used to describe system reliability. Reliability is usually defined as the probability that a system will perform a required function without failure under stated conditions for a given period of time.

Potential sources of unreliability include the following:

- Failure to recognize the use environment and the distribution of applied stresses
- Inadequate design margin (e.g. inadequate strength)
- Human error (in operation and maintenance)
- Low "extreme values" of strength (e.g. bimodal distribution) caused by material defects, design deficiencies or manufacturing induced faults
- Poor maintenance and inspection practices
- Abuse

One might define "life extension" in reliability terms by first identifying the life limiting failure modes in the context of the above sources of unreliability. Next, one would attempt to reduce the probability of these failure modes through some combination of redesign (including materials changes), change in manufacturing methods/processes, effective use of preventive maintenance techniques, or improved personnel training.

For an unchanged product, a plot of reliability vs time is always a monotonically decreasing function, unless a maintenance or replacement action has restored some useful life to the system. One minus the reliability is the cumulative distribution of the time-to- failure. The derivative of this cumulative distribution yields the probability density function of the time-to-failure of the system. This time-to-failure distribution contains very useful information about the reliability of a system. Extending life can be thought of as shifting this distribution to the right.

2. How is Reliability Engineering Changing?

Reliability Engineering, as a separate engineering discipline, originated in the aerospace industry during the 1950's, when the failure rates of military electronics systems resulted in greatly reduced system availability and increased life cycle costs. Early emphasis was on data collection, statistical analysis, reliability allocation and prediction, and reliability testing. Statistical methods for analyzing life

data advanced much more rapidly than did the basic understanding of the failure mechanisms and methods for systems reliability modeling.

Probabilistic risk assessment was a logical outgrowth of early reliability engineering efforts, designed to analyze the probability of failure of complex systems such as nuclear power plants. The extensive risk assessment efforts for nuclear power plants during the 1970's improved the event tree, fault tree, and risk-consequence techniques which have also been widely used by the chemical, aerospace and other industries. These efforts expanded earlier work concentrating on component reliability and the effects that component failure had on large systems.

The analysis of life data for individual components coupled with appropriate systems analysis techniques provides effective ways to highlight system vulnerabilities and present a sensitivity analysis of overall system reliability to the reliability of individual components. Thus the engineer is led to potential system improvement by providing redundant function at vulnerable locations or improving component reliability. Recent focus on reliability technology development has supported this later objective of component reliability improvement. It is now more fully appreciated that improvements must be built upon more thorough root cause analysis. One must understand the various failure mechanisms and their overall contribution to the total population of failures. To achieve high systems reliability, reliability engineering is shifting its emphasis to two specific areas: physics of failure; and concepts of robust design.

Understanding the physics of particular failure mechanisms, and the relationship between physics of failure and the time to failure distribution, are critical to achieving improved reliability and extending system life. In electronics, basic integrated circuits have proven themselves to be extremely reliable. Faults often occur in encapsulating the circuit, mounting it, and connecting it to the lead-out wires. The physics of failure approach thus includes a focus on thermal and mechanical stresses created during packaging and assembly, and the design of manufacturing and assembly methods to minimize these stresses. The University of Maryland has created a National Science Foundation sponsored research activity focusing on such problems. This is the Electronics Packaging Research Center.

The second area of significant recent emphasis in reliability engineering deals with concepts of robust design. A robust design, by definition, is one whose performance is minimally sensitive to uncontrolled environmental factors. To create a manufacturing process or a design which is "robust", one must have a clear understanding of which design parameters have a major effect on robustness, and which ones have minimal effect. Our abilities to understand complex competing failure modes frequently does not provide sufficient insight to identify these factors. Therefore, the key tool for creating robust designs is the statistical design of experiments. Such experiments may be run on actual hardware, or on a well designed computer simulation of the system. In the statistical design of the experiments, engineering

judgement is used to identify as many factors as possible which could affect the robustness of the design and then to design a factorial experiment. Using Analysis of Variance (ANOVA) techniques it is possible to identify the sensitivity of the design or process to each factor and interaction of factors. We believe that in many cases life extension can best be accomplished through a good understanding of the most significant failure mechanisms coupled with the design of experiments. The combination of the two should permit identification of the most significant factors affecting product life so that these factors can be suitably modified to improve reliability and meet life extension goals.

3. Designing for Reliability and Life Extension

Since deficiencies in basic design influence all items produced, the reliability of a product is, to a large extent, determined by the design process employed. A concurrent design/reliability analysis effort is critical to meet the challenge of long life. The design engineer and the reliability engineer must work hand-in-hand, with the design engineer asking: "How can I design to accomplish the required function?"; and the reliability engineering asking: "How can the system fail and how can I reduce the number of failure mechanisms and/or the probability of failure?"

Assuming that a reliability goal has been established for a system, one of the first tasks of the reliability engineer is to allocate reliability requirements to subsystems and components, and then to translate those requirements into information useful and relevant to the design engineer. Fundamental to this process, is the understanding of imposed stresses and resultant stress-life relationships. Finite element analysis methods have been sufficiently refined in recent years that they permit much more practical and effective analysis of thermal and mechanical stresses, provided that the operating environment for the equipment is sufficiently well understood. Then, stress life relationship curves must be developed as illustrated in Figure 1. At a particular stress as shown in the Figure 1, the life distribution must be estimated, from which lines of constant probability of failure can be drawn as shown by P_1, P_2 and P_3.

Another key task within a well structured reliability engineering program is failure logic analysis which involves identifying potential modes of failure (both hardware and software) and determining the effect of each failure mode on system or equipment operation. A complete understanding of the failure logic, provides the analyst with a methodical procedure for ranking the criticality of individual failure modes and identifying the most significant areas for redesign.

A system framework for developing failure (or success) logic is a "goal tree," which is a success-oriented logic structure with can be used to organize complex systems and their engineering knowledge into a format suitable for system reliability analysis[1][2]. A goal tree can be developed by first defining a reliability goal based on the overall objectives, and then identifying the subgoals which must necessarily

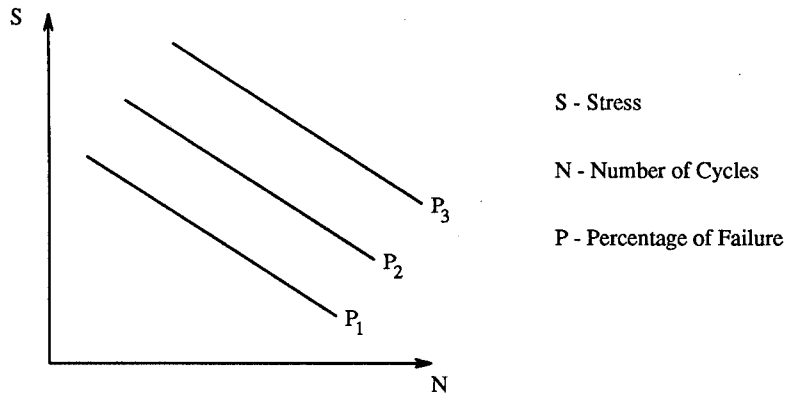


Figure 1: S-N-P Curves

achieved if the goal is to be satisfied. Such a successive decomposition of identified goals continues until the goal resolution reaches a level of engineering knowledge which will be further detailed by success trees. The goal tree/success tree is a useful tool for system reliability analysis because of the embedded cause-consequence information.

Two techniques are generally used for failure logic analyses. The first is failure mode, effects and criticality analysis (FMECA) a bottom-up analytical approach involving systematically identifying and tabulating failure modes at the lowest level of assembly, and determining their effects on higher levels of assembly. Ultimately, the effect on the overall equipment or system is addressed. The second approach is a "top-down" approach which represents the interrelationships between a component failure and the ultimate effects it has on the system. The logic however, is developed by identifying specific top level system faults of critical concern, and then tracing all the potential failures and combinations of failures which could result in that fault. A thorough and complete understanding of the failure logic of a system, coupled with relative information on probability of occurrence, provides much of the insight required to create ultra-reliable systems.

4. The Effects of Manufacturing Processes on Reliability and System Life

The analysis of system reliability histories has shown that failures are frequently induced as a result of manufacturing processes. This occurs in several different ways. First, it is not uncommon in electronic systems to subject components to higher stresses during subsequent manufacturing operations than will ever be seen in field service. An example is the temperature to which electronic components may be subjected during wave soldering. These temperatures frequently exceed the maximum temperature for which the system is designed, and can create stresses

product is fielded. Thermal gradients may be more important than the actual peak temperature. Thus, the analysis and understanding of system stresses should not be limited to operational stresses but should also consider possible stresses induced during manufacturing. Furthermore this may lead to the choice of manufacturing methods which tend to minimize such stresses.

A second way in which manufacturing methods can impact reliability is by the effect of tolerances on the variance of strength. That is, a basic design may be adequate on average to withstand design stresses, but variations in the manufacturing process may sufficiently increase the variance of strength so as to increase the probability of failure. More specifically, defects are frequently created during the manufacturing process which actually create a weak device subpopulation. This bimodal strength distribution results in premature or early life failures. A common technique which has been employed by the electronics industry to weed out this weak subpopulation, is to subject the system to Environmental Stress Screening (ESS). The idea of ESS is to subject the system to stresses generally induced through vibration and temperature cycling, of sufficient magnitude as to cause this weak subpopulation to fail. This obviously is a very inefficient way to do business. The preferred approach is to be certain the manufacturing processes are adequate and within sufficient control that a homogeneous strength distribution is produced with no outliers. If this is done, a crutch such as ESS is not required.

As a summary, the focus of reliability engineering shifts to control of the process through the well-planned and carefully executed quality inspections and reliability screening of incoming components, fabricated boards and assembled systems. Through these efforts, we can assure the inherent reliability potential provided by reliability design.

5. Reliability Testing and Growth Concepts

Reliability demonstration testing has often been required for military systems. However as reliabilities increased, test times correspondingly increased and demonstration testing became less practical. An alternative to demonstration testing is continuous reliability growth monitoring during the development process. During development, points of failure are identified and the design changed for improved reliability. Many reliability growth models have been formulated and are presented in the literature. The simplest and the most commonly use of these growth models is the "Duane" model which is illustrated in Figure 2 [3]. Reliability demonstration testing provides information before the production or deployment of an engineering system according to realistic test and assessment of equipment performance and reliability under given environmental conditions. Through reliability demonstration testing, we can get a measure of confidence that the equipment will perform as specified before actual operation.

Reliability growth is a process of testing prototype hardware to find and remove

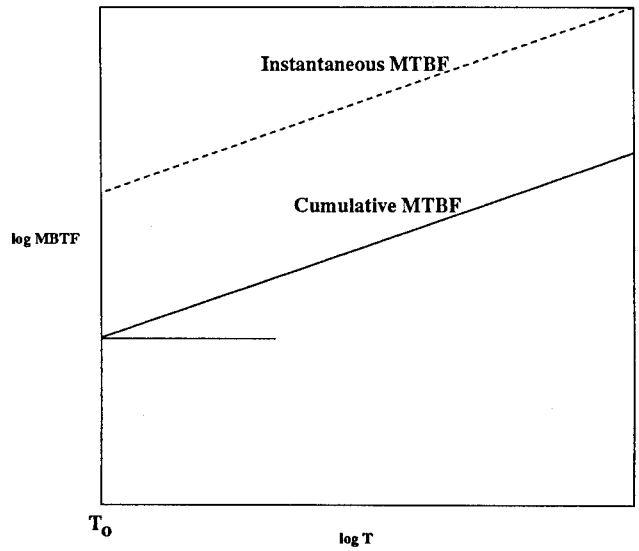


Figure 2: Duane Reliability Growth

misapplied parts, to correct design errors, and to eliminated workmanship or manufacturing process induced defects. Reliability growth can frequently be represented by a straight line on a plot of log failure rate (or MTBF) vs log time to failure. By plotting cumulative data one can obtain an estimate of cumulative MTBF which in turn can be translated to a current value as a function of the slope of the growth curve. Accumulating test data from different environments during the design process is not a simple task, but even crude measures developed in this way can provide useful insights into the current reliability development status of a system.

One of a the major problems of estimating reliability growth during factory development testing, is that failure modes resulting from the user interface in an operational environment are frequently not encountered. Thus, a second trend in industry, has been to subject systems to testing in the field environment much earlier in the development process then had previously been the norm.

6. Reliability Centered Maintenance and Its Relationship to Life Extension

One of the key elements of a successful reliability program is the establishment of effective (preventive and predictive) maintenance requirements and tasks. Reliability Centered Maintenance (RCM) is a concept which applies an analytical methodology or logic to set up specific (preventive and predictive) maintenance tasks for complex systems. Intrinsic to RCM is the identification of critical failure modes and deterioration mechanisms through engineering analysis and field

experience to determine the consequences and the most effective apportionment of maintenance activities. Thus, RCM follows logically from the reliability tools discussed previously and used during system design.

Extending the useful life of a system can often be accomplished by employing sensors identifying incipient failures and removing the implicated components prior to the occurrence of a failure. For example, life extension of a bridge is accomplished by monitoring cracks and riveting or welding reinforcement plates over cracks to prevent their propagation to failure.

7. The University of Maryland Reliability Engineering Program

Training engineers to design reliable systems and extend the life of existing systems is a great challenge, considering the diversity of tools and subjects which must be mastered. Recognizing the interdisciplinary nature of reliability engineering, the University of Maryland created an interdisciplinary, interdepartmental program in 1985 to train reliability engineers. Our academic and research programs are based upon the recognition that the performance of a complex system is affected by engineering inputs that begin at conception and extend throughout system life. Students may specialize in assessment, root cause failure analysis, probabilistic risk assessment, common cause failures, testing and operation, manufacturing, component and structural reliability, or electronic packaging reliability.

The Reliability Engineering Program is designed to meet industrial concerns and requirements. The courses are organized to provide today's engineers with the tools they will need to better understand the factors that cause components and systems to fail. Besides analytical courses presenting probabilistic methods to assess time to failure distributions, the following courses emphasize the modern reliability techniques, such as physics of failure, reliability design and life extension, robust design and laboratory tools for root cause analysis:

Advanced Reliability Engineering: Stress analysis methods to calculate loads; reliability statistics based on stress and strength distributions; basic electronic and mechanical failure mechanisms; principles of accelerated testing; and statistical design of experiments as a tool to create robust designs and reduce manufacturing variability.

Failure Mechanisms and Effects Laboratory: Techniques and equipment for conducting failure analysis of a wide spectrum of mechanical, electronic, and chemical components and systems including use of a TEM, SEM and an environmental SEM.

Reliable Design of Electronic Packages: Classification of PCB's for design; logic development and simulation; component selection; data base generation; placement, routing, thermal reliability analysis, thermal-mechanical analysis, and vibration analysis; manufacture; designing to system perfor-

mance requirements; designing to meet R&M requirements; Markov modeling of electronics; acceleration factors; derating factors.

Physics of Failure of Solid State Devices: Degradation mechanisms of silicon and GaAs devices and circuits and the study of dielectric integrity.

The Reliability Faculty closely interacts with over thirty companies nationally and with State of Maryland industries through the Engineering Research Center. Special courses and lectures are transmitted directly to industry by the University's Instructional Television System. The program has achieved national and international recognition through IEEE Reliability Society and ASQC Reliability Division awards.

The Reliability Engineering Program conducts ongoing research in the areas of:

Assessment: System Reliability Analysis, Human Performance Reliability Analysis, Root-Cause Failure Analysis, Probabilistic Risk Assessment, Common-Cause Failures

Test and Operation: Operator Advisory Systems, Life Testing, Accelerated testing, Environmental Evaluations

Manufacturing: Statistical Process Control, Improved Manufacturing Methods, Robust design for Manufacturing, Rapid Prototype Models

Failure Mechanisms: Generic Physics of Failure of Materials and Electronic Systems, Physics of Failure Computer simulation

Expert Systems: Integrated Reliability Predictors, Reliability Data Base

8. Summary

Reliability engineering as an engineering discipline has a more than 40 year history. Life extension is a word which has been increasingly used in recent years because economics frequently preclude the replacement of major infra-structure systems when they reach the end of their design lives. For example, nuclear power plants were initially licensed for 40 years but cost may make it unwise to exercise the complete replacement of a nuclear power plant at the end of that time. By the same token, highway and bridge construction has grown so in recent years that a combination of increased traffic and high loads has significantly reduced useful life. Again, since replacement cost are high, it is valuable to find ways to extend system life. The basic tools of reliability engineering which have been developed during the past 40 years are equally relevant and applicable to system life extension and to the design of new systems. In considerations of life extension the reliability engineer focuses both on identifying the system components of highest priority and on means for monitoring the condition of those components. Thus those of us who have been trained as reliability engineers must work increasing more closely with the engineering community concerned with life extension to build more reliable

systems and improve the useful life of existing systems. The University of Maryland welcomes opportunities to work more closely with industry in achieving these goals.

References

[1] M. L. Roush, M. Modarres, R. N. Hunt, D. Kreps, and R. Pearce, "Integrated Approach methodology: A Handbook for Power Plant Assessment," SAND-7138, Sandia National Labs. October 1987.

[2] M. Modarres, "What Every Engineer Should Know about Reliability and Risk Analysis," Marcel Dekker Publishing, New York, 1992.

[3] P. D. T. O'Connor, "Practical Reliability Engineering," John Wiley & Sons, Chichester, 1991.

[4] A. Christou (ed.), "Electromigration and Electronic Device Degradation," Wiley Publishing, 1994.

[5] A. Christou (ed.), "Reliability of Gallim Arsenide MMICs," Wiley Publishing, New York, 1992.

[6] M. Pecht, L. T. Nguyen and E. B. Hakim (ed.), "Plastic Encapsulated Microelectronics: Materials, Process, Quality, Reliability, and Applications", John Wiley & Sons, Inc., 1994.

[7] M. Pecht, "Integrated Circuit, Hybrid, and multichip Module Package Design Guidelines: A Focus on Reliability," John Wiley & Sons, Inc., 1994.

FAILURE ANALYSIS

Cochairmen: Nancey J. Maegerlein
Naval Surface Warfare Center/Crane

Robert A. Bayles
Naval Research Laboratory

FAILURE ANALYSIS OF A SELF-LOCKING NUT FROM AN ARMY ATTACK HELICOPTER

Marc Pepi
Victor Champagne
Scott Grendahl
U.S. Army Research Laboratory
405 Arsenal Street
AMSRL-MA-CB-292
Watertown, Massachusetts 02172-0001

Abstract: The U.S. Army Research Laboratory-Materials Directorate (USARL-MD) conducted a failure analysis of a self-locking nut 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 two fracture halves of the broken nut. The through crack which caused the nut to break apart originated at the root of the first thread. The morphology at the crack origin was intergranular and in some areas contained a thin film of cadmium as determined by an elemental X-Ray mapping of the entire fracture surface, utilizing energy dispersive spectroscopy (EDS). The fracture morphology beyond the intergranular fracture region transitioned to a mixed-mode consisting of intergranular decohesion and ductile dimples. The regions of final fracture displayed a pure ductile dimpled morphology. Metallographic examination of longitudinal and transverse cross sections of the nut material revealed internal cracking. Even specimens sectioned approximately 180° from the fracture contained internal cracks indicating the entire component was affected. Hardness results conformed to the governing specification. Chemical analysis did not match any of the seven alternative alloys specified, but most closely conformed to AISI 8740 steel. It was concluded that the nut failed as a result of hydrogen embrittlement (HE). The primary source of hydrogen would most likely be related to manufacturing as opposed to the service environment. Evidence substantiating this conclusion was the fact that many internal cracks were noted within the material, and the general lack of corrosion products on or near the fracture surface.

Key Words: Failure analysis, metallurgical examination, hydrogen embrittlement.

Background: The self-locking nut conforms to MS21042, *Nut, Self-Locking, 450° F, Reduced Hexagon, Reduced Height, Ring Base, Non-Corrosion Resistant Steel.* Two self-locking nuts secure an aft weight pack to the rotor blades of the aircraft, and failure of these nuts could cause instability during flight. The component was found to be broken in two pieces during a "phasing" of the helicopter blades. As a result of blade movement by the mechanic, a rattling noise was heard. The blade was subsequently removed, and the swept tip cap disassembled, whereupon the broken nut fell to the ground. The adjoining nut was still on the stud, but was split.

Visual Inspection and Light Optical Microscopy: The failed component in the as-received condition is shown in Figure 1. The nut was cadmium plated (electrodeposited) in accordance with QQ-P-416, Type II, Class 3. The nut was sectioned such that the fracture surfaces could be examined. The fracture origin was determined to be at the root of the first thread, as shown in Figure 2 (arrow denotes origin). This fracture half was designated half "A". Note the radial lines which emanate from the fracture origin. The cadmium plating adjacent to the fracture surface was worn and damaged, but no evidence of significant corrosion was observed. There was also no corrosion visible on the remaining fracture half. A shear lip region was observed extending around the opposite edge of the threads. In addition, beyond the fourth thread the fracture morphology revealed the direction of extrusion. This is evidenced by the sets of parallel lines running perpendicular to the threads.

Metallographic Examination: The nut was sectioned such that longitudinal (parallel to the thread direction) and transverse (perpendicular to the thread direction) specimens 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 9 micron diamond paste, and 0.5 micron alumina. Numerous internal cracks were noted within the transverse specimen, as shown in Figure 3 (etched with 1% nital). The longitudinal specimen showed no such anomalies. The internal cracking appeared intergranular in nature, as shown in Figure 4.

Hardness Testing: The mounted samples were subjected to Knoop hardness 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 geometry 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 ten readings were taken (five on each sample). A load of 500 gmf was utilized, at 20x magnification. The results are listed in Table I. The converted hardness values conformed to the governing specification.

Table I
Knoop Microhardness Testing
Longitudinal and Transverse Samples
500 gmf load

Sample	Knoop	Eq. HRC
Longitudinal	503	48
	523	49
	519	49
	499	47
	505	48
Transverse	506	48
	510	48
	510	48
	518	49
	515	48
MS21042		49 maximum

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 consisted of 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 8740, as shown in Table II. However, the chromium, nickel and molybdenum contents each exceeded the maximum limits set forth in AMS 6322. This nonconformity suggested a possible lack of melt control. The additional amounts of chromium, nickel and molybdenum would most likely have no adverse affects on mechanical properties, since each of these elements are generally added to provide increased hardenability, toughness and corrosion resistance.

Table II
Chemical Analysis
Weight Percent

Elem.	Nut	C1035	C1042	1050	4027	4037	8630	8740
C	0.40	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	0.98	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	0.33	0.15-0.35	0.30m	0.30m	0.15-0.35	0.15-0.35	0.15-0.35	0.15-0.35
P	0.008	0.040m	0.040m	0.040m	0.035m	0.035m	0.025m	0.025m
S	0.016	0.050m	0.050m	0.050m	0.040m	0.040m	0.025m	0.025m
Cr	1.86	*	*	*	*	*	0.40-0.60	0.40-0.60
Ni	1.51	*	*	*	*	*	0.40-0.70	0.40-0.70
Mo	0.37	*	*	*	0.20-0.30	0.20-0.30	0.15-0.25	0.20-0.30
Cu	0.11	*	*	*	*	*	0.35m	0.35m
Fe	rem.	rem.	rem.	rem.	rem.	rem.	rem.	rem.

* - Not specified.
m - maximum.

Scanning Electron Microscopy/Energy Dispersive Spectroscopy: The fracture morphology at the crack origin was intergranular (I) as shown in Figure 5. This SEM micrograph shows the traditional "rock candy" features, as well as secondary cracking. X-Ray mapping (utilizing energy dispersive spectroscopy) of the entire fracture surface revealed a thin layer of cadmium in some areas adjacent to the fracture origin. Two other regions which were located away from the initiation site also contained a thin film of cadmium and had also fractured by intergranular decohesion. The film of cadmium suggests the cracking most likely occurred within the plating bath. The intergranular region transitioned to a large mixed-mode region (combined intergranular and ductile dimples-D/I) covering nearly one-half of the fracture surface. This morphology is

exhibited in Figure 6. The mixed-mode morphology transitioned abruptly to ductile dimples (D) as shown in Figure 7. This abrupt change in morphology represents the area of fast fracture beyond the fourth thread. Figure 8 shows the ductile dimples aligned in the extrusion direction of the component. Shear lips (SL) were also noted, indicative of the last region to fail. Light optical and scanning electron microscopy did not reveal any gross anomalies at the fracture origin. The schematic shown in Figure 9 outlines the different morphologies noted on fracture half "A".

Stress Durability Testing: As a result of these findings, stress durability testing was conducted on two self-locking nuts which were previously in service. The nuts were fabricated from the same manufacturer, but due to a lack of traceability, the lot could not be identified. Testing was conducted in accordance with MIL-STD-1312-14 [3]. The nuts were torqued to a load which was 74-80 percent of the listed axial strength. One nut was subjected to a 23 hour test, while the other was tested for 168 hours. After the allotted duration had expired, the nuts were disassembled and examined through optical microscopy and magnetic particle inspection. No indications were noted as a result of this inspection.

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 on or near the fracture surfaces, and the presence of internal cracking away from the crack origin ruled out SCC as a failure mode. Therefore, the failure was attributed to hydrogen embrittlement. The most likely source of hydrogen in the fabrication of this component, was the cadmium plating process. Specification QQ-P-416 is an electrodeposition plating method, which contains nascent hydrogen which diffuses into the material during plating. A post-bake is recommended for high strength steels to prevent hydrogen embrittlement. Plating is generally carried out in alkaline cyanide baths prepared by dissolving cadmium oxide (CdO) in a sodium cyanide (NaCN) solution. Sodium hydroxide (NaOH) and sodium carbonate (Na_2CO_3) are formed by reactions and are also part of this bath [1]. However, in recent years the need to generate minimal hydrogen evolution and environmental concerns have led to the development of non-cyanide cadmium plating baths. These type of baths are preferred for high strength, quenched and tempered steels, because less hydrogen is produced, and the threat of hydrogen embrittlement is less. 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) not post-baked, b) post-baked after a prolonged period of time after the plating process, or c) post-baked at an incorrect temperature or bake time, or in an improper plating bath. It is well documented that the absorption of hydrogen into a steel results in a loss of ductility over sustained loads. However, tensile specimens could not be fabricated due to the small size of the failed component.

Conclusion: The nut under investigation had failed as a result of hydrogen embrittlement. The primary source of the hydrogen would most likely be related to a

manufacturing operation (electrodeposited cadmium plating) as opposed to the service environment. Evidence substantiating this conclusion is the fact that many internal cracks were observed within the material, and corrosion products were absent on or around the fracture surface. It is also important to note that QQ-P-416 does not recommend electroplating parts with a UTS above 240 ksi. Although similar nuts fabricated by the same manufacturer which were in service passed the stress durability testing, this does not rule out hydrogen embrittlement as the cause of failure, because specific manufacturing lots could not be identified.

Recommendations: The additions of increased amounts of strengthening elements (chromium, nickel and molybdenum) may have increased the hardness of this component (and other parts from the same melt) one or two HRC points, and subsequently raised the UTS above the 240 ksi threshold. Short-term suggestions included ensuring chemical conformance, as well as, proper electroplating techniques; from the bath to post-baking. In addition, MIL-STD-870 [5], *Cadmium Plating, Low Embrittlement, Electrodeposition,* may be used to replace QQ-P-416, since it boasts a lower embrittlement hazard for high strength steels over 180 ksi. Also increased stress durability testing of in-service parts may provide an understanding of the vastness of the problem. More aggressive, long-term recommendations were to replace the cadmium electroplating process with vacuum deposition (which introduces no hydrogen), as well as lowering the required maximum hardness from 49 HRC (246 ksi UTS), as listed on the current governing specification, to approximately 46-47 HRC (221-229 ksi UTS). These long-term recommendations should only be incorporated if cost and stress considerations dictate feasibility.

References:

1. Specification MS21042, *Nut, Self-Locking, 450° F, Reduced Hexagon, Reduced Height, Ring Base, Non-Corrosion, Resistant Steel.*

2. Specification QQ-P-416, *Federal Specification, Plating, Cadmium (Electrodeposited).*

3. Specification MIL-STD-1312, *Fastener Test Methods.*

4. Morrow, Hugh, *Cadmium Plating,* Metal Finishing Guidebook and Directory Issue '91, Metals and Plastics Publications Inc., Volume 89, Number 1A, pp. 175-178.

5. Specification MIL-STD-870, *Cadmium Plating, Low Embrittlement, Electrodeposition.*

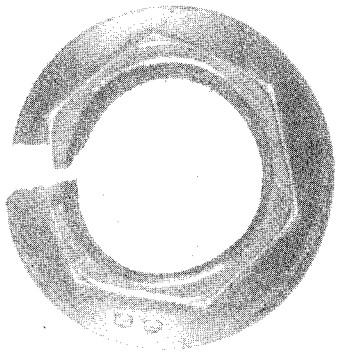


FIGURE 1 The failed self-locking nut in the as-received condition. Mag. ~3x.

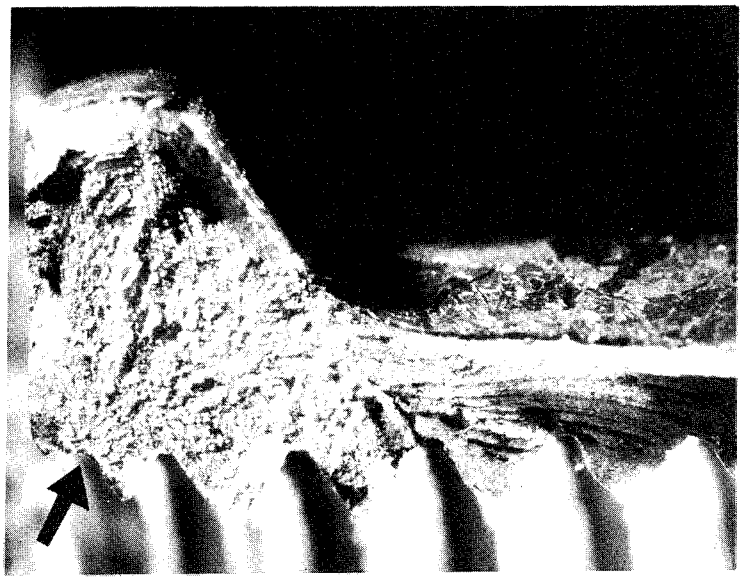


FIGURE 2 Light optical macrograph of one of the fracture surfaces of the failed nut. Radial lines emanate from the crack origin, denoted by arrow. Mag. 25x.

FIGURE 3 Macrograph of the etched transverse sample showing numerous internal cracks. Etchant: 1% nital. Mag. 50x.

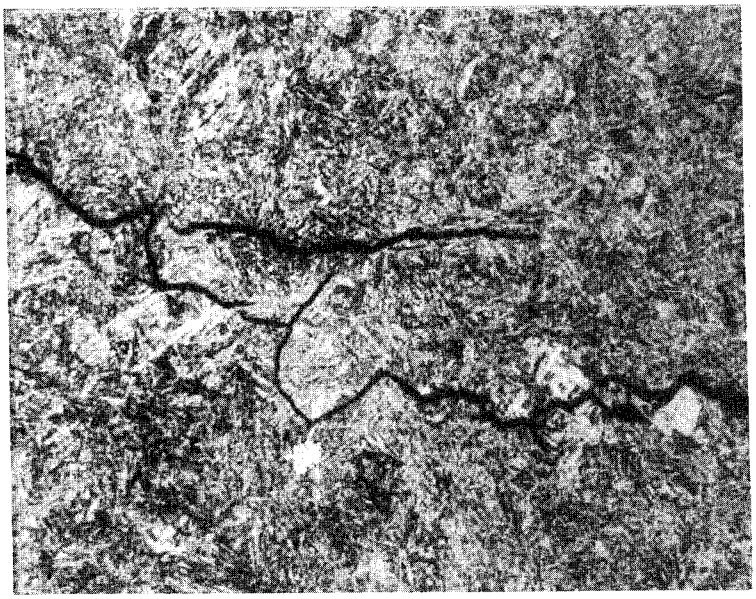


FIGURE 4 High magnification micrograph of an internal crack showing intergranular nature. Etchant: Vilella's reagent. Mag. 1000x.

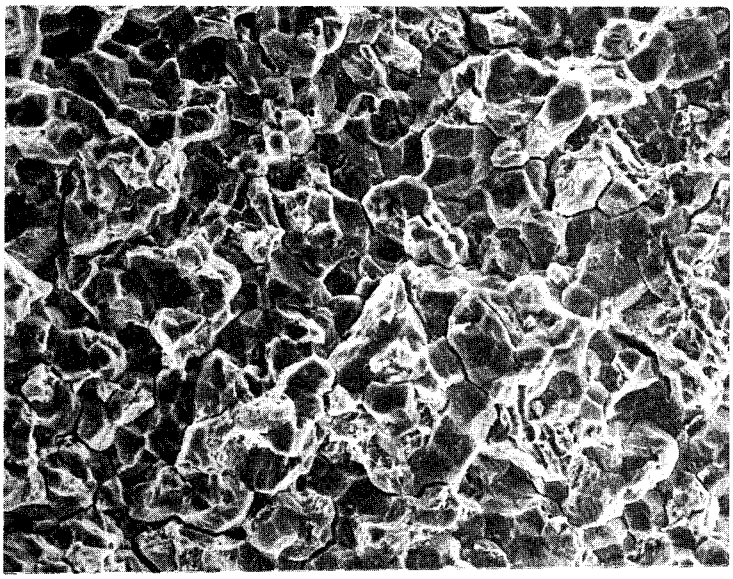


FIGURE 5 Scanning electron micrograph (SEM) of the intergranular fracture morphology at the crack origin. Note the traditional "rock candy" feature, as well as the secondary cracking, both typical of a hydrogen embrittlement failure. Mag. 1000x.

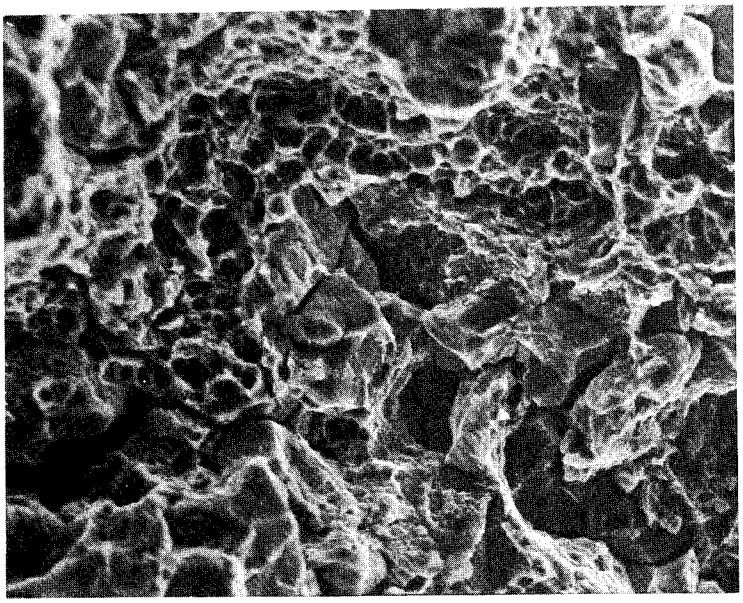


FIGURE 6 SEM of the mixed-mode (intergranular/ductile) morphology. Mag. 2000x.

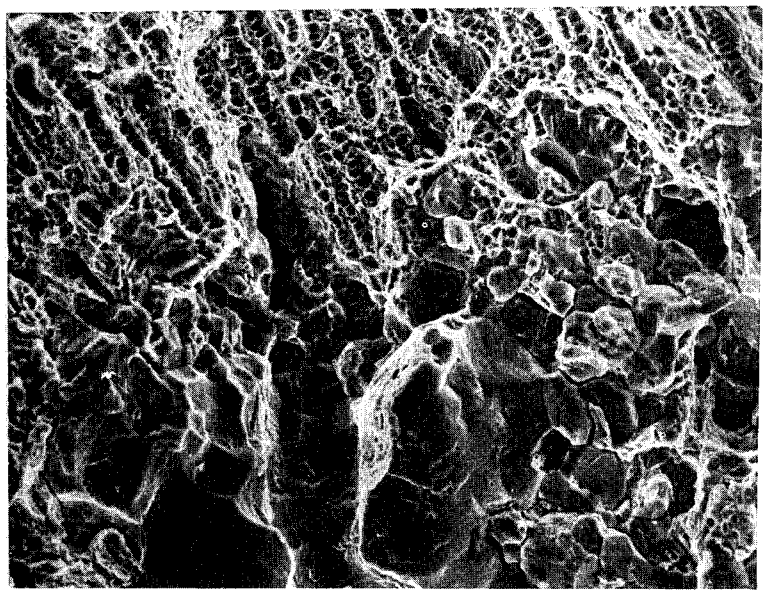


FIGURE 7 SEM of the abrupt transition between the mixed-mode and ductile regions. Mag. 1000x.

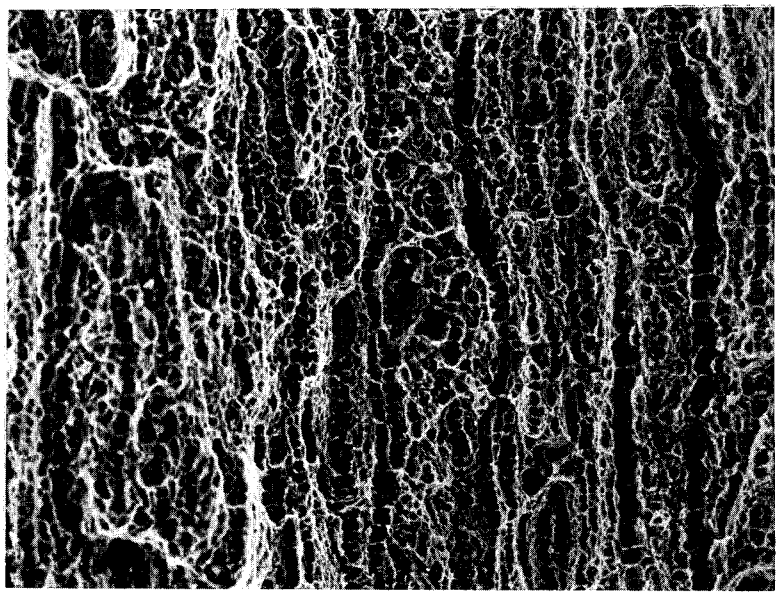


FIGURE 8 SEM of the ductile dimples oriented along the extrusion direction in the fast-fracture region. Mag. 1000x.

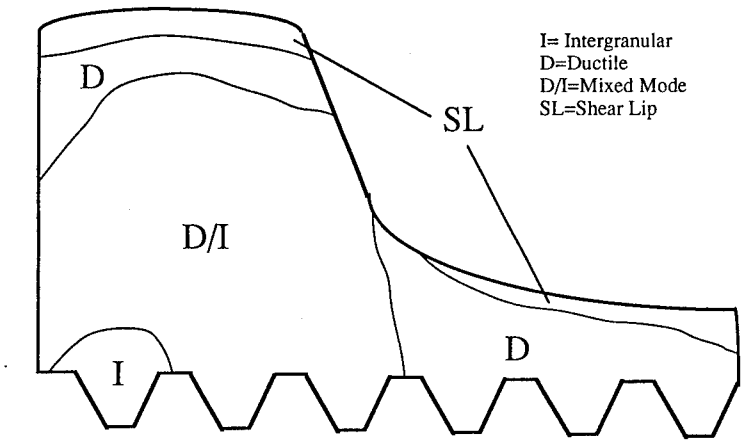


FIGURE 9 Schematic showing a map of the fracture morphologies, as determined through scanning electron microscopy.

FAILURE ANALYSIS OF A MAIN LANDING GEAR AXLE

K.Wolf and H.Richter, Wehrwissenschaftliches Institut für Materialuntersuchungen (WIM),
Landshuter Str. 70, 85435 Erding, Germany

ABSTRACT

During touchdown an aircraft encountered a failure of a Main Landing Gear Axle (MLGA). A routine visual examination revealed a corroded and discolored surface on the underside at the fitting position. After dismantling, it was found that the MLGA was broken at the midway point. Because of the severity of the incident a failure analysis was requested. The MLGA consisted of a high-strength, heat-treated 4340 steel with a chromium plated surface. A dye penetrant test revealed numerous surface cracks. Residual tensile stresses in circumferential position were detected by means of a Barkhausen noise analysis. The fracture surface of the MLGA showed a half-elliptical incipient crack through wall thickness (hollow shaft). Scanning electron microscopy of the fractured MLGA revealed that the initial region of the crack exhibited features consistent with corrosion products and locally those of intergranular fracture. The adjacent area showed intergranular fracture in contrast to the ductile structures of the forced fracture. The appearance of intergranular fracture seemed to be an indication of the influence of hydrogen at the early stage of fracture. Because of the difficulty in distinguishing between hydrogen induced stress corrosion cracking and hydrogen induced cracks, a number of tests were made. Both the hardness of the bulk material and the composition were within the requirements. Microhardness tests on deplated surface areas along a crack generating line revealed a drop in hardness. The hardness drop coincides with the area in which strip-type crack indications were found. These phenomena are considered to be due to bad surface processing (grinding) with local overheating.
Recommendations to the user included Non-destructive testing procedures for cracks, Barkhausen noise analysis to detect regions on the surface with high residual tensile stresses as well as replacement of the chromium plating in order to prevent the danger of cracks in the layer causing corrosion of the steel.

Keywords: Barkhausen noise analysis; dye penetrant testing; grinding cracks; high-strength heat-treated 4340 steel; hydrogen induced stress corrosion cracking; residual stresses;

INTRODUCTION

After several years in service one of the MLGA's of a transport aircraft was found to be broken. The fracture had occured in the bearing attachment area of the landing gear, but did not result in a dislocation of the MLGA. Because of the inaccessibility of the assembly it is difficult to discover a crack or fracture at an early stage. On the other hand, there are no records of previous failure of this kind. The intent of this investigation was to find the probable failure mechanism. Several mechanisms could have caused the failure of the MLGA: fatigue, over-loading, hydrogen embrittlement, stress corrosion cracking, manufacturing deficiencies or defects like quench cracks. WIM was tasked with making a failure analysis.

RESULTS

VISUAL INSPECTION

Figure 1 shows the fractured MLGA as well as the sampling points of specimens and areas subjected to non-destructive evaluation. The top of the MLGA in mounting position is defined to be at an angle of 0 degrees. Looking at the fractured surface of the RH part of the MLGA, the preinduced crack was at an angle of 200 degrees in flight direction that means from underneath. The fracture had started with a semielliptical crack of 9,5 mm in length and 1,8 mm in depth which was heavily covered with corrosion products. Subsequent to this incipient crack another crack of 78 mm length and 27 mm depth (i.e. full wall thickness of the MLGA), followed and resulted in the forced fracture (Figures 2, 3 and 4).

MATERIAL

The material the MLGA consisted of was SAE 4340 with all elements within specification MIL-S-5000 as shown in Table 1. In the bearing area where the fracture had occured, the MLGA was Cr-plated.

NON-DESTRUCTIVE EVALUATION

To check for additional cracks a **dye penetrant test** was performed at first on the Cr-plated surface adjacent to the preinduced crack. Despite the use of an ultrahigh sensitivity penetration medium only weak indications of cracks were found, which were mostly aligned axially (Figure 5). After the Cr-layer had been removed, stronger indications were detected by dye penetrant testing in the same area, which were mostly aligned axially and partly arranged in a strip-type manner (Figure 6). Deep grinding grooves as revealed by the optical microscope were attributed to these indication strips. The fracture line of the initial crack was located close to such grinding groove. **Magnetic particle inspection** confirmed the findings

gained by the dye penetrant testing. To check for residual stresses which might have been the reason for those tiny cracks with weak indication, a **Barkhausen noise analysis** was performed on the MLGA. On almost all Cr-plated surfaces at a preinduced crack circumferential position of 200±60 degrees, the border layer of the matrix was found to reveal residual tensile stresses or, as compared with the other border zones, at least reduced residual compressive stresses. These residual stresses were aligned circumferentially. Axial alignment of residual stresses, too, could be found, but with stress levels being markedly lower. In the border layer of a properly manufactured MLGA a residual compressive stress condition should be evenly balanced. Therefore, the above findings of the Barkhausen noise analysis suggest a manufacturing deficiency.

FRACTOGRAPHIC EXAMINATION

The surface of the preinduced crack was heavily coated with corrosion products on several locations (Figure 7). Despite the corrosion, intergranular microstructures were still visible in parts but failed to disclose further details. The elements S and Cl, usually indicative of environmental conditions developing corrosion were present on the fracture surface. Because of the corrosion layer the preinduced crack must have developed some time prior to the adjacent part of the crack which was almost free from corrosion products. The fracture surface of this secondary crack region revealed exclusively intergranular microstructures indicating exposed grain boundary cracks, pores and hair lines (Figure 8). The outward appearance points to a hydrogen induced crack [1 to 4]. In isolated places a thin layer of corrosion products could be identified which contained the element Cl. The forced fracture surface - the third stage of the fracture - showed distinct ductile fracture characteristics and was free from corrosion products (Figure 9). Even at the transition regions between the different crack areas there was no indication of a fatigue fracture. MLGA surface areas adjacent to the fracture surface angular position of 0 and 200 degrees were inspected for cracks both prior to and after removal of the Cr-layer by pickling. Cracks were found to have occured in the Cr-layer as well as in the matrix - i.e.additional net-like crack patterns which neither the optical microscope nor the afore mentioned non-destructive evaluations had revealed (Figures 10 and 11). However, cracks at the 0 degree position were finer and less frequent than those at the 200 degree position. A 0,5 mm deep crack had broken open and propagated roughly in parallel with the preinduced crack (Figure 12). The fractured surface showed an intergranular microstructur (Figure 13). It could be clearly discerned that there was a coating on the grain boundary surfaces. EDS analysis revealed that this coating consisted of siliciferous, sulphurous, and carboniferous substances which had been put in by grinding or shot

peening. Moreover, heat influence will have to be taken into consideration, too.

METALLOGRAPHY

The high-strength, heat-treated steel exhibited a very fine grained structure (Figure 14). Constituents promoting cracks like grain boundary precipitations and coarse martensite were not encountered. The required tensile strength of 260 to 280 ksi (1790 to 1930 N/mm^2) is equivalent to a hardness of 52 to 54 HRC. Results were between 53 and 54 HRC with only a slight variation of the values. Thus, the bulk matarial met the requirements. Micro- hardness tests on sections showed a local hardness decrease from about 630 HV 0,025 in areas more remote from the outer edge to 510 HV 0,025 at a distance of only 0,02 mm from the boundary line between Cr-layer and steel. In other places there was no drop in hardness near the surface. These findings were confirmed by a HV 10 hardness line which was additionally laid across the deplated surface along a generating line and started at the fracture line just behind the crack. Up to a distance of 8,5mm from the fracture line the average hardness was 429 HV 10 which is also markedly lower than in the adjacent area where hardness was up to 639 HV 10. This last mentioned area with a drop in hardness coincides with the area where strip-type crack indications were received. These phenomenas were considered to be due to surface finishing (grinding) and probably local overheating associated with it. The Cr-layer of the MLGA was between 18 and 34μm thick and thus did not meet the specification of 25μm to 51μm. Investigations were performed on two MLGA sections both taken radially in a longitudinal direction. Both sections revealed numerous cracks propagating throughout the Cr-layer and penetrating into the matrix up to 0,5 mm in depth (Figure 15). Moreover, the Cr-layer showed additional cracks, not all of which were found to advance to the surface. Crack surfaces partly ran in parallel to the preinduced crack. After removing the Cr-layer by pickling, the same specimen which already had been subjected to dye penetrant testing was slightly ground and polished. Then in the area next to the preinduced crack several cracks up to 2,6 mm in length with tree-like to net-like crack formations were detected with the crack path in some cases found to coincide with the circumferential direction. Thus the crack path partly ran in parallel with the preinduced crack and, because of its indented appearance on the surface, gave rise to the assumption of an intergranular crack propagation. Cracks often appeared to be filled or, at least, closed by a substance which could not be defined precisely (possibly corrosion products).

CORROSION TESTS

High-strength steels can fail due to stress corrosion if

corrosion produces hydrogen [5, 6]. The principal effects of hydrogen embrittlement in steels are a decrease in tensile strength and ductility, when tested under static loads or at low strain rates. To discern the differences in appearance between failures caused by stress corrosion cracking and hydrogen embrittlement initiated e.g. by the plating process several tests were made. Notch-bend specimens were loaded in a bending device to find out the flectional strength. All specimens failed after 12,6 turns of a screw which gave a measure of maximum loading condition. The stress durability tests were carried out in accordance with ASTM F 519 and applicable DIN standards. Hydrogen permeation into the specimens during bending resulted in fractur after a short time when only five turns had been chosen. The same number of turns were applied for the salt spray test and for condensation water alternating climate. In both cases the fractures occured within two days. There was no evidence of a ductile dimple topology as was the case with simple bending overloading. Fine indications, such as hairlines and more secondary cracking, were less pronounced with stress corrosion cracking than with hydrogen permeation. Intergranular fracture surfaces of the MLGA were neither similar to the fracture surface of the hydrogen permeated specimens nor to those of the stress corrosion tested specimens; maybe loading conditions had not been identical. But it was proved that this specific material is sensitive to stress corrosion.

CONCLUSION

Subject of this investigation was to find out which mechanisms had been responsible for crack initiation and propagation as observed on this MLGA. It could be proven that the reason for the MLGA failure was poor manufacturing: inexpert grinding of the MLGA surface prior to Cr-plating marks the beginning of a relatively complex failure history. The thick coating and the position of crack in an area which during the hardness test had revealed a marked drop in hardness on the surface and which during the non-destructive evaluation had shown a strip - like pattern due to the accumulation of crack indications testify to an older crack produced or induced by the grinding process. After the Cr-layer had been removed, numerous additional cracks of a net-like and tree-like patterns and some tenth of a millimeter deep were found on the steel surface of the MLGA. These cracks were more frequent and more distinct on the lower side of the MLGA. Initiated by linear corrosion attack due to cracks in the Cr-layer they may well have formed as a result of residual tensile stresses from fabrication processes and of superimposed operational bending stresses in connection with hydrogen. Corrosion attack was the result of fabrication induced cracks and probably of cracks in the Cr-layer formed during operation and which, can hardly be avoided considering the thickness of the Cr-layer. The rust

coloured deposit found on the Cr-layer, particulary in the area of the retaining parts and bolts, was a clear indication of corrosive action taking place underneath the Cr-layer. The crack area adjacent to the preinduced crack, indicative of a hydrogen induced brittle type crack, may have originated from the above corrosive action. The fact that this crack did not form until a long time after the axle had been installed may be explained by the fact that corrosion attack did not start until some time had elapsed. Microstructure indication of crack initiation and/or propagation due to cyclic stresses were not found. The forced fracture showed distinct ductile type fracture characteristics (dimples). Based on the foregoing investigations, analysis and tests, high-strength heat treated steel like 4340 appears to be susceptible to hydrogen embrittlement caused by manufacturing as well as to stress corrosion cracking influenced by environment. The characteristic material values of the MLGA examined were found to meet the required specifications. With the Cr-layer applied, magnetic particle inspections or dye penetrant tests make it rather difficult to distinguish between the fine cracks in the surface of the matrix and normal acritical noise indications such as resulting from grinding. After removal of the Cr-layer both test methods can be used for detecting unacceptable cracks, but probably not all of them will be identified. Therefore, by means of a spotcheck inspection at least a visual check for cracks and corrosion attacks should be made with an optical microscope. Areas with an irregular residual stress distribution and susceptible to producing cracks can be determined by means of the Barkhausen noise analysis without the need for prior removal of the Cr-layer.

RECOMMENDATIONS

WIM recommended a screening study by means of the Barkhausen noise analysis and, subsequent to it, additional NDT inspections for cracks to be performed especially on parts featuring a conspicuous residual stress distribution. Areas around the Cr-plated bearing showing corrosion products/discolorations of a rust-like appearance will have to be examined with particular care because it cannot be ruled out that hydrogen-induced brittle-type fracture is involved. Further measures which may be taken to prevent damages of this kind are: shot peening of steel surfaces to produce residual compressive stresses on and below the surface; replacement of the Cr-layer by another coating offering better protection of the steel against corrosion.

ACKNOWLEDGEMENT

This examination was carried out at the Institute for Material Research and Testing of the German Armed Forces (WIM). The author would like to thank the institute leader Prof.Dr.Guttenberger for supporting this work. The work

would not have been possible without the technical contributions and comments of Mr.Dietrich and Mr.Marold.

REFERENCES

1. VDEh, DVM, DGM, The Appearance of Cracks and Fractures in Metallic Materials, Verlag Stahleisen, Düsseldorf, 1983

2. V.J.Colangelo, F.A.Heiser, Analysis of Metallurgical Failures, J.Wiley&Sons, New York, 1987

3. H.Pircher, R.Großterlinden, Hydrogen Induced Corrosion of Low Alloy Steels in Hydrogen Media, Werkstoffe und Korrosion 38, 57-64 (1987)

4. W.Paatsch, Avoidance of Hydrogen Embrittlement of High Strength Steels During Electroplating Processes by Thermal Alloying, Werkstoffe und Korrosion 39, 167-172 (1988)

5. B.D.Craig, The Effect of Ni on Hydrogen Cracking Resistance in Low Alloy Steels, Corrosion-NACE, Vol.38, No.9, Sept.82

6. G.Lange, Die Gefährdung hochfester Stähle durch Wasserstoff, Radex-Rundschau, Heft 3/4 1972

Table 1: Chemical composition of the Main Landing Gear Axle

Samples or reference material	Element mass ratios in %								
	C	Si	Mn	P	S	Cr	Mo	Ni	Fe
Main Landing Gear Axle	0,41	0,29	0,80	0,009	0,002	0,84	0,21	1,98	BA
SAE 4340 MIL-S-5000 [1)]	0,38 to 0,43	0,20 to 0,35	0,65 to 0,85	≤ 0,025	≤ 0,025	0,70 to 0,90	0,20 to 0,30	1,65 to 2,0	SE

[1)] Material specified as per reference; values derived from MIL-S-5000

Section I (A.L.H.) Section II(A.R.H.)

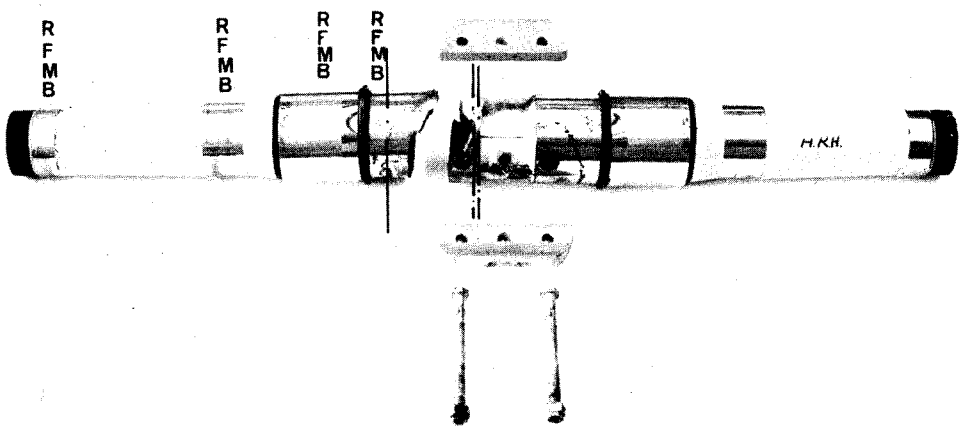


Figure 1 1:7,3
Overall view of the damaged Main Landing Gear Axle (MLGA); the dot-dash lines indicating the places where severance cuts for SEM-, metallographic specimens, and specimens were made for chemical analysis; places 1 to 4 indicating where radiographic stress measurements (R), dye penetrant testing (F), magnetic particle inspections (M), and Barkhausen noise analysis (B) were performed.

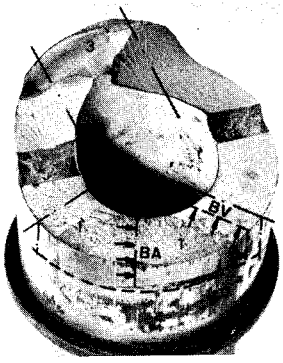


Figure 2 1:2
MLGA section I; arrows indicating direction of view on to section surface;
1' = SEM specimen
1 = specimen for sections BA/BV
3 = specimen for non-destructive testing

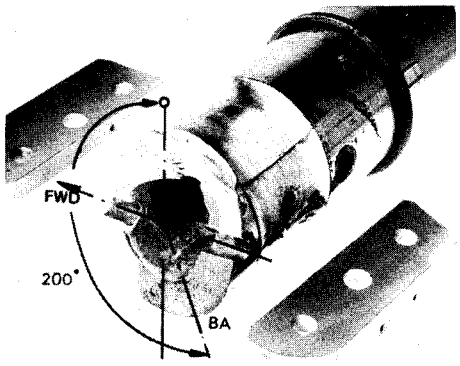


Figure 3 1:3,6
MLGA section II and retaining parts; the angle is indicative of the area subjected to non-destructive testing;
FWD = direction of flight
BA = preinduced crack in 200^0 position

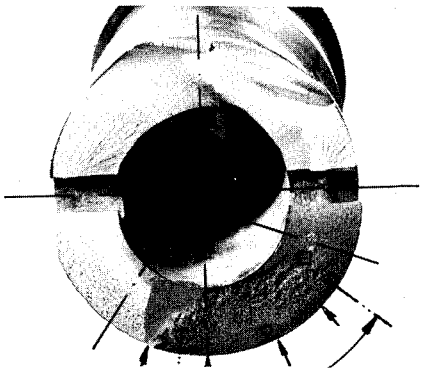


Figure 4 1:1,7
MLGA section II; arrows pointing at the examined fracture surface

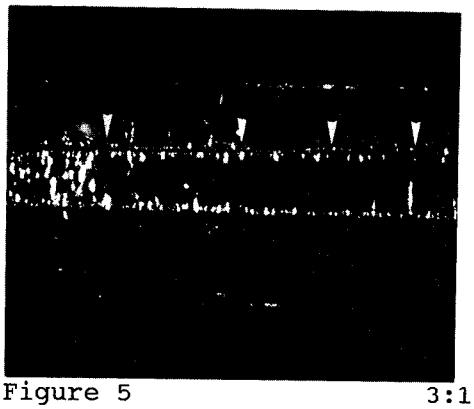


Figure 5 3:1
Dye penetrant indications in axle longitudinal direction around the preinduced crack; Cr-layer removed below the arrows

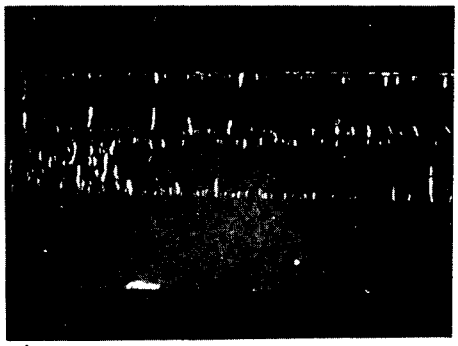


Figure 6 3:1
Same axle surface area as in Fig.5, but with Cr-layer completely removed

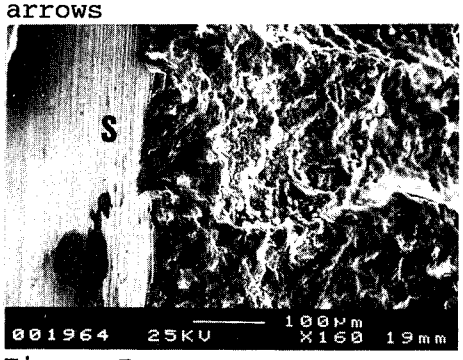


Figure 7
SEM micrograph showing preinduced crack area covered with corrosion products
S = surface

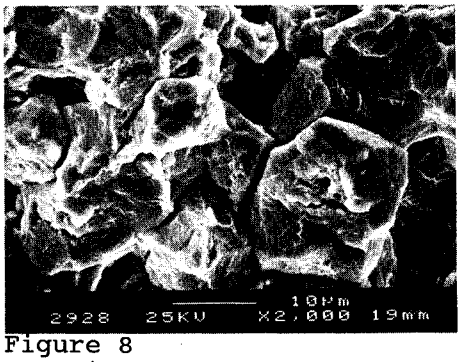


Figure 8
SEM micrograph showing part of non-corroded crack area with microcracks to be seen between grains, as well as pores, and hair lines on grain boundary surfaces

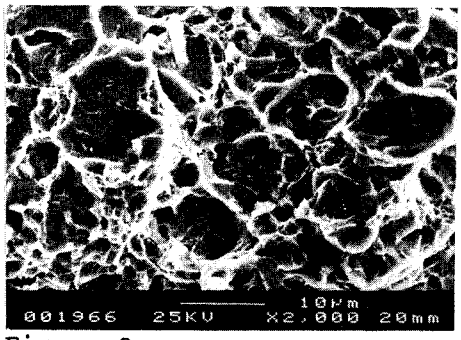


Figure 9
Forced fracture surface with transgranular microstructure

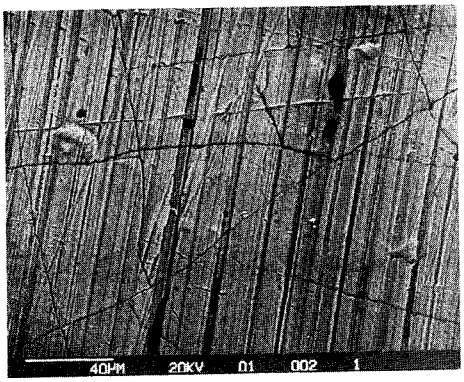


Figure 10
Net-like crack pattern on Cr-layer surface in axle section II about 15mm distant from point of preinduced crack

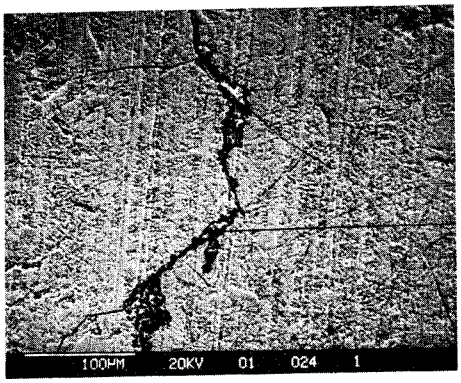


Figure 11
Net-like crack pattern on deplated surface in axle section II

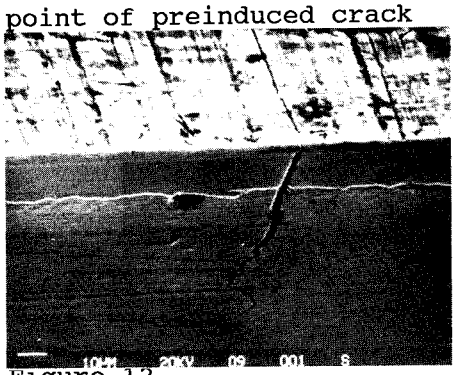


Figure 12
Section BA indicating a crack which extends trough the Cr-layer into the matrix and within the Cr-layer is congruent in sections with one of the processing furrows

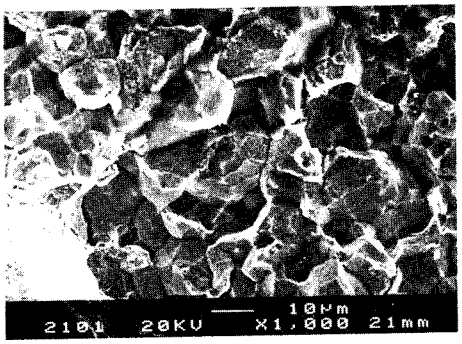


Figure 13
Crack shown in Fig.12 broken open in the laboratory and indicative of an intergranular microstructure with grain boundary surfaces covered by deposits

Figure 14
Structure in the far-surface area with martensitic texture and without special features

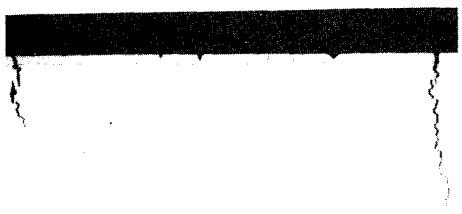


Figure 15 126:1
Cross section in 200^0 position of axle section I; the appearance of the cracks in steel is indicative of intergranular cracking

FAILURE ANALYSIS OF THE COBRA PITCH LINK SELF-LOCKING NUT

Victor K. Champagne
Scott Grendahl
Marc Pepi
U.S. Army Research Laboratory
405 Arsenal Street AMSRL-MA-CB-292
Watertown, MA 02172-0001

Abstract: The U.S. Army Research Laboratory Materials Directorate (USARL-MD) conducted a failure analysis of a pitch link self-locking nut which failed prior to pre-flight inspection. The part is used to secure the lower end of the pitch link on the Cobra helicopter and failure could cause instability of the rotor blade. Light optical microscopy of the fracture surface revealed that the cadmium plating was intact and no evidence of corrosion or pitting was observed at the crack origin. Fracture initiated from the outside diameter of the nut which was not a highly stressed region. A long narrow band of dark product was observed emanating from the crack origin. This feature was later determined to be a forming seam as confirmed by metallographic examination and electron microscopy. The seam depth was 0.020 inch and it contained a heavy dark iron oxide which formed during prior heat treatment. The chemical composition, hardness and surface finish all conformed to the governing specification. The failure of the nut was attributable to a forming seam. Evidence substantiating this conclusion is that the crack origin was located at the base of the seam, cadmium plating was found within the seam, the ductile mode of fracture and the heavy dark oxide on the surface of the seam.

Key Words: Failure Analysis, Forging Seam, Fractography.

Background: The Aviation and Troop Command (ATCOM) shipped a failed self-locking nut to the Army Research Laboratory-Materials Directorate (ARL-MD) for examination. The hexagon shaped, castellated self-locking nut (MS 17825-10) was attached to the lower end of the pitch link located on an AH-1S MOD army attack helicopter. The nut was fabricated from non-corrosion resistant steel and is to be used at temperatures up to 250°F. Upon routine inspection, it was noted that the nut was loose enough to allow the washer beneath it to spin. Further inspection revealed that the nut was cracked at one of the castellations (Reference QDR control # W41NB6940006). Figures 1 through 3 show the failed nut in the as-received condition.

Light Optical Microscopy: The locking nut received was split open approximately 1/8 of an inch along its longitudinal axis as shown in Figures 1-3. The cadmium plating was relatively intact with no evidence of general corrosion or pitting observed. Upon sectioning, it was determined that the crack originated from the O.D. of the nut as seen in Figures 4 and 5. The radial lines and chevron pattern converge to an area adjacent to the O.D. of the nut opposite the castellations. A few crack arrest marks were also observed emanating outward from this area. A long narrow band of dark oxide, such as that found after heat treating, was noted along the O.D. of the nut opposite the threads as shown in Figures 4 and 5. This region displayed a distinctly different morphology from the remaining fracture. This feature is shown enlarged in Figure 6. It was deduced that this artifact may be a forming seam. The forming seam encompassed the entire length of the nut. Figure 7 shows the seam and fracture surface in relation to one of the six castellations. Figures 8 and 9 depict the tail end of the seam and the castellation at higher magnification where extensive cracking within the cadmium plating is evident along the seam. This area represents that portion of the seam which did not fracture and was sectioned to determine the seam depth. The results can be found in the metallography section of this report. It should be noted that this seam was almost perfectly in alignment with an edge of a castellation (see Figures 7, 8, and 9). The fracture progressed under differential loading conditions as evidenced by several distinct crack arrest marks which indicate the origin area. The fracture surface morphology was generally ductile throughout with considerable inclusions aligned along the longitudinal axis, probably the extruding direction of the part.

Metallography: According to the governing specification, MIL-N-25027F, *Nut, Self-Locking, 250 Deg. F, 450 Deg. F, and 800 Deg. F*, the maximum allowance for a seam 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 acceptable as long as they do not exceed 1/2 the limits of those set forth in Table VII.

The failed part has a thread size of 0.625 inches and therefore falls under the above criteria. One half of the limit in Table VII is 0.017/2 or 0.0085 inches. A cross-section of the seam from the failed part can be seen in Figure 10. This section was taken from the area presented in Figures 8 and 9. The seam depth at this location is 0.020 inches as measured from the photomicrograph per ASTM-B-487. This is in violation of the limit set forth in MIL-N-25027F. The maximum depth of the seam was contained on the fracture surface and can be seen in the Scanning Electron Microscopy section of this report.

The microstructure of the nut was also investigated. There was no cleanliness rating or inclusion content limit found within the governing specification. However, the

inclusions within the nut matrix were large and numerous. Figures 11 and 12 show the microstructure of the nut in the as-polished condition in the longitudinal and transverse conditions, respectively. 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. The specimen microstructure can be seen in the etched state in Figures 13 and 14 for the longitudinal and transverse cross-sections, respectively. 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.

Table I
Chemical Composition of the Pitch Link Nut
Weight Percent

Element	Component	1137	11L37	4130	8740
Carbon	0.37	0.32-0.39	0.32-0.39	0.28-0.33	0.38-0.43
Manganese	1.45	1.35-1.65	1.35-1.65	0.40-0.60	0.75-1.00
Phosphorus	0.013	0.04 max.	0.04 max.	0.025 max.	0.025 max.
Sulfur	0.11	0.08-0.13	0.08-0.13	0.025 max.	0.025 max.
Silicon	0.16	*	*	0.15-0.35	0.15-0.35
Chromium	0.021	*	*	0.80-1.10	0.40-0.60
Nickel	0.006	*	*	0.25 max.	0.40-0.70
Molybdenum	0.0011	*	*	0.15-0.25	0.20-0.30
Lead	0.31	*	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
77
59
90
110
62
Longitudinal
108
51
80
57
72
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
65.1
65.6
65.9
66.2
66.2
Longitudinal
65.4
66.6
66.7
66.3
66.6
Average 66.1

Scanning Electron Microscopy: The fracture surface of the failed nut was characterized utilizing Scanning Electron Microscopy (SEM). The surface with the exception of the forming seam was composed entirely of ductile dimples with

elongated inclusions characteristic of an overload failure. Figure 15 shows the fracture surface morphology of the nut. The inclusions can be seen at higher magnifications in Figure 16. The inclusions were identified to be manganese-sulfide through Energy Dispersive Spectroscopy (EDS) in conjunction with the SEM (see Figure 17). Inclusions of this severity can be deleterious to the mechanical properties of the material. Another important finding through EDS was the fact that the forming seam contained cadmium (see Figure 18). This provided evidence that the seam was open to the surface and occurred during manufacturing as opposed to a service related anomaly. Figure 19 depicts a Backscatter Electron Image (BEI) of the crack origin area. EDS also helped identify that the nut contained lead as verified through chemical analysis as seen in Figure 20. The semi-circular crack arrest marks can clearly be seen emanating outward from the base of the seam. Figure 21 shows the featureless morphology within the forming seam which would be anticipated from such a manufacturing defect. Figure 22 is a representative EDS spectra from the seam revealing a high concentration of iron and oxygen. Figure 23 is a high magnification of the maximum depth of the forming seam. Per ASTM-E-487, the maximum depth of the seam was measured to be 0.0206 inches.

Failure Scenario: The MS 17825-10 self-locking nut that secures the lower end of the pitch link failed in overload. This failure was the result of a forming seam which was located exactly in line with an edge of one of the six castellations on the nut. The crack progressed by a few successive high stresses until final fracture. The failure was also aided by the materials overall low degree of cleanliness due to the numerous elongated inclusions within the matrix. The nut failed while under normal torquing and applied service loads due to the existence of the "out of specification" forming seam. Evidence substantiating this scenario is the fact that the crack origin was determined to be at the base of this seam, the cadmium found within the seam, dark heat treat oxide, the ductile mode of fracture, and the multiple successive crack arrest marks emanating outward from the bottom O.D. of the nut at the base of the seam.

Conclusion: The failure of the MS 17825-10 Pitch Link Self-Locking Nut that secures the lower end of pitch link on an AH-1S MOD army attack helicopter was the result of a forming seam. Cracking originated at a forming seam that ran along the O.D. of the nut adjacent to an edge of one of the six castellations or slots. The crack progressed by overload along the longitudinal axis of the nut. Crack propagation was aided due to a significant amount of Mn-S stringers within the material.

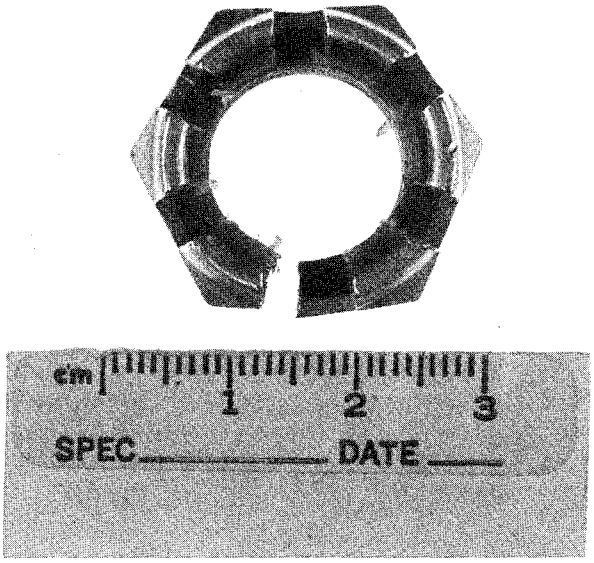


FIGURE 1. Top View of the failed Pitch Link Nut showing castellations and split opening width.

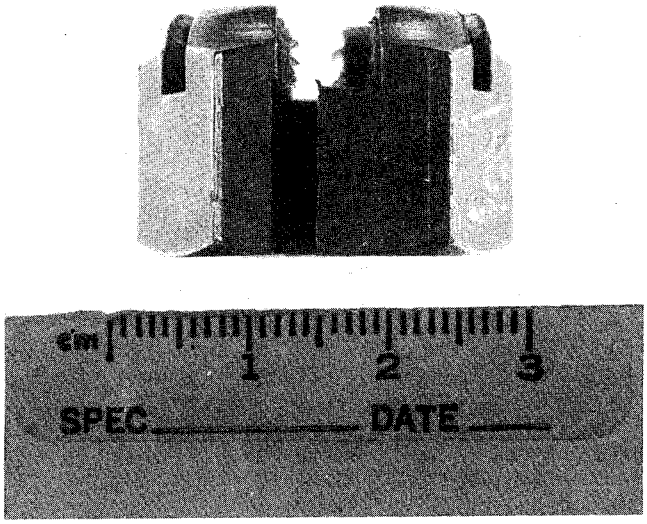


FIGURE 2. Side View of the failed Pitch Link Nut showing split opening width and fracture surface proximity to castellation edge.

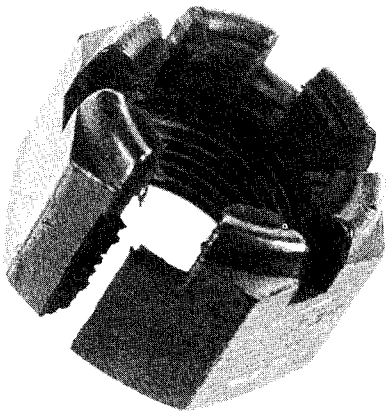
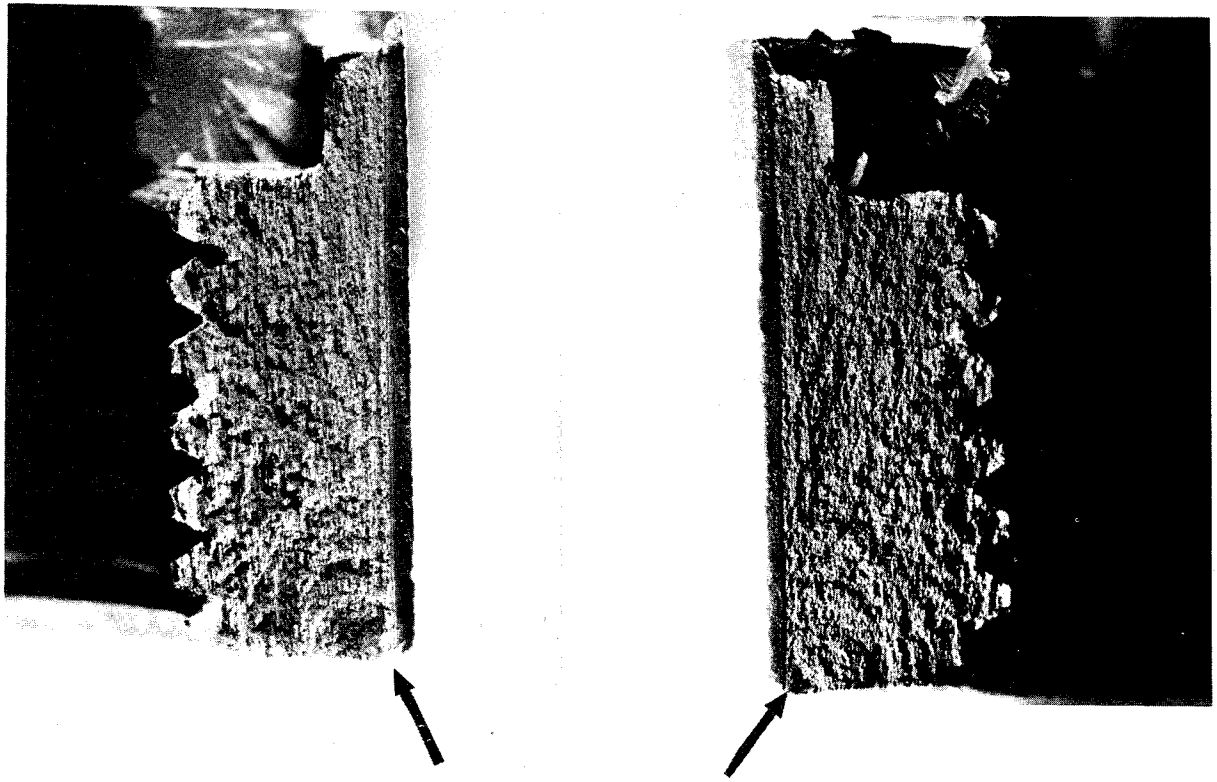


FIGURE 3. Off Axis View of the failed Pitch Link Nut.



FIGURES 4 and 5. Fractograghs showing both halves of the failed nut fracture surface. Mag. 10x.

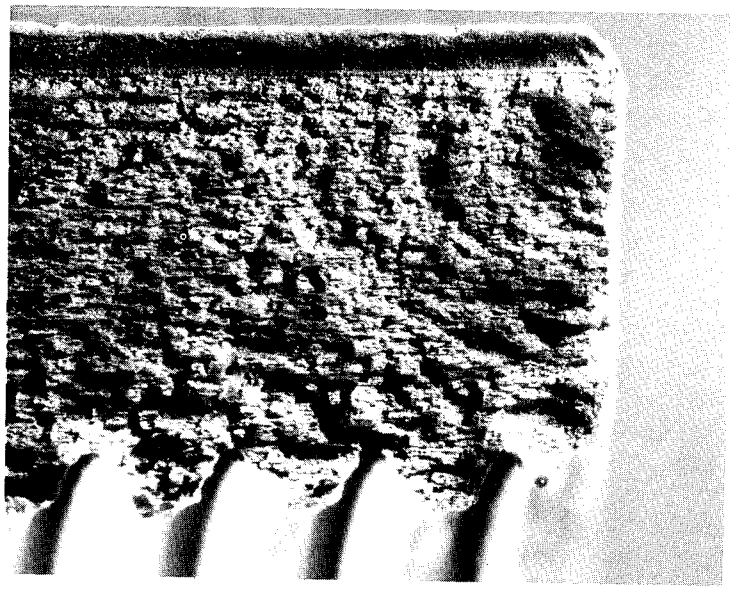


FIGURE 6. Fractograph showing crack arrest lines emanating from the base of the forming seam. Mag. 20x.

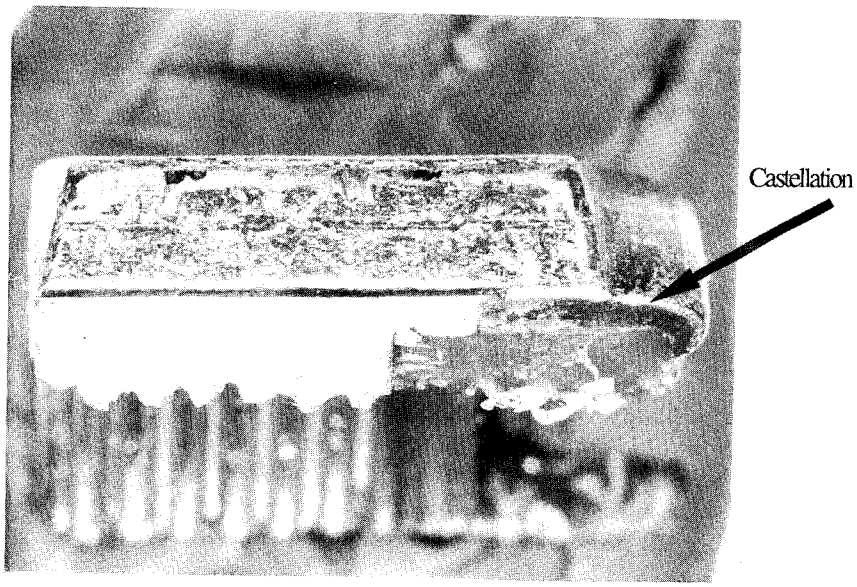


FIGURE 7. Fractograph showing the forming seam and fracture surface proximity to the castellation slot. Mag. 7.5x.

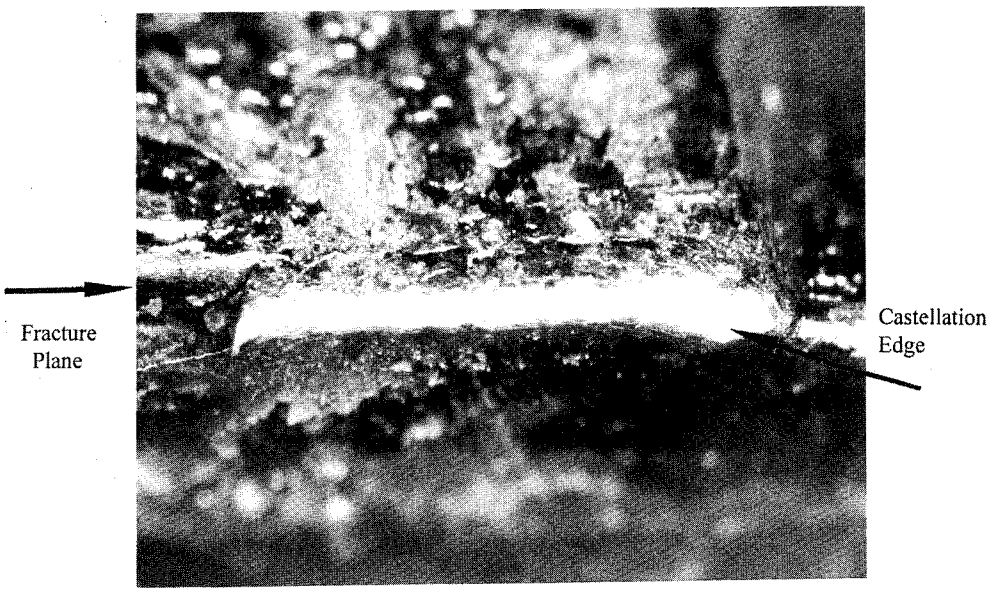


FIGURE 8. Detail of area in FIGURE 7 of surface cracking due to the presence of the forming seam. Mag. 40x.

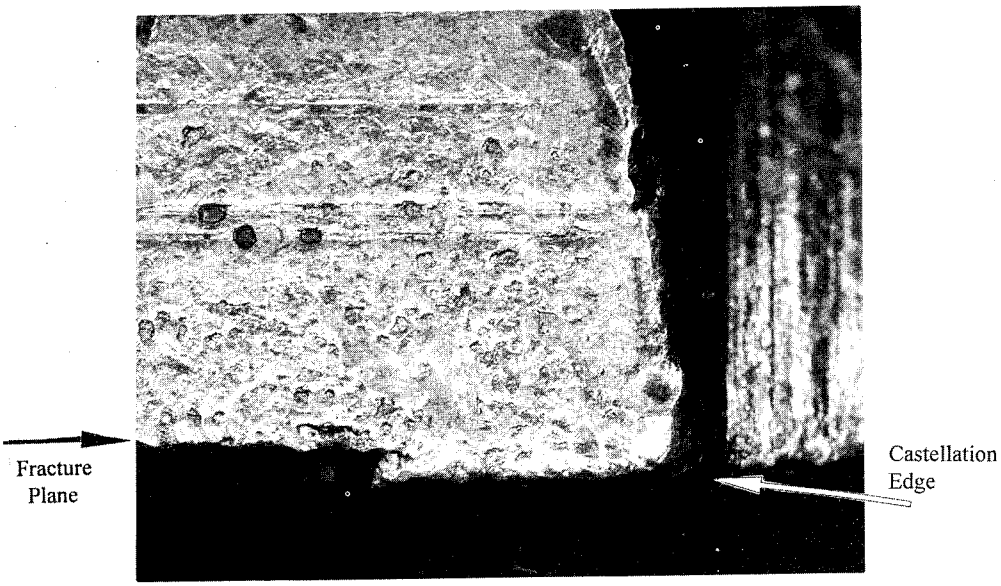


FIGURE 9. Detail of area in FIGURE 8 showing fracture surface proximity to edge of castellation. Mag. 12.5x

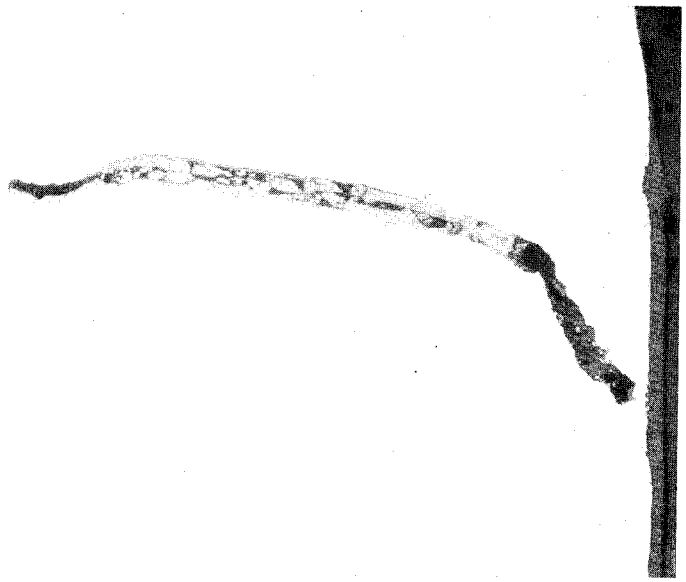


FIGURE 10. Micrograph of a cross-section of the forming seam taken through area in FIGURES 8 and 9. Mag. 200x.

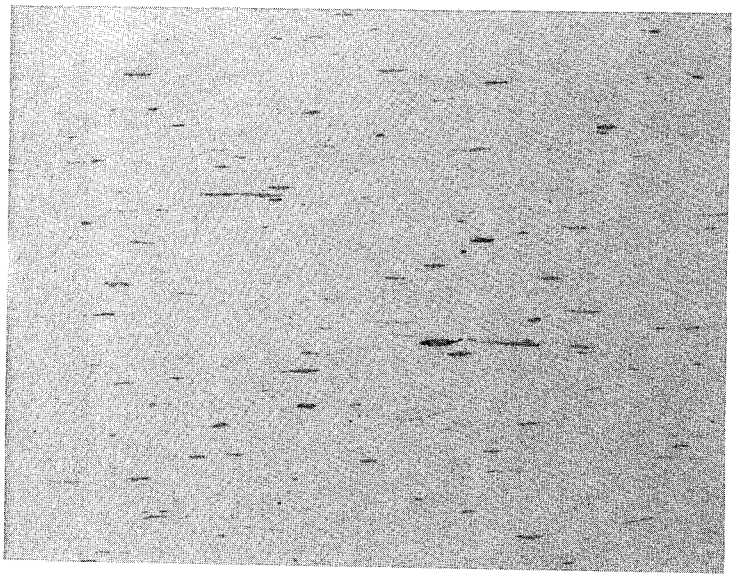


FIGURE 11. Micrograph of the as-polished longitudinal cross-section showing elongated inclusions. Mag. 100x.

FIGURE 12. Micrograph of the as-polished transverse cross-section showing the extent of the inclusions. Mag. 100X.

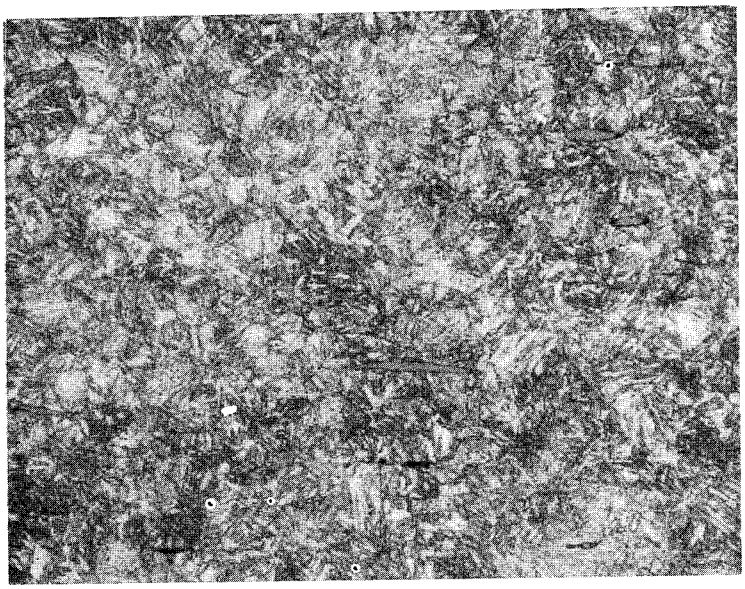


FIGURE 13. Micrograph of the longitudinal cross-sectional microstructure. 4% Nital Mag. 400x.

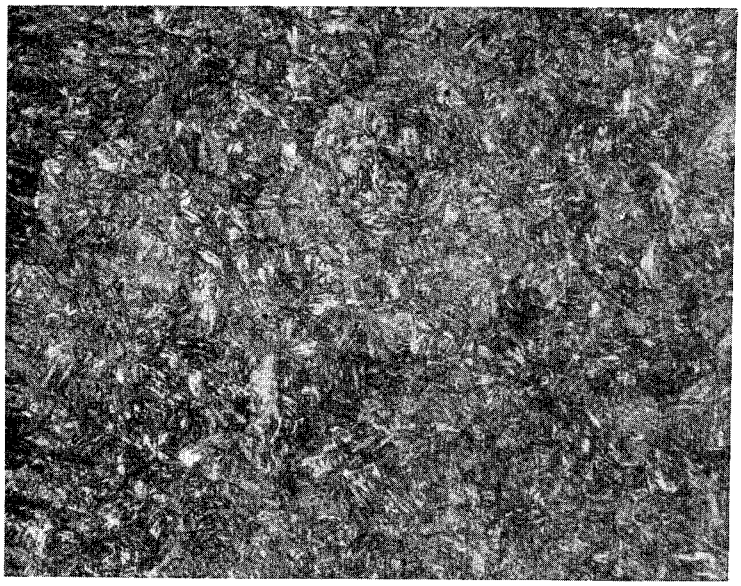


FIGURE 14. Micrograph of the transverse cross-sectional microstructure. 4% Nital Mag. 400x.

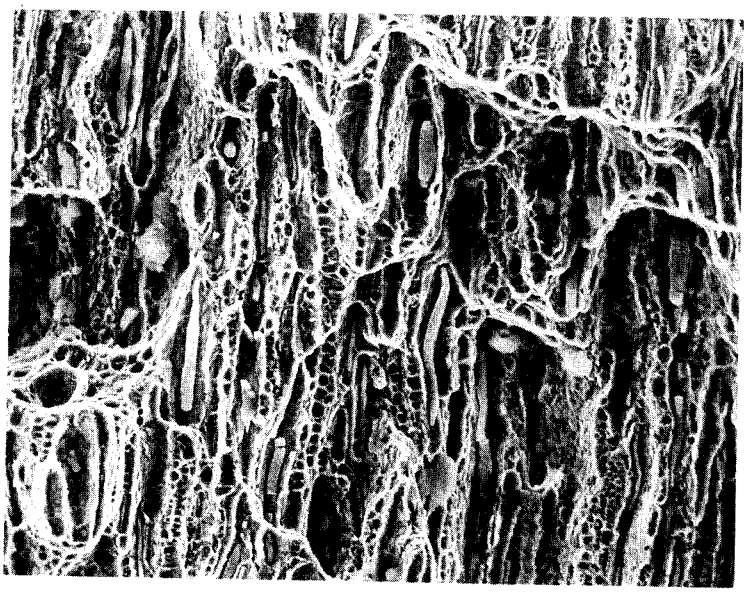


FIGURE 15. SEM fractograph showing the ductile dimpled morphology with numerous inclusions. Mag. 750x.

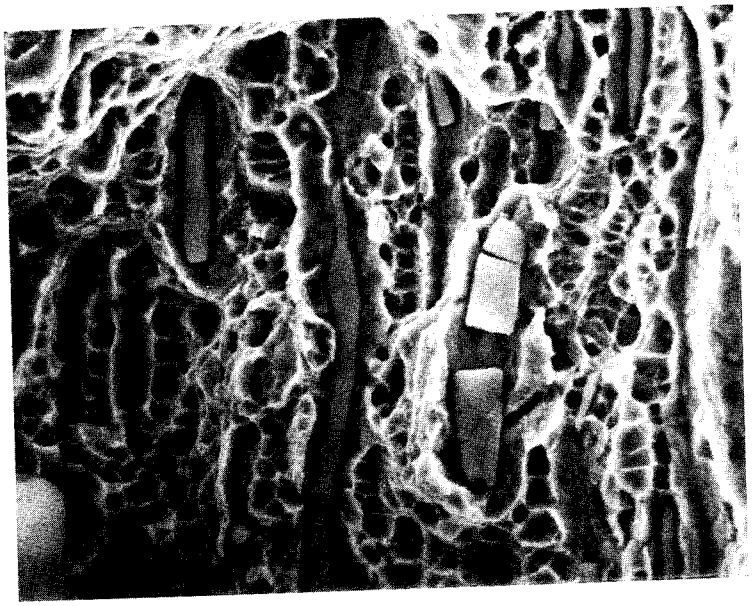


FIGURE 16. SEM fractograph showing the ductile dimpled morphology and inclusions in FIGURE 15 at higher magnification. Mag. 2000x.

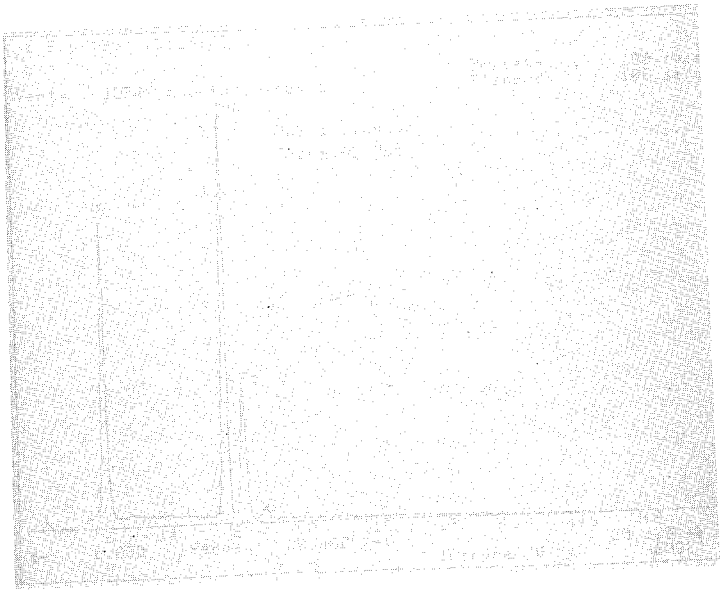


FIGURE 17. EDS spectrum of an inclusion determined to be Mn-S.

FIGURE 18. EDS spectrum of the forming seam containing cadmium.

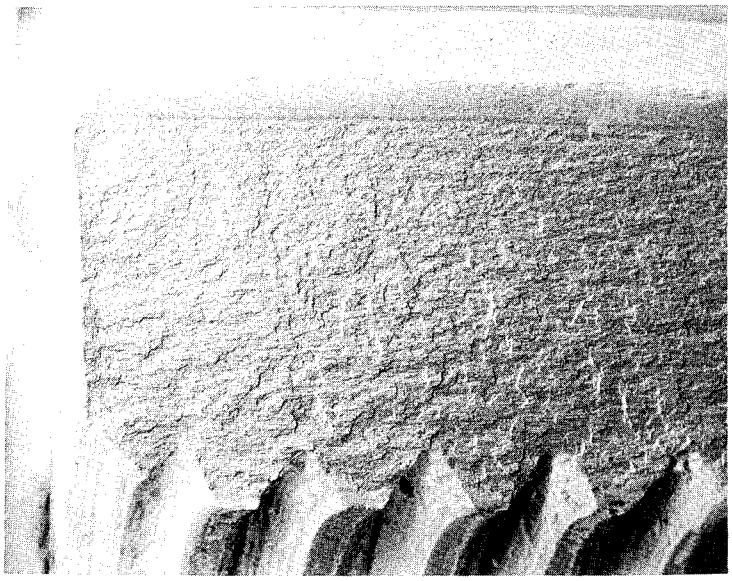


FIGURE 19. BEI fractograph showing the crack arrest marks radiating from the base of the forming seam. Mag. 12x.

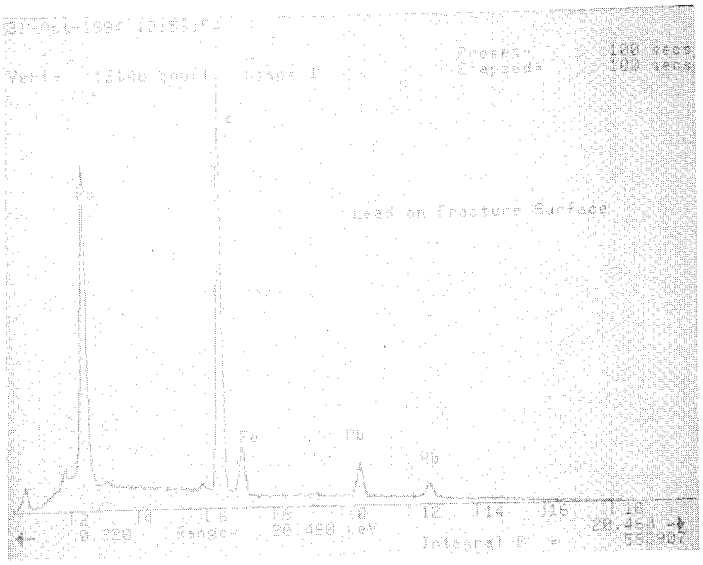


FIGURE 20. EDS spectrum showing lead on the fracture surface.

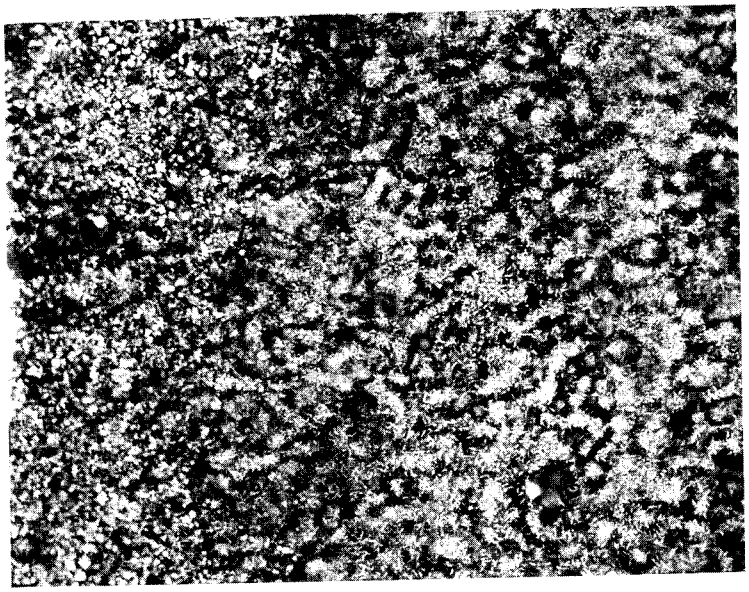


FIGURE 21. SEM fractograph showing the featureless morphology within the forming seam. Mag. 1500x.

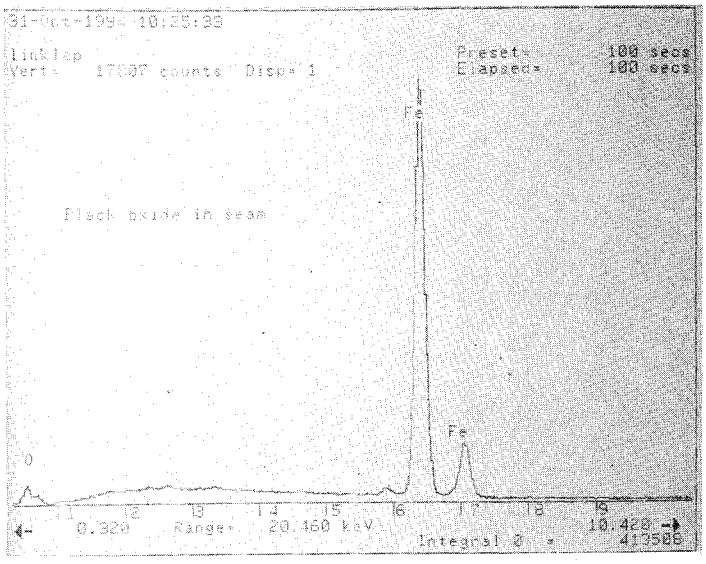


FIGURE 22. EDS spectrum within the seam revealing concentrations of iron and oxygen.

FIGURE 23. SEM fractograph of the maximum depth of the forming seam. Mag. 200x.

FAILURE ANALYSIS OF QUICK REACTION SATELLITE ANTENNA (QRSA) TRUNNION CASTINGS

Wego Wang and Martin G.H. Wells
U.S. Army Research Laboratory, Materials Directorate, Watertown, MA 02172

Abstract: Repeated failures of quick reaction satellite antenna (QRSA) trunnion castings in the field caused concerns about their durability. A number of possible causes for the failures were identified, ranging from improper component installation, unbalanced and/or excessive force resulting from poor design to insufficient material strength. Incipient melting during solution treatment was also suspected as one of the reasons responsible for cracking in the original aluminum castings. This study compared the microstructures of as-cast and heat-treated samples and concluded that incipient melting did not occur during solution treatment and thus was not the principal cause for cracking in the aluminum castings. Eventually, the light aluminum alloy was replaced by a heavier but stronger ductile iron. As a result, the durability of the QRSA trunnion casting has been significantly improved and the deficiency was remedied.

Key Words: Aluminum casting; failure analysis; incipient melting; trunnion casting

Historical Background: The quick reaction satellite antenna (QRSA) trunnion casting connects the antenna and the base as shown in Figure 1. It was originally a 771.0 aluminum alloy (Al-7Zn-0.9Mg) casting. The first cracking failure was reported in 1987. Material deficiency, improper heat treatment and irregularity of installation were cited as the possible major causes for these cracks. As a result, all fielded Al QRSA trunnion castings had to be replaced. Subsequently, the Government approved the implementation of an Engineering Change Proposal (ECP#62 - includes improved flatness tolerance) in 1988. Although the problem was considered solved, another failure of a modified Al QRSA trunnion casting (SN10A) was reported in early 1989. The improper implementation of the engineering change was cited as the cause. Cracks were later found in trunnion castings at Shaw Air Force Base (AFB); components were sent to the Naval Aviation Depot, Materials Engineering Laboratory (NMEL), Jacksonville, Florida for analysis. These cracks were partially attributed to stresses set up by surface mismatch between the trunnion casting and the pallet. The Naval Aviation Laboratory performed a finite element analysis that suggested an inadequate trunnion design and also experimentally found a latent flaw in the sample part; NMEL concluded that the casting might have been overheated during solution treatment causing incipient melting. In July, 1990, another crack in a trunnion casting (SN004A) was found at Patrick AFB. Three Al QRSA trunnion castings were sent by the Communication Electronics Command, Center for Space Systems (CECOM), Fort Monmouth, New Jersey, to the U.S. Army Research Laboratory (formerly the U.S. Army Materials Technology Laboratory), Watertown, Massachusetts for evaluation in October, 1990.

Cracks were hardly visible with the naked eye on the three Al QRSA trunnion castings when first received at the U.S. Army Research Laboratory-Watertown (ARL-Watertown) because surfaces were painted. For a better surface examination one Al casting was sent to CDS Group, Houston, Texas, for paint stripping. Figure 2 shows the bottom view of the stripped part and reveals cracks adjacent to the bolt hole marked "x".

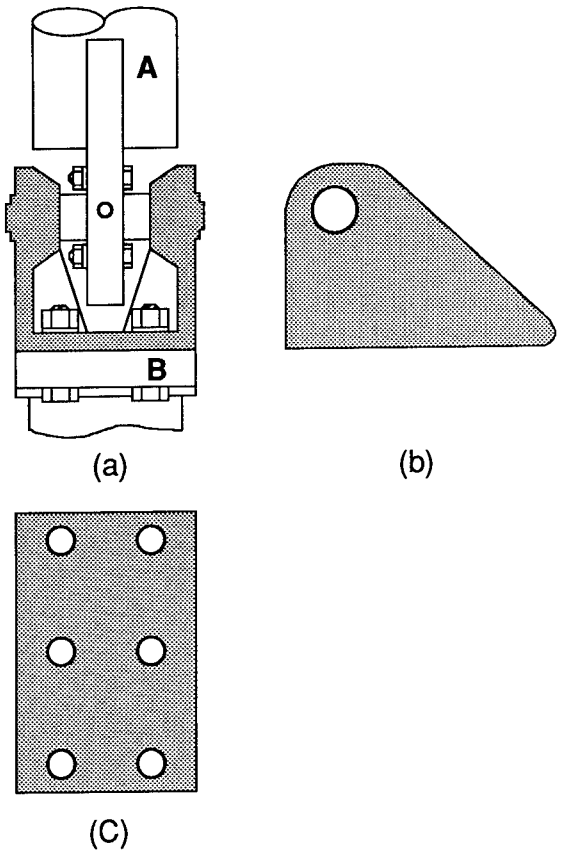


Figure 1. The quick reaction satellite antenna trunnion casting (shaded) showing (a) the overall view of the connection between the antenna (A) and the base (B), (b) side view and (c) bottom view.

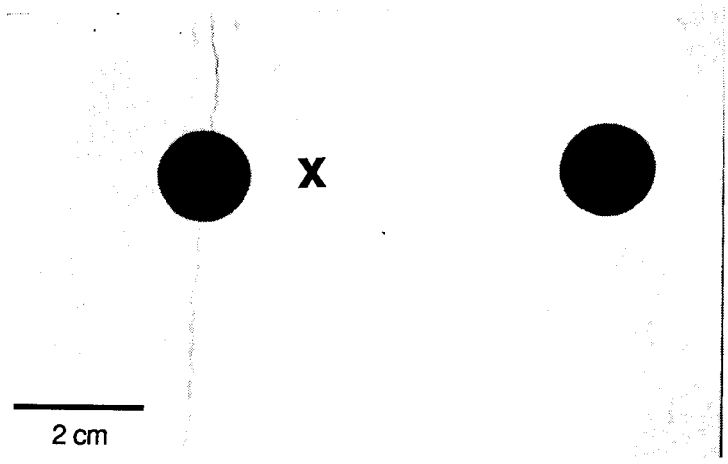


Figure 2. The bottom view of the stripped Al trunnion casting showing hair-line cracks adjacent to the bolt hole x.

To determine the heat treatment effect, one as-cast 771.2 aluminum alloy ingot was obtained from the contractor for microstructural examination and comparison. During the course of this evaluation, the Government approved the change of the cast material from 771.0-T6 aluminum alloy to ductile iron (grade 100-70-03). ARL-Watertown was also furnished with one sample of this ferrous casting. In this study, the following evaluation procedures were followed: part and material identification; paint removal; microstructural analysis; radiographic analysis; mechanical property evaluation; and durability analysis.

Part and Material Identification: The three aluminum trunnion castings had been coated with green paint and labeled with part numbers 80063-A3020139-00X (X's are numbers associated with individual parts). They had been removed from antennas in service. The original cast material was 771.0 Al alloy but was later changed to ductile iron to sustain a torque up to 40 ft-lbs. The aluminum alloys studied in this report were represented by two different numerical codes: 771.0 and 771.2. They belong to the same family; 771.0 usually refers to sand castings and 771.2 refers to ingots. The 771.2 has a similar chemical composition as that of 771.0 except less Si (0.10 versus 0.15%) and Fe (0.10 versus 0.15%) [1]. Because these cast aluminum alloys are not weldable [2], parts made of these alloys are usually joined by mechanical means. These QRSA trunnion castings are connected to the base with bolts. Any externally applied mechanical stress overloads at the joints might result in cracks. The 771.0 Al alloy under T6 temper condition also has only marginal resistance to general corrosion and to stress-corrosion cracking [3].

Ductile iron is a cast iron with all or most of its second phase graphitic carbon in the form of spheroids or nodules; this results from the addition of alloying elements, such as magnesium to the liquid melt. It is also known as spheroidal or nodular iron. The as-cast microstructure normally consists of graphite nodules surrounded by ferrite (bull's-eye structure) in a matrix of pearlite; it may also include some free cementite [4]. As specified by ASTM, the ductile iron of grade 100-70-03 has minimum ultimate tensile and yield strengths of 690 and 483 MPa (100 and 70 ksi), respectively, and a minimum tensile elongation of 3% [5].

The composition and properties of these materials are summarized in Table I. The reference property values were quoted from ASM and ASTM Handbooks [5 and 6]. The analyzed compositions were quoted from an earlier report [7] for 771.0 aluminum alloy and provided by the manufacturer for the ductile iron. Both typical and minimum values of mechanical property for 771.0 Al alloy are listed. They were determined using separately sand cast bars with a 1/2-in diameter. The effects of casting section thickness on properties were not considered. The minimum property values listed in Table I should not be interpreted as the conventional "minimum or limit values" used in mechanical design practices. However, this data serves as a useful reference for relative property comparison and quality control. 771.0 Al alloy is in a T6 condition and the ductile iron (grade 100-70-03) is in a tempered condition.

Microstructural Analysis: Metallographic studies by NMEL concluded that incipient melting was beginning to occur during the solution treatment of the Al castings [7]. Long and narrow secondary constituents, some having sickle or crescent shapes, were found at grain boundaries and the triple points of grain boundary intersections. These secondary phases were evidence of accidental heating above the eutectic temperature during the solution treatment. The poor temperature control was consequently cited as a possible cause for the failure of Al trunnion castings. For comparison and confirmation of this

Table I. Materials Characteristics [Ref. 5-7]

Property/Composition	771.0 Aluminum Alloy - T6		Ductile Iron (Tempered) Grade 100-70-03
Nominal Composition (%)	6.5-7.5 Zn, 0.8-1.0 Mg, 0.10-0.20 Ti, 0.06-0.20 Cr, 0.15 max Si, 0.15 max Fe, 0.10 max Cu, 0.10 max Mn, rem Al		N/A
Analyzed Composition (%)	6.7 Zn, 0.85 Mg, 0.18 Ti, 0.07 Cr, <0.04 Si, <0.02 Fe, 0.05 Cu, <0.03 Mn, rem Al		3.64 C, 2.59 Si, 0.62 Mn, 0.379 Cu, 0.134 Mo, 0.073 Cr, 0.051 Mg, 0.035 Ni, 0.011 S
Ultimate Tensile Strength, MPa (ksi)	Typical	Minimum	Minimum
	345 (50)	290 (42)	690 (100)
Yield Strength, MPa (ksi)	275 (40)	241 (35)	483 (70)
Elongation (%)	9.0	5.0	3.0
Hardness (Brinell)	N/A	90	N/A
Density, $lb/in^3 (g/cm^3)$	N/A	0.102 (2.823)	N/A

conclusion, NMEL attempted to obtain some as-cast materials and heat-treated (T6) specimens from the manufacturer and developer of the Al casting alloy Precedent 71 (AA771). Aluminum casting alloy Precedent 71 was the former name before the alloy was officially designated as 771.2 Al alloy. Neither the reference micrographs or the as-cast ingots were ever sent to NMEL. However, in November, 1990, the manufacturer furnished ARL-Watertown with one as-cast 771.2 Al casting for metallographic analysis. Extensive metallographic studies at ARL-Watertown showed that the as-cast and the heat-treated specimens had very similar microstructures. The structure of the heat-treated sample showed equiaxed grains with both angular and spherical cavities, and secondary constituents along the grain boundaries as shown in Figure 3. Since a similar microstructure of porosity and secondary constituents also existed in the as-cast sample before solution treatment, as shown in Figure 4, we concluded that these features resulted from the casting/solidification process rather than by the subsequent solution treatment. Any rounded cavities and secondary constituents were isolated and usually randomly distributed in the matrix and along the grain boundaries; the sickle- or crescent-shaped ones were often in clusters or networks along the grain boundaries, particularly at the grain boundary triple intersection points. We observed more dendritic cells in the as-cast material, but the size and volume fraction of porosity and secondary constituents were about the same in the heat-treated and the as-cast samples. No evidence indicated that cracks initiated at these cavities, secondary constituents or the grain boundary triple points. It was our conclusion that incipient melting did not occur during the solution treatment and was not the major cause for failure.

Figure 5 shows the nodular microstructure of the ductile iron casting that replaced the Al cast alloy. It has a typical bull's-eye structure with graphite nodules surrounded by ferrite in a pearlite matrix. The existence of any excessive amount of free carbide in ductile iron would impair its ductility and machinability. Since only a few scattered free carbide particles were observed in the sample material and a higher-than-specified elongation was obtained for this ductile iron (see later section - Mechanical Property Evaluation), their effects were rather insignificant. All bull's-eyes are about the same size and are evenly distributed in the matrix. As expected, there was no debonding or cracking between the boundaries of the graphite/ferrite or ferrite/pearlite. The presence of microporosity and carbides was also minimal. In summary, our examination found that the vendor produced a sound ductile iron casting.

Figure 3. Equiaxed microstructure of a heat-treated 771.0 Al sample showing both angular and spherical cavities and secondary constitutents.

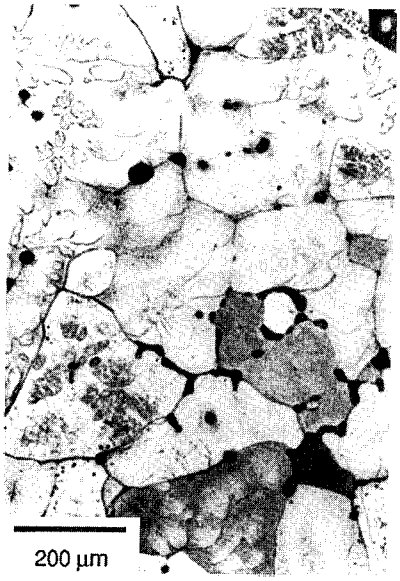


Figure 4. Microstructure of as-cast 771.2 sample which is similar to that of heat-treated material as shown in Figure 3.

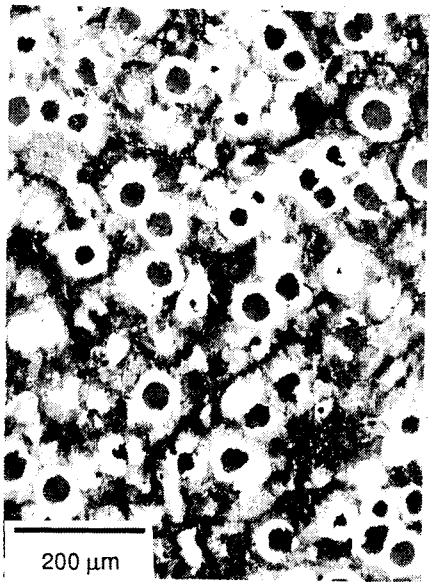


Figure 5. The nodular microstructure of the ductile iron showing graphite nodules surrounded by ferrite in a pearlite matrix.

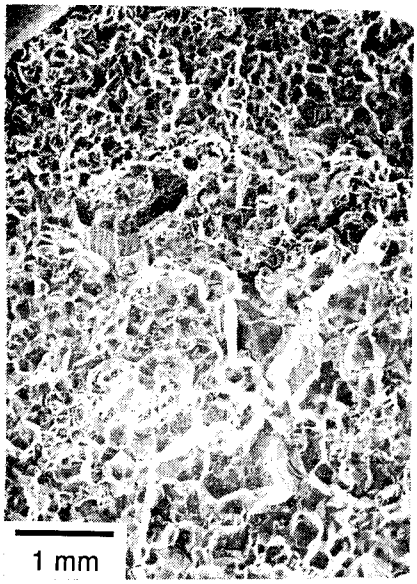


Figure 6. An overall tensile fracture surface of the Al trunnion casting sample indicating brittle fracture with secondary cracks.

NMEL found similar fracture surface appearances and topography (blocky or rock-candy type) for both samples that had been in service and those produced in the laboratory by overload. The blocky or rock-candy type fracture was evidence that the fracture mechanism was exclusively intergranular. Several inclusions and voids were observed in the casting, but no shrinkage porosity was discerned on the fracture surfaces of the existing cracks. It was concluded that in-service failures were induced by overloads based on the similarity with fracture surfaces of tensile bars pulled to failure [7]. A SEM fractographic study at ARL-Watertown arrived at similar conclusions. Figure 6 shows an overall fracture surface of the 771.0 Al trunnion casting tensile specimen indicating brittle fracture with secondary cracks. The sickle-shaped voids were clearly revealed at higher magnification in Figure 7. Intergranular fracture with rocky facets is illustrated in Figure 8. Residual dendritic cells and river-pattern fracture were also observed as shown in Figure 9. In a brittle coarse-grained material cracks usually propagate along the weakest crystalline planes or along the grain boundaries. As a result, a combination of cleavage brittle fracture with river patterns and intergranular fracture occurred in the cast 771.0 Al trunnion.

The tensile specimen of ductile iron tested at room temperature showed little or no necking and the fracture surface was flat. Any secondary (either longitudinal or transverse) cracks on the external specimen surface immediately underneath the fractured cross-sectional surface were also minimal. The overall tensile fracture fractograph of ductile iron is shown in Figure 10 indicating a brittle fracture. Minor secondary cracks were observed in the matrix of the fracture surface, however the interfaces between the nodular graphite and the matrix were mostly intact without interfacial debonding as shown in Figure 11 and in the back-scattered SEM photo of Figure 12. Closer examination showed fractured pearlite lamellae indicating that the transgranular cleavage mechanism predominated in the matrix as shown in Figure 13. Cracks were also observed in a graphite nodule in Figure 13. Many graphite nodules were left intact as spheroids but several others were sheared during crack propagation. Only limited stretching elongation was observed around most graphite nodule-bearing cavities due to the low ductility. However, most graphite nodules remained in place after fracture. Some cracks interconnecting adjacent graphite nodules were observed but not too frequently. In summary, there were three major operative fracture modes: first, brittle cleavage fracture with river pattern and plateau characteristics predominated in the pearlitic matrix; second, brittle fracture was observed around isolated graphite nodules because the surrounding ferrite rings were under severe mechanical restraint due to the harder and nondeformable pearlitic matrix; finally, relatively ductile tearing and microvoid coalescence occurred in areas clustered with closely-spaced graphite nodules and mild cup-cone type fracture was also observed.

Crack Detection and Radiographic Analysis: The previous investigation [7] reported that when Al QRSA trunnion castings in service at Shaw AFB were stripped of paint, liquid penetrant inspection revealed cracks in these castings. Some of the larger cracks were visible with the naked eye and open porosity of 1/8 inch in diameter was reported. These findings were later confirmed by NMEL and more cracks were also identified in the same castings by eddy current inspection.

Extensive radiographic inspection was conducted on a ductile iron casting at ARL-Watertown. In contrast to the aluminum castings reported in the previous investigation, neither hair-line cracks around the bolt holes nor larger cavities were observed in the ductile iron casting. Only minor shrinkage and porosity existed. These "defects" were considered within the acceptable range for normal castings. From the radiographic analysis it was concluded that there was no unusual inherent defect in the ductile iron trunnion resulting from the casting process.

Figure 7. Sickle-shaped intergranular cavities in the Al trunnion casting.

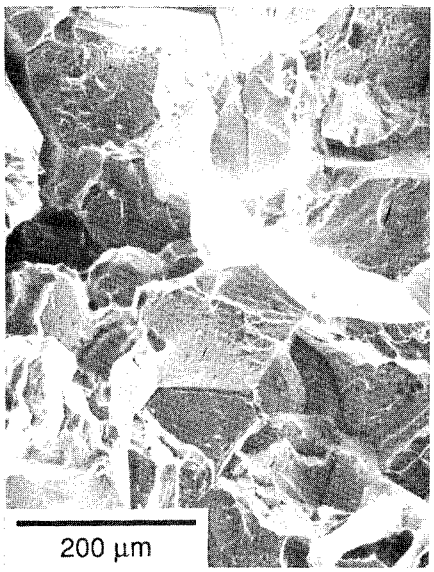


Figure 8. Intergranular rocky fracture observed in the Al trunnion casting.

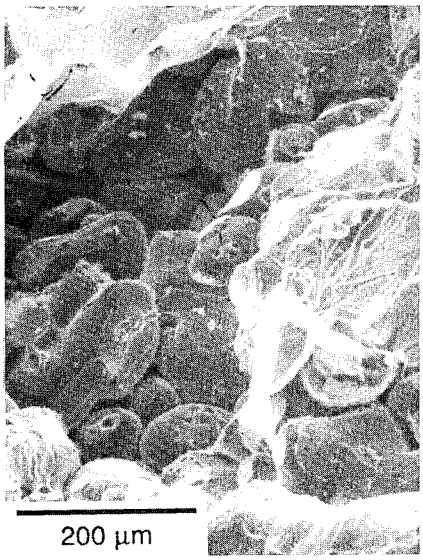


Figure 9. Residual dendritic cells and river pattern fracture observed in the Al trunnion casting.

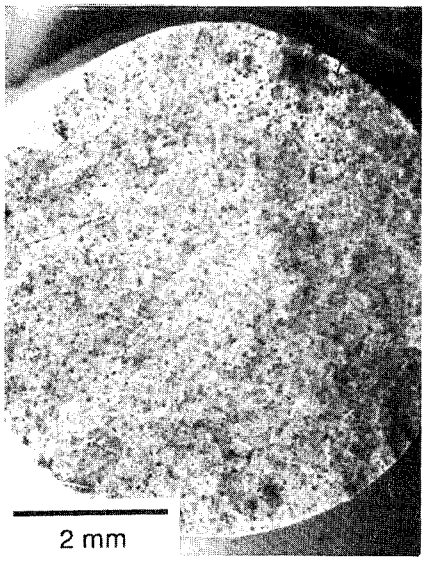


Figure 10. SEM fractograph of the ductile iron tensile specimen.

Figure 11. SEM fractograph of the ductile iron tensile specimen showing minor secondary cracks in the matrix.

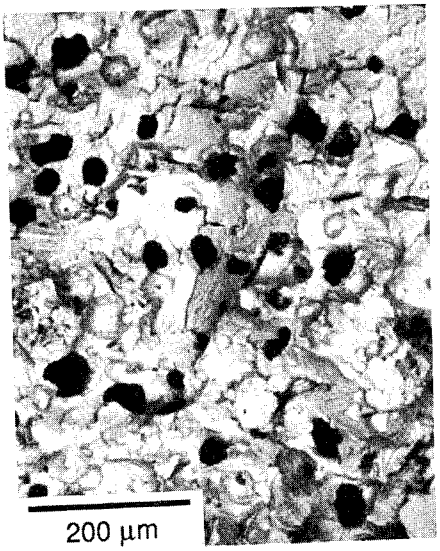


Figure 12. Backscattered SEM fractograph of the ductile iron tensile specimen.

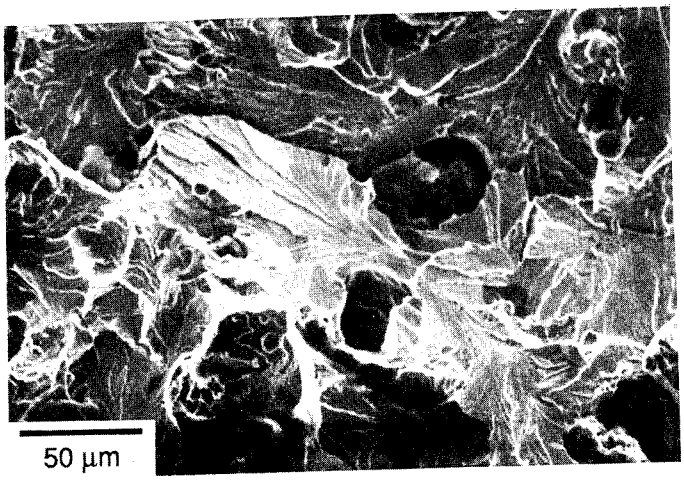


Figure 13. A higher magnification fractograph of the ductile iron tensile specimen showing transgranular cleavage fracture in pearlite.

Mechanical Property Evaluation: Compared with the minimum values of properties listed in Table I, the ARL-Watertown test results for Al castings, in Table II, are slightly higher for ultimate and yield tensile strengths; a lower ultimate tensile strength was reported by NMEL. Note, however, that a much lower elongation was measured for these castings, by at least a factor of 2 less than the specified values. According to the Federal Specification for Aluminum Alloy Sand Castings, QQ-A-601F, Section 3.3.2.1, "Unless otherwise specified, the average ultimate tensile strength and average elongation of test specimens cut from castings ..... shall be no less than 75 percent and 25 percent, respectively, of the values specified (in Table I) for separately cast bars." Thus, in general these Al castings have met the specification requirements for mechanical properties.

The cast iron mechanical property values reported by ARL-Watertown and the manufacturer are comparable with the reference values in Table I. The only significant difference is the higher elongation, 9.3 (or 10)% versus 3%, which is considerably above the requirement. The good elongation values are consistent with the yield strengths that are close to the required minimums. The tensile properties reported by ARL-Watertown are the average of 4 tests for Al casting and 3 for ductile iron, respectively. The Charpy impact value is the average of 6 tests and the hardness value is the average of 28 tests. The tensile data from NMEL is the average of 7 tests.

Table II. Measured Materials Properties

Property	771.0 Al Alloy		Ductile Iron Grade 100-70-03	
Source	ARL-Watertown	NMEL (7)	ARL-Watertown	Manufacturer
Ultimate Tensile Strength, MPa (ksi)	299 (43.3)	272 (39.4)	746 (108.2)	730 (105.9)
Yield Strength MPa (ksi)	285 (41.4)	N/A	441 (64.0)	474 (68.7)
Elongation (%)	~2	2.5	9.3	10.0
Reduction of Area (%)	<2	N/A	6.3	N/A
Charpy Impact (ft-lb)	N/A	N/A	2.67	N/A
Hardness (Brinell, Rockwell B or C)	N/A	55.1~59.3 (Rockwell B)	19.9 (Rockwell C)	241 (Brinell) 22~23 (Rockwell C)

Discussion and Conclusions: Based on the limited available sample material and information at that time, the conclusions in an earlier Naval report [7] included the following findings:

1. The material of the QRSA trunnion Al castings met all of the requirements of Federal Specification, QQ-A-601(F).
2. Microvoids with characteristics of both incipient melting and gas porosity were observed, however, their location and irregular shapes suggested incipient melting. The presence of crescent-shaped secondary constituents on the grain boundaries and grain boundary intersections also suggested incipient melting.

After more reference material was made available for evaluation the current study confirmed the finding of item 1 but concluded differently for item 2. NMEL did not receive the as-cast ingot at the time it issued the report on February 23, 1990. Thus, a direct comparison of metallographic features between the as-cast and the heat-treated (T6) specimens was not made. Based on their limited observations the beginning of incipient melting was strongly suggested. However, it was reported that there was not sufficient evidence to firmly support

or exclude either of the two mechanisms, gas entrapment or incipient melting, for forming the microvoids.

Incipient melting may occur at grain boundaries during solution treatment by heating the specimen above the local liquidus temperature. The resulting microstructural characteristics are irregular microvoids on grain boundaries and grain boundary intersections, and a lacy network of grain boundary secondary constituents; but these are not sufficient conditions for proof of the occurrence of incipient melting. All these features were observed by NMEL and ARL-Watertown for the heat-treated specimens, and a very similar microstructure was also observed for the as-cast ingot reference sample at ARL-Watertown. In particular, about the same amount of microvoids and secondary grain boundary constituents were found in both as-cast and heat-treated samples. It is concluded in this study that the microvoids and secondary constituents formed during the process of casting and not as a consequence of incipient melting during the subsequent solution treatment.

The trunnion casting is subject to dynamic loading in service and the exact stress cycles are not available. Arguments have been made on the various causes of repeated failures. The 771.0 Al alloy was originally chosen for its low density. Despite disagreement on the causes of failure, the repeated failures demonstrate that this Al components' load bearing capacity at best only provides a low margin of safety in service. The replacement of the aluminum trunnion casting with ductile cast iron has solved the original cracking problem and greatly improved service reliability.

Acknowledgments: This study was funded by Contract FY7620-90-00266 granted by Communication Electronics Command Center for Space Systems (CECOM) in Fort Monmouth, New Jersey. The authors are indebted to Mr. Paul J. Huang for his help with SEM and EDS analyses, Mr. John C. Beck for his help preparing the manuscripts and Mr. Thomas Harkins for radiographic analysis. The authors are also grateful to CDS Group, Houston, Texas, for paint removal from the cast sample.

REFERENCES

1. Registration Record of Aluminum Association Alloy Designations and Chemical Composition Limits for Aluminum Alloys in the Form of Castings and Ingot, pp.1-11, the Aluminum Association, Inc., Washington, D.C., 1989.
2. ASM Metals Handbook, v.2, 9th ed., Metals Park, Ohio, 1979, p.193.
3. ASM Metals Handbook, v.2, 9th ed., Metals Park, Ohio, 1979, p.210.
4. ASM Metals Handbook, v.9, 9th ed., Metallography and Microstructures, Metals Park, Ohio, 1985, p.245.
5. Annual ASTM Book of Standards, A536-84, v.1.02, "Standard Specification for Ductile Iron Castings," Philadelphia, PA, 1990, p.303.
6. ASM Metals Handbook, v.2, 9th ed., Metals Park, Ohio, 1979, p.178 and p.149.
7. F.W. Johnston, Materials Engineering Laboratory Report No. 341-3-90, QRSA Trunnion Casting #139 Failure Analysis, Naval Aviation Depot, Naval Air Station, Jacksonville, Florida, Feb., 1990.

FAILURE OF A MAIN ROTOR PITCH CONE BOLT FROM AN ARMY ATTACK HELICOPTER

Scott Grendahl
Victor K. Champagne
Gary Wechsler
Marc Pepi
Daniel J. Snoha
US Army Research Laboratory
405 Arsenal Street AMSRL-MA-CB-292
Watertown, MA 02172-0001

Abstract: The U.S. Army Research Laboratory-Materials Directorate (USARL-MD) conducted a failure analysis of a main rotor pitch cone bolt located on an Army attack helicopter. During a routine flight after replacement of the main rotor head, a severe vibration was felt by the pilot who responded by immediately landing the aircraft. Post flight inspection revealed both the upper and lower lead lag link bearings displaced on one of the four main rotor strap packs. Further inspection and disassembly revealed a broken bolt on the outboard end of the strap pack assembly which anchors the blade to the aircraft. The incident occurred in November of 1993. The disassembled strap pack, along with several other pieces of hardware were returned to the prime contractor of the helicopter for analysis. A report was issued in January of 1994 which concluded that the failure appeared to be the result of pure fatigue with a sub-surface (0.024 in.) initiation from below the bolt inner diameter. The prime contractor further reported that cracking progressed by fatigue until such time at which it underwent a transition to an overload mechanism. USARL-MD received the failed part and conducted a complete metallurgical investigation in May 1994. It was concluded that the failure was attributable to galling. A single crack front emerged from cracks initiated by galling and progressed a total distance equal to approximately 20° of the entire circumference of the bolt. At this point, final fracture occurred by a pure overload mechanism resulting in the bolt breaking into two mating halves.

Key Words: Failure analysis, galling, fatigue, PH 13-8 Mo.

Background: During a test flight after replacement of the main rotor head, a noticeable vibration was felt by the pilot. Post flight inspection revealed both the upper and back lead lag link bearings displaced on one of the four main rotor strap packs. Further inspection and disassembly revealed a broken bolt which secured the outboard end of the strap pack. The disassembled strap pack, along with several other pieces of hardware were returned to the prime contractor for analysis. They issued a report which concluded that the failure appeared to be the result of pure fatigue with a sub-surface (0.024 in.) initiation from the bolt inner diameter. It stated that cracking progressed by fatigue until undergoing a transition to an overload mechanism.

The USARL-MD failure analysis of the Pitch Cone Bolt was hindered due to the fact pieces of the broken bolt and the pin which were an integral part of the investigation were lost by the primary contractor. In addition, the specimens that were sent to USARL-MD for examination could not be easily traced back to their original locations on the failed component after being sectioned by the prime contractor. Therefore, valuable time was wasted attempting to reconstruct the bolt.

Visual Inspection and Light Optical Microscopy: The lower fracture half of the bolt (see Figure 1) contained very pronounced radial lines and chevron marks converging to a point located approximately 0.024 inches beneath the galled region, on the inside diameter which appeared to be the crack initiation site. Only two pieces of the upper bolt fracture half were provided to USARL-MD for examination (see Figure 2). These pieces formed approximately a 45° arc which contained the crack origin (see Figure

3). The white arrows in Figures 1 and 2, and black arrow in 3 designate the crack initiation site. The lower fracture half pieces containing the crack initiation site are shown at higher magnification in Figure 3. The fracture origin had been previously sectioned through by the prime contractor.

The fracture surface contained no discoloration or corrosion products. However, two lower fracture half pieces had been previously annealed by the prime contractor as identified in Figure 1 and therefore, had to be discounted for this examination. No evidence of fatigue beach marks were observed and the entire fracture surface appeared to be quite rough, except for a small 45° region surrounding the crack origin. Directly opposite the initiation site, at the outer diameter of the bolt, a raised lip encompassed an area of approximately 90° (see Figure 4). Closer examination of the outside diameter of the bolt revealed deep impressions on the surface which were caused by contact with the strap pack laminates (see Figure 5). Examination of the strap pack revealed deformation around the inside diameter of the laminates (see Figure 7) which were in contact with the outside diameter of the bolt.

A severely galled area was located adjacent to the crack initiation site on the inner diameter bolt surface (see Figure 6). According to the prime contractor a mirror image of that galled region was contained on the outside diameter of the pin (see Figure 7). The prime contractor had sectioned the galled areas of the pin and the bolt thereby preventing verification of this. However, after matching up the various left over pieces of both parts, it was concluded that both galled regions were most likely adjacent to each other. Secondary galled regions were located on each side of the central wear mark on the pin (see Figure 8). The pieces of the bolt which would have come into contact with these two wear marks were examined and found to have matching wear sites.

The pin also contained two heavy lines of wear running parallel and circumferentially through the central galled wear region. Since the prime contractor had sectioned through this area and misplaced or lost much of the upper half of the pin, Figures 9 and 10 are reproductions of figures from the prime contractor report for purposes of illustration. Remnants of these two wear lines were observed on the pieces examined. This wear pattern was probably caused by the edges of the two bolt fracture halves (see Figure 10).

Longitudinal wear marks were also observed on the inside diameter of the bolt and the outside diameter of the pin at the galled region (see Figures 6 - 8). They appear to have been caused when these surfaces were forced across another hard material. This may have occurred during assembly if the pin had been forced into the bolt and either surface contained debris or it could have been the result of disassembly after the failure. Random surface scratches and nicks were also found on all surfaces.

An additional pair of wear lines were noted on the pin, as reported by the prime contractor, and corresponded to the bottom and top edges of the bolt. Figure 8 shows the bottom wear line on the pin, however, the upper piece of the pin had been sectioned and misplaced by the prime contractor. Typically, as stated by the prime contractor, these areas were less pronounced when examining other fielded assemblies. The more severe wear observed in this case could be attributed to more severe contact of the bolt edges with the pin upon final fracture.

No apparent burn marks, as might be anticipated from an arc strike, were noted. However, slight discoloration on the inside diameter of the bolt, that had been burnished as a result of contact with the pin was observed. Burnishing has been known to occur in the center of the bolt on fielded components and at the top and bottom bolt edges. A schematic of the lower fracture half of the failed bolt containing all important macro fractographic features, and locations where specimens were sectioned for chemical analysis and mechanical testing, is shown in Figure 11.

Metallography: A representative longitudinal and transverse section of the pitch cone bolt were mounted and metallographically prepared. The specimens were mounted in phenolic powder (Bakelite) with edge retention, and rough polished with (180-600 grit) silicon carbide papers. Fine polishing was

accomplished with hand wheels utilizing 9μm and 5μm diamond suspensions. Final polishing was performed on a vibratory unit with 0.05 μm colloidal silica. The as-polished specimens exhibited no significant inclusions per ASTM E45, as referenced within HMS 6-1105. The polished specimens were subsequently etched with Fry's reagent, to reveal the resultant microstructure. At low magnification, no evidence of austenite banding was noted (banding is prohibited per 3.7.2 of HMS 6-1105). The longitudinal and transverse sections of this precipitation hardenable stainless steel exhibited a tempered martensitic structure. This structure was consistent with the prior heat treatment of solution annealing and aging. The grain size was determined to be between 7 and 8 from a grain size reticle fitted onto a metallograph and conforming to ASTM E112. This conformed to section 3.7.1 of HMS 6-1105, which states "Unless otherwise specified, grain size in accordance with ASTM E112 shall be 5 or finer, with an occasional 3."

Figure 12 is a cross sectional view through the crack origin on the lower fracture half. The fracture surface at the outer diameter of the bolt was deformed by the force and motion of the strap pack laminates. Figure 13 shows the resulting grain flow. The inner diameter of the bolt at high magnification exhibited a lightly etched region. This lightly etched region is the galled bolt material adjacent to the origin (see Figure 14). The prime contractor had also observed a lightly etched layer within the galled region. This light layer reacted differently to the etchant as a result of cold working and should be harder than the base material.

Mechanical Properties:

Tensile Testing: A sub-sized threaded tensile rod was machined from a block of material sectioned from the failed bolt (see Figure 11). The specimen was subsequently subjected to tension testing according to ASTM E8, and section 3.6.2 of the governing specification HMS 6-1105. The tensile specimen was tested on a 20 Kip capacity Instron Universal electromechanical tensile test machine. Per 3.6.2 of HMS 6-1105, a crosshead speed of 0.005 inches per minute was utilized until the yield point was reached, whereupon the crosshead speed was increased to 0.05 inches per minute. The purpose of this test was to determine the representative mechanical properties of the pitch cone bolt, and to obtain a fracture surface which could subsequently be examined utilizing the scanning electron microscope (SEM) then compared to the fracture surface under investigation (see Scanning Electron Microscopy section). Table 1 lists the results of tensile testing. The acquired properties of the test specimen met the requirements specified within HMS 6-1105.

Table 1. Mechanical Properties

Property	Component (Actual)	HMS 6-1105 (requirement)
UTS, ksi	220.4	220.0
0.2% Y.S., ksi	208.3	205.0
% Elongation	13.1	10
% Red. of Area	57.4	45

Knoop Microhardness Testing: The representative longitudinal and transverse sections of the pitch cone bolt which were mounted and metallographically prepared were utilized for microhardness testing. The as-polished samples were subjected to Knoop microhardness tests utilizing a 500 gram load, and 20X objective. The required (converted from HRC) Knoop hardness of the pitch cone bolt is 480-576 HK. Table 2 shows the results of microhardness testing. The average of the ten readings for each section were converted to HRC values and met the requirements set forth in HMS 6-1105.

Table 2. Knoop Microhardness Testing - Pitch Cone Bolt

	Longitudinal (HK)	Transverse (HK)
	496 HK 507	499 HK 497
	507 496	497 503
	515 511	484 486
	499 512	504 491
	511 503	490 513
Average	506	496
Equiv. HRC	48	47
HMS 6-1105	46-52 HRC	46-52 HRC

Rockwell Macrohardness Testing: Rockwell "C" scale readings were measured directly upon the surfaces of the mounted cross sections of the pitch cone bolt. A total of five readings were taken on each section. Testing was performed on a Wilson hardness testing machine, incorporating a major load of 150 kg. The values of this test are listed in Table 3. In addition, macrohardness testing was conducted on a sectioned surface of the pin. The pin is required to exhibit a strength level of 260 ksi, which converts to approximately 50.5 HRC. The Rockwell macrohardness measurements of the pin are listed in Table 4. The results of this testing revealed that the hardness of the bolt compared favorably to the microhardness results, and that both components conformed to the required hardness range.

Table 3. Rockwell Macrohardness Testing - Pitch Cone Bolt (HRC)

	Longitudinal	Transverse
	47.4	47.0
	47.8	47.6
	48.1	47.9
	47.8	47.1
	47.7	47.7
Average	47.8	47.5
HMS 6-1105	46-52	46-52

Table 4. Rockwell Macrohardness Testing - Pin (HRC)

	53.4	52.8
	53.5	53.3
	53.1	53.2
	52.9	53.5
	53.2	53.8
Average	53.3	Required 50.5 min.

Chemical Analysis: The pitch cone bolt was specified to be fabricated from PH 13-8 Mo stainless steel bar stock according to the requirements contained in HMS 6-1105. The section of material utilized for chemical analysis is shown in Figure 11. The pin associated with the failed bolt was specified to be fabricated of MP35N (AMS 5844) according to the requirements contained in HMS 6-1105. Atomic Absorption (AA) and Inductively Coupled Plasma (ICP) were used to determine the chemical composition of the materials. The carbon and sulfur content of both samples were analyzed by the

LECO Combustion Method. The weight percent chemical composition of the bolt and pin materials are listed in Tables 5 and 6 respectively, and compare favorably to the requirements as shown.

Table 5. Chemical Composition of Bolt

Element	Failed Bolt	HMS 6-1105
Carbon	0.030	0.05 max.
Manganese	0.034	0.10 max.
Phosphorus	0.001	0.010 max.
Sulphur	0.003	0.008 max.
Silicon	0.054	0.10 max.
Chromium	12.9	12.25 - 13.25
Nickel	8.14	7.50 - 8.50
Aluminum	1.14	0.90 - 1.35
Molybdenum	2.04	2.00 - 2.50
Nitrogen	0.003	0.01 max.
Iron	Remainder	Remainder

Table 6. Chemical Composition of Pin

Element	Pin	AMS 5844
Carbon	0.006	0.025 max.
Manganese	0.001	0.15 max.
Phosphorus	0.010	0.015 max.
Sulphur	0.005	0.010 max.
Silicon	0.057	0.15 max.
Chromium	19.7	19.00 - 21.00
Nickel	35.3	33.00 - 37.00
Molybdenum	9.79	9.00 - 10.50
Iron	0.36	1.00 max.
Titanium	0.89	1.00 max.
Cobalt	Remainder	Remainder

Scanning Electron Microscopy: The four pieces of the bolt which contained the primary galled region were subsequently pieced together (see Figure 15). The top two pieces represent the upper fracture half while the lower pieces, the bottom fracture half. The upper right piece contains the undisturbed crack origin. The mating crack initiation site located directly beneath, had been sectioned through by the prime contractor. The damage which resulted from point contact between the inner diameter of the bolt and the outer diameter of the pin was severe. Figure 16 is an enlarged view of the wear adjacent to the crack origin. Note the tearing and smearing which occurred. Secondary cracks were observed within the galled region.

From what remained of the fracture surface, important fractographic features relating to the cause of failure were identified. Very distinct radial marks converged to the crack origin located 0.024 inches below or adjacent to the galled region on the inside diameter of the bolt (see Figure 17). The radial marks emanated outward across the thickness of the bolt and faded away. Also, classic chevron patterns appeared on both sides of the fracture origin and worked their way outward. However, as a result of the load distribution and the part geometry, crack propagation did not proceed equally on both sides

of the initiation site. One crack front progressed approximately three-fourths of the entire circumference of the bolt before meeting up with the second crack front (see Figure 11). Since the two crack fronts progressed on different planes there was a ridge of material where they finally met, signifying final fracture.

The crack morphology was primarily transgranular within the confines of the V-shaped origin area and then transitioned to a mixed-mode of fracture. Ductile dimples gradually emerged within the transgranular fracture surface as the crack progressed farther from the origin. The amount of dimples increased as the fracture extended around the circumference of the bolt. Figure 18 represents an area on the galled surface which extended approximately 0.0017 inches into the fracture. This region contained a dimpled morphology surrounded by smeared metal along one side and transgranular fracture on the remaining edge. The ductile fracture was most likely the result of adhesive wear between the pin and the bolt. Attractive forces at the interface caused material from one surface to be pulled away when the two parts were separated. Figure 19 shows the typical transgranular mode of fracture found at and near the origin. Figure 20 is representative of areas as close as 0.25 inches or 19° away from the crack origin. These regions displayed significant ductility as evidenced by the ductile dimples. The prime contractor had previously determined that this area had fractured as a result of fatigue. Figures 21 and 22 show the transition in fracture 90° from the origin, and 180° away from it, respectively. In addition, there existed a shear lip along the inside and outside diameter of the fracture surface. The morphology of the shear lip was dimpled, as shown in Figure 23.

X-Ray Mapping and Energy Dispersive Spectroscopy: Energy dispersive spectroscopy (EDS) and x-ray mapping were utilized to examine the fracture origin of the bolt and adjacent areas for machining debris or an inclusion which may be associated with the failure. X-ray maps were collected with a KEVEX EDS system. The x-ray maps were confined to the following eight elements; Fe, Cr, Mo, Co, Mn, Ni, Si, and S. Only elements associated with the bolt material were found.

The primary galled region on the bolt at the initiation site is shown in Figure 15. It should be noted that the fracture surfaces shown in this SEM photograph were sectioned by the prime contractor prior to this investigation. The corresponding x-ray map of the primary galled area on the bolt was acquired with the EDS system. A high concentration of cobalt was detected on the periphery of the wear site indicating that material transfer from the pin had taken place. No evidence of surface anomalies were noted in the region.

The corresponding galled region of the pin was examined (Figure 24) and found to contain high concentrations of bolt material. Iron was revealed at high concentrations while cobalt, molybdenum, nickel, and chromium were depleted in this area. It should be noted that the pin was also sectioned prior to this investigation by the prime contractor and only half could be analyzed. The two secondary galled areas on the pin were analyzed and found to be similar in composition to that of the primary galled region on the pin with high concentrations of elements associated with bolt material and depletions of pin constituents. All of these x-ray maps were reviewed for the presence of inclusions or debris that could have initiated the wear, i.e. oxides or machining burrs. Neither silicon, oxygen nor localized sulfur, which could be signs of these inclusions or other particulates, were found in the galled wear regions or on the fracture surfaces.

X-Ray Diffraction Residual Stress Analysis: The results of x-ray diffraction residual stress characterization on the I.D. surface of the failed strap pack bolt are listed in Table 7. Axial and hoop direction measurements were performed at four circumferential orientations at a location of 1/2" from the crack initiation site toward the flange end of the bolt.

Table 7. Residual Stress Measurements of Bolt

Orientation	Axial Direction, ksi	Hoop Direction, ksi
0°	-85.4 ±5.0	-77.3 ±1.9
90°	-113.8 ±3.0	-90.9 ±2.1
180°	-119.8 ±4.1	-87.6 ±1.9
270°	-107.5 ±4.9	-71.1 ±2.3

The axial direction values agree reasonably well with the Technology for Energy Corp. Report R-94-002 for the primary contractor as-measured surface residual stress of -120.6 ±4.5 ksi. TEC did not report hoop direction residual stress data.

Discussion:

Crack Origin (Wear vs. Sub-Surface Defect): The radial lines, observed on the fracture, which identified the crack origin, seemed to converge to a site located approximately 0.024 inches below the surface of the inside diameter of the pitch cone bolt. The morphology of the crack origin displayed no evidence of an internal defect or inclusion which may have caused a sub-surface crack to form. Whether or not the crack origin was located sub-surface had little bearing on the failure mechanism. The metallurgical evidence, including fractographic examination, light optical microscopy and energy dispersive spectroscopy indicated that surface galling caused premature crack initiation.

The wear pattern at the crack origin was very severe and a series of competing cracks can be observed extending from the surface into the material. Closer examination of this region revealed that several of the cracks seemed to converge beneath the surface before forming a single crack front. This made the crack initiation site appear to be sub-surface. The cracks initially propagated on different planes resulting in an irregular surface topography and an outcrop of fractured material. The prime contractor ruled out wear as a possible crack initiation mechanism. When a cross-section of the galled region was prepared metallographically and examined, by the prime contractor, the region below the surface which was affected by the wear (as evidenced by plastic deformation within the microstructure) measured only a few thousands of an inch, whereas the supposed crack origin was located much deeper, 0.024 inches. Therefore, it was concluded by the prime contractor that wear did not play a role in the failure. However, this logic ignored the fact that the depth of galling measured as much as 0.0026 inches and the compressive stress profile from the inner diameter surface of the bolt to 0.002 inches dropped significantly as reported by the prime contractor (approximately 120 ksi to 11 ksi compressive). Below 0.002 inches, the residual stress leveled out at approximately 8 ksi compressive. Low sub-surface compressive stresses could have easily been overcome by a high stress concentration caused by surface galling. In fact, numerous cracks did form as a result of galling. Therefore, the fact that plastic deformation stopped short of the supposed crack origin, 0.024 inches beneath the surface, did not eliminate the possibility that wear caused cracking which initiated failure.

In addition, if the wear observed on the inside diameter of the bolt lead to seizing or as electron microscopy revealed, adhesive wear (cold welding, see Figure 18) between the bolt and the pin, then the stress concentration region could have been shifted a distance below the surface causing a sub-surface crack initiation while showing no subsurface defects. This would occur because the contact area between the outside diameter of the pin and the inside diameter of the bolt would now act as a weld. Scanning Electron Microscopy revealed that several areas of the galled site adjacent to the crack origin displayed a dimpled topography. Beyond this region the crack origin fracture surface abruptly became transgranular and a distinct demarkation line was observed separating the two fracture zones. This information suggested that localized areas of the contact surfaces between the bolt and the pin

experienced adhesive wear[1] in service. Unfortunately, the specimen containing the mating area on the pin could not be examined for the same features because the prime contractor had lost or misplaced it. Even in the absence of seizing, the region identified as the sub-surface crack origin may have been formed when several surface cracks caused by galling converged at this location. The individual crack fronts could combine to form a single crack and proceed in one plane.

The bolt material (PH 13-8 Mo stainless steel) was reported as having no internal defects or anomalies by the prime contractor. This information was confirmed. If a sub-surface defect caused the fracture in this relatively "clean" material, then the critical flaw was located directly below an area which experienced the most severe wear observed on any pitch cone bolt examined to date. In addition, it was at the most outboard (12 o'clock) position and at the center of the laminate stack. Therefore a sub-surface defect, if it existed, was aligned perfectly at the center of the strap pack laminate stack-up, at the 12 o'clock position, and directly beneath a severe wear site. In summary, whether the crack origin was at the galled surface of the bolt inner diameter or 0.024 inches below this area is of no consequence to the cause of fracture which has been attributed to galling.

Fatigue: The primary issues that were addressed regarding fatigue involve determining the size of the fatigue crack region, if one existed, and the number of cycles before the onset of overload. It was difficult to examine the lower fracture half of the bolt since it had been sectioned previously by the prime contractor into seven separate pieces. In addition, most of the upper fracture half had been lost or misplaced by the prime contractor and could not be examined. However, with the use of oblique lighting and electron microscopy, USARL-MD was able to resolve enough important fractographic features to draw several conclusions relating to fatigue. Also, a general understanding of how the part was loaded in service and the circumstances leading to the failure aided in this analysis.

There were no classic beach marks or fatigue striations observed within the crack origin region. Most of the fracture surface contained very pronounced chevron marks and radial lines which were flanked on the inside and outside diameter of the bolt by shear lips. However, the shear lips located directly opposite the crack origin, on the outside diameter of the bolt had been almost completely obliterated by several laminates which had forced their way into the opened crack during the failure and smeared this edge. The shear lips displayed a purely dimpled topography. Shear lips were not observed within a 45° circumferential area on the inside diameter bolt edge where fracture initiated (see Figure 11). Scanning Electron Microscopy of the crack origin area revealed a transgranular fracture surface. Beyond this region a mixed mode of fracture was observed, and consisted of ductile dimples within transgranular fracture. The fracture gradually transitioned into a more dimpled morphology further away from the initiation site. No distinct demarkation line separated the transgranular morphology from that of the mixed-mode fracture.

The slight variation in morphology between the crack origin (transgranular) and the remaining fracture surface (transgranular-dimpled) may be attributed to overload or seizing between the pin and the bolt. This same phenomena may have also occurred if several cracks caused by galling converged to a point below the surface and formed a single crack front. A subsequent high load would result in a difference in crack propagation rate. Initially, the high load and stress intensity would result in a more brittle transgranular fracture but as the bolt material absorbed more of the force, the crack rate would decrease and transition into a more ductile fracture.

Fatigue striations in high strength steel alloys are sometimes difficult to resolve, but published research of the PH 13-8 Mo steel and conversations with producers of this material, ARMCO and Carpenter

[1] Rabinowicz, E. *Friction and Wear of Materials*. "Adhesive Wear." John Wiley and Sons, Inc. New York, 1965, p. 126.

Steel, indicate that fatigue striations should be discernable on the fracture surface.[2,3,4] Fatigue fractures of PH 13-8 Mo stainless steel are characterized by numerous narrow plateaus that align themselves parallel to the crack propagation. The H-1000 heat treat is depicted in both HCF and LCF in Figures 25 and 26, respectively. These figures were taken from the SEM/TEM Fractography Handbook.[5] Although fatigue striations are not easily seen at the crack origin area, representative of fractographs 25 and 26, they are readily discernable away from the crack origin area (see Figure 27). USARL-MD fatigued three samples of the PH 13-8 Mo stainless steel utilizing two different specimen geometries. Two beam specimens from another fielded pitch cone bolt were machined and then fatigued according to ASTM E 812 with R = 0.4. The two bend specimens failed at 25,800 and 26,700 cycles with a maximum stress of 16,000 and 17,300 psi, respectively. A sub-sized, 0.0845 inch diameter, threaded axial fatigue specimen was also machined from the failed part as seen in Figure 11 and fatigued according to ASTM E 606 with R = 0.4. The axial test specimen failed at 26,000 cycles with a maximum stress of 110,000 psi. The characteristic fatigue fracture zones of the USARL-MD specimens can be seen in Figure 28. The morphologies were all extremely similar. Fracture initiated from multiple origins and converged to a single crack front that had a relatively smooth featureless texture with no traditional fracture direction markings. The fatigue cracks then propagated until final fracture. As the bend beam fatigue specimens approached their final cycles, there was a noticeable change in morphology. The last few fatigue cycles before failure caused a transition to transgranular fracture with a small degree of ductility. A small portion of the axial fatigue specimen exhibited similar behavior. The published data for this material does not mention this morphology in its discussion of fatigue. The fracture surfaces of these specimens were then compared to the failed component fracture surface as well as the published data of PH 13-8 Mo. The failed bolt fracture surface did not contain the morphology characteristic of the published data on fatigue or the fatigue regime on the three USARL-MD specimens. However, it was noted that the transgranular morphology of the failed component was present in each of the three USARL-MD fatigue specimens but only in their respective overload regimes. More specifically, the regions in which the very last cycles before catastrophic failure occurred. This area is distinguished by general transgranular fracture with a small degree of ductility as can be seen in Figure 29, from a fatigue specimen, and compared with Figures 19, 20, 21, and 22 from the failed part. If an extensive fatigue regime were present on the failed part some fatigue striations would be present. Striations were found in the published data as seen in Figure 27 and can be found in the USARL-MD fatigued specimens as seen in Figure 30. Striations were even discernable within the regime characterized by numerous narrow plateaus, the traditional LCF and HCF morphology, as seen in Figure 31 which is from an USARL-MD fatigue specimen.

Based upon the macro- and microscopic features of the fracture surface there was no significant evidence of fatigue. However, one area which may have experienced fatigue lies beneath the galled surface of the bolt inner diameter. This area lies 0.024 inches below the galled surface and may represent the critical flaw size of the bolt. This region fractured primarily in a transgranular fashion. Figure 11 illustrates this small area of transgranular fracture. Shear lips were observed along most of the bolt inner diameter (except along the edge where the crack originated) and all of the outer diameter as Figure 11 depicts, thereby making fracture by traditional fatigue highly unlikely.

The two parallel wear lines observed on the pin which were caused by the fracture edges of the broken bolt may provide some insight to the issue of fatigue. The lines were very distinct and were separated by a small uniform gap. The surface of the pin between the wear lines seemed undisturbed and therefore, it was deduced that they were probably formed after the bolt had cracked. A slowly

---

[2] McIntyre, D. "Fractographic Analysis of Fatigue Failures." *Transactions of the ASME*. July 1975, pp. 197-198.

[3] Private conversation with ARMCO 5/19/94.

[4] Private conversation with Carpenter Steel 5/19/94.

[5] McDonnell Douglas Astronautics Company, SEM/TEM Fractography Handbook. December 1975, pp 358-373.

propagating fatigue crack would leave behind a pair of wear lines which would most likely be very close together initially and gradually grow further apart as the crack opened up, until final fracture occurred at which time the wear lines would be at their furthest distance from each other.

Point Contact: The galled region at the crack initiation site was the result of a point contact between the pin and the bolt. The circumstances which caused such a condition have been investigated. Three possibilities have been presented to explain how this may have taken place.

A) The bolt may not have been fully torqued to the specified value. The bolt break-away torque values were measured and recorded for a number of strap packs. The results indicated that not all conformed to the requirements. In fact, one bolt was only torqued to 30% of the specified value. Such a low torque would result in a non-uniform load distribution during service allowing point contact in the area where galling was observed. This area is the point of highest stress when inadequate torque is applied.[6] Under this condition, individual laminates would experience some degree of freedom because of the insufficient torque, and thus low clamp-up load, thereby acting independently instead of as a single unit.

The contact area between the bolt and the pin, defining the regions of high stress for low torque, correspond in location to the centerline of the laminate stack as well as the top and bottom edges of the bolt. These three zones are typically found to be burnished when examining fielded bolts and pins. This slight wear is largely the result of centrifugal forces acting on the bolt and the pin in service. However, in the case of the failed part, non-uniform loading caused by insufficient bolt torque could create a more significant stress concentration at these three locations, especially at the crack origin. Galling eventually took place here and numerous cracks were formed ultimately leading to failure. In addition, the bolt not being fully torqued and containing a crack, would have a slight degree of movement during service, causing the top and bottom edges of the bolt to come into contact with the pin. The two parts would rub against one another forming a more pronounced circumferential wear pattern.

Further non-uniform loading would likely result in an indistinct line of corrosion on the outermost laminates beneath the spaces of the strap pack assembly.[7] Crevice corrosion occurs between the 301 stainless steel spacers and the AM-355 laminates in service. The attack is much more pronounced between the laminates and the outside diameter of the spacers forming a sharp line of corrosion. However, in the case of the failed part, if the bolt had not been fully torqued the laminates and spacers would be allowed to shift slightly, forming a larger crevice which would extend more toward the inside diameter due to a decrease in clamp down force. This in turn, would result in a less distinct line of corrosion. This pattern of corrosion was observed on the outermost laminates of the failed bolt strap pack assembly. The corrosion did not form a distinct line beneath the outside diameter of the spacer, but was noticeably different. The corrosion in this area was broader, less distinct and extended more toward the inside diameter indicating that movement of the spacer and/or laminates may have taken place.

B) An obvious condition which would also cause point contact between the inside diameter of the bolt and outside diameter of the pin would be a lack of dimensional control of one or both of these parts. According to the engineering drawings of the bolt and the pin the gap between the parts when installed is 0.001 - 0.0015 inches for the Basic Strap Pack Assembly. The modified (-5) design allows for a much wider gap ranging from 0.006 to 0.00775 inches. Table 8 lists the minimum and maximum dimensions of the two parts as specified on their respective engineering drawings. The minimum and maximum difference between the dimensions of the inside diameter of the bolt and the outside diameter of the pin have also been calculated.

---

[6] Private communications, ATCOM Engineering (Carlton Smith)

[7] Private communications, ATCOM Engineering (Carlton Smith)

Table 8. Pin and Bolt Diameter Tolerances (inches)

	Basic			-5	
	Pin O.D.	Bolt I.D.		Pin O.D.	Bolt I.D.
max	1.4990	max 1.5015	max	1.4990	max 1.5140
min	1.4985	min 1.5010	min	1.4985	min 1.5110
	minimum difference = 0.0020			minimum difference = 0.0120	
	maximum difference = 0.0030			maximum difference = 0.0155	

The modified (-5) design allows a gap of only 0.006 - 0.00775 inches, therefore, any slight deviation from the specified dimensions along the longitudinal axis of either part could cause a point contact. In addition, any misalignment of the threaded bolt could contribute to this condition.

C) The remaining scenario which may have caused the pin and the bolt to come into contact at a single point would be trapped machining debris which could have become lodged between the pin and the bolt during installation. While in service, the centrifugal force acting on the assembly would pull the parts into one another. This would cause a point contact at the site of the debris leading to wear and eventually galling. Lines of wear were observed extending longitudinally on the surface of the pin as well as across the galled region on the bolt and the pin. These scratches were aligned parallel to one another suggesting the parts had been forced into or across a hard surface. This could have occurred if burrs were not removed on the inside diameter of the bolt or the outside diameter of the pin after machining, and the pin was later forced into the bolt during assembly. However, these marks could also simply be due to a post-fracture incident which occurred during disassembly or handling.

X-ray mapping and energy dispersive spectroscopy (EDS) were performed to determine if foreign debris had become lodged between the pin and the bolt at the galled region. Both galled areas on the pin and the bolt made available to USARL-MD were examined (refer to EDS analysis). The results showed no evidence of any elements foreign to the two materials being analyzed, MP35N and PH 13-8 Mo stainless steel. In addition, high concentrations of silicon and oxygen, as would be anticipated from a sand particle, were not detected. The spectra obtained did show a transfer of material between the pin and the bolt indicating that both materials experienced wear. It is important to note, however, that a machining burr would not be able to be detected by EDS analysis because of the extensive galling that took place and the fact that there would be no extraneous elements to detect if that burr came from the bolt or the pin.

Conclusion: The failure of the Main Rotor Strap Pack Outboard Pitch Cone Bolt was attributable to galling. Galling occurred as a result of high contact stress between the outer diameter of the pin and the inside diameter of the bolt initiating cracks. Since the source of the high contact stress could not be precisely determined three hypotheses are forwarded: a) lack of required torque on the bolt during installation resulting in low clamp-up load which allowed some relative motion between the pin and the bolt, b) the dimensional tolerance of the pin and/or bolt were not within specification, c) the presence of wedged in debris (possibly machining burrs) during the insertion of the close tolerance pin into the bolt. Regardless, a single crack front emerged approximately 0.024 inches below the bolt inner diameter, from the cracks initiated as a result of galling, and then progressed through the cross-sectional thickness of the part and on each side of the origin area a total distance equal to approximately 20° of the entire circumference of the bolt. Beyond this point, final fracture occurred by a pure overload mechanism until the bolt broke in two pieces. There was no evidence of fatigue observed on the fracture surface. The material was relatively "clean" and a specific inclusion or inherent material defect suggesting sub-surface crack initiation was not detected.

Failure Scenario: Point contact occurred between the inner diameter of the bolt and the outer diameter of the pin, as a result of machining debris, failure to meet the dimensional tolerances, or because the bolt had not been torqued to the required level. Wear took place at this contact point which eventually lead to galling. The galling became very severe, initiating cracks and forming cold welds, as evidenced by secondary cracks and ductile fracture within the galled region of the bolt.

The crack initiation site most likely formed when several cracks caused by galling converged approximately 0.024 inches below the inner diameter surface of the bolt. However, the crack origin could have also formed 0.024 inches below this surface as a result of highly localized stresses from point contact and the adhesive wear between the bolt and the pin. Regardless of the exact location of the crack origin, the important issue is that fracture was initiated by galling. The depth of galling coincided with a point within the bolt at which a significant decrease in residual compressive stress was reported by the prime contractor.[8] Therefore, the high stresses, which would accompany instances of severe galling, initiated cracks as observed on the inner diameter of the failed bolt. A single crack front formed beyond this point. The crack front progressed by overload either all at once or in a few steps. The absence of any macroscopic or microscopic evidence which could be attributed to high cycle or low-cycle fatigue affirm this conclusion. This is further substantiated by the similarity of the fractographic features observed between the fast fracture region of the fatigue specimens and the fracture surfaces of the failed bolt.

[8] Prime Contractor Report.

Annealed
Pieces

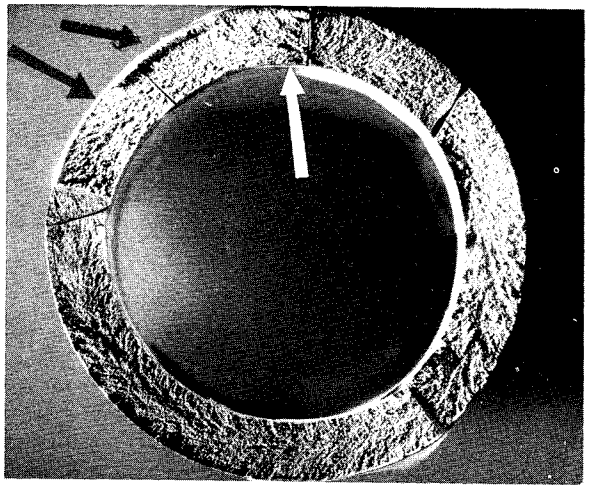


FIGURE 1 Macrograph of the lower fracture half of the pitch cone bolt as reconstructed by USARL-MD. Mag 3.5x.

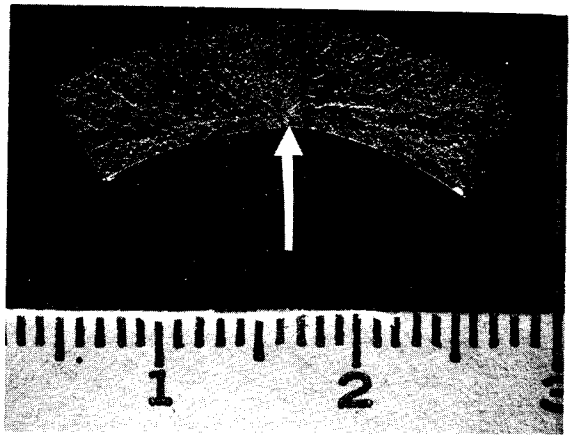


FIGURE 2 Macrograph of the two upper fracture half of the pitch cone bolt Mag. 4x.

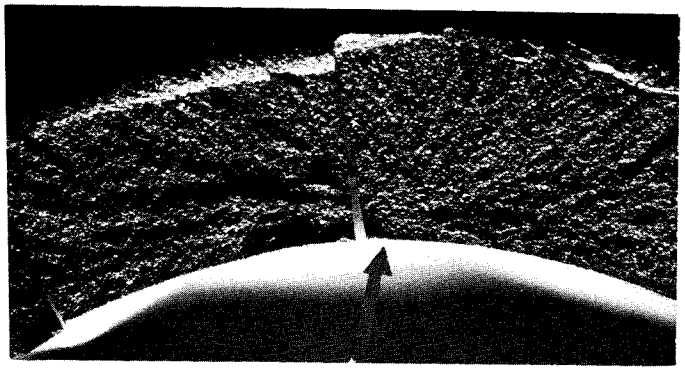


FIGURE 3 Macrograph of the lower fracture half region containing the crack origin. Mag 7.5x.

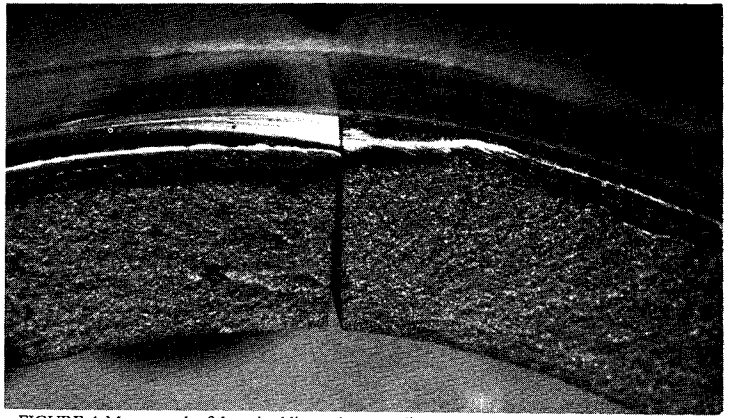


FIGURE 4 Macrograph of the raised lip on the outer diameter on the lower fracture half. The initiation site is on the inner diameter directly across from the raised lip. Mag. 7.5x.

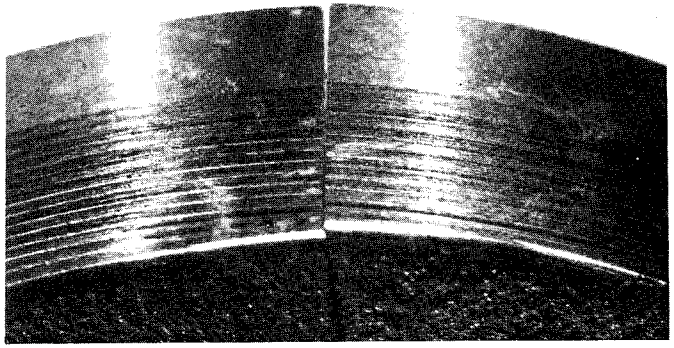


FIGURE 5 Macrograph of the deep impressions on the outer diameter bolt surface Mag 7.5x.

FIGURE 6 Macrograph of the severly galled region and vertical wear marks adjacent to the crack initiation site on the inner diameter bolt surface. Mag. 15x.

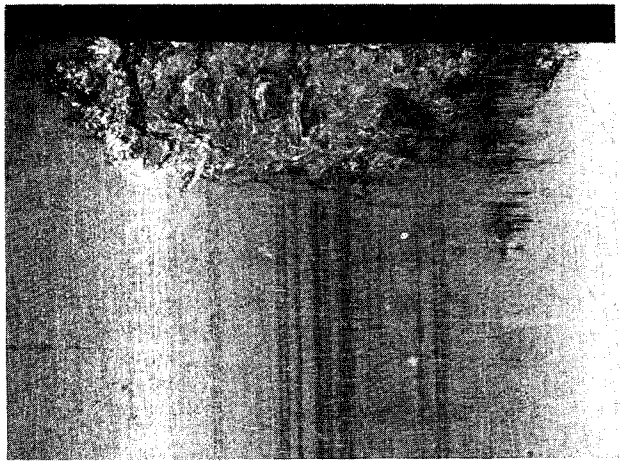


FIGURE 7 Macrograph of the galled region and vertical wear marks on the outer diameter of the pin. Mag. 12.5x.

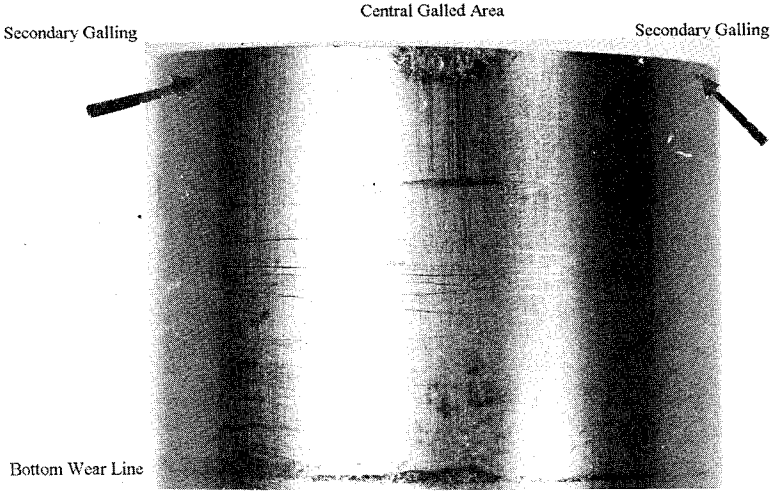


FIGURE 8 Macrograph of the galled regions on either side of the central wear mark on the pin. Mag. 3.5x.

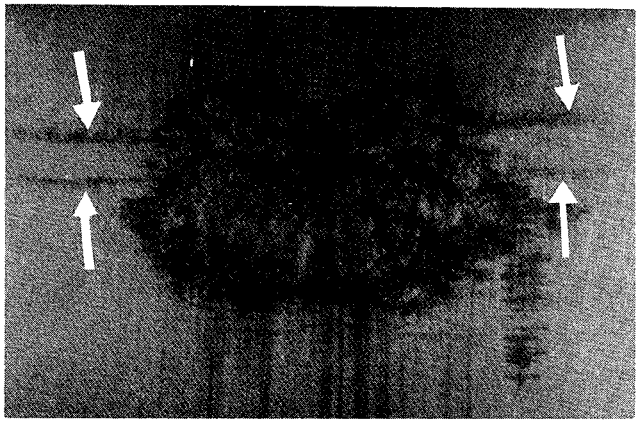


FIGURE 9 Reproduction from the prime contractor report. Arrows show two heavy lines of wear running circumferentially through the central galled wear region of the pin. Mag. 6x.

FIGURE 10 Reproduction from the prime contractor report. Arrows show knife edge of bolt fracture that caused circumferential wear lines on the pin. Mag. 6x.

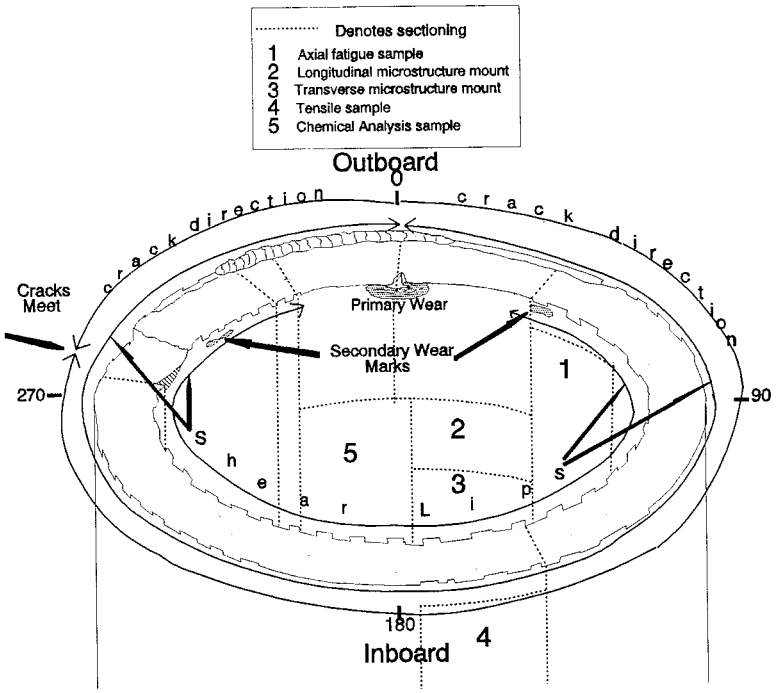


FIGURE 11Schematic of the lower fracture half containing all important fractographic features of the failed bolt.

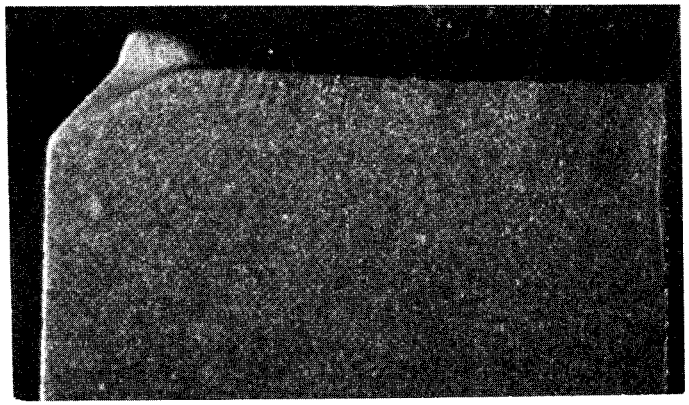


FIGURE 12 Cross-sectional micrograph through the origin of the lower fracture half. Mag. 1.5x.

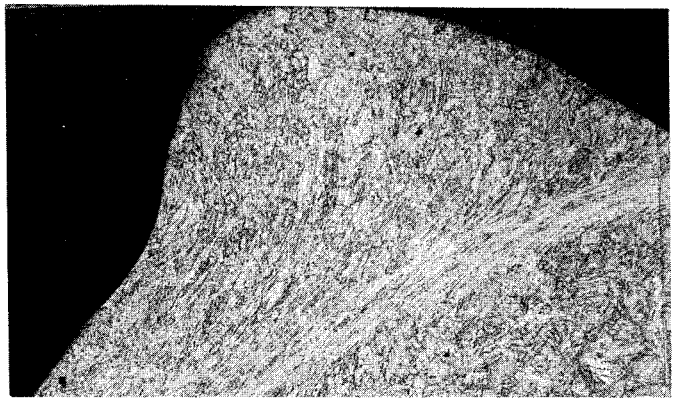


FIGURE 13 Micrograph of the deformed structure caused by the strap pack laminates. Mag. 100x.

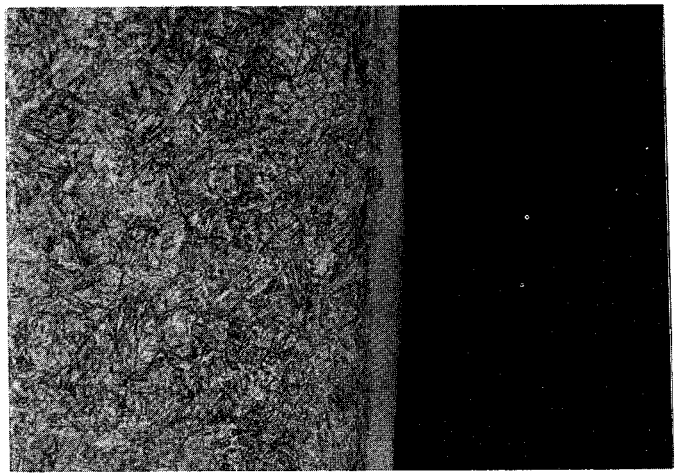


FIGURE 14 Micrograph of the lightly ecthed layer caused by galling. Mag. 150x.

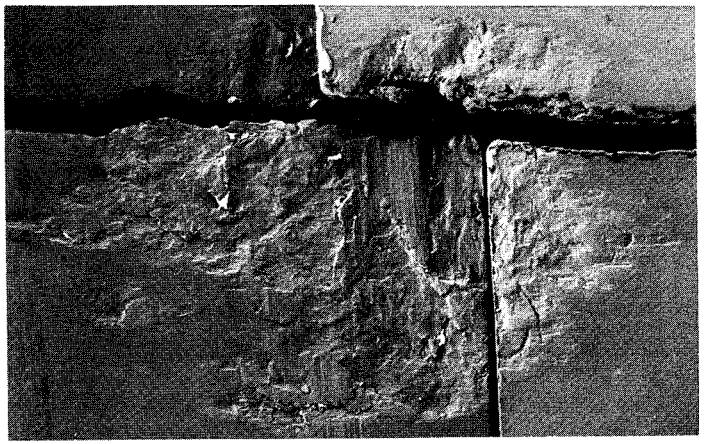


FIGURE 15 SEM macrograph of the bolt pieces containing the primary galled region. Mag. 20x.

FIGURE 16 SEM macrograph of the wear adjacent to the crack origin. Mag.75x.

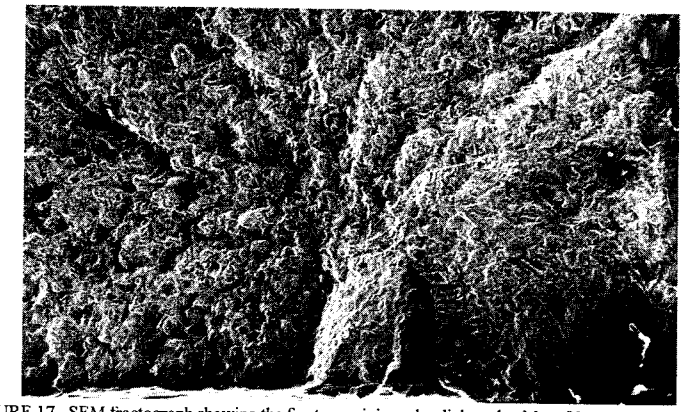


FIGURE 17 SEM fractograph showing the fracture origin and radial marks. Mag. 50x.

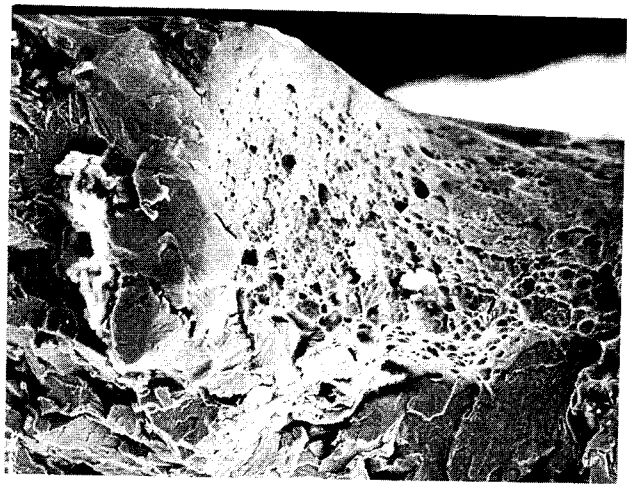


FIGURE 18SEM macrograph of the dimpled morphology on the galled surface. Mag. 500x.

FIGURE 19 SEM fractograph of the trangranular fracture morphology near the origin. Mag. 1,500x.

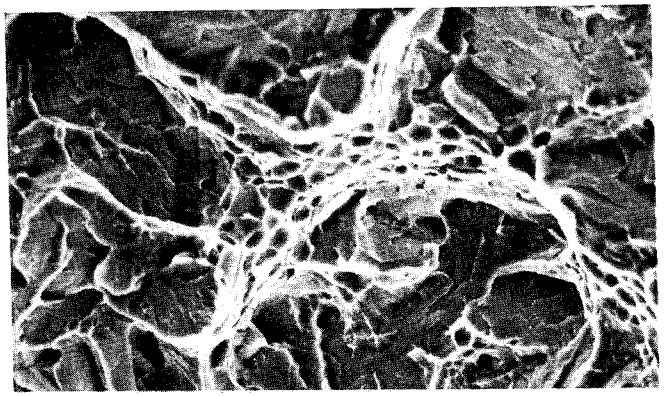


FIGURE 20 SEM fractograph showing mixed mode morphology0.25 inches or 19 degrees from the origin. Mag. 1.500x.

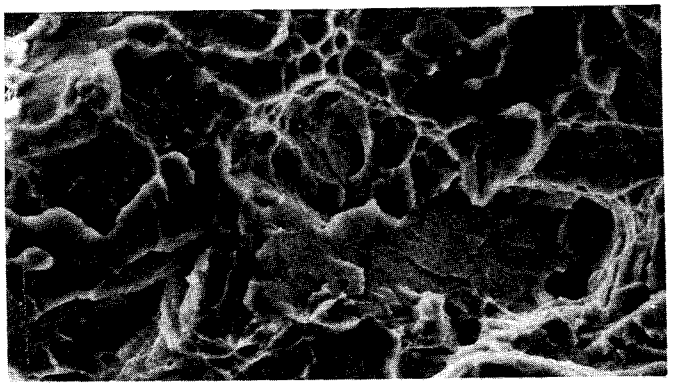


FIGURE 21 SEM fractograph showing the mixed mode morphology 90 degrees from the origin. Mag. 1,500x.

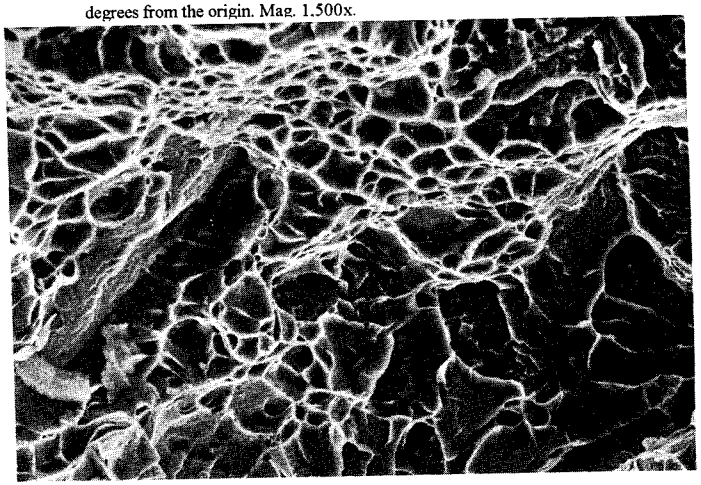


FIGURE 22 SEM fractograph showing the mixed mode morphology 180 degrees from the origin. Mag. 1,500x.

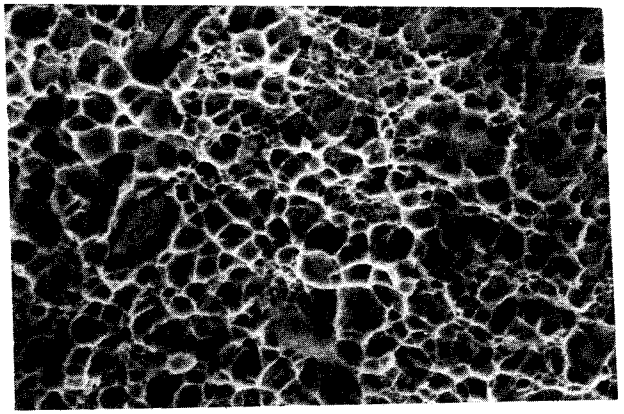


FIGURE 23 SEM fractograph of the shear lip showing the ductile dimpled morphology. Mag. 1,500x.

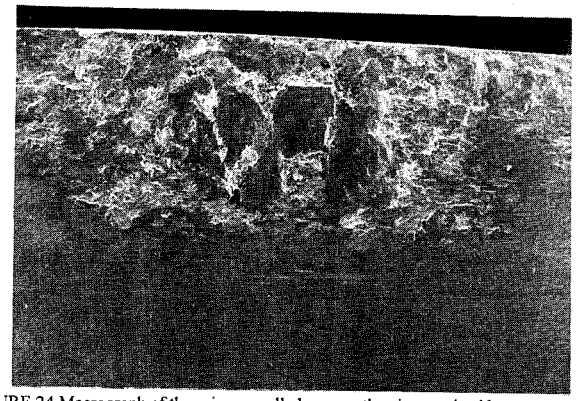


FIGURE 24 Macrograph of the primary galled area on the pin examined by x-ray mapping.Mag .15x.

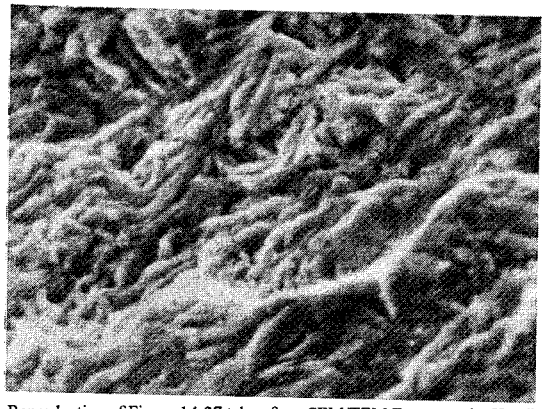


FIGURE 25 Reproduction of Figure 14-27 taken from SEM/TEM Fractography Handbook showing a representative high cycle fatigue region in PH 13-8 Mo stainless steel. Edge of specimen is at right. Mag. 2000x.

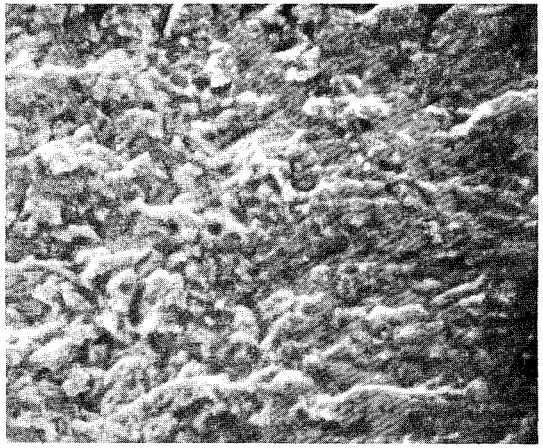


FIGURE 26 Reproduction of Figure 14-51 taken from SEM/TEM Fractography Handbook showing a characteristic low cycle fatigue region in PH 13-8 Mo stainless steel. Edge of specimen is at right. Mag. 2000x.

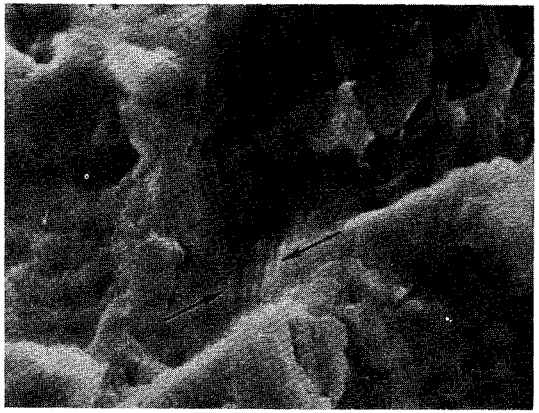


FIGURE 27 Reproduction of Figure 14-42 taken from SEM/TEM Fractography Handbook showing a patch of striation from a HCF specimen. The arrows bracket 5 striations. Mag. 5000x.

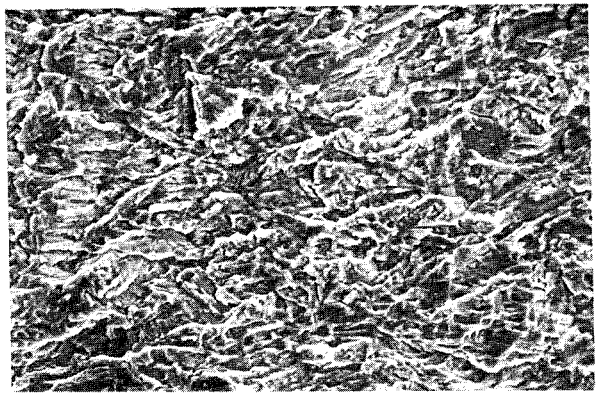


FIGURE 28Fractograph depicting the characteristic fatigue fracture surface observed on the ARL-MD bend beam specimens. Mag. 1,000x.

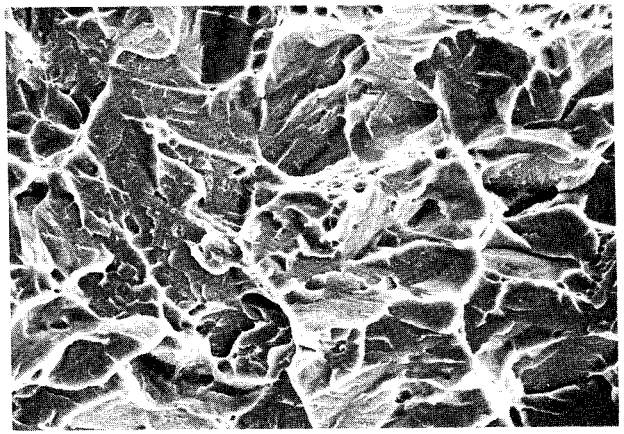


FIGURE 29 SEM fractograph of transgranular / ductile region on USARL-MD fatigue specimens. Mag. 1,500x.

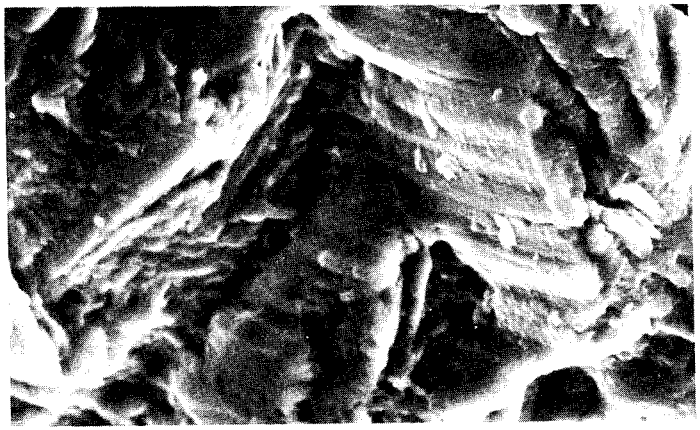


FIGURE 30 SEM fractograph showing striations found on the axial fatigue specimen. Mag. 5,000x.

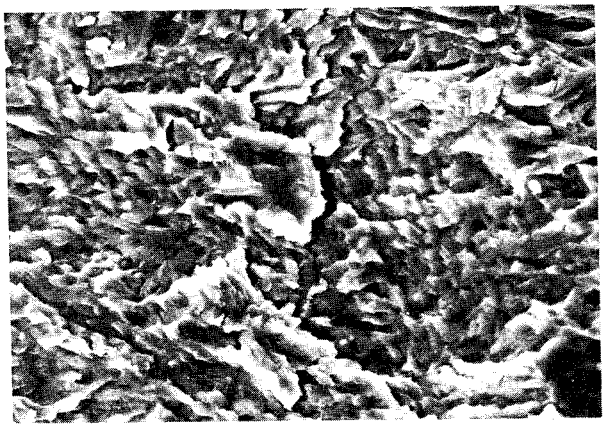


FIGURE 31 SEM fractograph showing striations near the crack origin (at left) on USARL-MD fatigue specimens. Mag. 1,000x.

DIAGNOSTICS I

Cochairmen: Mark L. Hollins
Naval Air Weapons Center

David Markley
Radian Corporation

DETECTING GEAR TOOTH FRACTURE IN A HIGH CONTACT RATIO FACE GEAR MESH

James J. Zakrajsek
National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135

and

Robert F. Handschuh, David G. Lewicki, and Harry J. Decker
U.S. Army Research Laboratory
Lewis Reseach Center
Cleveland, Ohio 44135

ABSTRACT

This paper summarizes the results of a study in which three different vibration diagnostic methods were used to detect gear tooth fracture in a high contact ratio face gear mesh. The NASA spiral bevel gear fatigue test rig was used to produce unseeded fault, natural failures of four face gear specimens. During the fatigue tests, which were run to determine load capacity and primary failure mechanisms for face gears, vibration signals were monitored and recorded for gear diagnostic purposes. Gear tooth bending fatigue and surface pitting were the primary failure modes found in the tests. The damage ranged from partial tooth fracture on a single tooth in one test to heavy wear, severe pitting, and complete tooth fracture of several teeth on another test. Three gear fault detection techniques, FM4, NA4*, and NB4, were applied to the experimental data. These methods use the signal average in both the time and frequency domain. Method NA4* was able to conclusively detect the gear tooth fractures in three out of the four fatigue tests, along with gear tooth surface pitting and heavy wear. For multiple tooth fractures, all of the methods gave a clear indication of the damage. It was also found that due to the high contact ratio of the face gear mesh, single tooth fractures did not significantly affect the vibration signal, making this type of failure difficult to detect.

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 number of fault detection methods have been applied to spur gear fatigue data [2] and spiral bevel gear fatigue data [3], with gear tooth surface pitting as the primary failure mode.

This paper extends the research by applying gear fault detection methods to fatigue data from high contact ratio face gears in a test rig. The methods applied to the face gear experimental data include method FM4, developed by Stewart [4] to detect isolated damage on gear teeth, and methods NA4* and NB4, both recently developed at NASA Lewis [2,3,5] to detect general damage on gear teeth. Verification of these detection methods with experimental face gear fatigue 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 view of the above, it becomes the object of the research reported herein to determine the relative performance of the detection methods as they are applied to experimental data from a face gear fatigue rig at NASA Lewis. The vibration signal from four face gear fatigue tests were monitored and recorded for gear diagnostics research. Gear tooth bending fatigue and surface pitting were the primary failure modes found in the tests. The damage ranged from partial tooth fracture on a single tooth in one test to heavy wear, severe pitting, and complete tooth fracture of several teeth in another test. Results of each method are compared, and overall conclusions are made regarding the performance of the methods.

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 face gear. The averaged data was then resampled by linear interpolation to provide exactly 1024 samples over two complete revolutions of the face 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 three diagnostic methods discussed below.

FM4 was developed to detect changes in the vibration pattern resulting from damage on a limited number of teeth [4]. 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 (non-dimensional). 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 [2,3]. Similar to FM4, a residual signal is constructed by removing regular meshing components from the original signal, however, for NA4, the first order sidebands stay in the residual 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*, a modified version of NA4, allows the parameter to continue to grow further by "locking" the value of the averaged variance when the instantaneous variance exceeds predetermined statistical limits [5]. As with FM4, NA4 and NA4* are dimensionless, with a value of 3, under nominal conditions.

NB4 is another parameter recently developed at NASA Lewis. 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 (primary) meshing frequency. For the face gear tests, a bandwidth of +/- 50 sidebands was used. 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 unlike the normal tooth load fluctuations, and thus be observed in the envelope of the signal. NB4 is also dimensionless, with a value of 3 under nominal conditions.

APPARATUS AND GEAR DAMAGE REVIEW:
The damage on the face gears shown in figures 1 through 4 were a result of a series of face gear fatigue tests conducted on a gear test rig at NASA Lewis Research Center. The face gear fatigue tests were part of an Advanced Rotorcraft Transmission (ART) program that was initiated to develop advanced transmission technologies for future military and civil rotorcraft [6]. The overall objective of the face gear tests was to determine the feasibility of using face gears in aerospace applications. Each test was allowed to progress beyond the pitting and heavy wear stages until a tooth fracture occurred. The load in some tests, after running at 100 % load for a period of time, was gradually increased to a maximum of 200% load, in order to precipitate failure. During the tests, vibration data from an accelerometer mounted on the pinion 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 107 tooth face gear meshes with a standard 28 tooth spur pinion rotating at a nominal speed of 19,107 rpm. This transmits 136 kW (182 Hp) at 100 % design load to the face gear, rotating at 5,000 rpm. The face gear/ pinion mesh has an effective contact ratio of 2.1, meaning that at least two gear teeth are in contact at all times.

Figure 1 illustrates the tooth damage resulting from face gear run #1. During the last 7 hours of the test all of the teeth on the face gear experienced heavy wear damage, with some developing severe pitting and surface fatigue cracks as well. One tooth broke off at approximately 18 minutes before the end of the test, and the test was terminated when the second tooth broke off. Both were complete tooth fractures, as shown in the figure.

Figure 2 illustrates the damage resulting from face gear run #2. After about 25 hours into the run, the load was incrementally increased over a period of 5 hours to 200%. The test continued at 200% load until 3 teeth broke off. Almost one complete tooth broke off along with a majority of an adjacent tooth, as seen in this figure. The third tooth, positioned approximately 120 degrees from the two shown in figure 2, also experienced a partial tooth fracture of over 50% of the tooth.

Figure 3 illustrates the tooth damage resulting from face gear run #3. During the last 3 hours of the run several teeth on the face gear experienced gradual pitting damage on the tooth surface. This resulted in the pitting-induced single tooth fracture at the end of the test. As seen in figure 3, nearly 75% of the tooth broke off. In addition, the tooth adjacent to the fractured tooth has extensive pitting damage, and a fatigue crack across most of the tooth width.

Figure 4 illustrates the damage resulting from face gear run #4. As seen in this figure, the damage was limited to a single tooth fracture. Approximately 75% of the tooth broke off during the test.

DISCUSSION OF RESULTS:
Results of applying the various diagnostic methods to face gear run #1, 2, 3, and 4 are illustrated in figures 5, 6, 7, and 8, respectively. In addition, figures 9, 10, 11, and 12 plot the time synchronous averaged vibration signal for the test face gear at the start (a) and end (b) of face gear run #1, 2, 3, and

4, respectively. The time averaged vibration signals shown in these figures are for two complete revolutions of the face gear.

Based on the results, all of the gear diagnostic methods investigated reacted significantly to the multiple tooth fracture damage experienced in face gear runs #1 and 2. As seen in figure 5, all of the parameters clearly indicated the multiple tooth fracture damage at the end of run #1. One tooth, or a portion of a tooth, broke off at approximately 18 minutes before the end of the test, which may account for the "knee" seen in the resulting plots at approximately the same point in time. Table 1 lists the results of the methods for the last 30 minutes of run #1. From the graphs in figure 5, it appears that the indications from the methods are nearly instantaneous. However, as seen in table 1 there is a gradual increase in FM4, NA4*, and NB4 over the last 24 minutes. At the end of run #1, FM4 reaches a value of 7.0, NA4* reaches 549, and NB4 reaches 1213, all relative to the nominal value of 3.0. Similarly, as seen in figure 6, all of the parameters clearly reflect the multiple tooth fracture damage at the end of run #2. Table 2 lists the results of the methods for the last 24 minutes of run #2. Again, the damage indications from the methods are not instantaneous, but gradually increases over the last 21 minutes of the test. As seen in this table, at the end of run #2, FM4 reaches a value of 6.7, NA4* reaches a value of 142, and NB4 reaches a value of 369.

Single tooth damage, as experienced in face gear runs #3 and 4, was more difficult to detect using the methods investigated. As seen in figure 7, only FM4 and NA4* reacted to the gradual pitting damage and single tooth fracture experienced in run #3. The final values of these two parameters are much lower than that experienced in the runs with multiple tooth fracture damage. NB4 did not show any indications of the damage in this run. As seen in figure 8, none of the parameters give any indications of the single tooth fracture damage experienced in run #4. For both run #3 and run #4, the tooth fracture damage was similar, i.e. approximately 3/4 of one tooth broke off, however only run #3 had some damage detection success. This may be due to the fact that the tooth adjacent to the fractured tooth in face gear run #3 had extensive pitting damage and a large fatigue crack. Even though this adjacent tooth did not break off in run #3, it was damaged enough to act similarly as if it were partially fractured.

The difficulty in detecting single tooth fracture in a high contact ratio gear mesh is a direct result of the nature of high contact ratio gearing. The contact ratio of the test face gear and pinion mesh in the tests was 2.1. Because the contact ratio is greater than 2, at least two pairs of teeth will be in contact at all times. Thus if one tooth becomes damaged, another tooth is available to carry the total mesh force. Theoretically this should change the vibration pattern, however, it will not change it as much as if the same damage was on a tooth in a low contact ratio gear mesh. For a contact ratio greater than 1 but less than 2, each tooth carries the total mesh load for a portion of the mesh cycle, and thus no other tooth is available to take up the load if that tooth is damaged. As seen in figures 9 and 10, the time averaged vibration signal at the end of runs #1 and 2 clearly indicate an impulsive event at the point of multiple tooth fracture. This impulsive signal is easy to detect with the methods used. As seen in figures 11 and 12, the time averaged vibration signal indicates very little impulsive event for the single tooth damage in run #3 (figure 11), and no real change in the signal for the single tooth damage in run #4 (figure 12). It is apparent from these results that single tooth damage is difficult to detect in a high contact ratio gear mesh.

Method NA4* was able to detect more than just pitting and tooth fracture damage. As seen in figure 5b, only NA4* reacted significantly to the heavy wear damage experienced by the face gear during the last 7 hours of face gear run #1. As in previous tests [2,3], NA4* was shown to be capable of reacting to a number of different failure modes. In these series of face gear tests, NA4* was able to detect gear tooth pitting, heavy wear, and fracture damage. NA4* is however, sensitive to load and

speed changes. For example, as the load was gradually changed from 100% to 200% of design load between the run time of 26 and 31 hours during run #2, NA4* gave a false indication of damage, as seen in figure 6b. These false reactions by NA4*, in this case, are much lower in magnitude than its reaction to the actual tooth fracture damage at the end of run #2.

Although method NB4 gave no reaction to the runs with single tooth damage, it did give the most robust reaction to the runs with multiple tooth fracture damage (runs #1 and 2). As seen in figure 5c and table 1, NB4 stays near the nominal value of 3.0 until near the end of run #1, where NB4 increases to over 400 times the nominal value. Similarly, as seen in figure 6 and table 2, NB4 stays relatively close to the nominal value until near the end of run #2, where NB4 increases to over 120 times the nominal value. In addition, NB4 does not appear to be as sensitive to load and speed changes as method NA4*.

SUMMARY AND CONCLUSIONS:

A study was conducted in which three different vibration diagnostic methods were used to detect gear tooth fracture in a high contact ratio face gear mesh. The NASA spiral bevel gear fatigue test rig was used to produce unseeded faults, natural failures of four face gear specimens. Based on the results of applying the diagnostic methods FM4, NA4*, and NB4 to the vibration signals from the fatigue tests, the following conclusions can be made.

1) All of the methods investigated gave strong reactions to multiple tooth fractures. The impulsive behavior of multiple tooth fractures dominate the time synchronous vibration signal of the damaged gear.

2) Single tooth fractures are difficult to detect in a high contact ratio gear mesh. Due to the nature of high contact ratio gearing, single tooth fractures may not significantly affect the vibration signal.

3) Of all the methods, NA4* is the only one capable of detecting gear tooth fractures, surface pitting, and heavy wear. Method NA4*, however, is affected by load and speed changes to a higher degree than methods FM4 and NB4.

4) Method NB4, although not able to detect single tooth fracture, does give the most robust reaction to multiple tooth fracture damage.

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) Zakrajsek, J.J.;Townsend, D.P.; and Decker, H.D.: 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, (NASA Technical Memorandum 105950).

3) Zakrajsek, J.J.;Handschuh, R.F.; and Decker, H.D.: 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, (NASA Technical Memorandum 106467).

4) 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.

5) Decker, H.J.; Handschuh, R.F.; Zakrajsek, J.J.: An Enhancement to the NA4 Gear Vibration Diagnostic Parameter. Proceedings of the 18th Annual Meeting of the Vibration Institute, Hershey, PA, 1994, pp. 259 - 268, (NASA Technical Memorandum 106553).

6) Heath, G.; Bossler, R.: Advanced Rotorcraft Transmission (ART) Program - Final Report. NASA CR 191057, ARL-CR-14, January, 1993.

TABLE 1: Last 30 minutes of face gear run #1.

Run Time (Hour)	FM4	NA4*	NB4
17.30	3.6	64.	3.1
17.35	3.7	61.	2.5
17.40	3.8	66.	4.0
17.45	3.9	115.	18.
17.50	4.2	130.	58.
17.55	4.6	153.	78.
17.60	4.7	162.	78.
17.65	4.7	147.	42.
17.70	4.8	133.	37.
17.75	5.1	223.	81.
17.80	7.0	549.	1213.

TABLE 2: Last 24 minutes of face gear run #2.

Run Time (Hour)	FM4	NA4*	NB4
46.35	2.6	15.	4.4
46.40	2.7	20.	7.0
46.45	3.6	36.	24.
46.50	4.0	48.	44.
46.55	4.7	67.	79.
46.60	5.5	98.	222.
46.65	5.8	113.	277.
46.70	6.7	142.	369.

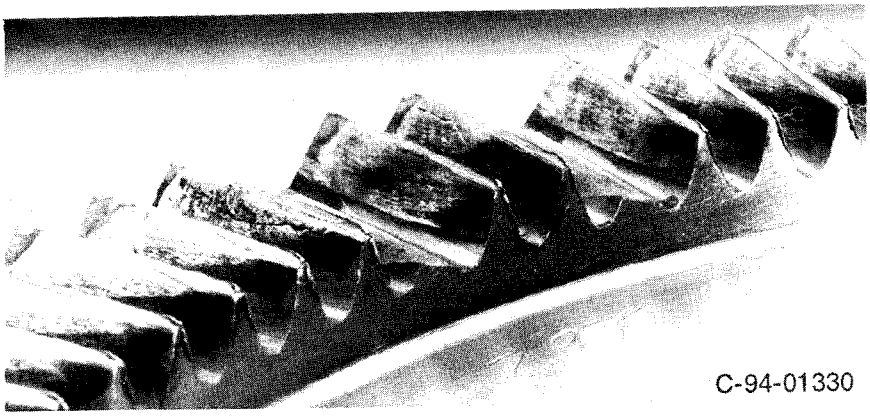


Figure 1.—Gear tooth damage at end of face gear run #1.

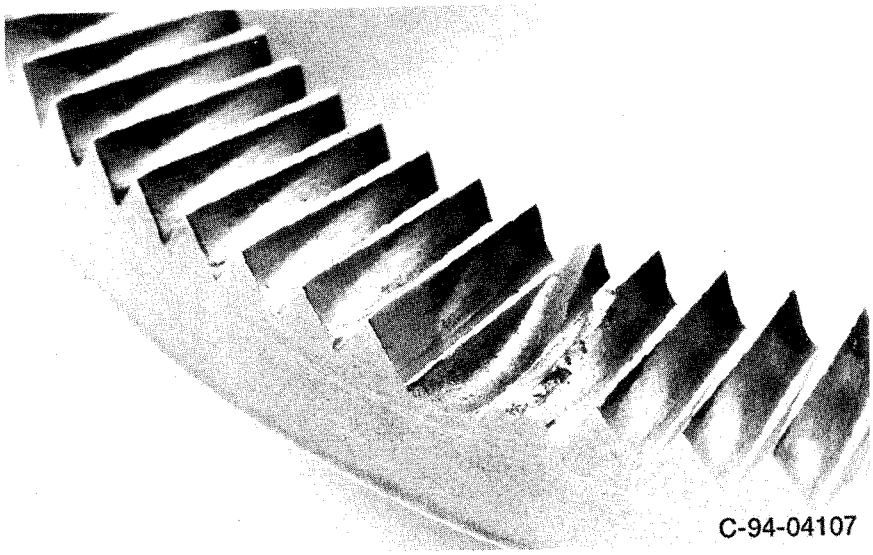


Figure 2.—Gear tooth damage at end of face gear run #2.

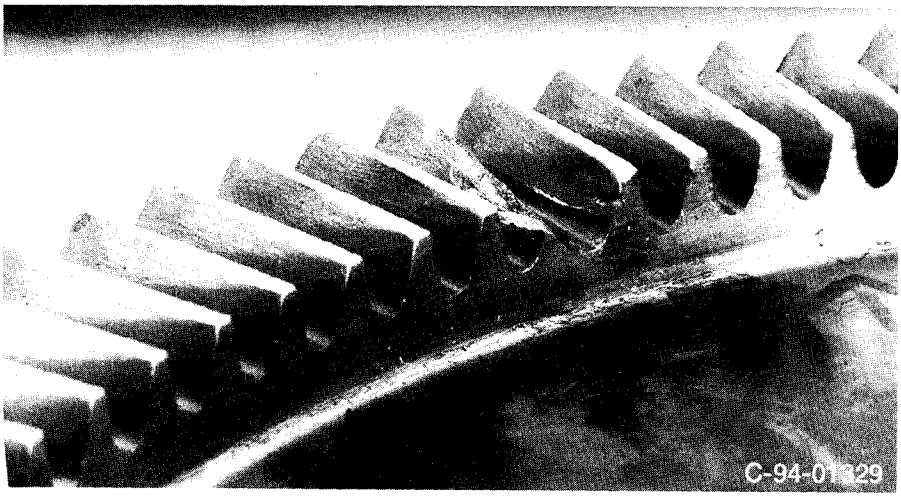


Figure 3.—Gear tooth damage at end of face gear run #3.

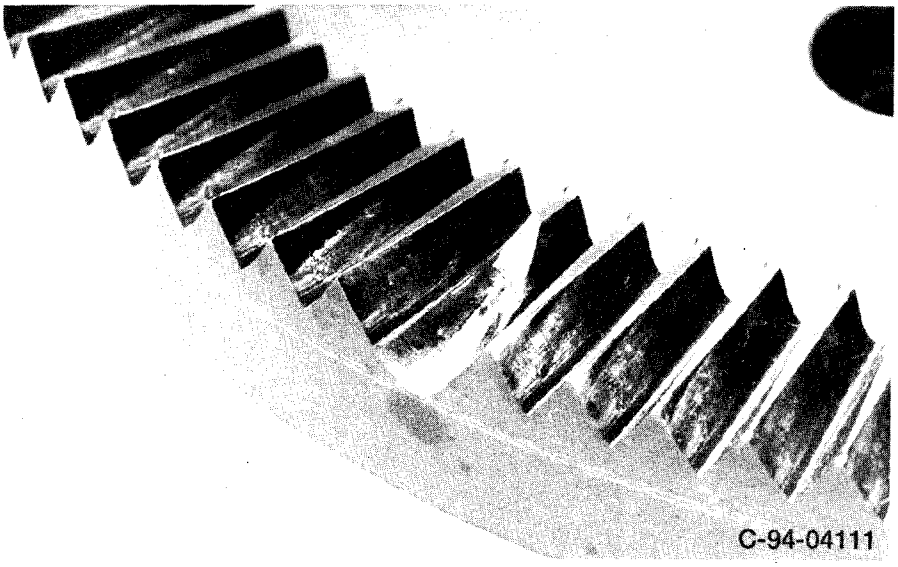


Figure 4.—Gear tooth damage at end of face gear run #4.

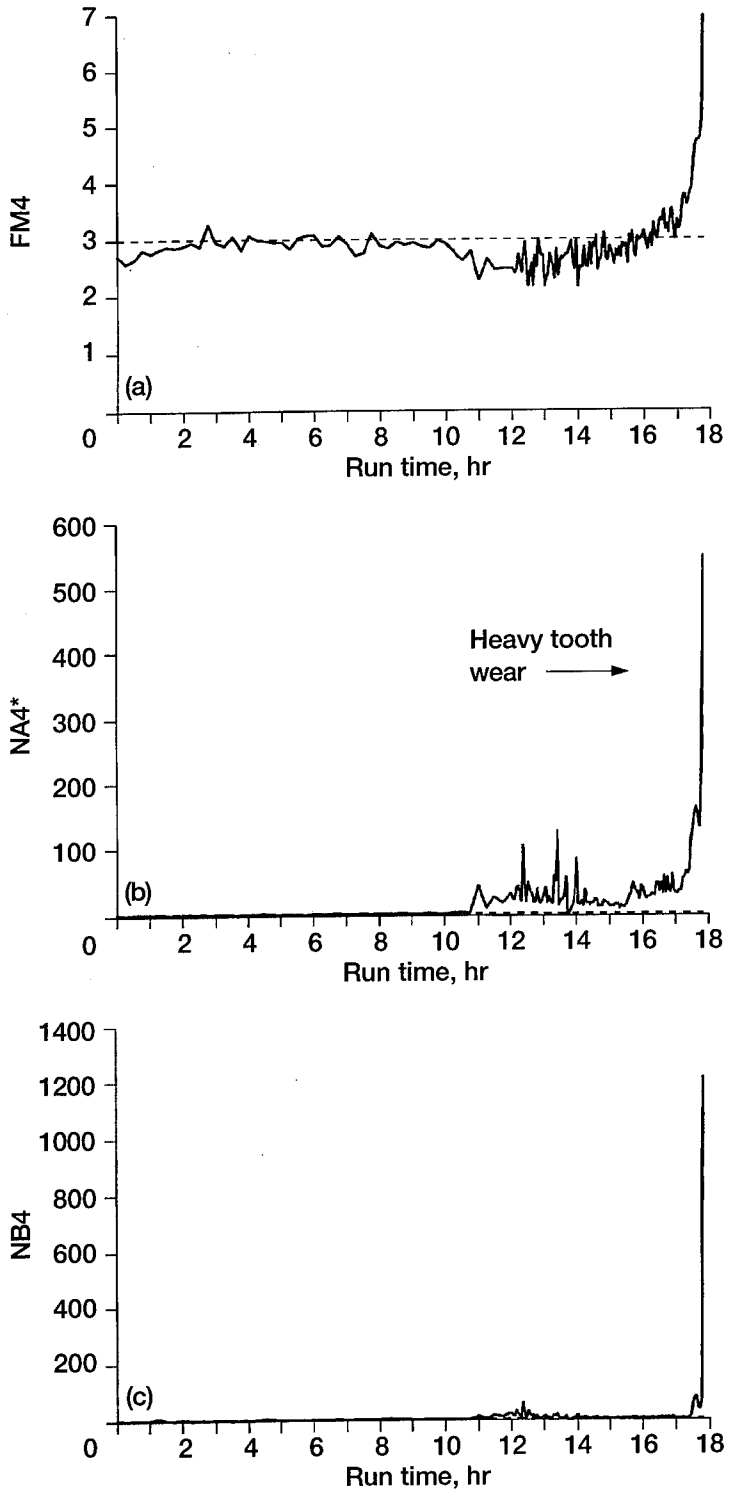


Figure 5.—Face gear run #1 results. (a) FM4. (b) NA4*. (c) NB4.

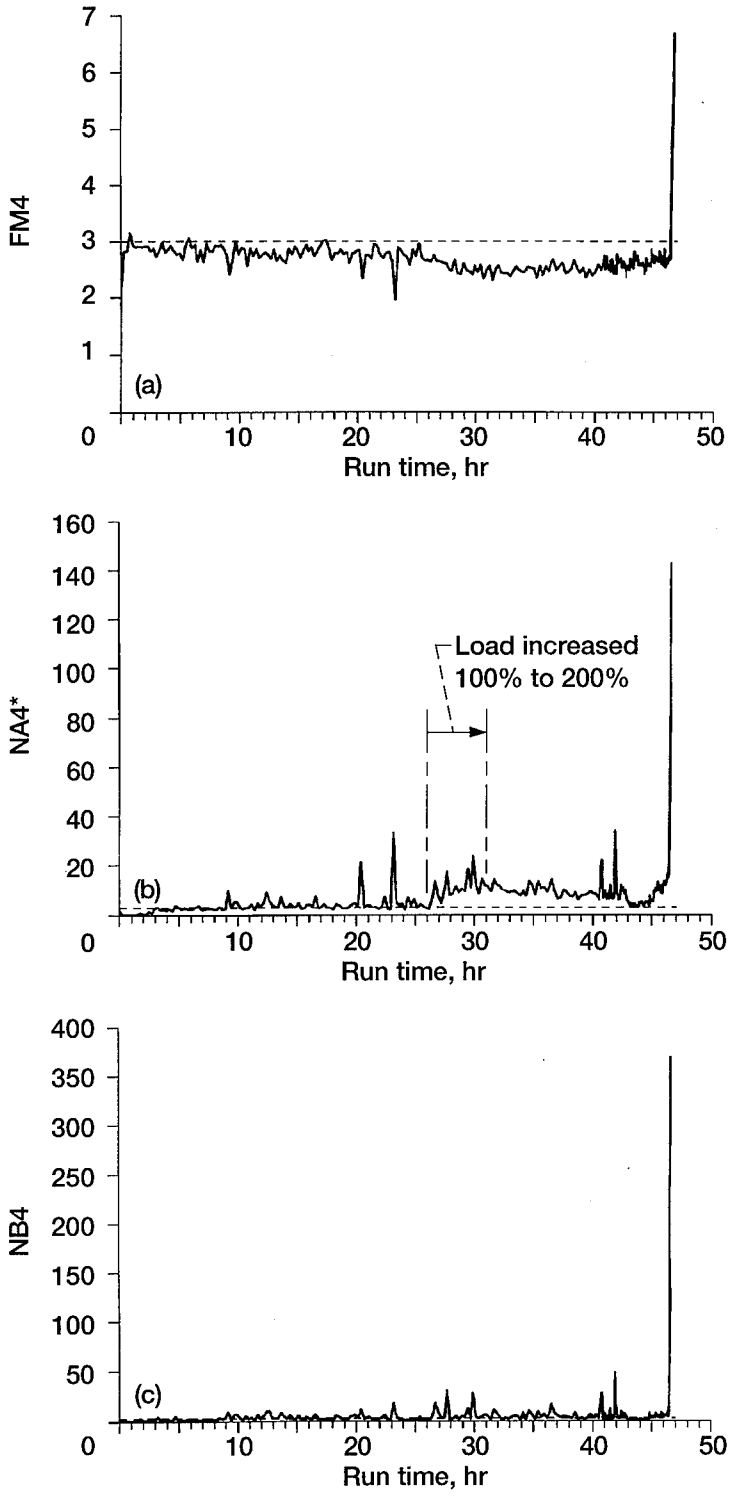


Figure 6.—Face gear run #2 results. (a) FM4. (b) NA4*. (c) NB4.

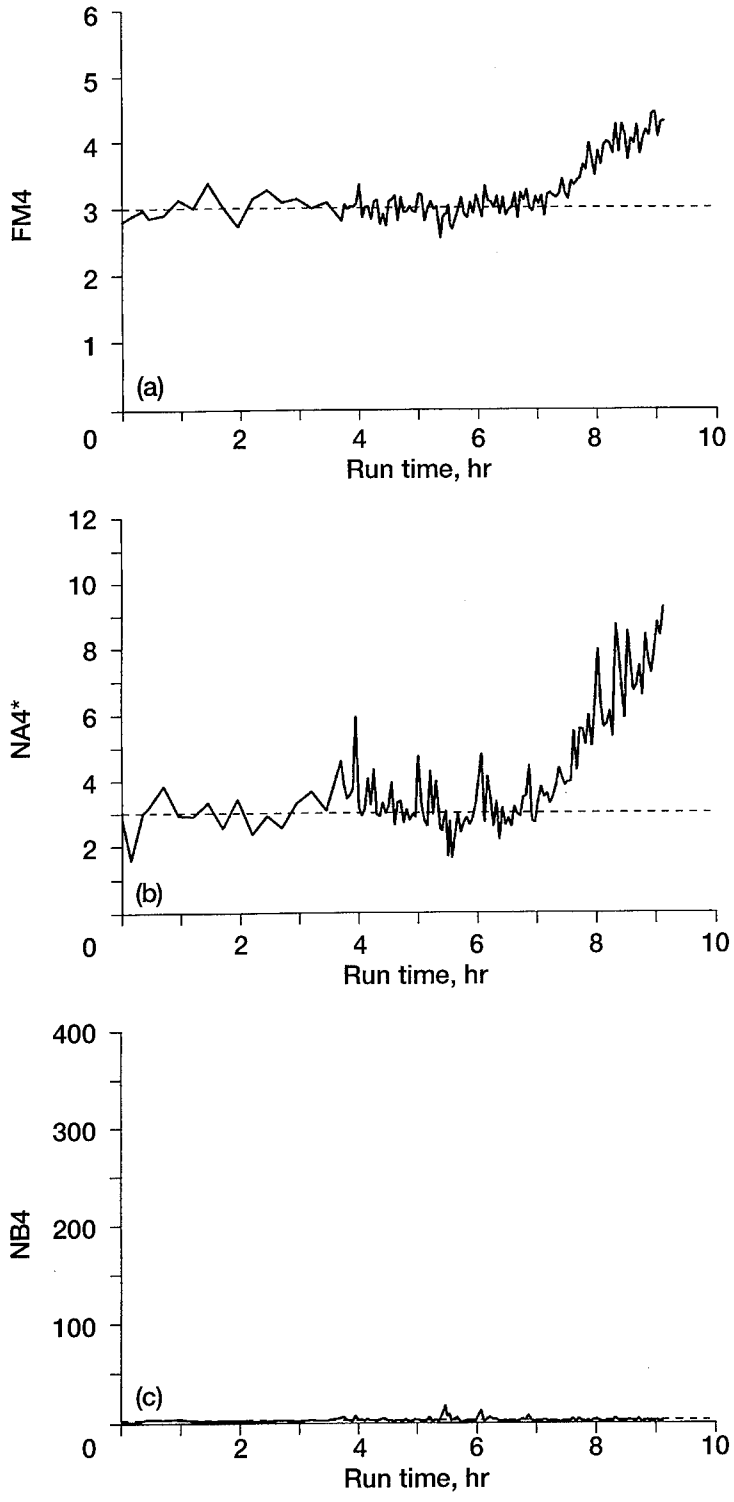


Figure 7.—Face gear run #3 results. (a) FM4. (b) NA4*. (c) NB4.

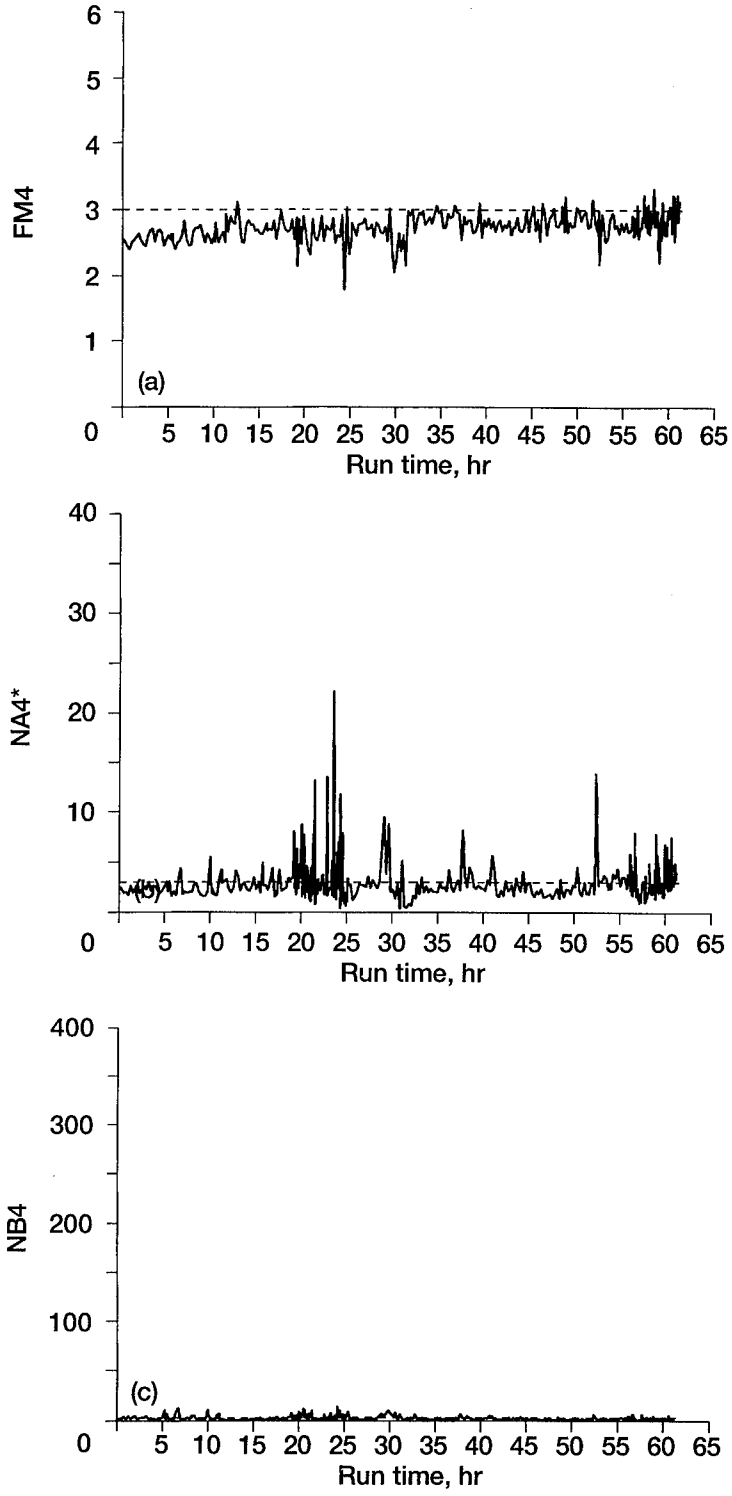


Figure 8.—Face gear run #4 results. (a) FM4. (b) NA4*. (c) NB4.

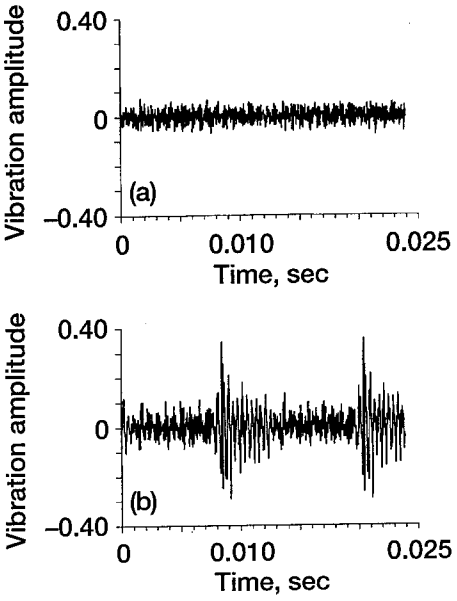


Figure 9.—Time averaged vibration signal for face gear run #1.
(a) At start. (b) At end of test.

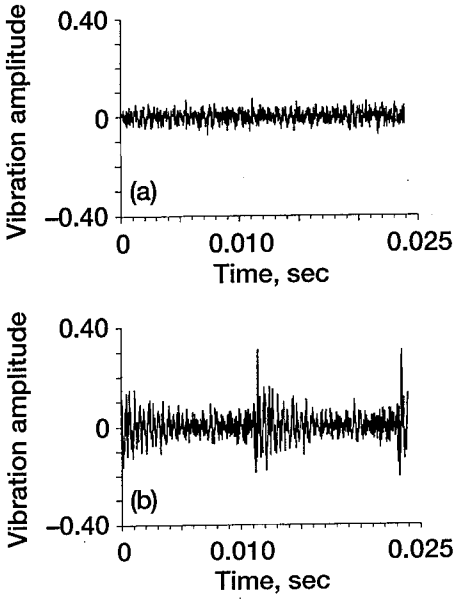


Figure 10.—Time averaged vibration signal for face gear run #2.
(a) At start. (b) At end of test.

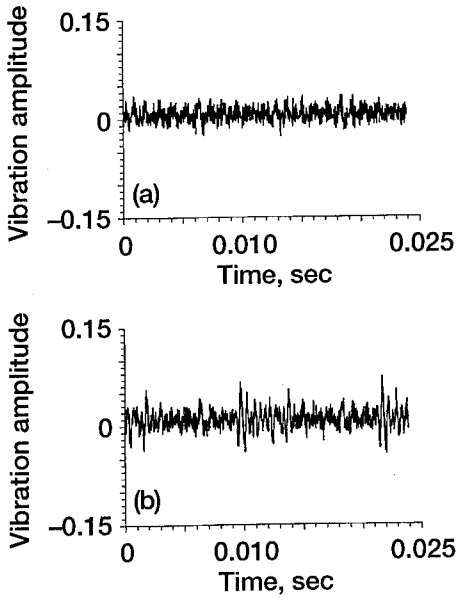


Figure 11.—Time averaged vibration signal for face gear run #3.
(a) At start. (b) At end of test.

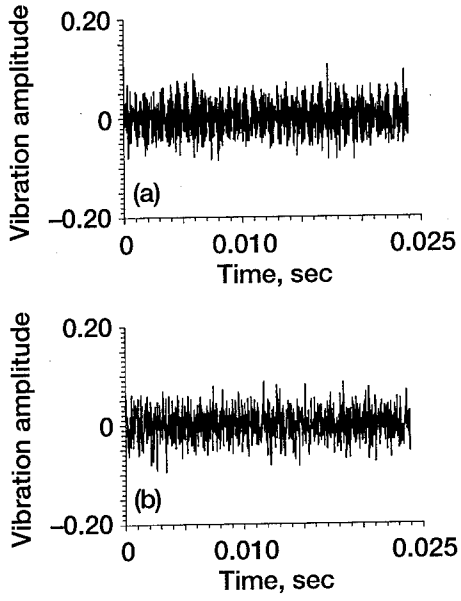


Figure 12.—Time averaged vibration signal for face gear run #4.
(a) At start. (b) At end of test.

GEAR PITTING ASSESSMENT VIA TIME-FREQUENCY ANALYSIS AND DEMODULATION--A CASE STUDY

C. J. Li, J. Limmer, J. Yoo
Department of Mechanical Engineering
Rensselaer Polytechnic Institute, Troy, New York 12180

Abstract:

The objective of this paper is to establish the utility of advanced signal processing algorithms for gear pitting assessment. Vibration measurements have been taken from a 350hp, 4-square gear testing setup. These measurements contain vibration data obtained at different times during a 2000-hour gear test in which frosting, i.e., micro-pitting develops as documented by digital images of the pinion tooth surface. Examining digital images of all the teeth, we found that the development of frosting is distributed throughout the pinion and the extent of frosting grows with testing time.

To enhance the signal to noise ratio, synchronized averaging was used. This technique attenuates signals whose frequencies are not correlated to the rotational speed and its harmonics. However, these data were taken with a fixed temporal sampling interval instead of a spatial one which is the most convenient for synchronous averaging. A preprocessing algorithm has been established to perform signal segmentation and interpolation so that synchronous averaging can be performed on these data.

This study tried to establish the utility of several signal processing algorithms for the detection and assessment of gear pitting. These algorithms include spectrum analysis, narrow and wide band demodulations, the Choi-Williams time-frequency distribution, and Figures of Merit (FM). Among them, only the algorithm based on the wide band demodulation scheme provides a diagnostic variable that grows with the extent of frosting.

Key Words: diagnostics, gear, pitting, frosting, signal processing

1. Introduction

Developing a way to monitor the condition of power transmission equipment has been a research imperative for some time. An Army report (U.S. Army, 1976) indicates that 68% of flight safety incidents related to failures in mechanical systems and 58% of direct maintenance costs involved power plants and drive systems. Because gearboxes are the basic element of power transmission systems they have received considerable attention in the field of condition monitoring and fault diagnosis. A Navy study found that a condition-monitoring maintenance system would allow a 12% crew reduction and 15% increase in available volume in modern combat ships.

In other research, localized gear defects have been extensively studied. One study (Allianz, 1978) found that 60% of gear box damage is due to faults in the gears and about 90% of gear faults were due to localized gear defects such as forced fractures, fatigue fracture, and incipient cracks. If these defects can be detected as they are forming, the gear can be replaced before catastrophic failure. Several methods have been proposed for detecting localized defects and imminent gear failure, including, e.g., narrow and wide band demodulation (McFadden, 1986; Ma and Li, 1994), cepstrum analysis (Aatola and

Leskinen, 1990), and time-frequency distribution (Forrester, 1990, Wang and McFadden, 1993).

However, very little work has been performed to evaluate what is perhaps the most important precursor of the aforementioned gear faults, i.e., tooth surface pitting. Among localized defects, gear surface pitting differs greatly from other kinds and therefore perhaps is the most difficult kind to detect. Pitting is a "distributed" localized defect which spreads around the gear while the aforementioned localized defects, such as tooth fracture, are usually limited to a single tooth or a few teeth. Consequently, a pitted tooth does not stand out in the way a chipped tooth would. What is more important, while other localized defects such as chipping usually change the conformity between meshing teeth significantly and therefore generate a large signal, the effect of pitting is much more subtle due to its small size.

In the following, we will describe the signature generating mechanism of gear pitting. Let us consider a pair of perfect mating gears whose teeth are rigid with exact involute profile and are equally spaced. Such a pair of gears would transmit exactly uniform angular motion in the absence of distributed defects such as runout, imbalance, and misalignment. Any deviation from this ideal situation will cause variations in both angular displacement and velocity, i.e., the transmission error, which in turn results in variations in the force transmitted between the meshing teeth. Vibrations will then be generated and transmitted everywhere in and on the gearbox through the gear-bearing-shaft-bearing-casing path.

Then consider a pair of gears whose teeth are not rigid but otherwise the same as the aforementioned perfect gears, meshing under constant load at constant speed. Since the contact stiffness varies periodically with the number of teeth in contact and with the contacting position on tooth surface, vibration will be excited at the tooth meshing frequency. In the form of a Fourier series, the vibration may be approximately represented as a weighted sum of sine/cosine functions at the tooth meshing frequency and its harmonics:

$$x(t) = \sum_{m=0}^{M} X_m \cos(2\pi m T f_s t + \phi_m) \tag{1}$$

where $x(t)$ is the vibration signal, M is the number of harmonics used to describe the signal, X_m and ϕ_m are the magnitude and phase of the harmonic m, T is the number of teeth on the gear, and f_s is the shaft rotation frequency.

Gear tooth surface pitting reduces the meshing stiffness and therefore produces changes, i.e., modulations, in the amplitude and phase of the vibration at the meshing frequency and its harmonics as the teeth go through the meshing. Taking this into consideration, (1) becomes

$$x(t) = \sum_{m=0}^{M} X_m (1 + a_m(t)) \cos(2\pi m T f_s t + \phi_m + b_m(t)) \tag{2}$$

where $a_m(t)$ and $b_m(t)$ are the amplitude and phase modulations of the signal.

If the amplitude and phase modulating functions can be extracted from the gear vibration, then we will be able to evaluate the gear pitting. Through a case study, this paper evaluates the effectiveness of several advanced signal processing methods for the evaluation of tooth pitting. These algorithms include the energy operator, narrow and wide band demodulation, figures of merit, and time-frequency distribution. A brief

summary of each of them is given in the following section.

2. Algorithm Summary

Energy operator: In communication theory, the energy operator is used to decode the information on an amplitude modulated or frequency modulated signal (Maragos, et. al., 1993). In a similar way, the energy operator can be used to extract the amplitude or frequency modulation from a vibration signal. In the discrete form, the energy operator is defined as

$$\Psi(s(n)) = s^2(n) - s(n+1)s(n-1) \tag{3}$$

If the signal has the form of a single harmonic,

$$s(t) = a(t)\cos(\phi(t)) \tag{4}$$

then the energy operator can be shown to be related to the amplitude and frequency as follows:

$$\Psi(s(n)) = a^2(t)\dot{\phi}^2(t) \tag{5}$$

Obviously, modulations in amplitude and/or frequency will be revealed in the energy operator.

Note that this algorithm requires narrow band filtering of the vibration signal so that the signal can be approximated by (4).

Demodulation: McFadden (1986) proposed a simple algorithm to recover the amplitude and frequency modulations from gear vibration. By assuming the sidebands of different meshing harmonics do not interfere significantly with each other, the vibration signal was band-pass filtered about the most prominent harmonic of the gear meshing frequency. Only the sidebands of the single harmonic were considered. The resulting expressions for amplitude and phase modulations are:

$$\begin{aligned} a_m(t) &= |c_m(t)|/X_m - 1 \\ b_m(t) &= \arg(c_m(t)) - (2\pi m T f_s t + \phi_m) \end{aligned} \tag{6}$$

where $c_m(t)$ is the analytic signal of the result of band-pass filtering $x(t)$. Note that this algorithm would perform poorly if the aforementioned assumption is not met.

Ma and Li (1994) created an iterative demodulation algorithm that can determine amplitude and phase modulation from a wide-band signal, i.e., a signal includes several meshing harmonics and their sidebands. No bandpass filtering is needed and the interactions among side bands of neighboring meshing harmonics are accounted for. The algorithm has been shown to be more sensitive to gear tooth chipping than McFadden's demodulation algorithm.

Figures of merit: The various figures of merit, FM0, FM2, and FM4 (Stewart, 1977; Zakrajsek, et. al., 1993), were designed to detect certain kinds of gear fault from the vibration of a gear. For example, FM4 indicates the amount of localized damage, such as pitting or small cracks on one or two teeth, with a single number. Using FM4, the total amount of localized damage can be tracked over the lifetime of the gear without having to

save the entire vibration signal at each point in time. FM4 is found by removing the tooth-meshing harmonics and their first order sidebands from the vibration signal and taking the normalized fourth statistical moment (i.e., the normalized kurtosis) of the signal that remains.

Time-Frequency distributions: Time-Frequency distributions (Cohen, 1989) have been used to analyze gear vibrations, particularly for finding single-tooth gear faults (Forrester, 1990, Wang and McFadden, 1993), since the power spectrum for the time during which a faulty tooth is in mesh will differ from that when a normal tooth is in mesh. The Choi-Williams distribution (Choi and Williams, 1989) is a particular Cohen class time-frequency distribution that clearly represents the time-varying power spectrum of a signal while reducing spurious cross-term noise. This distribution has been successfully used by the authors to identify single-tooth gear faults.

3. Experimental Setup and Data

The test rig consists of two gear boxes in parallel driven by an electric motor (see Figure 1). The two gear boxes are arranged in a four-square configuration (Shipley, 1958) which enables the loading of gears by simply twisting one of the shafts between the two boxes. The test gear and pinion have 55 and 25 teeth, respectively. The gear was driven at approximately 1000 RPM.

Vibration was measured by an accelerometer mounted on the test gear box along with the output of a tachometer consisting of a 72 tooth gear on the pinion shaft and a proximity sensor. The lubricant temperature is also regulated by modulating the oil flow rate between the gearboxes and the radiator. Vibration Data and digital images of the pinion teeth were taken during the test, at approximately 0 hours (after break-in), 300 hours, 513 hours and 1010 hours.

To reduce noise, synchronized signal averaging (Braun and Seth, 1979), which has the effect of a comb filter, was employed for each data set. In general, vibrations of the gear of interest are sampled at the same angular positions for a number of rotations and samples corresponding to the same angular position are then averaged over the rotations. However, our data was sampled with a constant sampling interval in time. Samples corresponding to the same time in different rotations are not at the same angular position unless the speed is constant, which is unlikely due to modulations.

An algorithm was developed to transform the raw data to make it suitable for synchronized averaging. Using the output of the tachometer, the vibration data is segmented into pieces that correspond to consecutive rotations. The number of data points are different from one rotation to the next due to speed variation. Assuming a constant speed within a rotation, the number of data points is made the same for all the rotations using linear interpolation. Then synchronized averaging is carried out.

Because of the uncertainty in the load applied, the synchronized average of each data set was normalized so that the total energy in each signal is equal. Each data set will be compared without respect to the relative amplitudes of the vibration signals. Pitting will show up in the redistribution of vibration energy instead of its amplitude.

4. Results

The tested gears are fully hardened and therefore develop frosting, i.e., micro-pitting, as they go through meshing cycles under load. Frosting is found throughout the pinion, however, with some tooth-to-tooth variation. Fig. 2 shows the pictures taken of the same pinion tooth at the aforementioned times during the test. The size of the white area, which

indicates the micro-pitting, grows with time.

Figure 3 shows the FM4 for each data set. Figure 5 shows the synchronized average of each of the four data sets. (In the testing rig, there was no once-per-revolution marker or other way of identifying individual teeth, so that the plots for different data sets are not synchronized with each other, i.e., tooth 3 for a plot of the data at 300 hours is not necessarily the same tooth on the pinion as tooth 3 for a plot of the data at 1010 hours.) Their spectra are shown in Fig. 6. Note that no progressive, consistent change appears in the spectrum. These synchronized averages were processed using the energy operator, McFadden's narrow band demodulation, the Choi-Williams distribution, and Ma's wide band demodulation as shown in Fig. 7-10, respectively. Figure 4 shows the rms values of the amplitude modulations shown in Fig. 10. Recall that both the energy operator and McFadden's demodulation algorithm require band-pass filtering to include only the sideband frequencies about the largest harmonic peak for that data set.

Although previously these methods have been shown useful for detecting localized gear defects, they have difficulty measuring the micro-pitting in the gear data studied. The energy operator and single-band demodulation seemed to offer no clear indication of the steady increase of micro-pitting. The Choi-Williams distribution also failed to give any consistent pattern indicating the extent of micro-pitting. However, the wide band demodulation scheme does provide a general trend from low levels of pitting to high levels; this trend can be clearly seen in Fig. 4. However, note a sharp rise at 513 hours which will be explained below.

By examining the experimental log, it was found that the data taken at 513 hours was taken just after the machinery was started up cold after being shut down temporarily to correct elevated lubricant temperature, while the other data was taken when the setup was warm. This expains why this particular run is rougher than others as indicated in its modulating signal shown in Fig. 10.

5. Conclusions

This study tried to establish the utility of several signal processing algorithms for the detection and assessment of gear frosting. These algorithms include spectrum analysis, narrow and wide band demodulations, the Choi-Williams time-frequency distribution, and figures of merit. Among them, only the algorithm based on the wide band demodulation scheme provides a diagnostic variable that grows with the extent of pitting.

It is obvious that micro-pitting is a very difficult gear fault to detect and assess. Its small fault signature could be easily masked by variations due to a small change of operating conditions such as temperature. The authors will try to obtain a higher level of performance with more data sets taken from this setup, with data from a new setup, and with more advanced gear diagnostic algorithms.

Acknowledgment: The authors would like to express their gratitude to Mr. Richard Smith and Mechanical Technology, Inc., of Albany, New York for their technical assistance and data.

References

Aatola, S. and Leskinen, R., 1990, "Cepstrum Analysis Predicts Gearbox Failure," Noise Control Engineering Journal, Vol. 34, No. 2, pp. 53-59

Allianz Versicherungs-AG, 1978, *Handbook of Loss Prevention*, Springer-Verlag, Berlin

Braun, S., and Seth, B. B., 1979, "On the Extraction and Filtering of Signals Acquired from Rotating Machines," *Journal of Sound and Vibration*, Vol. 65, No. 1, pp. 37-50

Choi, H-I. and Williams, W. J., 1989, "Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernels," IEEE Transaction of Acoustics, Speech and Signal Processing, Vol. 37, No. 6, pp. 862-871

Cohen, L., 1989, "Time-Frequency Distributions--A Review," Proceedings of the IEEE, Vol. 77, No. 7, pp. 941-981

Forrester, B. D., 1990, "Analysis of Gear Vibration in the Time-Frequency Domain," Proceedings of the 44th Meeting of the Mechanical Failures Prevention Group, pp. 225-234

Ma, Jun and C. James Li, 1994, "A New Approach to Demodulation of Gear Vibrations and its Application to Defect Detection," Proc. of 48th general meeting of Mechanical Failures Prevention Group, pp. 43-55

Maragos, P., Quatieri, T. F., and Kaiser, J. F., 1993, "On Amplitude and Frequency Demodulation Using Energy Operators," IEEE Transactions on Signal Processing, Vol. 41, No. 4, pp. 1532-1550

McFadden, P. D., 1986, "Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration," Journal of Vibration, Acoustics, Stress, and Reliability in Design, Vol. 108, No. 2, pp. 165-170

Shipley, G., 1958, "Twelve Ways to Load-Test Gears," Product Engineering, Jan. 6, pp. 77-82

Stewart, R. M., 1977, "Some Useful Data Analysis Techniques for Gearbox Diagnostics," Report MHM/R10/77, Machine Health Monitoring Group, Institute of Sound and Vibration Research, University of Southampton

US Army, 1976, "Update to Reliability and Maintainability Planning Guide for Army Aviation Systems and Components," USAAVSCOM Technical Report 77-15

Wang, W. J. and McFadden, P. D., 1993, "Early Detection of Gear Failure by Vibration Analysis--I: Calculation of the Time-Frequency Distribution," Mechanical Systems and Signal Processing, Vol. 7, No. 3, pp. 193-203

Zakrajsek, J. J., Townsend, D. P., and Decker, H. J., 1993, "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, p. 199-208

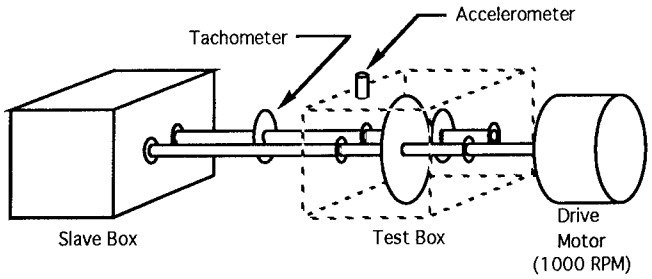
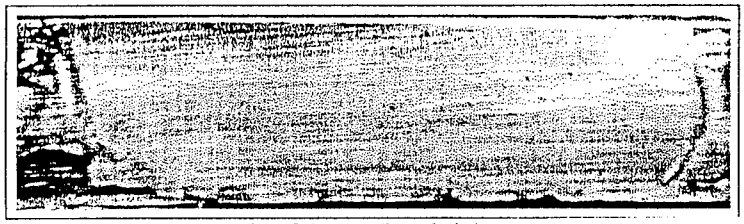
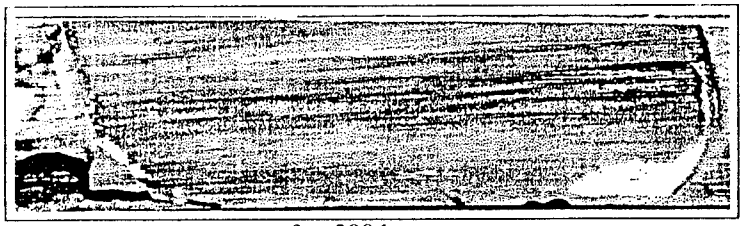


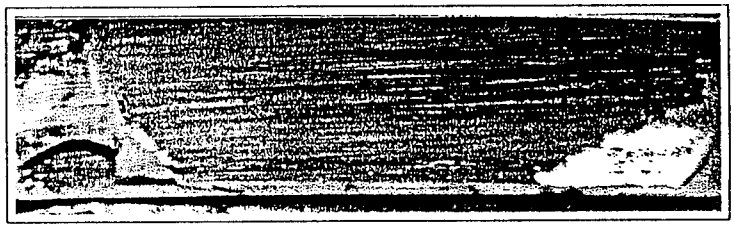
Figure 1. Gear Test Rig



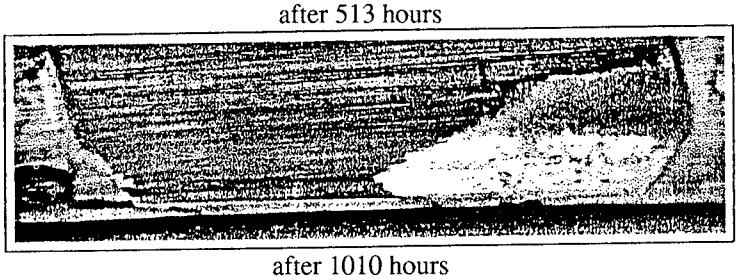
after 0 hours (break-in)



after 300 hours



after 513 hours



after 1010 hours

Figure 2: Micro-pitting on a typical tooth

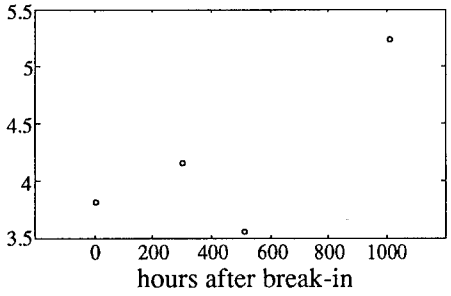


Figure 3: Fourth order Figure of Merit (FM4) for each data set

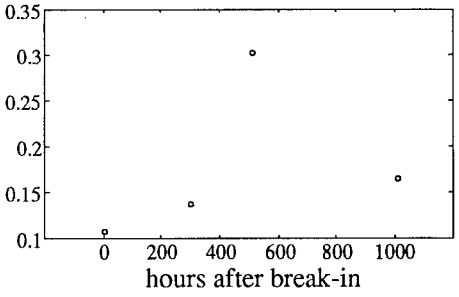
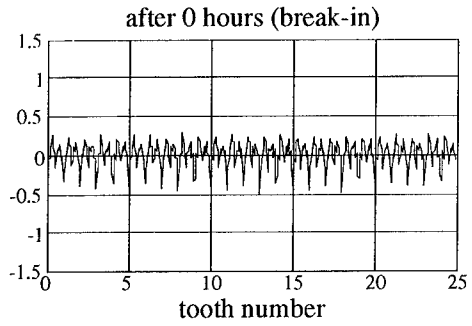
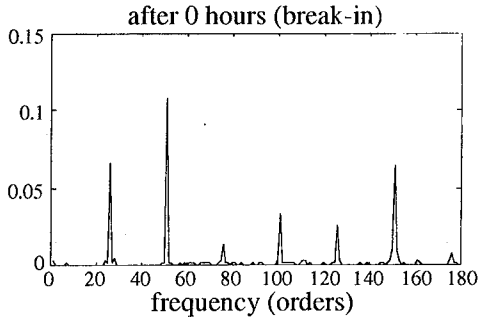
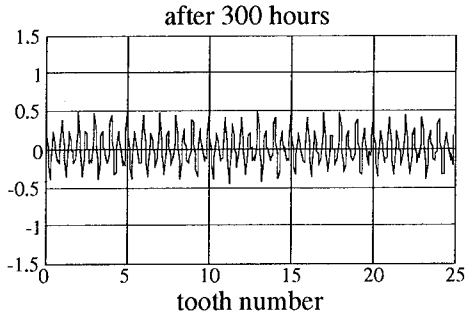
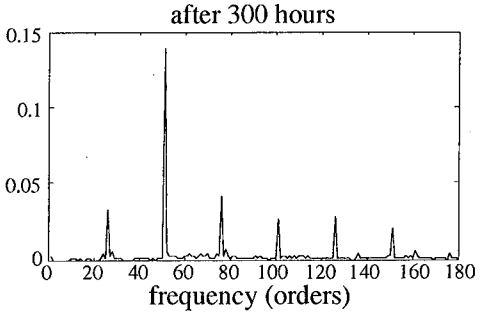


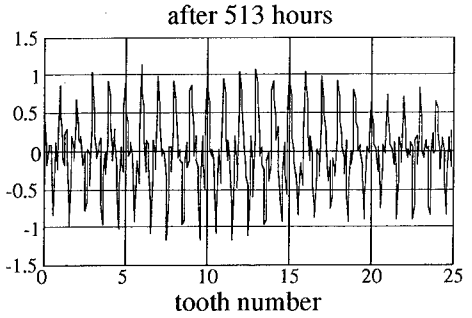
Figure 4: RMS values of the amplitude modulation signals shown in Fig. 10

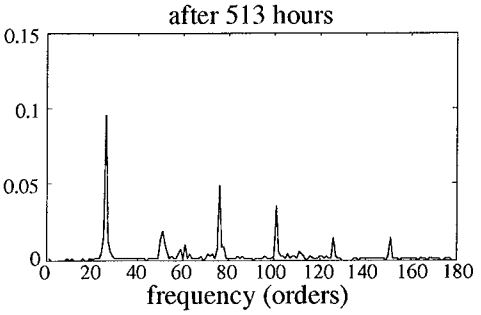


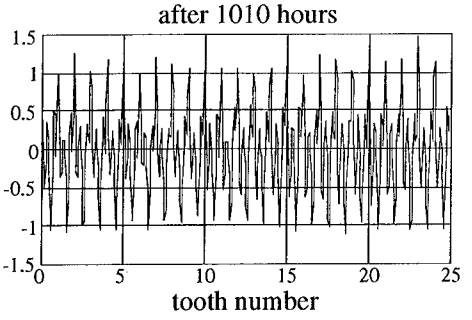












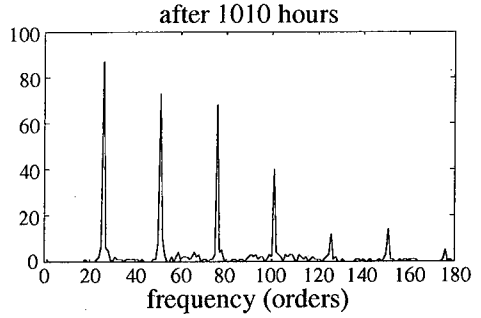
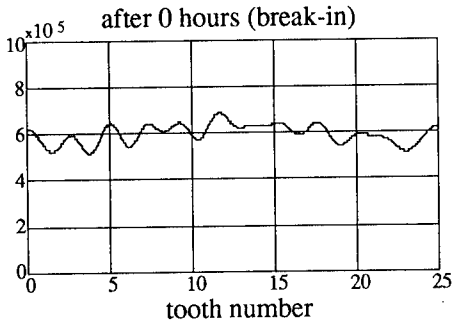
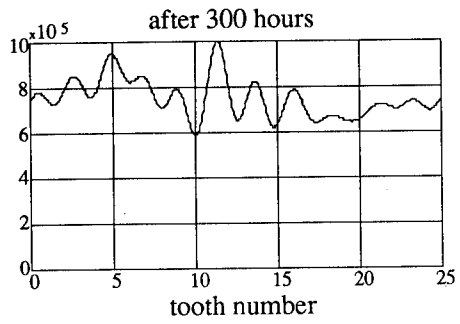
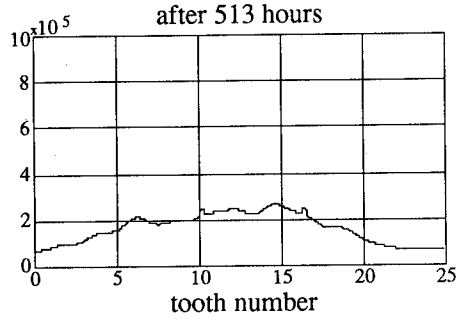


Figure 5: Synchronized average of vibration data

Figure 6: Normalized magnitude spectra of vibration data







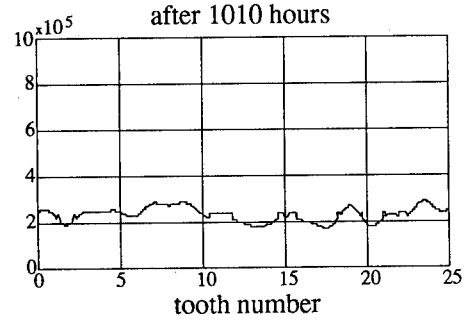
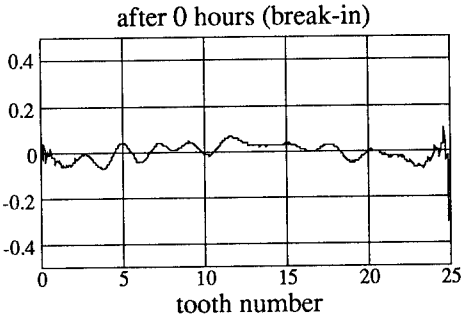
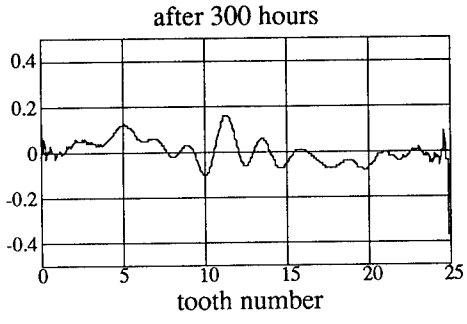
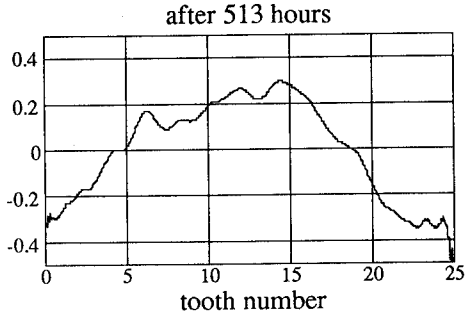


Figure 7: Amplitude modulation from energy operator







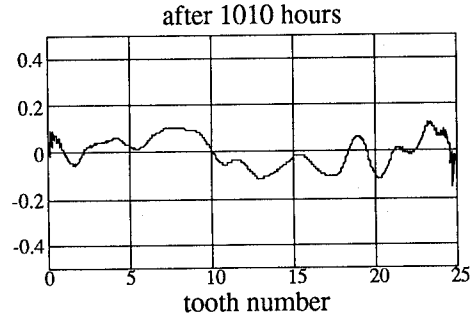
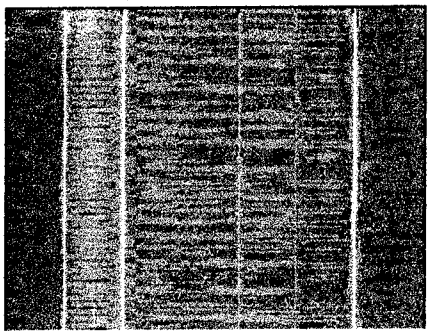
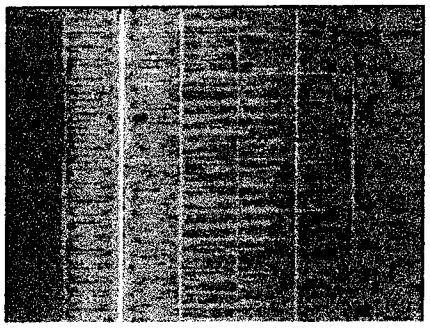
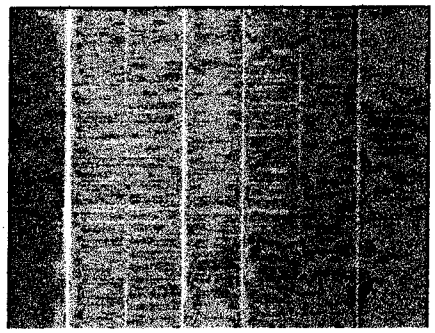


Figure 8: Amplitude modulation from single-band demodulation







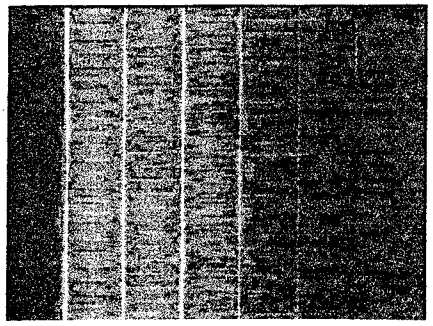
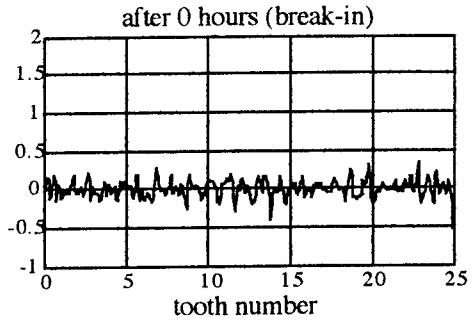
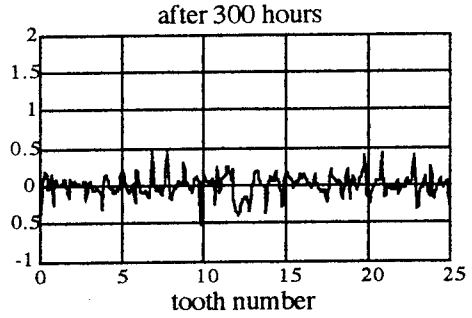
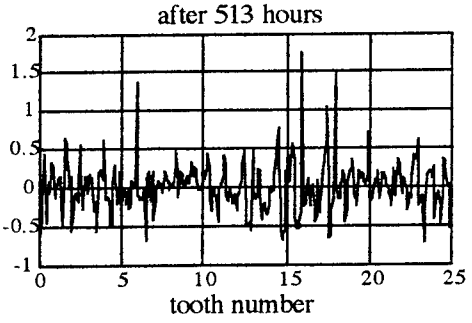


Figure 9: Choi-Williams distribution. (Vertical lines occur at meshing harmonics)







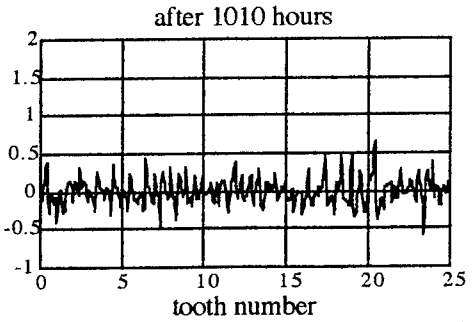


Figure 10: Amplitude Modulations obtained by wide-band demodulation

SYNCHRONOUS AVERAGING OF MULTIPLE GEAR SHAFTS USING AN INTERPOLATING FILTER

Dr. George P. Succi

Technology Integration
54 Middlesex Turnpike
Bedford Ma. 01730

Abstract

Here we discuss a device using a microprocessor for order tracking of rotating machinery. The goal is to synchronously average a signal. The technique is to sample both the signal and tachometer at a fixed rate and store them in memory. We apply the method to data from an accelerometer mounted on a gearbox housing.

Key Words FIR, Gear, Synchronous Average, Interpolation

1.0 Introduction

When testing rotating machinery it is useful to consider the order analysis of the signal. By order analysis we mean the discrete Fourier transform of the signal based on the fundamental period of the signal. An example where such is such information is useful is a gearbox. A damaged tooth on a gear will generate a signal with a period equal to the period of its shaft. There are also other periods based on the time it takes for the assembly of gears and shafts to return to the initial position.

A direct method to sample at the appropriate times is to mount an optical or mechanical encoder on the gear shaft. This method gives a fixed number of pulses for each rotation of the shaft. Data is sampled at each pulse. This method depends on direct access to the shaft of interest. It is difficult to use when an intermediate shaft is located entirely within the gearbox.

An alternate method is to use a phase locked loop. Here an encoder gives a one pulse per shaft rotation. The duration of the next pulse is predicted from previous pulses and the data is sampled at the appropriate fraction of this interval. The drawback is that the duration of the pulse may differ from the prediction because of speed variation. One purpose of synchronous sampling is to gain a coherent signal from one rotation to the next. The fundamental problem with the phase locked loop method is that the period of rotation may vary from one cycle to the next. Because of this variation it is better to sample at a fixed rate and then interpolate the signal at the desired times .

Here we explore a method to directly sample both the signal and tachometer at a fixed rate unrelated to the rotation rate. The data is stored in memory and interpolated at the desired rate. The same data can be used to analyze faults on gears on different shafts. The rate is deduced from the number of tachometer pulses per rotation and the ratio of the desired shaft period to the tachometer shaft period. Errors come from three sources: the elasticity in the shaft, the measurement of the tachometer period, and the interpolation of the signal. We will not discuss problems associated with the elasticity of the shaft. Here we assume all shafts are rigid. We will discuss errors due to the tachometer measurement and to interpolation.

A similar method is discussed by Potter [1990] and by Potter and Gribler [1989]. They review a method for order tracking during acceleration. They include a discussion of interpolation filter, but do not reveal the detail design of the filter. The general problem of reconstructing a signal from a bandwidth limited digitally sampled signal was discussed by Shannon [1949]. Oppenheim and Wilsky [1983] review Shannon's principal results. The problem of interpolation and decimation, and computer programs to implement the method, is discussed by Crochiere [1979]. The methods discussed by Crochiere are limited to cases where the ratio of the number of interpolations to the number of samples is a rational number. Because our sample rate is not related to the shaft rate, the ratio of interpolations to samples is not necessarily a rational number.

We will discuss the problem of interpolation and the errors due to the tachometer measurement. We will apply the method to a helicopter gearbox.

2.0 Interpolation

Shannon, [1949] observed that a signal can be uniquely specified by a sequence of equally spaced samples. Oppenheim [1983] states "if a bandwidth limited signal x(t) is amplitude modulated with a periodic pulse train, corresponding to extracting equally spaced time segments, it can be recovered exactly by low pass filtering if the fundamental frequency of the modulating pulse train is greater than twice the highest frequency present in x(t)." The ideal low pass filter passes all frequencies below the Nyquist frequency (equal to one half the sample rate) and rejects all frequencies above the Nyquist frequency. It is the low pass filter which allows us to interpolate between points.

The low pass filter can be done in the time or frequency domain. Assume we sample at frequency $f_s = 1/T$ and filter at the Nyquist frequency $f_c = f_s/2$. Low pass filtering is done in the time domain by convolution. The expressions can be normalized by the sampling interval T. We find

$$x(t)=\sum_{n=-\infty}^{\infty} x_n H(t-n) \qquad (1)$$

where

$$H(t)=\frac{\sin(\pi t)}{\pi t} \qquad (2)$$

It is always possible to reconstruct the signal between the sampled points. However, the accurate reconstruction of the signal requires an infinite number of terms. To avoid the infinite summation we use an alternate function, h(t), that approximates the desired low pass filter characteristics. These filters take the form

$$h(t)=\begin{cases} H(t) & |t|\le m \\ 0 & |t|>m \end{cases} \qquad (3)$$

The simplest of these is linear interpolation. Here

$$h(t)=|1-t| \qquad t\le 1 \qquad (4)$$

Linear interpolation is a convolution between two nearest points. The problem is that the Fourier transform of the linear interpolation function only roughly approximates the ideal low pass filter.

In all cases, the data must be bandwidth limited at half the sampling frequency. Shannon's interpolator is designed to pass all frequencies below the Nyquist frequency. In this sense it is a low pass filter. It is possible to modify the form of the Shannon interpolator to filter out some frequencies below the Nyquist frequency. We do not consider these modifications. We will propose a modified form of Shannon's interpolator that we call a truncated low pass filter. Like the original interpolator, it will be designed to pass all frequencies in the bandwidth limited signal.

2.1 Interpolation with a truncated low pass filter

We propose another filter. The filter is based on the ideal filter. It differs in two ways: it is truncated after 10 terms and is multiplied by a window function that drives it to zero at |t| = m. Here

$$h(t)=\left(\frac{\sin(\pi t)}{\pi t}\right)\frac{1}{4}\left(1+\cos\left(\frac{\pi t}{m}\right)\right)^2 \qquad (5)$$

In figure 1 we illustrate the performance of the truncated low pass filter. Figure 1A gives the filter characteristics versus frequency. The ideal filter is rectangular in shape with the upper frequency at $f_s/2$. The proposed filter approximates this ideal. Figure 1B is similar to 1A but with the vertical axis in decibels. Figure 1C is the plot of the interpolation function versus time. In figure 1D we plot the interpolation error. This error is defined as the maximum difference between the interpolated value based on sampled data and the original sine wave. The error increases with frequency

We can estimate the allowable error by considering the digital error on the A to D converter. A twelve bit A to D converter has an error equal to one part in 4096 (2^{12}) or 72 dB. The interpolation error in figure 1 is less than 82 dB up to $f_s/4$. Thus the data is reliably reconstructed up to one quarter the sampling frequency. We have not explored the leakage of a signal between one quarter and one half f_s into the frequencies below one quarter f_s. We suggest that the way to avoid the problem is to bandwidth limit the signal at one quarter the sampling frequency. This procedure is called over-sampling by a factor of two.

A similar calculation can be done for a sixteen bit A to D converter. Here the error is one part in 65356 (2^{16}) or 96 dB. To meet the criteria of an interpolation error less than one bit we would need more than ten terms.

The filter meets two criteria:

1) accurate interpolation for frequencies below $f_s/4$,
2) a small number of terms in the convolution.

By accurate, we mean an interpolation error smaller than 72 dB. The filter itself is composed of two parts: a part identical to the Shannon interpolation function, and an envelope equal to $(1+\cos(..))^n$, where the exponent n is equal to 2. The larger the value of n the better the low frequency performance ($f/f_s \approx 0$) and the worse the high frequency ($f/f_s \approx 1/2$) performance. Using the Shannon interpolation function alone, truncated after ten terms, gives a filter whose peaks are nearly equal in magnitude. Adding the envelope function suppresses the low frequency peaks and increases the high frequency peaks.

This approach is similar to Potter [1990]. Potter recommends,as we do, that the sample be bandwidth limited at one quarter the sampling frequency. His filter has a pass band approximately equal to unity between 0 and $f_s/4$ and uses the interval up to the Nyquist frequency as a transition band.

2.2 Interpolation with "extended" digital filters

There are other ways to construct a low pass filter. Much of the literature in filter theory concerns digital filters. These filter coefficients are defined at discrete times. They are used to low pass and band pass digitally sampled data by convolution. Two filters in wide use are the Butterworth filter and the Chebyschev filter. Parks and Burrus [1987] review the methods used to construct these filters.

Use of these filters in decimation and interpolation is discussed by Crochiere [1979]. To use them in our application requires that we generalize them to non-integer values of time. There are four types of digital filters, each of which has different symmetry properties. The type 1 digital filter is best suited for low pass filter applications. The filter is specified as

$$h(t) - \frac{1}{N}\left\{a_0 + \sum_{k=1}^{m} a_k \cos\frac{2\Pi(n-m)K}{N}\right\} \qquad (6)$$

where N = 2m+1. These filters are symmetric about n = m+1. The problem in digital filter theory is to construct the coefficients a_k. The phase shift is due to the factor m in the argument of the cosine function.

The extension of equation 6 from discrete to continuous time is what we mean by "extended " digital filters. We can construct a continuous function using the coefficients a_k. The phase shift can be removed by replacing m with 0 in the argument of the cosine function. This phase shift is due to the causality restriction in linear finite impulse response filters (FIR). The FIR filter replaces a data point with the weighted sum of the preceding points. Centering the filter (replacing m with 0) replaces a data point with the weighted sum of its nearest neighbors. The filter is no longer causal when it's centered because some of the points in the weighted sum occur later in time. The centered filter can be generalized to arbitrary times by replacing n with t. Thus

$$h(t) - \frac{1}{N}\left\{a_0 + \sum_{k=1}^{m} a_k \cos\frac{2\Pi tK}{N}\right\} \qquad |t|<m \qquad (7)$$

The restriction that |t| <= m implies that this function is no longer periodic. We can construct the continuous Fourier transform analytically. Here

$$H(f) = \int_{-m}^{m} e^{2\pi i f t} H(t)\, dt = \frac{1}{N} \sum_{-m}^{m} a_k S(f+k/N) \qquad (8)$$

$$S(f+k/N) = \begin{cases} \dfrac{\sin(2\pi m(f+k/N))}{\pi(f+k/N)} & (f+k/N) \neq 0 \\ 2m & (f+k/N) = 0 \end{cases} \qquad (9)$$

where $a_{-k}=a_k$. The result follows form expressing the cosine in exponential form and integrating from -m to +m.

We construct the optimal 10 term Chebyschev digital filter using the Remes exchange algorithm as described by Parks and Burrus [1987]. The filter is illustrated in figure 2. Figure 2A gives the filter characteristics versus frequency. Figure 2B is similar to 2A but with the vertical axis in decibels. Figure 2C is the plot of the function versus time. In figure 2D we plot the interpolation error. This error is greater than the error for the proposed low pass filter illustrated in Figure 1.

We repeated the procedure for the maximally flat Butterworth filter. Again, the interpolation error for an extended ten term Butterworth filter is greater than the error for the proposed low pass filter.

The truncated low-pass filter gives a lower error than the extended digital filter.

2.2 Interpolation by Fourier transform

One other possibility is to compute the Fourier transform of the signal. Given these coefficients we can interpolate to any time t. The problem is that the number of terms between tachometer pulses is, in general, not a power of two. In such cases an efficient routine, such as an FFT, cannot be used to construct the Fourier coefficients. The coefficients can be calculated, but the processing time is too long for practical use. Once the Fourier coefficients are calculated, the set of N sine and cosine function must be computed for each of the interpolation points. Such a procedure is useful when we only need to include the influence of first few harmonics.

3.0 The influence of the tachometer

Tachometer errors are independent of interpolation errors.

To interpolate we ask given a signal sampled at times mT, how do we properly reconstruct the signal at any intermediate time t. The tachometer allows us to specify the time t at which we do the interpolation. Our goal is to sample the signal at equal increments in phase. These equal increments in phase occur, in general, at unequal increments in time.

The rule is to assign a phase for each of the tachometer arrival times. At each of the arrival times T_n we assign a phase Φ_n. We interpolate between these fixed values of Φ_n to find 2^N values of the signal at F_i and the corresponding time t_i.

The procedure is illustrated in figure 3. Each time a tachometer pulse arrives we assign an angle phi. We plot the measured tachometer time versus phi. The interval between two angles phi is equally subdivided. If the shaft is accelerating, equal intervals in phi will yield unequal intervals in time. To accommodate the acceleration we use cubic splines.

An error arises because of the finite accuracy with which we measure the tachometer. This problem exists even in the absence of acceleration. Suppose we have a signal p(t) that is periodic with period T_0 and angular frequency ω_0. The exact expression for p is given by

$$p(t) = \sum_{k=-N}^{N} a_k e^{2\pi i \omega_0 t} \qquad \textbf{(10)}$$

Now let us assume that, due to errors in the measurement of the tachometer arrival time, we estimate the period as T_1 and the angular frequency ω_1. We then approximate the signal as

$$p(t) = \sum_{k=-N}^{N} b_k e^{2\pi i \omega_1 t} \qquad \textbf{(11)}$$

The coefficients b_k are only approximately equal to the exact coefficients a_k. We can calculate the coefficients b_k by substituting Eq. 11 into Eq. 10 and using the orthogonality properties of the exponential. Thus

$$b_m = \sum_{n=-N}^{N} \frac{\sin((n\omega_0 - m\omega_1)T_1/2)}{(n\omega_0 - m\omega_1)T_1/2} \qquad \textbf{(12)}$$

The coefficients b_m equal the coefficients a_m in the limit as T_1 tends to T_0. In the presence of an error in the measurement of the period there is "leakage" into the nearby harmonics. The leakage diminishes with separation between the harmonics.

The leakage depends on both the harmonic and the ratio of the measured to the actual period. When the ratio is one, there is no leakage. When the ratio is other than one the leakage is finite. In table 1 we give the calculated leakage from the tenth to eleventh harmonic as a function of accuracy. Notice, to keep the leakage below 72 dB we need an accuracy of one part in 10^5.

Table 1
Effect of errors in the measurement of the period

Accuracy 1 per 10_n n	Ratio T_0 / T_1	Leakage Db
1	0.9	0.00
2	0.99	-19.66
3	0.999	-39.92
4	0.9999	-59.99
5	0.99999	-80.00
6	0.999999	-100.00

Errors due to the tachometer are independent of errors due to interpolation. However, if we estimate the number of points needed per cycle, then we can compare the rate at which we must sample the tachometer. Let's say we need approximately 1000 points per cycle, thus we sample the signal at 10^3 f_0. To minimize leakage we need to sample the tachometer at 10^5 f_0, or two orders of magnitude faster than the signal.

The surprising result is that the tachometer needs to be sampled so much faster than the signal. Consider the case in which there is only one tachometer pulse per cycle. The simple approach of sampling and storing the tachometer at a rate of 10^5 per cycle yields only one piece of information for every 10^5 data points. A better method is to store the tachometer arrival times using specialized hardware.

4.0 A gearbox example

Here we show an example of synchronous averaging applied to a helicopter gearbox. The goal of synchronous averaging is to reduce all signals not occurring at multiples of the shaft rate. Suppose the signal is due to two parts, one at the shaft rate at amplitude A_1 and a second at a non integer multiple of this rate at amplitude A_2. The ratio of the levels of these two signals is

$$dB_1 = 10Log(A_1^2/A_2^2) \qquad (13)$$

Now let us synchronously add N samples of the signal. Because the signal is coherent, the amplitude at integer multiples of the shaft rate is $N*A_1$ and the level is proportional to $(N*A_1)^2$. The incoherent signal increases in

level as N * A_2^2. Thus the ratio of the levels of these two signals is

$$dB_N = 10Log(N) + dB_1 \qquad \textbf{(14)}$$

Thus the signal to noise ratio is increased by 10 log(N).

We apply our method to the main transmission of a helicopter gearbox. This gearbox is installed at a test facility. A set of accelerometers is mounted on the external housing of the gearbox. A tachometer with sixty pulses per cycle is mounted on the tail rotor take off shaft. Tachometer (125 Khz) and signal (62.5 Khz) data are collected simultaneously. All data channels have an anti-aliasing filter set at 20 kHz. For a 20 kHz filter sampling at 40 kHz corresponds to and over-sampling factor of 1.0. Thus the tachometer is over-sampled by a factor of 3.13 and the signal by a factor of 1.56

In figure 4 we select three of the principal shafts to illustrate the relation between shaft frequencies and gear mesh tones. The tachometer yields a signal at frequency f_t. The tail rotor take off shaft rotates at a rate $f_0 = f_t / 60$. If a single tooth is damaged then there is a disturbance at this rate. Even if no tooth is damaged there is still a disturbance each time a gear tooth meshes with the main bevel gear. This frequency is known as the gear-mesh frequency and is equal to the number of teeth times the shaft rate.

The ratio of shaft speeds is found by noting the gear mesh frequency of mating gears is the same. Thus the gear mesh tone between the main bevel gear and the tail rotor take off shaft is

$$GM_1 = 40\ f_1$$
$$85\ f_2$$

which gives

$$f_2 = (40/85)\ f_1.$$

In figure 5 we show the results of synchronous averaging for the tail rotor take off shaft. The data and the Fourier transform are shown for 1, 10 and 100 averages. The amplitude at the gear mesh frequency is the same for 10 and 100 averages. The amplitudes at other frequencies are diminished.

5.0 Conclusions

We develop a method for synchronously averaging data. Signal data and tachometer data are taken at a fixed rate unrelated to the shaft rate. Date are interpolated using a truncated low pass filter to provide samples at fixed intervals in shaft phase. Errors due to interpolation and inaccuracies in

the tachometer are discussed. The method is applied to the analysis of accelerometer data mounted on the housing of a helicopter gear box. The example shows the theoretical synchronous average gain in signal to noise is achieved.

6.0 References

Potter,"A new order tracking method for rotating machinery,Sound and Vibration,Sept,1990, pp 30-34.

Potter and Gribler,"Computer order tracking obsoletes older methods",SAE Noise and vibration conference, May 16-18,1989,pp 63-67.

Shannon, "Communications in the presence of noise", Proceedings of the IRE, January, 1949,pp 10-21.

Oppenheim and Wilsky,Signals and Systems,Prentice Hall, 1983,pp 514-526.

Crochiere, "Interpolation and Decimation", chapter 8,Programs for digital signal processing,IEEE Press, New York, 1979.

Parks and Burrus,Digital filter design,section 3.3,"Chebyshev approximation", John Wiley and Sons, New York, 1987,pp 83-106.

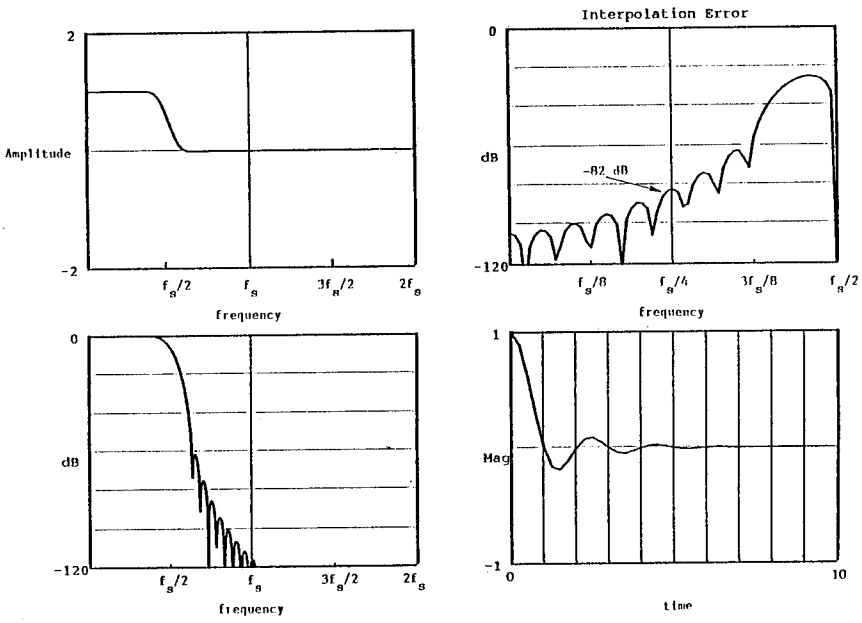


Figure 1. The truncated low pass filter

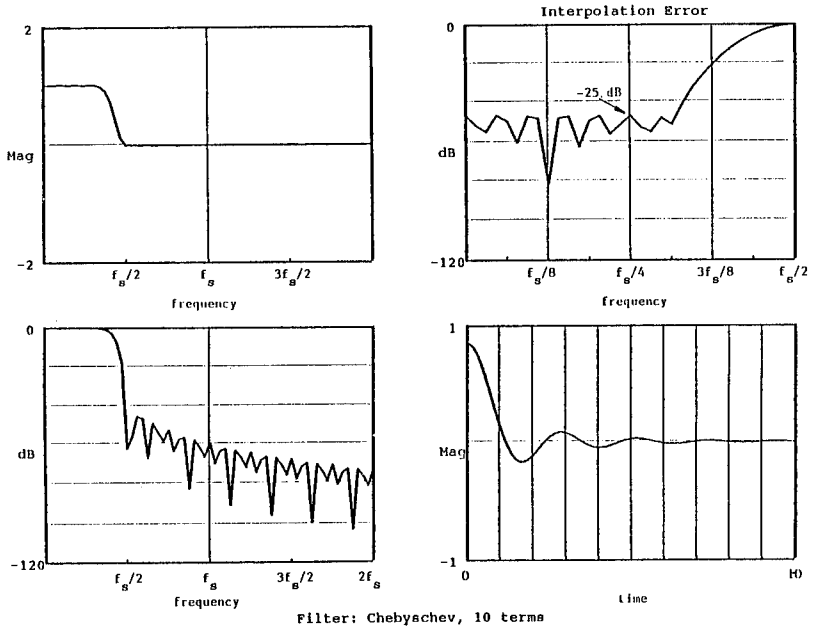


Figure 2. An "extended" digital low pass filter

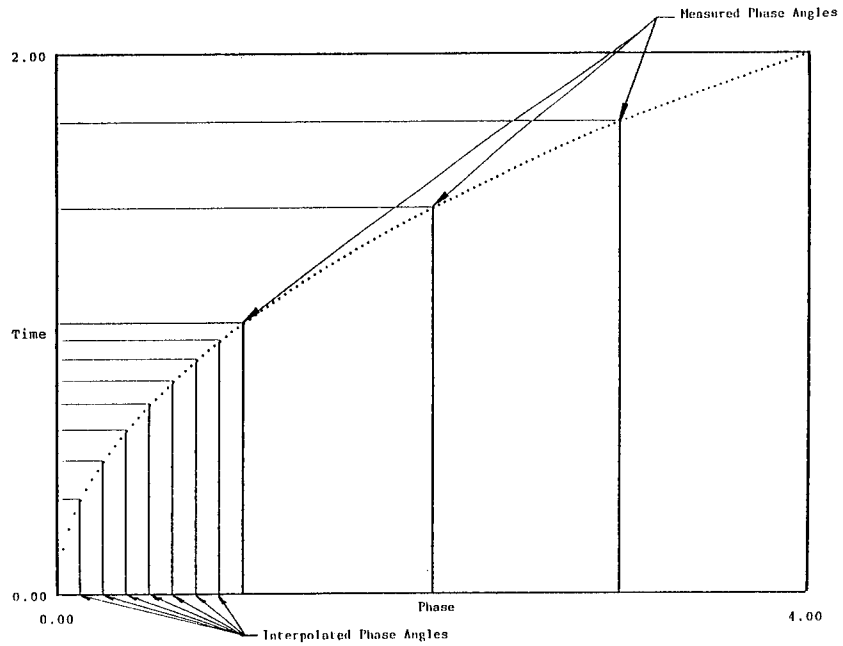
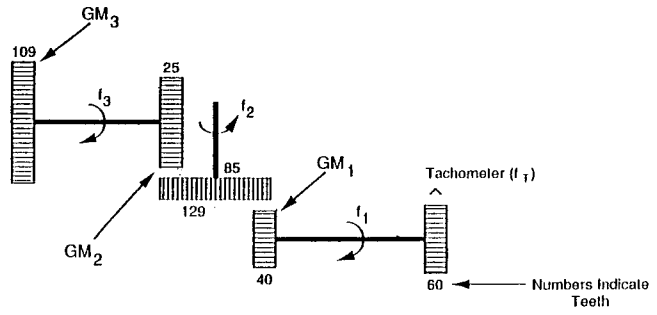


Figure 3. Method for determining equal increments in phase



RPM	Shaft Name	Shaft Frequencies	Gear Mesh Frequencies
3030	Main Input	$f_1 = f_T / 60$	$GM_1 = 40 \cdot f_1 = 129 \cdot f_2$
939	Main Bevel Gear	$f_2 = (40 / 129)\, f_1$	$GM_2 = 85 \cdot f_2 = 25 \cdot f_3$
3195	Tail Rotor Take Off	$f_3 = (85 / 25)\, f_2$	$GM_3 = 109 \cdot f_3$

Figure 4. Relation of gear-mesh tones to shaft frequencies

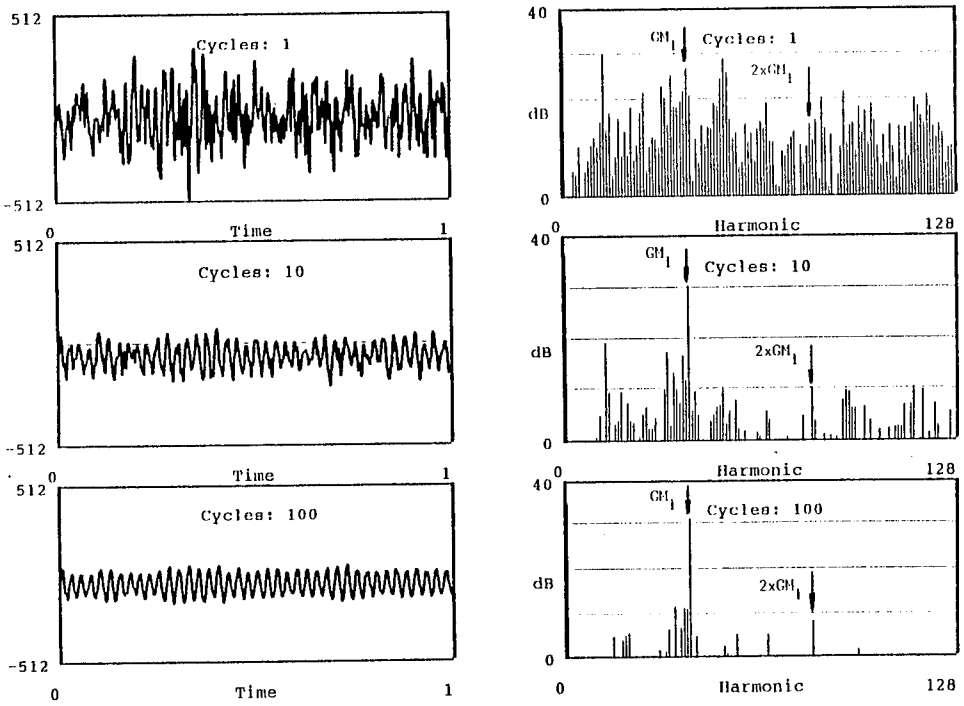


Figure 5. Synchronous averaging at the tail rotor take off shaft frequency

ENERGY OPERATOR AND OTHER DEMODULATION APPROACHES TO GEAR DEFECT DETECTION

Jun Ma

Department of Mechanical Engineering
New York Institute of Technology
Old Westbury, NY 11568

Abstract: *Defects on a gear alter the gear's vibration through amplitude and phase modulation. This provides opportunities for detection of gear defects without interfering with the normal operation of gearboxes, and considerable research has focused on uncovering modulation information from measured vibrations. In this paper, a new approach employing energy operator to characterize modulation energy is investigated for gear defect detection. Theoretical background of the energy operator is briefly reviewed, followed by an analysis of its application to gear vibration average. The effectiveness of the proposed approach is compared to other demodulation techniques, including residual signals, demodulation of bandpassed vibration, full-spectrum demodulation, and time-frequency analysis.*

Key Words: *Condition monitoring; Demodulation of gear vibrations; Energy operator; Gear defect detection; Signal analysis; Signal processing; Vibration analysis.*

INTRODUCTION: Prevention of mechanical failure has grown in importance as safer operation, lower manufacturing costs, and quality products have become essential. Recently, condition monitoring of gears has received considerable attention since a large percent (60%) of gearbox damages are attributed to gear faults, according to Allianz Versicherungs-AG (1978). In particular, gear localized defects have been extensively studied because gear faults are mostly initiated by localized defects. As revealed by the same study, three types of gear localized defects, namely, forced fracture, fatigue fracture, and incipient cracks, were respectively responsible for 56%, 17%, and 16% of gear faults found in gearboxes.

Defects on a gear alter amplitudes and phases of the gear's vibration. This provides opportunities for detection of gear defects without interfering with the normal operation of gearboxes, and thus has motivated most research to focus on vibration analysis techniques. Among all the vibration analysis methods for gear defect detection, demodulation approaches that uncover defect related modulation information from measured vibrations are regarded most effective. In this paper, traditional vibration analysis techniques for gear defect detection are summarized; the modulation phenomenon in gear vibration average is examined. Then, a new approach to gear defect detection employing energy operator to characterize modulation energy is investigated. Theoretical background of the energy operator is briefly reviewed, followed by an analysis of its application to gear vibration average. The effectiveness of the proposed approach is compared to other demodulation techniques, including residual signals, demodulation of bandpassed vibration, full-spectrum demodulation, and time-frequency analysis.

Traditional Methods: Vibrations externally measured on a gearbox have been used to monitor the health condition of the gearbox and diagnose the fault without interfering with the normal operation. The most common method employed for examining mechanical vibration is spectral analysis, by which defects such as eccentricity or local tooth damage are expected to be identified by increases of modulation sidebands in the spectrum. These sidebands are located on both sides of gear tooth meshing frequency and its harmonics, and are separated by integer multiples of gear rotational frequency. For example, the ratio of the sideband power to the carrier (tooth meshing frequency) power, called SBratio, was investigated (Dousis, 1986). However, the recognition of modulation sidebands is difficult due to the large number of gears rotating at different speeds in a gearbox. Moreover, the sensitivity of spectral sidebands to localized defects is low because modulations produced by a localized defect are transient events (Randall, 1982) and consequently, are inherently unsuitable for spectral analysis which assumes stationarity of signals.

In addition to spectral analysis, some statistical parameters, such as non-dimensionalized sixth moment (Astridge, 1986) and some other mean value based indices (Rose, 1990), have been established to assess the condition of gears from their vibrations. This type of technique usually has to be applied together with a trend analysis and thus the whole history of parameters must be monitored. Unfortunately, as the modulation of short duration does not change statistical properties of the overall vibration by much, they are frequently insensitive to localized defects, especially to incipient ones. Furthermore, most of these statistical parameters are dependent on the operation conditions such as load and speed, and consequently have limited practical uses.

Vibration Average: To a large extent, vibrations produced solely by a gear and its carrying shaft exhibit a repetitive pattern from one rotation to the next. Thus, there is a strong correspondence between time domain features of the vibration signal and angular positions of the rotating shaft. To obtain this angular position dependent signal from noisy measurements, the synchronized signal averaging (Braun and Seth, 1979), which has the effect of a comb filter, maybe employed. Vibrations of the gear of interest are sampled at the same angular positions for a large number of rotations, and samples corresponding to the same angular position are then averaged over the rotations. This widely adopted pre-processing procedure offers noise reduction and removal of interferences from components rotating at different speeds. The result, called signal average, is a function of angular position that shows the repetitive vibration pattern of tooth meshing, including any modulation, over one rotation.

For advanced gear defects, a simple visual inspection on the signal average maybe sufficient to detect the damage. Fig. 1 shows the signal average that is to be used throughout this paper as a standard signal for evaluation of the energy operator method and for comparison with other demodulation methods. This signal average was obtained over 200 rotations of a 20-tooth input pinion in a three-stage reduction gearbox. There are two artificially seeded defects of different size at teeth 11 and 18 of the pinion, respectively, simulating fractured teeth at different stage. In Fig. 1 it is plotted as a function of angular position (tooth number), together with its spectrum indexed by the rotational frequency (order). It is obvious that the advanced fracture on tooth 11 can be immediately detected from the signal average even by naked eyes. The larger amplitude of vibration around that tooth indicates reduced meshing stiffness and thus more violent vibration. However, the smaller defect on tooth 18 has little noticeable effect on the signal average, registering the challenge of detection of such defects.

The detection of defects at a very early stage usually requires sophisticated signal detection and processing techniques to enhance the defect information contained in the signal average. Considering the modulation nature of defect related vibration, a number of gear defect detection techniques use amplitude modulation and/or phase modulation as defect signature. In general, methods looking into modulations are most sensitive to gear localized defects. In the following, the modulation phenomenon in gear vibration is examined and modeled. Then, energy operator is introduced to track modulation energy. Its effectiveness in gear localized defect detection is examined and compared to other demodulation schemes, using measured vibration average shown in Fig. 1.

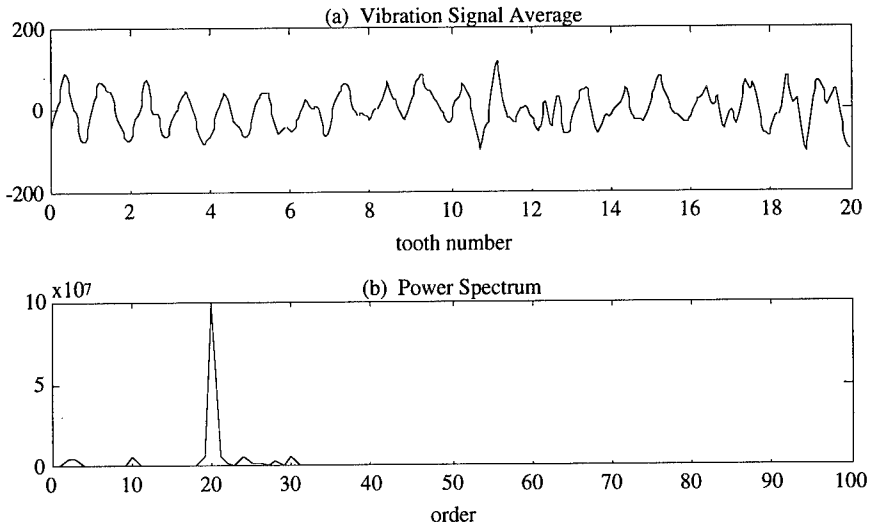


Fig. 1 Vibration signal average and its spectrum

MODULATIONS IN GEAR VIBRATIONS: Imagine a pair of perfect mating gears whose teeth are equally spaced and rigid with exact involute profile. Such a pair of gears would transmit exactly uniform angular motion in the absence of distributed defects such as runout, imbalance, and misalignment. Any deviation from this ideal situation will cause variations in both angular displacement and velocity, i.e., the transmission error, which in turn results in variations in the force transmitted between the meshing teeth. Vibrations will then be generated and transmitted everywhere in and on the gearbox through the gear-bearing-shaft-bearing-casing path.

Now consider a pair of gears whose teeth are not rigid but otherwise the same as the aforementioned perfect gears, meshing under constant load and running at constant speed. Since the contact stiffness varies periodically with the number of teeth in contact and with the contacting position on tooth surface, vibration will be excited at tooth meshing frequency and its harmonics. After being synchronously averaged, the vibration of this pair of gears may be approximately represented in terms of tooth meshing frequency f_m and its harmonics:

$$x_t(t) = \sum_{k=0}^{K} X_k \cos(2\pi k f_m t + \phi_k) \qquad (1)$$

As mentioned in the previous section, the vibration average, or the vibration of a single gear, can be viewed as a function of angular position θ. Therefore, we shall express the vibration of a gear with N teeth as:

$$x(\theta) = \sum_{k=0}^{K} X_k \cos(kN\theta + \phi_k) \tag{2}$$

Then let us take into account the tooth profile error, tooth spacing error, and defects. All these will vary contact stiffness and therefore produce changes in the amplitude and phase of the vibration at the meshing frequency and its harmonics as they go through the meshing. These changes are described by the amplitude and phase modulating functions ($a_k(\theta)$ and $b_k(\theta)$, for the k-th meshing harmonic), then the modulated vibration, or the measured vibration average, is modeled as

$$y(\theta) = \sum_{k=0}^{K} [1 + a_k(\theta)] X_k \cos[kN\theta + \phi_k + b_k(\theta)] \tag{3}$$

Based on this modulation model of vibration average, various methods have been developed to detect gear defects. It can be seen that each tooth meshing harmonic functions as a carrier in both amplitude and phase modulations; the measured vibration average is a superposition of modulated tooth meshing harmonics. Without any model simplification, it is impossible to recover modulation signals $a_k(\theta)$ and $b_k(\theta)$ from the vibration average $y(\theta)$, even from a single component AM-FM signal (no superposition). Some demodulation schemes (e.g., McFadden, 1986; Ma and Li, 1994) were designed based on certain assumptions about the modulation signals that led to simplified versions of the general model (3). However, modulations vary instantaneous frequency contents and/or energy concentration of gear vibrations. Therefore, localized defects can be detected if one can find any localized signs of modulation, namely, short lasting high level modulations. The time-frequency analysis (e.g., Forrester, 1990; Wang and McFadden, 1993; Ma, 1994) was such an attempt, and was successful. The energy operator method presented here is another effective solution, but simpler.

ENERGY OPERATOR APPROACH: Energy operator was developed by Teager (1980) during his work on speech production modeling and was first introduced systematically by Kaiser (1990a and 1990b). It is a simple nonlinear "energy-tracking" operator, given in its discrete form for sampled signals $y(n)$ as

$$\Psi_d[y(n)] = y^2(n) - y(n-1)y(n+1) \tag{4}$$

or in its continuous form for continuous-time signals $y(t)$ as

$$\Psi_c[y(t)] = \left(\frac{dy(t)}{dt}\right)^2 - y(t)\left(\frac{d^2y(t)}{dt^2}\right) \tag{5}$$

Such an operator, when applied to AM-FM signals, can approximately track the product of their amplitude and their instantaneous frequency (Maragos, et al, 1993). Specifically, in discrete-time form,

$$\Psi_d \approx [1+a(n)]^2 \cdot \sin^2\left[\omega + \frac{db(n)}{dn}\right] \tag{6}$$

for $$y(n) = [1+a(n)]\cos[\omega n + b(n)] \tag{7}$$

or, in continuous-time form,

$$\Psi_c \approx [1+a(t)]^2 \cdot \left[\omega + \frac{db(t)}{dt}\right]^2 \tag{8}$$

for $$y(t) = [1+a(t)]\cos[\omega t + b(t)] \tag{9}$$

This analysis is for single component AM-FM signals, thus only provides a theoretical basis for applying energy operator to bandpassed vibration average while vibration average of a gear is a sum of many AM-FM signals, containing a family of harmonics that are amplitude and phase modulated by the gear's imperfections including defects. Bandpassed vibration average, however, has lost its high frequency part of modulation signal to which the energy operator is more sensitive. While it is intricate to analytically express and study the energy operator for multicomponent modulated signals, a direct application of the energy operator to signals from a gear of known condition is a simple and pragmatic way to explore its effectiveness in condition monitoring.

Both the vibration average shown in Fig. 1 and its bandpassed version were used in the following. The energy operator was first applied to the bandpassed vibration average, which had been bandpass filtered about the tooth meshing frequency. The result is presented in Fig. 2. The defect on tooth 11 was marked by high modulation "energy". For the smaller defect, the energy operator did indicate an increase in modulation "energy" around tooth 18, but to give a warning merely according to this would yield a high rate of false alarm (how about tooth 1, for example).

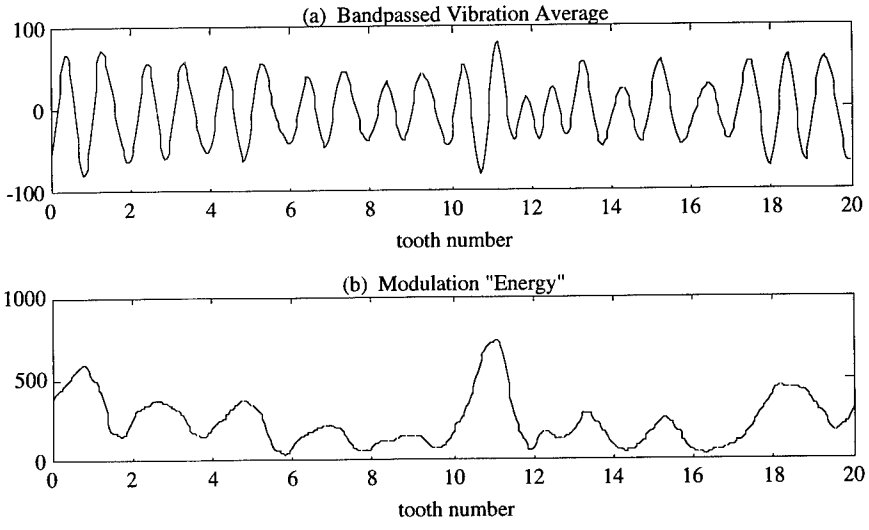


Fig. 2 Bandpassed vibration average and its modulation energy by energy operator

When directly applied to the vibration average, the energy operator gave a better result as expected. A careful inspection on the result shown in Fig. 3 reveals that the modulation "energy", as assessed by energy operator, fluctuated at tooth meshing frequency, which is believed attributed mostly to the amplitude modulation. The abrupt rise of modulation "energy" at around tooth 11 and tooth 18 signified strong modulation of short-lasting nature, suggesting the existence of two localized defects.

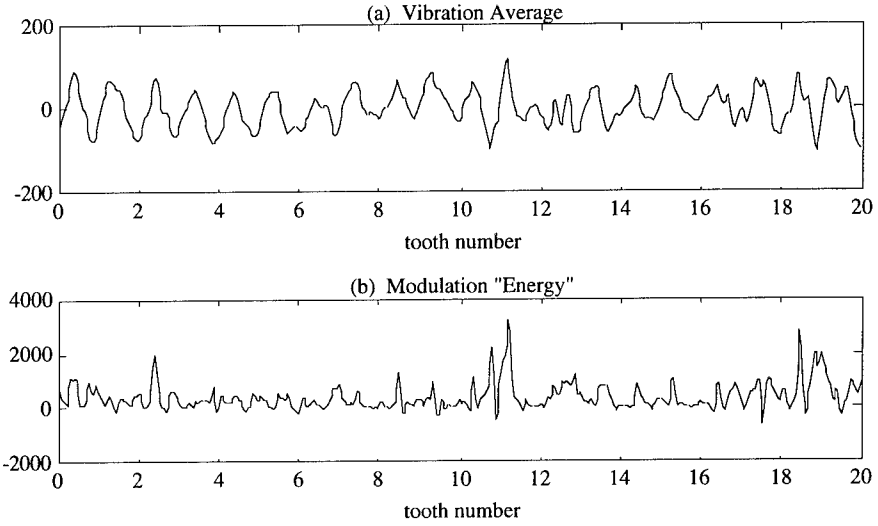


Fig. 3 Vibration average and its modulation energy by energy operator

OTHER DEMODULATION APPROACHES: In the following, other demodulation approaches are briefly reviewed, including the principles and procedures. Their effectiveness in gear localized defect detection is also evaluated using the same vibration average for comparison.

Residual Signal: The spectrum of the unmodulated vibration as expressed in Eq. (2) consists of a series of tooth meshing harmonics only. Once modulated (see Eq. (3)), these harmonics are surrounded by modulation sidebands. Since each harmonic acts as a modulation carrier and thus has its own sidebands, there will be interference between sidebands of neighboring meshing harmonics. In addition, meshing harmonics themselves are also affected by the high-order sidebands of adjacent meshing harmonics. If the vibration average is bandpass filtered about one of the meshing harmonics (usually the largest one, say, the k-th), provided that the amplitude of that harmonic is much greater than its neighbors, the interference from the sidebands of the neighboring harmonics will be negligible (McFadden and Smith, 1985). The filtered signal can be approximated by

$$q_k(\theta) = \left[1 + a_k(\theta)\right] X_k \cos\left[kN\theta + \phi_k + b_k(\theta)\right] \tag{10}$$

Further, McFadden and Smith (1985) showed that if $|a_k(\theta)| < 1$, $|b_k(\theta)| << \pi$, and $|a_k(\theta) b_k(\theta)| << 1$, then the residual signal, obtained by eliminating the meshing harmonic from the filtered signal, may be approximated by

$$r_k(\theta) = X_k\sqrt{a_k^2(\theta)+b_k^2(\theta)}\cos\left[kN\theta+\phi_k-\tan^{-1}\frac{b_k(\theta)}{a_k(\theta)}\right] \tag{11}$$

The envelope of this signal

$$s_k(\theta) = X_k\sqrt{a_k^2(\theta)+b_k^2(\theta)} \tag{12}$$

is directly related to both amplitude and phase modulation; it represents the total modulation "strength" (amplitude and phase) as a function of time (or angular position of the gear).

Fig. 4 shows the residual signal and its envelope. The defect on tooth 11 is obviously detectable using either of them, indicated by the singular high level around tooth 11 in the residual signal and its envelope as well. However, it is quite ambiguous whether there is a defect on tooth 18 as both signals appeared at tooth 18 in a similar fashion as at neighboring teeth.

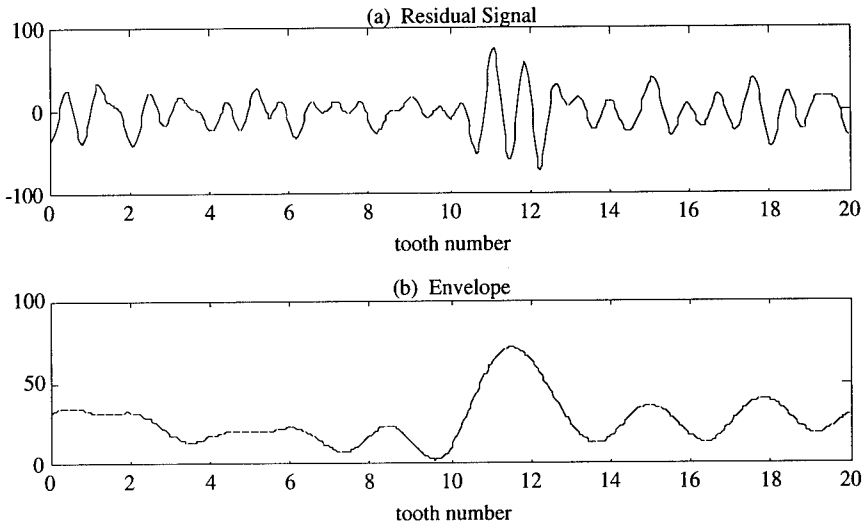


Fig. 4 Residual signal and its envelop

Demodulation of Bandpassed Signal Average: A further study by McFadden (1986) tried to demodulate the signal average around the largest meshing harmonic to produce amplitude and phase signals. Again, the signal average is first bandpass filtered about the largest meshing harmonic. This filtered signal, expressed in Eq. (10) as a cosine function, can be viewed as the real part of a complex function $c_k(\theta)$, known as the analytic signal and defined by:

$$c_k(\theta) = q_k(\theta) - j\mathbf{H}(q_k(\theta)) \tag{13}$$

where j is the complex numeric, and $\mathbf{H}(q_k(\theta))$ is the Hilbert transform of $q_k(\theta)$. When $a_k(\theta)$ and $b_k(\theta)$ are narrow-band signals containing no components with frequency

higher than the tooth meshing frequency, this complex signal can be well approximated by (Rihaczek, 1966)

$$c_k(\theta) = [1 + a_k(\theta)] X_k \exp\{j[kN\theta + \phi_k + b_k(\theta)]\} \tag{14}$$

Noting that the magnitude and phase of $c_k(\theta)$ are dependent on $a_k(\theta)$ and $b_k(\theta)$, respectively, estimates of $a_k(\theta)$ and $b_k(\theta)$ can be obtained by

$$a_k(\theta) = \frac{\|c_k(\theta)\|}{X_k} - 1 \tag{15}$$

$$b_k(\theta) = \arg[c_k(\theta)] - (kN\theta + \phi_k) \tag{16}$$

This method of demodulation is also refereed as Hilbert transform approach. Theoretically, X_k and ϕ_k, respectively the Fourier magnitude and phase of the *unmodulated* meshing harmonic, should be used in Eqs. (15) and (16) to estimate $a_k(\theta)$ and $b_k(\theta)$. However, the magnitude and phase of a modulation carrier are altered by phase modulation (Ma and Li, 1994). Therefore, one has to know $b_k(\theta)$ to determine X_k and ϕ_k from the measured vibration average while $b_k(\theta)$ is what one wishes to know and yet to be determined. An iterative procedure may be developed to resolve this conflict, starting with an initial guess. Nevertheless, it is not a crucial issue in this case since X_k and ϕ_k are two constants and do not change the shape of $a_k(\theta)$ and $b_k(\theta)$. We used the Fourier magnitude and phase of the *modulated* meshing harmonic to approximate the true values.

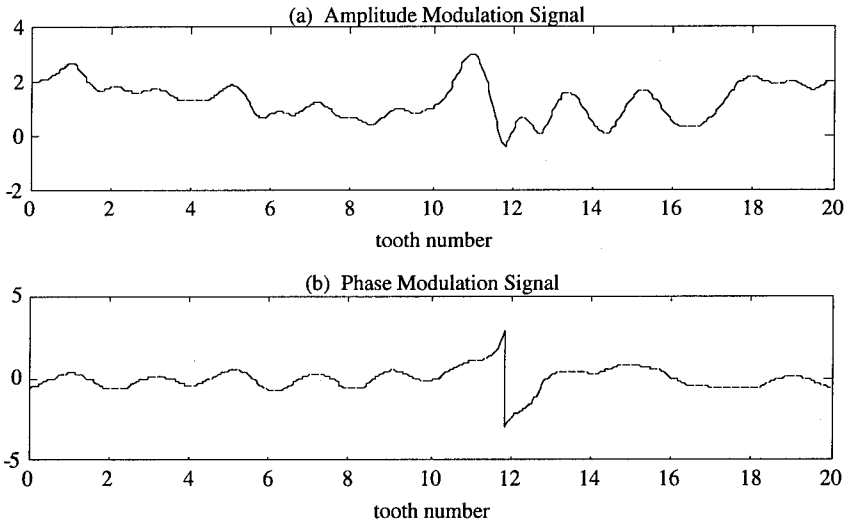


Fig. 5 Results by bandpass-demodulation method

Fig. 5 shows the amplitude and phase modulation signals determined using the bandpass demodulation method. Again, the defect on tooth 11 is detected, as indicated by the strong local modulation in both amplitude and phase. The abrupt change in phase modulation signal, limited within $[-\pi, \pi]$ in the plot, seems closer to tooth 12 than to tooth 11. Such an error in defect location in the phase modulation signal is likely induced by the use of incorrect ϕ_k, that results in the 360-degree shift point slightly to the right. Further, this method also failed to detect the smaller defect on tooth 18.

Full-spectrum Demodulation: To make full use of information and to achieve a better detection sensitivity, the full-spectrum demodulation method (Ma and Li, 1994), as opposed to the narrowband approaches, takes all tooth meshing harmonics into account as modulation carriers, and recovers amplitude and phase modulation signals from the wide band signal. The modulation model of gear vibration described in Eq. (3) was simplified based on the following considerations. First, the amplitude modulating signal was assumed independent of tooth meshing harmonics since it modifies the envelope of $x(\theta)$. Second, the phase modulating signal at any given instant should produce a same time delay in all tooth meshing harmonics, thus can be modeled as linear with respect to the harmonic number k. Therefore the vibration average of the gear of interest can be expressed as

$$y(\theta) = [1 + a(\theta)] \sum_{k=0}^{K} X_k \cos[kN\theta + \phi_k + kb(\theta)] \tag{17}$$

Based on this simplified modulation model, the full-spectrum demodulation method employs an iterative procedure to uncover from $y(\theta)$ the amplitude and phase modulation signals. For narrowband $a(\theta)$ and $b(\theta)$, the analytic signal of $y(\theta)$ is approximated by

$$y_a(\theta) = [1 + a(\theta)] \sum_{k=0}^{K} X_k \exp\{j \cdot [kN\theta + \phi_k + kb(\theta)]\} \tag{18}$$

Let

$$z(\theta) = \sum_{k=0}^{K} X_k \exp\{j \cdot [kN\theta + \phi_k + kb(\theta)]\} \tag{19}$$

Then Eq. (18) may be rewritten for the amplitude modulation signal as

$$a(\theta) = \frac{\|y_a(\theta)\|}{\|z(\theta)\|} - 1 \tag{20}$$

This rearrangement separates $a(\theta)$ and $b(\theta)$ because $b(\theta)$ can be first estimated using the condition that phase angles of $y_a(\theta)$ and $z(\theta)$ are equal. Since $z(\theta)$ is dependent on $b(\theta)$, both directly and through X_k and ϕ_k, the Fourier magnitudes and phases of unmodulated meshing harmonics, an iterative procedure has to be utilized. Specifically, with an initial guess of phase modulation signal $b(\theta)$, X_k and ϕ_k are reconstructed, and so is $z(\theta)$. Then an optimization subroutine is commanded to

search for a better $b(\theta)$ that reduces the difference between the phase angles of $y_a(\theta)$ and $z(\theta)$. Once a satisfactory $b(\theta)$ is found, $a(\theta)$ can be obtained using Eq. (20).

Fig. 6 shows the result of the full-spectrum demodulation scheme applied to the gear vibration average. The initial $b(\theta)$ employed was the result of bandpass demodulation method (see Fig. 5b). To reduce the computational burden, only low frequency part of the phase modulation signal was included (see Ma and Li, 1994 for details), and consequently, the amplitude modulation signal became a better indication of gear condition. In fact, gear defects on teeth 11 and 18 were obvious as the amplitude modulation signal appeared strong on these two teeth.

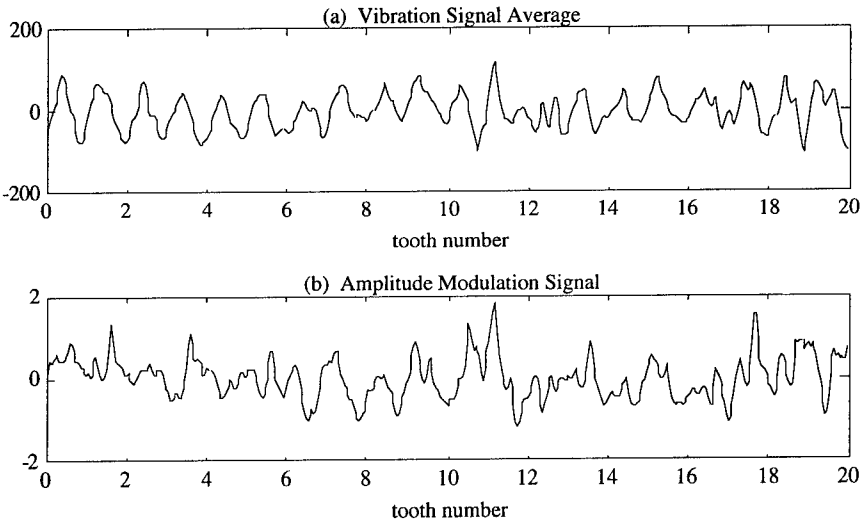


Fig. 6 Vibration signal average and amplitude modulation signal

Time-frequency Analysis: Imperfections on a gear vary instantaneous energy and frequency contents of the gear's vibration through modulations; for localized defects, the resulting variations are local in time, i.e., they are short-lasting and not a common phenomenon to normal teeth of the gear. Spectral analysis can detect energy and frequency fluctuations by the presence of sidebands, but can not tell when they occurred or whether the sidebands are just marks of other additive, stationary components at frequencies slightly different from meshing harmonics. On the other hand, time-frequency analysis is an ideal tool for analysis of signal components with time and/or locality.

High resolution in both time and frequency domain is required for the detection of gear localized defects, especially incipient ones because they usually do not result in significant fluctuations in instantaneous energy and frequency contents of vibrations. This limits the application of spectrogram or short-time Fourier analysis, where time and frequency resolutions are uniquely bounded by widow size. Different time-frequency distributions have been applied to gear vibrations (e.g., Forrester, 1990; Wang and McFadden, 1993; Ma, 1994). While they are capable of revealing the distribution of signal energy in the joint time and frequency domain, the inherent interfering cross terms may become very disturbing if one wishes to use time-

frequency features for pattern recognition and classification. This is especially true in detection of incipient gear defects where cross terms may dominate the distribution and overshadow defect signature.

By using an exponential kernel function, Choi and Williams (1990) proposed a distribution that minimizes the spurious cross terms while inducing desirable properties. The so-called Choi-Williams distribution is given by

$$D_y(t,\omega)=\frac{1}{2\pi}\iiint e^{-j(\theta t+\tau\omega-\theta u)}K(\theta,\tau)y^*(u-\frac{\tau}{2})y(u+\frac{\tau}{2})dud\tau d\theta \qquad (21)$$

where the kernel function is

$$K(\theta,\tau)=e^{-\theta^2\tau^2/\sigma} \qquad (22)$$

with scaling factor $\sigma>0$. Large scaling factor gives better resolution (suitable for signals with fast changing amplitude and frequency) while small scaling factor results in greater cross term attenuation (suitable for signals with slowly changing amplitude and frequency). More properties of Choi-Williams distribution, as well as comparison with others, can be found in review papers (e.g. Cohen, 1989; Hlawatsch and Boudreaux-Bartels, 1992).

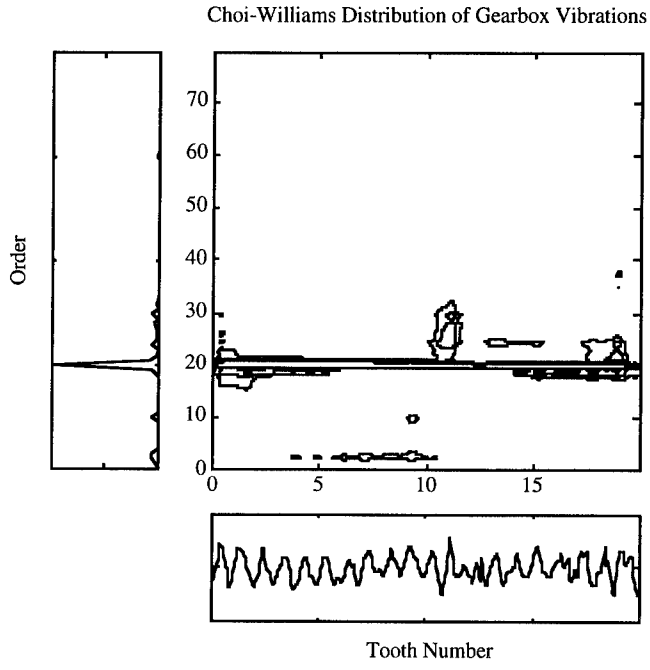


Fig. 7 Choi-Williams Distribution of Gear Vibration Average

The Choi-Williams distribution of the vibration average was calculated (σ=1.0) and the contour of the distribution is presented in Fig. 7. The signal average and its spectrum are also shown in the figure (below and on the left, respectively). It can be

observed from either the spectrum or the Choi-Williams distribution that vibration energy concentrated mostly at the tooth meshing frequency (the 20th order), but the spectrum could not tell whether the sidebands were due to stationary components or marks of defects. The time-frequency distribution, however, revealed that energy and frequency of the vibration around teeth 11 and 18 were more deviated from the tooth meshing frequency, indicating the existence of the localized defects in corresponding locations. The size of defects may also be inferred from the degree of energy and frequency deviation.

COMPARISON AND CONCLUSIONS: In this paper, the energy operator is employed to evaluate the modulation "energy" in gear vibrations for gear localized defect detection. It is applied to a vibration average from a 20-tooth gear, containing two defects of different size, as well as the bandpassed version of the vibration average. The vibration average and its power spectrum are plotted in Fig. 1. The larger defect, on tooth 11, simulates a defect of advanced stage and the one on tooth 18 is considered of early stage. Other demodulation approaches to gear localized defect detection are then reviewed and applied to the same vibration average. All results are summarized in Table 1.

Table 1 Comparison of Demodulation Methods

Methods	*Signal Preprocessing*	*Defect(s) Detected*	*Computing Power Required*
Energy Operator	B.P. filtering	the large one	low
Energy Operator	none	both	low
Residual Signals	B.P. filtering	the large one	low
Demod. of B.P. Signal	B.P. filtering	the large one	low
Full-Spectrum Demod.	none	both	high
Time-Frequency Analysis	none	both	high

From the comparison, it is clear that all methods that use bandpassed vibration signal average yield low sensitivity to the small defect — none of them gave an unambiguous sign about the small defect. This is obviously due to the information loss or misinterpretation caused by bandpass filtering that removes high frequency part of amplitude and phase modulation signals. In addition, energy operator requires much less computation power than do the full-spectrum demodulation and time-frequency analysis, which also successfully detected both defects. As a matter of fact, the implementation of energy operator is simple and straightforward (see Eq. (4)), making it a perfect tool for on-line applications.

The modulation "energy" by energy operator is generally consistent with the amplitude modulation signal estimated using the bandpass demodulation method (see Fig. 2b and Fig. 5a for comparison), or that estimated using full-spectrum demodulation (see Fig. 3b and Fig. 6b for comparison). Although it may be premature to say that the amplitude modulation reflects the gear condition better than phase modulation signal does, it is certainly the case here. Further, since both amplitude and phase modulations contribute to the modulation "energy", energy operator yields more discriminating power to gear imperfections, thus provides a clearer indication of localized defects.

In summary, this study has found that energy operator has a successful application in gear localized defect detection. Compared to other demodulation methods, energy operator is sensitive to small defects, computationally simple, and more suitable for on-line implementation.

REFERENCES

Allianz Versicherungs-AG, 1978, *Handbook of Loss Prevention*, Springer-Verlag, Berlin

Astridge, D. G., 1986, "Vibration Health Monitoring of the Westland 30 Helicopter Transmission — Development and Service Experience," *Detection, Diagnosis and Prognosis of Rotating Machinery to Improve Reliability, Maintainability, and Readiness through the Application of New and Innovative Techniques,* Proceedings of the 41st Meeting of the Mechanical Failures Prevention Group, edited by T. R. Shives and L. J. Mertaugh, Cambridge University Press, New York, pp. 200-215

Braun, S., and Seth, 1979, "On the Extraction and Filtering of Signals Acquired from Rotating Machines," *Journal of Sound and Vibration*, Vol. 65, No. 1, pp. 37-50

Choi, H.-I., and Williams, W. J., 1989, "Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernels," IEEE *Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 6, pp. 862-871

Cohen, L., 1989. "Time-frequency Distributions—A Review," *Proceedings of the IEEE*, Vol. 77, No. 7, pp. 941-981

Dousis, D. A., 1986, "Gear Failure Analyses in Helicopter Main Transmissions Using Vibration Signature Analysis," *Detection, Diagnosis and Prognosis of Rotating Machinery to Improve Reliability, Maintainability, and Readiness through the Application of New and Innovative Techniques,* Proceedings of the 41st Meeting of the Mechanical Failures Prevention Group, T. R. Shives and L. J. Mertaugh ed., Cambridge University Press, New York, pp. 133-144

Forrester, B. D., 1990, " Analysis of Gear Vibration in the Time-Frequency Domain," *Current Practices and Trends in Mechanical Failure Prevention*, Proceedings of the 44th Meeting of the Mechanical Failures Prevention Group, edited by H. C. Henry and S. C. Henry, Vibration Institute, Willowbrook, NJ, pp. 225-234

Hlawatsch, F., and Boudreaux-Bartels, G. F., 1992, "Linear and Quadratic Time-frequency Signal Representations," *IEEE Signal Processing Magazine*, pp. 21-67, April 1992

Kaiser, J. F., 1990a, "On a Simple Algorithm to Calculate the 'Energy' of a Signal," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, NM, pp. 381-384

Kaiser, J. F., 1990b, "On Teager's Energy Algorithm and Its Generalization to Continuous Signals," *Proceedings of 4th IEEE Digital Signal Processing Workshop*, Mohonk (New Paltz), NY

Ma, J., 1994, *Advanced Signal Processing Approaches to Condition Monitoring of Gearboxes*, Ph.D. thesis, Columbia University, New York, NY

Ma, J. and Li, C. J., 1994, "A New Approach to Gear Vibration Demodulation and Its Application to Defect Detection," *Advanced Materials and Process Technology*

for Mechanical Failure Prevention, Proceedings of the 48th Meeting of the Mechanical Failures Prevention Group, Compiled by H. C. Pusey and S. C. Pusey, Vibration Institute, IL, pp. 43-55

Maragos, P., Kaiser, F., and Quatieri, T. F., 1993, "On Amplitude and Frequency Demodulation Using Energy Operators," *IEEE Transactions on Signal Processing*, Vol. SP-41, No. 4, pp. 1532-1550

McFadden, P. D., 1986, "Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration," *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 108, No. 2, pp. 165-170

McFadden, P. D. and Smith, J. D., 1985, "A Signal Processing Technique for Detecting Local Defects in Gear from the Signal Average of the Vibration," *Proceedings of Institute of Mechanical Engineers*, Vol. 199, No. C4, pp. 287-292

Randall, R. B., 1982, "A New Method of Modeling Gear Faults," *Journal of Mechanical Design,* Vol. 104, pp. 259-267

Rihaczek, A. W., 1966, "Hilbert Transforms and the Complex Representation of Real Signals," *Proceedings of IEEE*, vol. 54, pp. 434-435

Rose, H. J., 1990, "Vibration Signature and Fatigue Crack Growth Analysis of a Gear Tooth Bending Fatigue Failure," *Current Practices and Trends in Mechanical Failure Prevention,* Proceedings of the 44th Meeting of the Mechanical Failures Prevention Group, H. C. Pusey and S. C. Pusey ed., Vibration Institute, Willowbrook, IL, pp. 235-245

Teager, H. M., 1980, "Some Observations on Oral Air Flow During Phonation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-28, No. 5, pp. 599-601

Wang, W. J., and McFadden, P. D., 1993, "Early Detection of Gear Failure by Vibration Analysis — I. Calculation of the Time-Frequency Distribution," and "— II. Interpretation of the Time-Frequency Distribution Using Image Processing Techniques," *Mechanical Systems and Signal Processing*, Vol. 7, No. 3, pp. 193-203, and pp. 205-215

AUTOMATED BEARING FAULT DETECTION WITH UNSUPERVISED NEURAL NETS

Hsinyung Chin
Technology Integration, Inc.
54 Middlesex Turnpike
Bedford, MA 01730

Abstract: This paper presents the results of a newly developed fault detection method for early detection of bearing failures for the Space Shuttle Main Engine (SSME) turbomachinery. The proposed method, which uses an *Adaptive Feature Map (AFM)* as its detection model, has the ability to detect subtle changes in a signal. The AFM method automatically performs pattern classification on a set of input data and updates its weights without supervision (i.e., training is not required). This is particularly important for cases where a large amount of experimental data is not available and/or the operating condition is not constant. The proposed method can also be used to determine the region of a signal where most of the changes occur. Conventional high frequency envelope analysis can then be applied to reveal bearing tones once such region is identified. Results of this study show that the proposed AFM method is quite effective in detecting bearing faults at their earliest stages for a number of SSME test firings.

Key Words: Detection; diagnosis; pattern classification; space shuttle main engine; unsupervised learning; vibration signal processing

Introduction:

Early detection of bearing faults is important for preventing catastrophic breakdowns of rotating machinery, especially for critical components like the Space Shuttle Main Engine (SSME) turbopumps. Mechanical failure of these components must be detected at the earliest stages of development such that immediate actions can be taken to prevent catastrophic failures during pre-flight testing or during an actual space shuttle flight.

Among the various components in the SSME turbopump, the bearings used to support the turbopump shaft are usually more prone to failures due to high rotational speeds of the shaft. The predominant mode of turbopump bearing failure is generally believed to be ball wear which should be promptly detected so as to prevent further bearing deterioration. Since the SSME turbopump is operated under extreme conditions (i.e., high temperatures, fluid pressures and rotational speeds), bearing fault detection is further compounded by the severe noise and vibration contamination from the fluid flow, acoustic and structural resonances and combustion [8].

In general, traditional bearing fault detection is achieved by monitoring the vibrations generated from a rotating machinery and by analyzing the vibration signals

with techniques such as spectrum analysis. However, the vibration energy of bearing elements is usually lower than those produced by gears, shafts, and sometimes noise. As such, bearing faults cannot be readily detected through inspection of bearing tones within a spectrum [12].

Since bearing faults such as spalling usually produce time domain impulses which modulate the shaft speed over a wide range of frequencies, there are features of high frequency vibration that would reflect bearing faults. Envelope analysis [5] which is designed to enhance bearing tones in the high frequency region is used to detect the periodic impulses by demodulating the vibration in a narrow frequency band about a structural resonance. The time period between periodic impulses gives an indication of the location of the initial bearing damage by relating it back to the shaft speed and the bearing geometry. One difficulty of envelope analysis is that although it can enhance bearing tones, it is usually quite time consuming to perform continuous enveloping on a signal. Furthermore, it usually produces false alarms and/or undetected faults when the actual location of structural resonance is not available or inaccurate.

Advanced signal processing techniques such as *Wavelet Transform* [11] and *Hyper-Spectrum Analysis* [9] have been applied to bearing fault detection. However, these techniques usually require a long processing time and careful interpretation of the results. As such, they are usually not suitable for in-flight application.

Recently, techniques based on pattern classification have been applied to fault detection and diagnosis [3,4]. Among the various pattern classifiers used for detection, artificial neural nets are the most notable due to their nonparametric nature (independence of the probabilistic structure of the system), and their ability to generate complex decision regions and to perform non-linear interpolation [7]. The application of neural nets to SSME fault detection/diagnosis has recently been proposed [6]. In [6], a supervised neural net trained with residuals generated from a simulation model was used to detect various SSME faults. Although this approach is quit effective in detecting various SSME valve failures, it is at the mercy of the simulation model to generate a large amount of residuals for training. In cases where a reliable model is not available to generate training data, one should resort to unsupervised neural nets.

In this paper, a new automated bearing fault detection method for SSME turbopumps is introduced. This method, which is based on pattern classification of vibration data, uses an *Adaptive Feature Map (AFM)* as its detection model and has the ability to detect subtle changes in a signal. The proposed AFM method which relies on a fast unsupervised learning algorithm to adapt its detection model, does not require any training. This is particularly important for cases where a large amount of experimental data is not available and/or the operating condition is not constant. In this method, time domain vibration data is first transformed to frequency spectra via Fast Fourier Transform (FFT). These spectra are then used by the AFM as inputs for bearing fault detection. The proposed method is ideally suited to in-flight application, due to its simple detection strategy and fast learning.

To test the applicability of the proposed AFM method in bearing fault detection, vibration data collected from two newly designed SSME turbopumps during various test firings at NASA Marshall Space Flight Center was used. Results of this application show that the AFM method provides early and reliable detection of bearing faults for both cases. For comparison, envelope analysis was also performed on the same sets of data. It was shown that the proposed AFM method is in fact more sensitive to early bearing faults than envelope analysis is. The AFM method also identified the frequency bandwidth of 14-20 kHz as the region containing most of the bearing fault information for the two SSME turbopumps.

SSME Turbopump Experiments:

The SSME turbopump rotates at high speeds of up to 500 Hz on two sets of bearings, one at the pump-end and one at the turbine-end (see Fig. 1). The actual bearing design has undergone numerous changes over a number of years trying to extend the bearing life, and new bearing designs are still being tested. It has been found that the majority of bearing failures are caused by the ball wear of pump-end bearings, due to the turbopump mechanical condition.

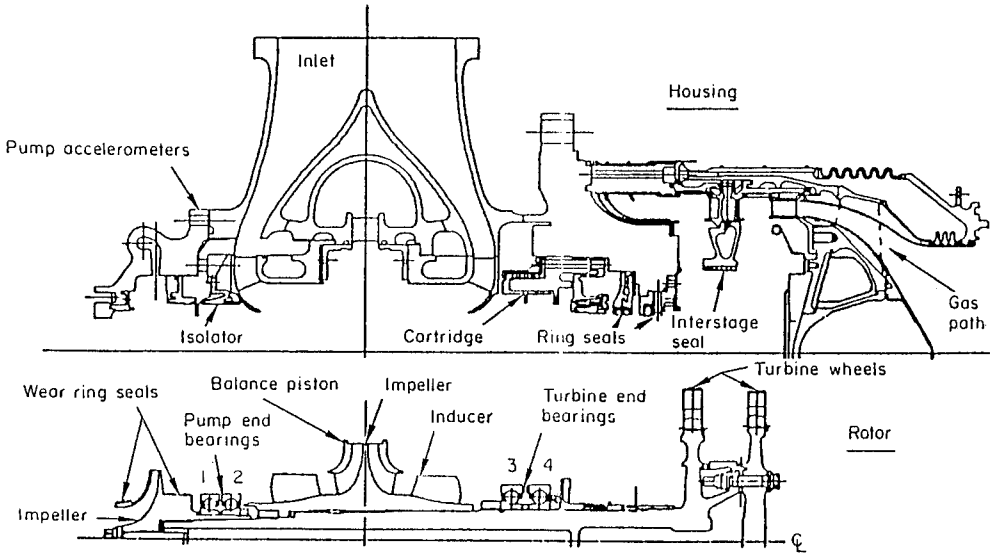


Figure 1: Configuration of a typical SSME turbopump.

Vibration data was collected at NASA Marshall Space Flight Center during several test firings of two newly designed turbopumps: *PW8003R2* and *PW8105R2*. This particular turbopump design has one pump-end and one turbine-end bearing. The vibration signals were measured from two external accelerometers located at the vicinity of the pump-end and turbine-end bearing. The vibration data from these accelerometers was recorded on 1" magnetic tapes during the tests, using IRIG IB at 60 in/sec giving an effective recorder bandwidth of 20 kHz with a signal to noise ratio of 51 dB RMS. The external accelerometers was lowpass-filtered at 5 or 10

kHz prior to recording, to eliminate much of the high frequency noise from various extraneous sources [8].

The total test time for PW8003R2 was 1560 seconds (520 seconds each of three firings), and was 2080 seconds for PW8105R2 (520 seconds each of four firings). The specific details of both tests are listed in Tables 1 and 2, respectively. Note that since each test firing consisted of three stages: start-up, throttle-down, and shut-down, the shaft speed was not constant through out a test. At the end of each tests the bearings were disassembled and checked. A severe ball wear of 3 mils was found at the pump-end bearing of PW8003R2, and a light ball wear of 0.25 mils was found also at the pump-end bearing of PW8105R2.

Table 1: Test series using PW8003R2.

Test No.	Duration	Cumulative Time	Bearing Condition
A904-150	520 secs	520 secs	New
A904-151	520 secs	1040 secs	
A904-154	520 secs	1560 secs	Severe Ball Wear

Table 2: Test series using PW8105R2.

Test No.	Duration	Cumulative Time	Bearing Condition
A904-152	520 secs	520 secs	New
A904-153	520 secs	1040 secs	
A901-717	520 secs	1560 secs	
A901-718	520 secs	2080 secs	Light Ball Wear

Adaptive Feature Map (AFM):

The proposed *Adaptive Feature Map (AFM)* method is specially designed for any mechanical failure that requires immediate actions, such as the SSME turbopump. The AFM method, which is based on unsupervised pattern classification, does not require any pre-training.

The proposed AFM method uses a hybrid learning algorithm that takes the advantage of both *Kohonen's feature mapping* algorithm [10] and *Adaptive Resonance Theory (ART)* [1]. Similar to these two unsupervised learning algorithms, the AFM method utilizes continuous-valued input patterns (vectors) which are presented sequentially to a feature map without specifying the desired outputs.

In the AFM method, it is assumed that the first input vector represents the normal (no-fault) case and the weight vector in AFM is an exact copy of the first input

vector. Note that initially there is only one weight vector in the AFM representing the prototype vector of the normal case. The subsequent input vectors are then presented to AFM and compared to the weight vector by computing the Euclidean distance between the two vectors. If the computed Euclidean distance exceeds a vigilance factor, then a new category (e.g., faulty case) is formed, with the current input vector as the prototype vector (weight vector) of the new category. If the computed distance does not exceed the vigilance factor, then the first weight vector is adjusted so as to account for the current input vector. After enough input vectors have been presented, the weight vectors of the AFM will become the centers of these input vectors (clusters). As such, the point density function of these clusters approximates the probability density function of the input vectors. In addition, the weights will be organized in such a way that nodes which are close together in the feature map are sensitive to input vectors that are physically similar.

In the AFM method, the input vectors $\mathbf{X} \in \mathcal{R}^n$ are used to sequentially train a feature map with n input nodes and initially one output node ($j = 1$) representing the normal case, as illustrated in Fig. 2. The AFM method then determines whether $\mathbf{X}$ belongs to the normal case (represented by the vector $\mathbf{W}_{j=1}$) by computing the Euclidean distance d_1 between $\mathbf{X}$ and $\mathbf{W}_1$, that is,

$$d_1 = \mid \mathbf{W}_1 - \mathbf{X} \mid . \tag{1}$$

The computed Euclidean distance d_1 is then compared to a vigilance factor ρ to determine whether the input vector $\mathbf{X}$ is close enough to $\mathbf{W}_1$ according to the relationships

$$\begin{cases} \mathbf{X} \in \mathbf{W}_1 & \text{if } d_1 < \rho \\ \mathbf{X} \in \mathbf{W}_{j\neq 1} & \text{otherwise} \end{cases} \tag{2}$$

where $j = 1, 2, \ldots$ represents all the possible categories. If $\mathbf{X} \in \mathbf{W}_{j\neq 1}$, a new category is formed with a new weight vector $\mathbf{W}_{j=2} = \mathbf{X}$. The same process continues for both $j = 1$ and $j = 2$. Learning is performed so as to adaptively adjust the weight vectors according to the recursive relationship

$$\Delta W_j = \frac{1}{N_j}(X - W_j) \tag{3}$$

for all j, where N_j denotes the number of input vectors belonging to the jth category.

Results:

To test the applicability of the AFM method in SSME turbopump bearing fault detection, the vibration data collected from Tests A904-154 and A901-718 was used since they both contain bearing faults (see Tables 1 and 2). The vibration data from both tests (520 second worth of data per test) was digitized at 125 kHz and lowpass-filtered at 50 kHz. Fast Fourier Transform (FFT) was then used to convert the 520 second data into 520 1024-point spectra (i.e., one spectrum per second) for bearing fault detection.

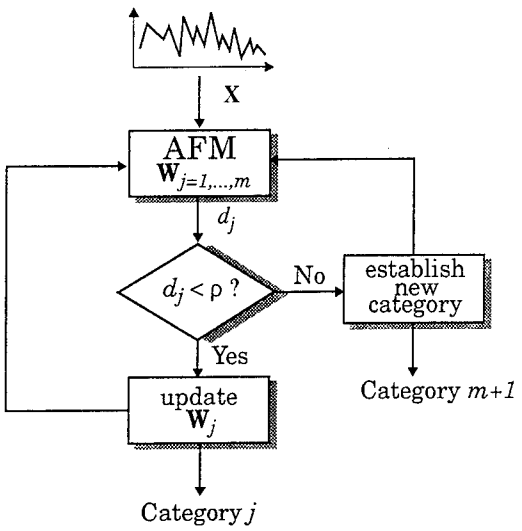


Figure 2: Schematic of the Adaptive Feature Map (AFM) method.

As mentioned earlier, the effect of bearing faults is usually masked in the lower frequency region by other rotating elements such as shafts. As such, it is better to consider only the components in the high frequency region where the effect of shaft speed and its harmonics is minimal. In this study, frequency components below 10 kHz were not used as they contain strong shaft speed and its harmonics. On the other hand, the frequency components above 20 kHz were not considered since the recording bandwidth is 20 kHz.

The frequency components in 10-20 kHz were then divided into two regions: 10-16 kHz and 14-20 kHz for bearing fault detection. Note that it is also possible to use all the components in 10-20 kHz, but it is better to use only a subset of the components for better detectability and faster detection. The 10-16 kHz region contains 98 components, whereas the 14-20 kHz region contains 99 components. These components were then used by the AFM method as the elements of input vectors for fault detection. Table 3 shows the detection results obtained for both Tests A904-154 and A901-718.

The results in Table 3 indicate that the AFM method was able to detect bearing faults in both tests, and that it produced earlier indication of faults with the components from 14-20 kHz as input. Specifically, the AFM method detected the bearing faults 25 and 9 seconds earlier in that bandwidth for both Tests A904-154 and A901-718, respectively. The vigilance factor for each case was dynamically determined by statistical analysis on the no-fault portion of the vibration data (see Table 3).

For comparison purposes, envelope analysis was also performed on both sets of

Table 3: Detection results for Tests A904-154 and A901-718 using the AFM method.

Test No.	Bandwidth	Vigilance Factor	Estimated Fault Occurrence
A904-154	10-16 kHz	0.07	145th sec
	14-20 kHz	0.05	120th sec
A901-718	10-16 kHz	0.5	467th sec
	14-20 kHz	0.1	458th sec

data. For envelope analysis, each of the one-second data was first bandpass-filtered with respect to a given bandwidth. For this study, the bandwidth of 14-20 kHz was used since this region contains most of the changes produced by bearing faults, as identified by the AFM. The bandpass-filtered signal was then Hilbert transformed to an analytical signal which was then converted to an amplitude envelope. An envelope spectrum and several envelope indicators were then obtained from this amplitude envelope [2].

Figures 3 and 4 show the envelope spectrum waterfall plots obtained for Tests A904-154 and A901-718, respectively. It can be seen from Fig. 3 that at approximately the 120th second, cage frequency (F_c) and cage modulated shaft frequency (F_s) appeared, which were the fist signs of ball wear. As further into the test, F_s and its harmonics appeared, and the energy associated with F_c and F_c modulated F_s became much higher, suggesting the ball wear was much worse with possible inner-race damage. At around the 200th second, F_c started to increase till around the 300th second. One explanation of this F_c increase is that the cage started to rub against the shaft and finally lock on to the shaft, which in terms increased F_c.

From Fig. 4, it can be seen that at approximately the 490th second, cage frequency (F_c) and shaft frequency (F_s) appeared, which were the fist signs of ball wear. In addition to F_c and F_s, both outer-race frequency (F_o) and inner-race frequency (F_i) appeared in the waterfall plot, suggesting a possible ball wear.

Several bearing fault indicators were also obtained through envelope analysis. Among then, *Bearing Distributed Fault (BDF)* index, a proprietary bearing fault indicator developed by the author, is one of the most effective indicators for early bearing fault detection [2]. Figure 5 shows the results of the BDF index for both Tests A904-154 (upper) and A901-718 (lower). From the upper part of Fig. 5, it can be seen that the BDF index increased drastically at approximately the 120th second from below 0.1 to around 0.24, an indication of the progression of bearing wear. The BDF index shown in the lower part of Fig. 5 also increased drastically at approximately the 470th second from below 0.08 to around 0.125, also an indication of bearing wear.

As shown above, envelope analysis is also quite effective in detecting bearing faults.

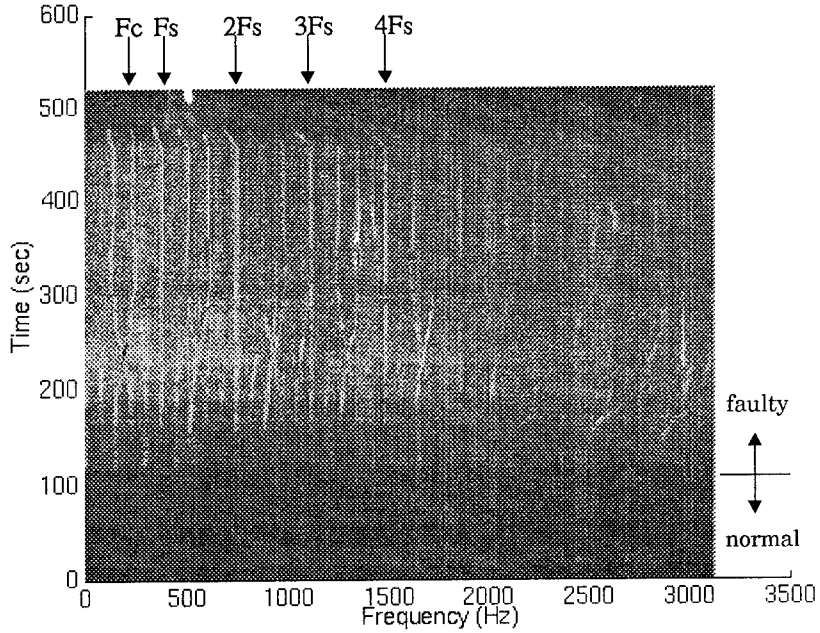


Figure 3: Envelope spectrum waterfall plot of Test A904-154.

However, the computational effort involved in each envelope analysis is usually quite cumbersome. The AFM method, on the other hand, performs fault detection more quickly and effectively due to its simple detection strategy and fast learning. For example, the AFM method detected the light bearing wear in Test A901-718 much earlier (32 seconds early) than envelope analysis did. Based on this comparison, the AFM method is more sensitive to subtle changes produced by early bearing faults. Envelope analysis, which takes a much longer processing time, can be performed later to interpret the nature of the bearing faults.

Conclusions:

A new fault detection method based on pattern classification has been introduced and applied to early bearing fault detection for the SSME turbopump. This method, which uses an Adaptive Feature Map (AFM) as its detection model, has the ability to detect subtle changes in a signal. Vibration data collected from two newly designed SSME turbopumps was used to test the performance of the proposed AFM method. Results of this study show that the AFM method was quite effective in detecting early bearing faults for both cases, when the frequency components within 14-20 kHz were used. The results also show that the AFM method is more sensitive to early bearing faults than envelope analysis is.

Acknowledgements: The author would like to express his gratitude to Mr. Preston Jones and Mr. Robert Smith of NASA Marshall for their support of this project and for providing the experimental data. This work was supported in part by NASA Marshall under a Phase II SBIR NASA contract No. NAS8-38955.

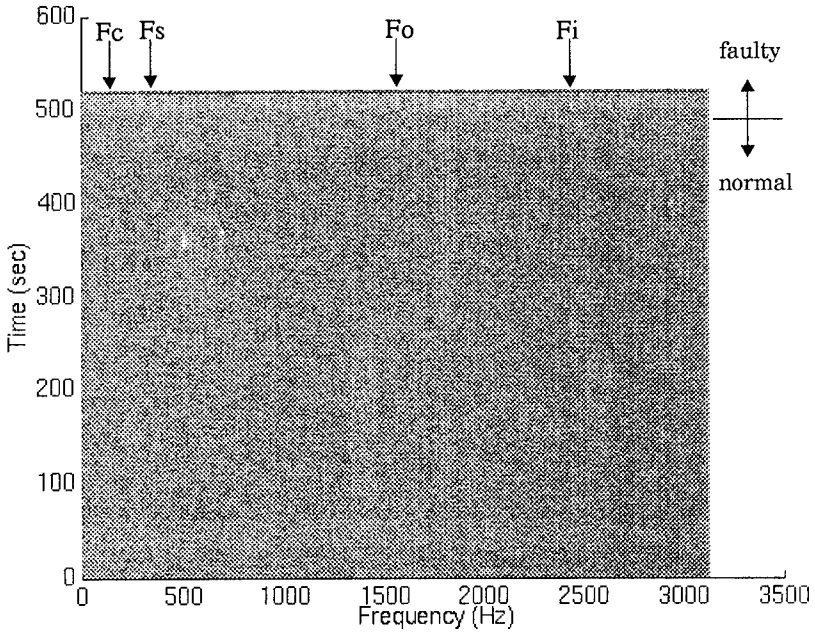


Figure 4: Envelope spectrum waterfall plot of Test A901-718.

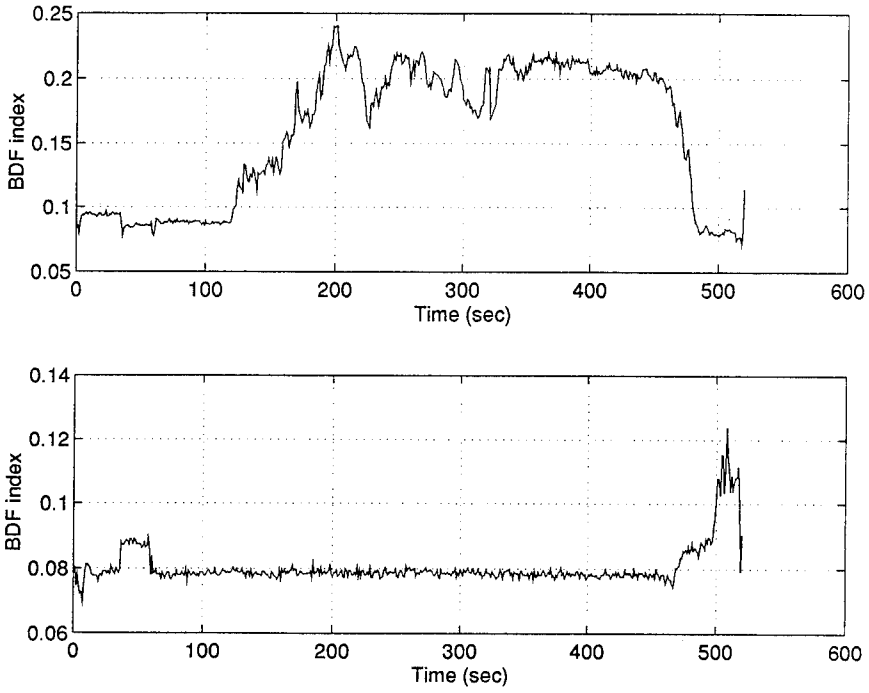


Figure 5: BDF index obtained from Test A904-154 (upper) and Test A901-718 (lower).

References:

[1] Carpenter, G. A., and Grossberg, S., 1988, "The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network", *Computer*, March, pp. 77–88.

[2] Chin, H., 1994, *Space Shuttle Main Engine Bearing Fault Detection with Envelope Analysis*, Technical Report R9410-002-RD, Technology Integration, Inc., Bedford, MA.

[3] Chin, H., and Danai, K., 1992, "Improved Flagging for Pattern Classifying Diagnostic Systems", *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 23, No. 4, pp. 1101–1107.

[4] Chin, H., Danai, K., and Lewicki, D., G., 1992, "Pattern Classifier for Fault Diagnosis of Helicopter Gearboxes", *Control Engineering Practice*, Vol. 1, No. 5, pp. 771–778.

[5] Courrech, J., and Gaudet, M, *Envelope Analysis - the Key to Rolling-Element Bearing Diagnosis*, Bruel & Kjaer Application Notes.

[6] Duyar, A., and Merrill, W., 1992, "Fault Diagnosis for the Space Shuttle Main Engine", *J. of Guidance, Control, and Dynamics*, Vol. 15, No. 2, March-April, pp. 384–389.

[7] Hertz, J., Krogh, A., and Palmer, R. G., 1991, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA.

[8] Howard, I. M., 1993, *Signal Processing Techniques for Mechanical Diagnosis of Space Shuttle Main Engine Turbomachinery Using Vibration Analysis*, Technical Report R9301-002-RD, Technology Integration, Inc., Bedford, MA.

[9] Jong J., Jones, J., Jones, P., Nesman, T., and Zolads, T., 1994, "Nonlinear Correlation Analysis for Rocket Engine Turbomachinery Vibration Diagnostics", *Proc. of the 48th Mechanical Failure Prevention Group (MFPG) Meeting*, April, pp. 379–389.

[10] Kohonen, T., 1989, *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, Germany.

[11] Li, C. J., and Ma, J., 1992, "Bearing Localized Defect Detection Through Wavelet Decomposition of Vibrations", *Sensors and Signal Processing for Manufacturing*, ASME PED Vol. 55, pp. 187–196.

[12] Pratt, J. L., 1986, "Engine and Transmission Monitoring - A Summary of Promising Approaches", *Proc. of the 41th Mechanical Failure Prevention Group (MFPG) Meeting*, Oct., pp. 229–236.

A STOCHASTIC MODEL FOR CRANKSHAFT MOTION IN A COMPRESSOR

Bruce Friedman

Naval Surface Warfare Center, Code 612
Carderock Division Annapolis Detachment
3A Leggett Circle
Annapolis Maryland 21402-5067

Abstract: A simple stochastic model of compressor crankshaft motion is presented for application to the statistical analysis of the compressor operation. From the Langevin equations for crankshaft motion a Fokker-Planck (FP) equation is obtained. A solution of the FP equation yields the transition probability density for crankshaft travel and angular velocity for short times. For the marginal crankshaft travel and angular velocity densities expressions are found for the means and confidence limits of the initial values.

Key words: Fokker-Planck equation; compressor; crankshaft; stochastic model

Introduction: The United States Navy is undertaking a research and development effort in the area of condition-based maintenance (CBM) in order to enhance the cost-effectiveness of naval operations. The implementation of CBM would allow more dependable diagnoses and prognoses to be made concerning mechanical machinery performance.

A high pressure air compressor (HPAC) is being used as a test bed in the CBM program.[1] Pressure and temperature sensors are mounted at various points of the compressor and their measurements are recorded. It is noted that under presumably steady-state conditions, the readings derived from the sensor outputs fluctuate with a random component. So as to allow a greater understanding of the stochastic elements of these sensor readings and of the operation of the compressor, a very simplified stochastic model of compressor crankshaft motion will be presented. This model would be of assistance in the statistical analysis of the compressor operation. Also, the model would be useful in constructing fuzzy set and logic techniques for diagnosis and prognosis, since there already exists a body of research that has discovered relationships between probability, possibility, fuzzy sets and statistics. The use of fuzzy sets and logic for mechanical diagnostics has been incorporated into the CBM program[2]

Approach: Chlumsky gives the approximate equation (3)

$$x = r\left(1 - \cos\alpha \pm \frac{\lambda}{2}\sin^2\alpha\right) \qquad (1)$$

for the description of the travel of the crankshaft connecting rod where $\lambda \equiv r/\ell < 1$ is the ratio of the radius, r, of the crank circle (crank arm) to the length, ℓ, of the connecting rod. The approximation becomes more exact the smaller λ becomes. The angle α is the crank angle. The negative sign before λ is associated with ℓ for travel from the outer dead center of the cylinder.

Upon differentiation with respect to α, eq. (1) becomes:

$$\frac{dx}{d\alpha} = r(1 + \lambda_a \cos\alpha)\sin\alpha \qquad (2)$$

with the notation λ_a, $a = 1,2$; $\lambda_1 = \lambda$, $\lambda_2 = -\lambda$ replacing the $\pm\lambda$ notation of Chlumsky.[3] In principle, eq.(1) can be considered to be an equation for α in terms of x.. Eq.(3) is obtained from eq. (1).

$$\lambda_a^2 r^2 \sin^4\alpha + 4\left[1 + (1-x)\lambda_a\right] r\sin^2\alpha + 4\left[(r-x)^2 - r\right] = 0 \qquad (3)$$

Eq.(3) is a biquadratic equation with respect to $\sin\alpha$ and has four roots denoted by $\sin\alpha = f(x_i), i = 1,2,3,4$. At this juncture, the nature of these solutions is irrelevant so long as at least one of these solutions is physically meaningful. As a function of x, eq. (2) becomes

$$\frac{dx}{d\alpha} = r f_i(x)\left(1 + \lambda_a\sqrt{1 - f_i^2(x)}\right). \qquad (4)$$

Although in the case of deterministic motion $dx/d\alpha$ can be taken to be equivalent to dx/dt where t denotes elapsed time this equivalence no longer holds for the stochastic case. Since

$$\frac{dx}{d\alpha} = \frac{dx}{dt}\frac{dt}{d\alpha}, \qquad (5)$$

where $\frac{d\alpha}{dt} = \omega$ is the angular velocity of the connecting rod, for the case of deterministic motion, the situation is described by the solutions of the equations

$$\frac{dx}{dt} = \omega f_i(x)\left(1 + \lambda_a\sqrt{1 - f_i^2(x)}\right), \qquad (6a)$$

$$\frac{d\omega}{dt} = 0, \qquad (6b)$$

since the angular velocity remains constant.

In actuality, the deterministic description is less adequate than a stochastic description. Fluctuations in the cylinder pressure and sticking and slipping of the piston give rise to a component of the connecting rod motion which may be depicted as being random in

nature. Eqs. (6a)-(6b) are converted from deterministic equations to stochastic Langevin equations through the addition of random noise terms on their right sides.[4] These equations then become

$$\frac{dx}{dt} = \omega f_i(x)\left(1+\lambda_a\sqrt{1-f_i^2(x)}\right)+K_1^{1/2}\Gamma_1(t), \qquad (7a)$$

$$\frac{d\omega}{dt} = K_2^{1/2}\Gamma_2(t), \qquad (7b)$$

where $K_1^{1/2}\Gamma_1(t)$ and $K_2^{1/2}\Gamma_2(t)$ are the noise functions and $K_1^{1/2}$ and $K_2^{1/2}$ are constants representing the noise strengths.

There exist standard approaches to obtain the Fokker-Planck (FP) equation associated with the eqs. (7a)-(7b).[4] The solution of the FP equation yields the transition probability density of x and ω as a function of t.

A system of nonlinear Langevin equations for N variables, x_i, is given by

$$\frac{dx_i}{dt} = h_i(x_1,\ldots,x_N;t)+\sum_{j=1}^{N} g_{ij}(x_1,\ldots,x_N;t)\Gamma_j(t), \qquad (8a)$$

$$\langle\Gamma_i(t)\rangle = 0, \quad \langle\Gamma_i(t)\Gamma_j(t')\rangle = 2\delta_{ij}\delta(t-t'), \qquad (8b)$$

for $i,j = 1,\ldots,N$ where $\langle\Gamma_i(t)\rangle$ is the expectation (average) value of the random function of time $\Gamma_i(t)$ and $\langle\Gamma_i(t)\Gamma_j(t')\rangle$ is the expectation value of the product $\Gamma_i(t)\Gamma_j(t')$.[4] The FP equation for N variables is given by

$$\frac{\partial P(x_1,\ldots,x_N;t|x'_1,\ldots,x'_N;t')}{\partial t} =$$

$$\left(-\sum_{j=1}^{N}\frac{\partial}{\partial x_i}D_i(x_1,\ldots,x_N;t)+\sum_{j=1}^{N}\frac{\partial^2}{\partial x_i\partial x_j}D_{ij}(x_1,\ldots,x_N;t)\right)$$

$$\bullet P(x_1,\ldots,x_N;t|x'_1,\ldots,x'_N;t'), \qquad (9a)$$

where P is the transition probability density and the D_i and D_{ij} are Kramers-Moyal coefficients determined by reference to eq. (8a) so that

$$D_i(x_1,\ldots,x_N;t) = h_i(x_1,\ldots,x_N;t)$$

$$+\sum_{j,k=1}^{N} g_{kj}(x_1,\ldots,x_N;t)\frac{\partial g_{ij}(x_1,\ldots,x_N;t)}{\partial x_k}, \qquad (9b)$$

$$D_{ij}(x_1,\ldots,x_N;t) = \sum_{k=1}^{N} g_{ik}(x_1,\ldots,x_N;t)g_{jk}(x_1,\ldots,x_N;t), \qquad (9c)$$

where the h_i and g_{ij} functions are the same as those found in eq (8a).

Applying eqs (9b) - (9c) to eqs. (7a)-(7b) gives

$$D_1(x,\omega,t) = \omega f_i(x)\left(1+\lambda_a\sqrt{1-f_i^2(x)}\right),$$
$$D_2(x,\omega,t) = 0,$$
$$D_{12}(x,\omega,t) = 0,$$
$$D_{21}(x,\omega,t) = 0,$$
$$D_{11}(x,\omega,t) = K_1,$$
$$D_{22}(x,\omega,t) = K_2. \qquad (10)$$

(It should be noted that the transition probability $P(x_1,\ldots,x_N;t|x'_1,\ldots,x'_N;t')$ is the distribution $W(x_1,\ldots,x_N;t)$ for the special initial condition $W(x_1,\ldots,x_N;t') = \delta(x_1-x_1')\delta(x_2-x_2')\ldots\delta(x_N-x_N')$. This initial condition is adopted for this investigation).[4]

The FP equation associated with eqs (7a)-(7b) is then

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\left(\left[\omega f_i(x)\left(1+\lambda_a\sqrt{1-f_i^2(x)}\right)\right]P\right)$$
$$+K_1\left(\frac{\partial^2 P}{\partial x^2}\right)+K_2\left(\frac{\partial^2 P}{\partial \omega^2}\right), \qquad (11)$$

Results: The general solution of eq.(11) is difficult to obtain. However, a solution for small times can be found through the application of the expression [4]

$$P\left(x_1,\ldots,x_N;t+t\middle|x'_1,\ldots,x'_N;t\right) = (2\sqrt{pt})^{-N}\left[\mathrm{Det}\left\{D_{mn}\left(x'_1,\ldots,x'_N;t\right)\right\}\right]^{1/2}$$
$$\bullet\exp\left(\sum_{j,k=1}^{N}\left\{-\frac{1}{4t}\left[D^{-1}_{jk}\left(x'_1,\ldots,x'_N;t\right)\right]\left[x_j-x'_j-D_j\left(x'_1,\ldots,x'_N;t\right)t\right]\left[x_k-x'_k-D_k\left(x'_1,\ldots,x'_N;t\right)t\right]\right\}\right). \qquad (12)$$

For the connecting rod problem it is found that

$$P(x,\omega,t+\tau|x';\omega';t) = \left(4\pi\tau\sqrt{D_{11}D_{12}}\right)^{-1}$$
$$\bullet\exp\left(-\frac{1}{4\tau}\left(\frac{1}{D_{11}}\left(x-x'-D_1\tau\right)^2+\frac{1}{D_{22}}\left(\omega-\omega'-D_2\tau\right)^2\right)\right), \qquad (13)$$

which upon substituting from eq.(10) becomes

$$P(x,\omega,t+\tau|x';\omega';t) = \left(4\pi\tau\sqrt{K_1K_2}\right)^{-1}$$
$$\bullet\exp\left(-\frac{1}{4\tau}\left(\frac{1}{K_1}\left(x-x'-\omega f_i(x')\left(1+\lambda_a\sqrt{1-f_i^2(x')}\right)\tau\right)^2+\frac{1}{K_2}\left(\omega-\omega^1\right)^2\right)\right), \qquad (14)$$

From eq.(14) the two marginal transition probability densities for x and for ω are obtained. These marginal densities [5] are defined by

$$p(x,t+\tau|x';t) \equiv \int_0^\infty p(x,\omega,t+\tau|x',\omega';t)d\omega, \tag{15a}$$

$$p(\omega,t+\tau|\omega';t) \equiv \int_0^\infty p(x,\omega,t+\tau|x',\omega';t)dx, \tag{15b}$$

for x and for ω, respectively.

For the x marginal density, it is found that [6]

$$p(x,t+\tau|x;t) = \left(\frac{1}{4\tau}\sqrt{\frac{A_0}{\pi K_1 K_2}}\right)\left[\exp\left(A_0A_1^2 - A_2\right)\right]\left[1-\Phi\left(A_1\sqrt{A_0}\right)\right], \tag{16a}$$

where $\Phi\left(A_1\sqrt{A_0}\right)$ is the probability integral such that $\Phi(y) = \frac{2}{\sqrt{\pi}}\int_0^y e^{-z^2}dz$ and

$$A_0 \equiv \tau\left[f_i^2(x')\left(1+\lambda_a\sqrt{1-f_i^2(x')}\right)^2\left(\frac{\tau^2}{K_1}\right)+\frac{1}{K_2}\right]^{-1}, \tag{16b}$$

$$A_1 \equiv \frac{1}{2\tau}\left[-\left(\frac{\tau}{K_1}\right)(1-x')f_i(x')\left(1+\lambda_a\sqrt{1-f_i^2(x')}\right)-\frac{\omega'}{K_2}\right], \tag{16c}$$

$$A_2 \equiv \frac{1}{4\tau}\left[\left(\frac{1}{K_1}\right)(x-x')^2+\frac{(\omega')^2}{K_2}\right], \tag{16d}$$

For the ω marginal density, it is found that [6]

$$p(\omega,t+\tau|\omega;t) = \left(\frac{1}{4\tau}\sqrt{\frac{B_0}{\pi K_1 K_2}}\right)\exp\left(B_0B_1^2 - B_2\right)\left[1-\Phi\left(B_1\sqrt{B_0}\right)\right], \tag{17a}$$

where $\Phi\left(B_1\sqrt{B_0}\right)$ is again the probability integral function and

$$B_0 \equiv \tau K_1, \tag{17b}$$

$$B_1 \equiv \frac{1}{2\tau K_1}\left[-x'-\omega f_i(x')\left(1+\lambda_a\sqrt{1-f_i^2(x')}\right)\tau\right], \tag{17c}$$

$$B_2 \equiv \frac{1}{\tau}\left[\frac{(x')^2}{4K_1}+\frac{\omega^2}{4K_1}f_i^2(x')\left(1+\lambda_a\sqrt{1-f_i^2(x')}\right)^2\tau^2\right]$$
$$+\frac{1}{\tau}\left[\frac{x'\omega}{2K_1}f_i(x')\left(1+\lambda_a\sqrt{1-f_i^2(x')}\right)\tau+\frac{1}{4K_2}(\omega-\omega')^2\right]. \tag{17d}$$

The $p(x,t+\tau|x',t)$ distribution for a given τ value is proportional to a normal distribution which is truncated for $x < 0$. The $p(\omega,t+\tau|\omega;t)$ distribution for a given τ

value is proportional to a normal distribution in the limiting case $f(x') \rightarrow 0$ and which is truncated for $\omega < 0$.[7]

Both x' and ω' are initial values for x and ω, respectively. By computing the expectation values for x and ω with respect to the corresponding marginal probability density functions, relationships between the average $\bar{x}$ and x' and ω' and between the average $\bar{\omega}$ and x' and ω' are obtained.

The mean $\bar{x}$ is given by [5]

$$\bar{x} = \int_0^\infty x p(x, t+\tau | x', t) dx \qquad (18a)$$

which is calculated to be[6]

$$\bar{x} = \left(\frac{1}{4\tau}\sqrt{\frac{A_0}{\pi K_1 K_2}}\right)\left[\exp\left(A_0 A_1^2 - \frac{(\omega')^2}{4\tau K_2}\right)\right]\left[1 - \Phi\left(A_0\sqrt{A_0}\right)\right]$$

$$\bullet\left\{2\pi K_1 \exp\left(\frac{-(x')^2}{4\tau K_1}\right) + x'\sqrt{\pi\tau K_1}\left[1 - \Phi\left(\frac{-x'}{2\sqrt{\tau K_1}}\right)\right]\right\}. \qquad (18b)$$

The mean $\bar{\omega}$ is given by [5]

$$\bar{\omega} = \int_0^\infty \omega p(\omega, t+\tau | \omega', t) d\omega \qquad (19a)$$

which in the limiting case $f(x') \rightarrow 0$ is calculated to be [6]

$$\bar{x} = \left(\frac{1}{4\sqrt{\pi K_2 \tau}}\right)\left[1 - \Phi\left(\frac{-x'}{2\sqrt{K_1 \tau}}\right)\right]$$

$$\bullet\left\{2 K_1 \tau \exp\left(\frac{-(\omega')^2}{4\tau K_2}\right) + \omega'\sqrt{\pi\tau K_2}\left[1 - \Phi\left(\frac{-\omega'}{2\sqrt{\tau K_2}}\right)\right]\right\}. \qquad (19b)$$

Suppose that the initial values x' and ω' are considered to be unknown parameters. Point estimates for the values of these two parameters can be obtained. A point estimate is itself a random variable which is distributed about the true although unknown value of the parameter. The introduction of the concept of confidence limits or intervals allows a technique to be utilized which yields an interval within which there is reasonable confidence that the true value of the parameter is to be found. [8]

For the marginal x distribution the upper and lower confidence limits, x'_{up} and x'_{lo} respectively, for the true value x'_{tr} of x' are calculated from the equations [8]

$$\int_0^{x'_{est}} p\left(x, t+\tau \,|x'_{lo}, \tau\right)dx = P_2, \qquad (20a)$$

$$\int_0^{x'_{est}} p\left(x, t+\tau \,|x'_{up}, \tau\right)dx = P_1, \qquad (20b)$$

where x'_{est} is the estimated value for x'. The probability $P\left(x'_{lo} < x'_{tr} < x'_{up}\right)$ that x'_{tr} lies between x'_{up} and x'_{lo} is equal to $P_2 - P_1$.

Upon performing the integration of eqs.(20a)-(20b) it is found that [6]

$$\frac{1}{2}\sqrt{K_1\tau}\left[1-\Phi\left(\frac{\omega'}{2\sqrt{K_2\tau}}\right)\right]\left[\Phi\left(\frac{x'_{est}-x'_{lo}}{4K_1\tau}\right)-\Phi\left(\frac{-x'_{lo}}{4K_1\tau}\right)\right]=P_2, \qquad (21a)$$

$$\frac{1}{2}\sqrt{K_1\tau}\left[1-\Phi\left(\frac{\omega'}{2\sqrt{K_2\tau}}\right)\right]\left[\Phi\left(\frac{x'_{est}-x'_{lo}}{4K_1\tau}\right)-\Phi\left(\frac{-x'_{up}}{4K_1\tau}\right)\right]=P_2, \qquad (21b)$$

For the marginal ω distribution in the limiting case $f_i(x') \to 0$ the equations corresponding to EDS (20a) -(21b) are [6], [8]

$$\int_0^{\omega'_{est}} p\left(\omega, t+\tau \,|\omega'_{lo}, \tau\right)dx = P_2, \qquad (22a)$$

$$\int_0^{\omega'_{est}} p\left(\omega, t+\tau \,|\omega'_{up}, \tau\right)dx = P_1, \qquad (22b)$$

$$\frac{1}{2}\sqrt{K_2\tau}\left[1-\Phi\left(\frac{x'}{2\sqrt{K_1\tau}}\right)\right]\left[\Phi\left(\frac{\omega'_{est}-\omega'_{lo}}{4K_2\tau}\right)-\Phi\left(\frac{-\omega'_{lo}}{4K_2\tau}\right)\right]=P_2, \qquad (23a)$$

$$\frac{1}{2}\sqrt{K_2\tau}\left[1-\Phi\left(\frac{x'}{2\sqrt{K_1\tau}}\right)\right]\left[\Phi\left(\frac{\omega'_{est}-\omega'_{up}}{4K_2\tau}\right)-\Phi\left(\frac{-\omega'_{up}}{4K_2\tau}\right)\right]=P_1. \qquad (23b)$$

CONCLUSIONS: The validity and significance of the stochastic model depends upon the acquisition of sufficient experimental data to provide verification of the probabilistic and statistical relations derived in this investigation.

REFERENCES

1. Henry R. Henna, John P. HUD, and Michael H. Porricelli, Condition-Based Maintenance (CBM) System For High Pressure Air Compressor (HPAC), Technical Report, Engineering Computer Optecnomics, Inc., Annapolis, Maryland, 28 Feb 1991.
2. B. Friedman, E.A. Schroeder and G.W. Nickerson, "Reasoning from Uncertain Data for Mechanical Diagnostics" Proceedings of the Workshop on Artificial

Intelligence for Customer Service and Support, Monterey, California, March 3, 1992, Eighth IEEEE Conference on Artificial Intelligence Applications, pp. 50-59.
3. Vladimir Chlumsky, Reciprocating and Rotary Compressors, (SNTL-Publishers of Technical Literature, Prague, 1960: E&FN Spon Ltd., London, 1965)
4. H. Risken, The Fokker-Planck Equation: Methods of Solution and Applications, (Springer-Verlag, Berlin, 1989, second edition).
5. Athanasios Papoulis, Probability, Random Variables, and Stochastic Processes, (McGraw-Hill Book Company, New York, 1984, second edition)
6. I.S. Gradshteyn and I. M. Ryzkik, Table of Integrals, Series and Products, (Academic Press, New York, 1980, corrected and enlarged edition).
7. A. Hald, Statistical Theory with Engineering Applications, (John Wiley & Sons, Inc., New York, 1952).
8. K.A. Brownlee, Statistical Theory and Methodology in Science and Engineering (John Wiley & Sons Inc., New York, 1965, second edition).

LIFE EXTENSION

Cochairmen: Michael Roemer
Stress Technology, Inc.

Reginald G. Mitchiner
Virginia Polytechnic Institute

DESIGN HEALTH MONITORING AND LIFE EXTENSION OF MECHANICAL DYNAMIC STRUCTURES

Robert S. Ballinger
Assistant Professor

David W. Herrin
Research Assistant

Department of Mechanical Engineering
University of Kentucky
Lexington, Kentucky 40506-0108

ABSTRACT

This research combines analytical and experimental modal analysis techniques to verify the structural integrity or monitor the "health" of a dynamic structure. This paper outlines the general procedure and documents the results of two case studies. Essential to the procedure is the development of a baseline *dynamic fingerprint* model of the structure. The baseline structural model is then excited and the structural response is measured. The structure is then placed in service and, periodically, the structural response due to the same excitation is measured. The response motion of the in-service structure is enforced on the baseline *dynamic fingerprint* structural model. Significant differences in the normalized modal or operating strain energy density identify structural regions where a loss of stiffness, weakening of the structure, and/or damage has occurred. This identification of a potentially "unhealthy" structural region allows a quick visual inspection of the region or further analytical and/or experimental submodelling of the area to precisely identify the damage. The method is ideally suited to CAE application. The method is demonstrated analytically and experimentally for two case studies: an eight-bay cantilevered truss structure and a free-free rectangular plate.

Key Words: Damage detection; health monitoring; fatigue; nondestructive evaluation; preventive maintenance; structural dynamics; systems; testing; vibration.

1.0 INTRODUCTION

Health monitoring or damage detection of both new and aging structures is important in anticipating and preventing failures in the field. Current efforts in damage detection involve the generation of a baseline analytical model or a model synthesized from experimental measurements. Model correlation is then performed in order to match the modal parameters of the model with the measured modal parameters of the candidate structure. When the structure is placed into service, deviations between the analytical response and the measured response are used to correct the model resulting in a standard strategy where the change between models is minimized to obtain a match [1-4].

2.0 GENERAL PROCEDURE

The basis for this research is that when a structure becomes damaged, the damage manifests itself in localized loss of stiffness which results in an increase in the global response (of the fundamental modes) due to the application of forces [5]. This behavior is true for structures in both static and dynamic operation. This paper will be concerned with the increased dynamic response and how it can be used to identify area(s) in a structure where a loss of stiffness has occurred. The process can be implemented in the design phase during accelerated life-cycle testing to verify the structural integrity of the prototype design or the procedure can be employed for long-term health monitoring of mechanical dynamic structures in the field. The procedure requires an accurate baseline *dynamic fingerprint* model of the structure. This model can be created using finite element techniques in conjunction with subsequent model correlation and updating to match experimental test data, or exclusively from experimental modal analysis techniques.

Once the *dynamic fingerprint* of the structure is obtained, the measured response displacements (usually a subset of the total number of degrees-of-freedom), for a chosen excitation, are used with Guyan expansion to calculate the full degree-of-freedom response set of the structure. The full displacement set is used to determine the normalized modal strain energy density for each element of the structure. The element strain energy for the ith element in mode r, SE_{ir}, is given by:

$$SE_{ir} = \frac{1}{2}\{\phi_r\}^T[K_i]\{\phi_r\} \tag{1}$$

where $\{\phi_r\}$ is the eigenvector (full expanded response set) for the rth mode and $[K_i]$ is the stiffness matrix for element i. The strain energy density for the ith element, SED_{ir}, is then calculated by dividing the element strain energy, SE_{ir}, by the element volume, v_i.

$$SED_{ir} = \frac{SE_{ir}}{v_i} \tag{2}$$

Finally, the normalized modal strain energy density for the ith element in mode r is found by dividing the element strain energy density, SED_{ir}, by the total element strain energy density in mode r.

$$\overline{SED}_{ir} = \frac{SED_{ir}}{\sum_{i=0}^{n} SED_{ir}} \tag{3}$$

where n is the total number of elements.

The structure is then placed into "field service" and the response is monitored periodically to check for damage. The remeasured displacements, due to the same

initially chosen excitation, are used to again calculate the full degree-of-freedom response set of the structure. The full displacement set is used to determine the normalized modal strain energy density for each element of the structure. By examining the differences in the normalized element modal strain energy densities, a determination regarding damage can be made. If the differences in the element normalized modal strain energy densities are not significant (to within the measurement error of the displacements) then the structural response is the "same" as the initial baseline and there is no measurable damage. However, if the difference in normalized modal strain energy density for certain element(s) has increased significantly, then the areas corresponding to these element locations should be visually checked for damage or modeled further to quantify the extent of the damage. The increase in normalized modal strain energy density is a result of enforcing an unnatural response displacement (due to damage) on the baseline *dynamic fingerprint* model of the structure. The high strain energy density is concentrated at the localized area of the damage.

3.0 ANALYTICAL SIMULATION

This method has already been verified for a cantilever beam structure subjected to transverse dynamic loading [5]. The following sections discuss the analytical application of the method to an eight-bay cantilevered truss structure and a "free-free" rectangular plate structure. The eight-bay truss structure is the same truss for which a complete set of experimental data has been acquired [6]. Experimental data was generated for the rectangular plate.

3.1 TRUSS STRUCTURE REDUCED DEGREE-OF-FREEDOM MODEL

As a case study of an analytical structure, an eight-bay cantilevered aluminum truss structure is modeled [6]. The eight-bay truss structure is depicted in Figure 1. The truss contains 36 nodes and 109 elements. Four nodes are restrained in all six degrees of freedom. From the remaining 32 nodes, sixteen nodes from alternating bays are chosen to define a reduced degree-of-freedom set for the baseline dynamic fingerprint model. These degrees-of-freedom are in the transverse directions (global x- and z-directions). The first five modes of the eight-bay structure are listed in Table 1. The parenthetic (a) and (b) designate the two principal axes of bending of the truss.

A 10 lbf harmonic excitation force is applied at node 4 in the x-direction to generate a forced response. A 0.01 viscous damping ratio is assumed for the forced response determination. The magnitudes of the response displacements at the resonant frequencies are used to create an enforced displacement load case for use in Guyan expansion to calculate the full displacement set and the element strain energies. Five damage cases are considered for evaluation. Each of the damage cases of the eight-bay truss structure are simulated by removing one member of the truss. Forced responses are calculated, displacements enforced on the undamaged structure, Guyan expansion performed, and the normalized modal strain energy densities calculated. Again, the difference in normalized modal strain energy density identifies the damaged location of the truss structure. Figure 2 depicts the difference in normalized element modal strain energy densities for damage case A (element 46 removed).

Table 1: Normal Modes of the Eight-Bay Cantilevered Truss Structure

Mode	Description	Frequency (Hz)
1	First Bending (a)	14.02
2	First Bending (b)	14.54
3	First Torsion	49.52
4	Second Bending (a)	65.98
5	Second Bending (b)	70.65

Mode 2 for the undamaged truss structure becomes mode 1 when element 46 is removed (lowest natural frequency). Figure 3 depicts the difference in normalized element modal (mode 3) strain energy densities for damage case I (element 99 removed). Note that for both of these damage cases two elements are identified as possible damage areas. In both cases the element removed to simulate damage is one of the two elements identified. The second element is an adjacent element.

3.2 PLATE STRUCTURE REDUCED DEGREE-OF-FREEDOM MODEL

As a second case study, a rectangular steel plate is analyzed before and after a saw-cut was made in the structure to simulate a fault. The structure measures approximately 12 inches by 22 inches and has a thickness of 0.375 inches. Damage is simulated by making a one inch saw-cut in the structure as shown in Figure 4 (free edge between elements). Table 2 lists the modal parameters of the plate before and after the saw-cut.

An impact test is simulated choosing the reference location at node 5 of the structure. For the forced response calculation, a damping ratio (viscous) was chosen to be 0.001. The response displacements due to the impact force are calculated at 45 nodal locations on the plate. The magnitudes of the response displacements are used with Guyan expansion to determine all displacements. The displacements are used to calculate the element strain energy densities for each of the first twelve modes. The element strain energy densities are then normalized to create the baseline dynamic fingerprint of the structure. This process is then repeated for the structure with the damage simulated by the saw-cut (free edge between elements).

It should be noted that the total element strain energy in a damage case is generally greater than the total element strain energy in the baseline case. This is because the undamaged structure is enforced to deform in an unnatural manner. By forcing the baseline structure to deform in this unnatural manner, high strain energy density results in the element(s) where the damage exists. The element strain energy densities are then normalized and compared to the normalized element strain energy densities of the undamaged baseline dynamic fingerprint structural model. Figure 5 depicts the difference in normalized element strain energy density for the first 12 modes of the saw-cut damage case. The unshaded or clear elements have a decrease in normalized strain energy density. Figure 6 is a nodal strain energy contour plot for the same normalized element strain energy density as in Figure 5. element in the plate.

Table 2: Modal Parameters of Rectangular Plate Before and After Damage Simulation

Mode	Description	Undamaged (Hz)	Damaged (Hz)
1	First Bending (a)	163.464	162.949
2	First Torsion	184.301	183.599
3	First Antisym. Bending (a)	409.682	407.910
4	Second Bending (a)	451.263	448.333
5	First Bending (b)	553.911	553.885
6	First Antisym. Bending (b)	671.366	669.958
7	Second Antisym. Bending (a)	722.182	717.620
8	3rd Bend (a), 1st Bend. (b)	885.140	880.536
9	Oil Can Mode	967.033	963.541
10	Third Antisym. Bending (a)	1153.765	1146.421
11	Second Oil Can Mode	1337.758	1333.070
12	Fourth Bending (a)	1490.515	1488.228

By plotting the difference in normalized element strain energy density, both of these figures correctly identify the location of the saw-cut.

4.0 EXPERIMENTAL RESULTS

Experimental data for the eight-bay truss structure was also available in the form of frequency response functions from baseline and damage simulation testing using dual input random excitation [6]. These frequency response functions are used with both full and reduced degree-of-freedom set (y-direction now included) to identify the damage locations.

4.1 EIGHT-BAY TRUSS STRUCTURE FULL DEGREE-OF-FREEDOM MODEL

All damage cases are clearly identified using the full measured degree-of-freedom data, except damage case F. Case F could not be identified because there is minimal strain energy in this element for each of the first five modes.

4.2 TRUSS STRUCTURE REDUCED DEGREE-OF-FREEDOM MODEL

Figures 7 through 10 depict the difference in normalized modal strain energy for the first five modes for the reduced degree-of-freedom eight-bay truss structure. The reduced degree-of-freedom set is the same set used in the analytical damage investigation of the eight-bay truss (see Section 3.1). These figures show the difference in normalized

element modal strain energy densities for the damage cases C, I, L, and N. Damage case L is clearly identified using the reduced degree-of-freedom data. Damage cases C, I, and N are identified to within two elements, one being the correct element and the other, an adjacent element. Other damage cases not presented here displayed similar trends. As expected, damage case F could not be identified with the reduced degree-of-freedom set (using the first five modes). It should be noted that this reduced degree-of-freedom set was the only set utilized. Other sets, perhaps having fewer degrees-of-freedom, could be more adept in identifying damage.

4.3 PLATE STRUCTURE REDUCED DEGREE-OF-FREEDOM MODEL

An impact modal test is conducted for the "free-free" rectangular plate before and after damage. Damage to the plate is in the form of a "crack" simulated by making a one inch saw-cut at a perimeter location 7 inches from corner node 5 on the long side of the plate (see Figure 4). A reference accelerometer is located at node 5 and frequency response functions are measured across the plate (see Figure 4) for both the undamaged and damaged plate. From the data, twelve deformation modes are identified in the measured frequency band of 0 to 1500 Hz. These undamaged and damaged deformation mode shapes are enforced on the FEM model of the undamaged and a full displacement set is calculated using Guyan expansion. From the full displacement set, normalized modal strain energy densities are calculated for each element for both the undamaged and damaged conditions. Figure 11 depicts the difference in normalized element strain energy density for the first twelve modes for the saw-cut damage case. The unshaded or clear elements have a decrease in normalized strain energy density. Figure 12 is a nodal strain energy contour plot for the same normalized element strain energy density as in Figure 1. Both figures, again, correctly identify the location of the saw-cut.

5.0 CONCLUSIONS AND RECOMMENDATIONS

This research establishes a procedure for identifying damaged locations in a structure by enforcing displacement data from a damaged structure on the undamaged baseline *dynamic fingerprint* model of the structure. The method assumes that damage is manifested as a localized loss of stiffness with no loss of mass. The physical equivalent of this assumption could be a fatigued member, loosened bolt(s), weakened weld(s), or delaminated composite material. The important modes for locating damage are the modes in which the damaged member participates (possesses significant strain energy) in the response. The procedure is illustrated in the frequency domain, but might also be applicable in the time domain or statically. The expansion technique does not involve the mass and inertia properties of the structure for the full degree-of-freedom displacement calculation. This should be included for a more accurate displacement determination. The method can be extended to not only identify, but quantify, the amount of stiffness lost in a damage case by using a correlation procedure such as the error vector method, the stiffness error matrix method, or the force balance method.

ACKNOWLEDGEMENTS

The authors would like to sincerely acknowledge the support of Dr. Suzanne W. Smith for her assistance in acquiring and interpreting the experimental test data.

REFERENCES

1. Hajela, P. and Soeiro, F.J., "Structural Damage Detection Based on Static and Modal Analysis," *Proceedings 30th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Mobile, Alabama, 1989.

2. Smith, S.W. and Hendricks, S.L., "Damage Detection and Location in Large Space Truss Structures," *Proceedings of the AIAA SDM Issues of the International Space Station*, Williamsburg, Virginia, 1988.

3. Zimmerman, D.C. and Smith, S.W., "Model Refinement and Damage Detection for Intelligent Structures," book Chapter in Intelligent Structural Systems (H.S. Tzou, editor), Kluwer Academic Publishers, 1992.

4. Zimmerman, D.C. and Kaouk, M., "Structural Damage Detection Using Measured Modal Data and No Original Analytical Model," *Proceedings of the 12th International Modal Analysis Conference*, Honolulu, Hawaii, 1994.

5. Ballinger, R.S. and Herrin, D.W., "Design Health Monitoring of Mechanical Dynamic Structures," *Proceedings of the 13th International Modal Analysis Conference*, Nashville, Tennessee, February, 1995.

6. Kashangaki, T.A. "Ground Vibration Tests of a High Fidelity Truss for Verification of on Orbit Damage Location Techniques," NASA-TM-107626, May 1992.

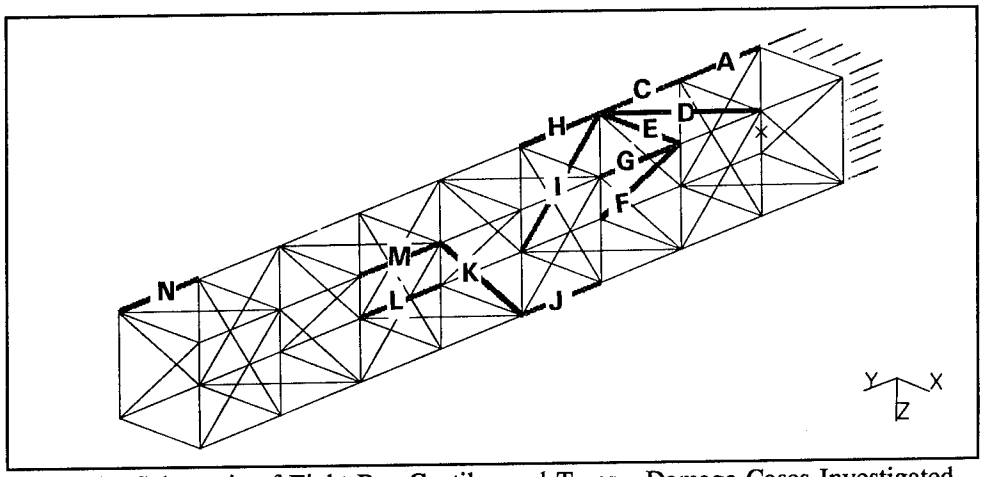


Figure 1: Schematic of Eight-Bay Cantilevered Truss - Damage Cases Investigated

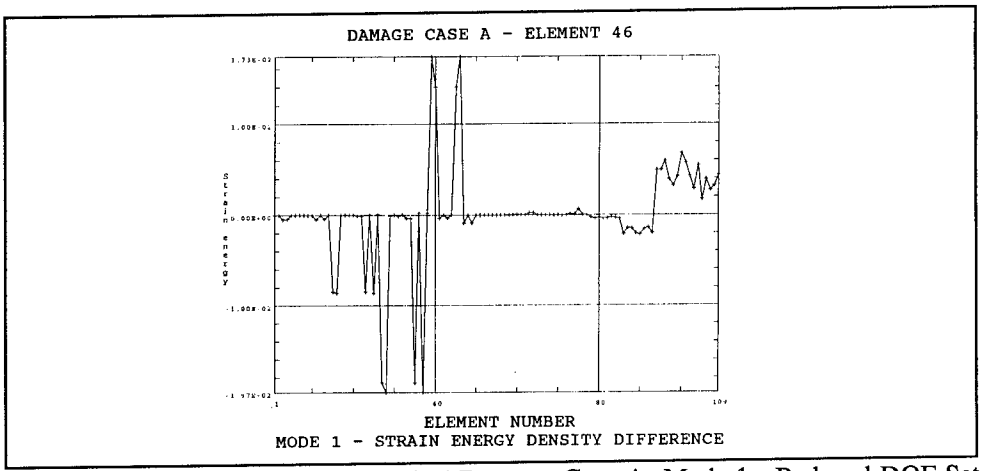


Figure 2: Eight-Bay Truss - Analytical **Damage Case A**, Mode 1 - Reduced DOF Set

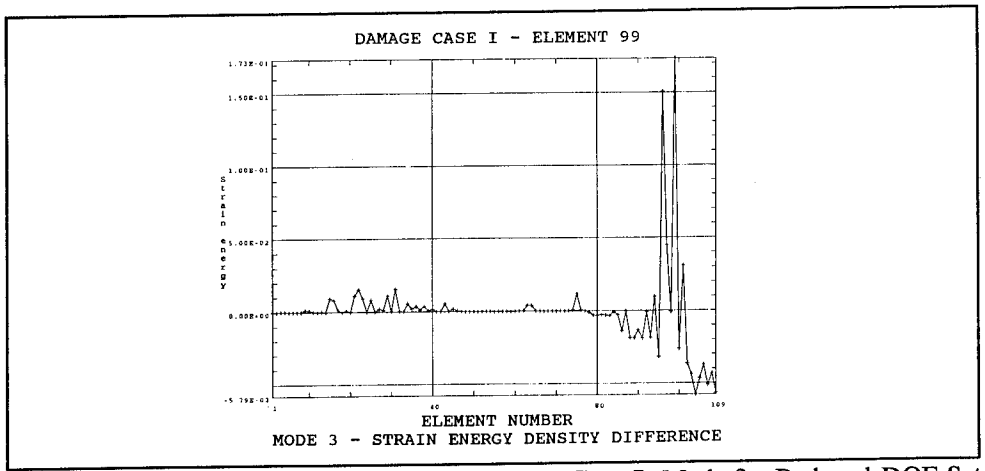


Figure 3: Eight-Bay Truss - Analytical **Damage Case I**, Mode 3 - Reduced DOF Set

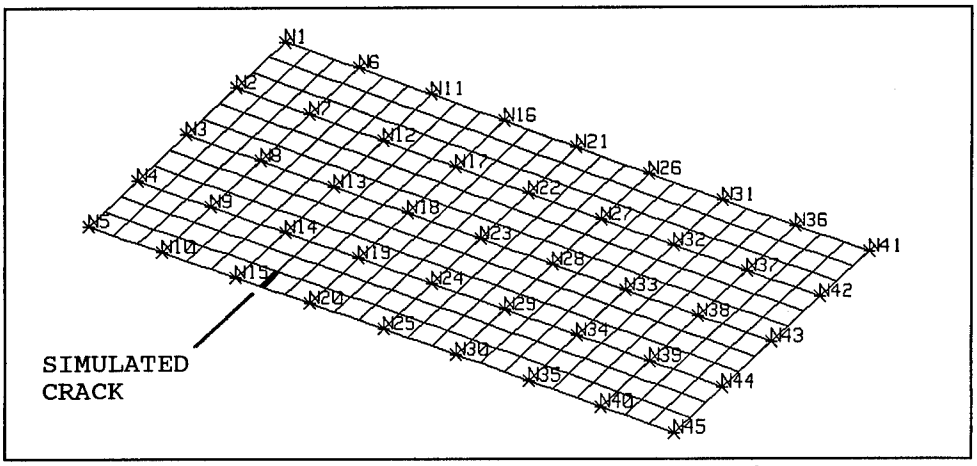


Figure 4: Plate FEM Model, Test Points Grid and Damage Location

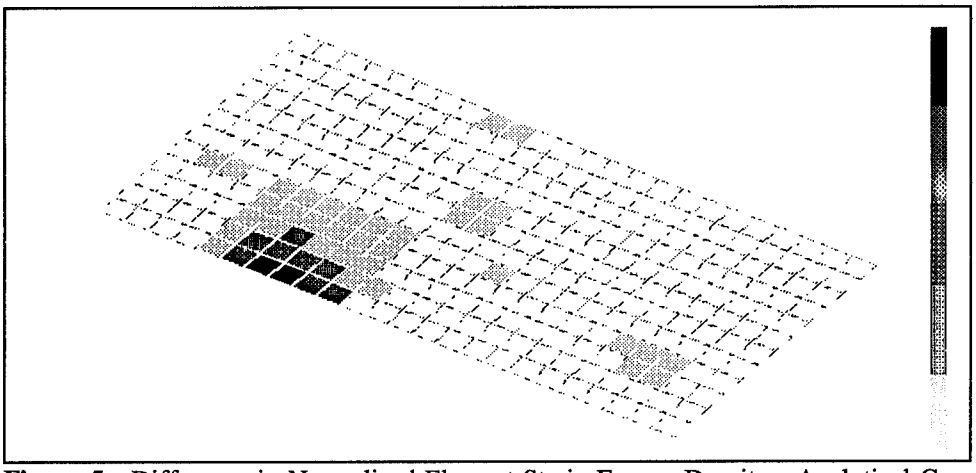


Figure 5: Difference in Normalized Element Strain Energy Density - Analytical Case

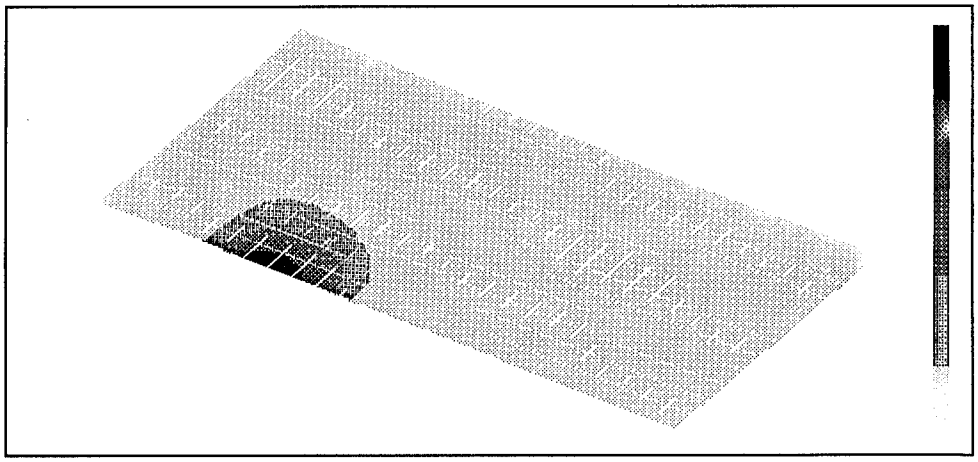


Figure 6: Difference in Normalized Strain Energy Density - Analytical Case

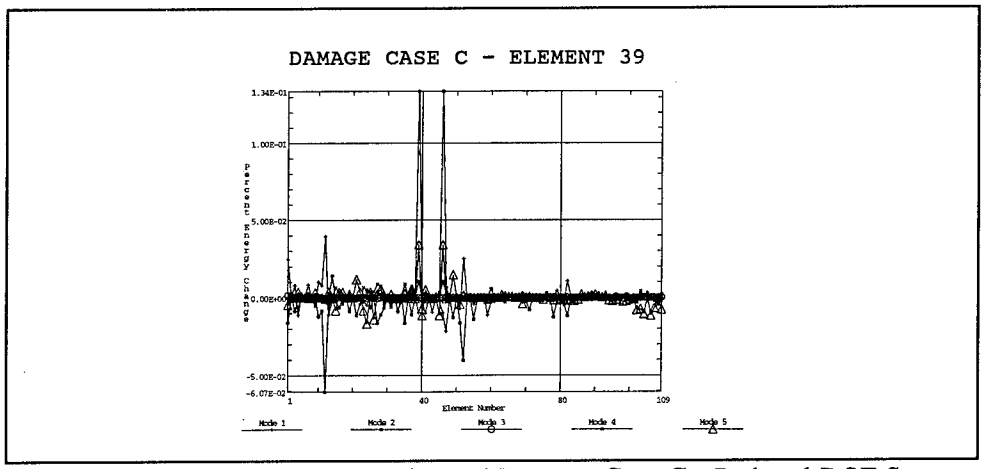


Figure 7: Eight-Bay Truss - Experimental **Damage Case C** - Reduced DOF Set

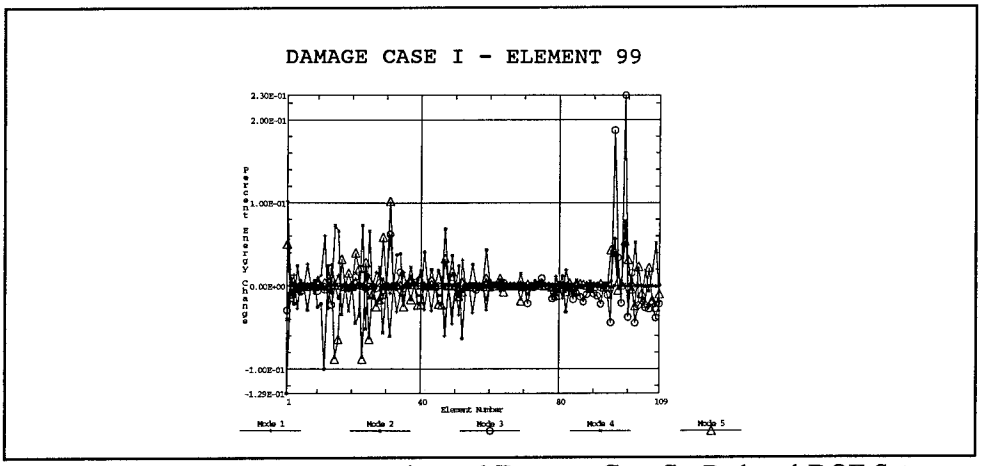


Figure 8: Eight-Bay Truss - Experimental **Damage Case I** - Reduced DOF Set

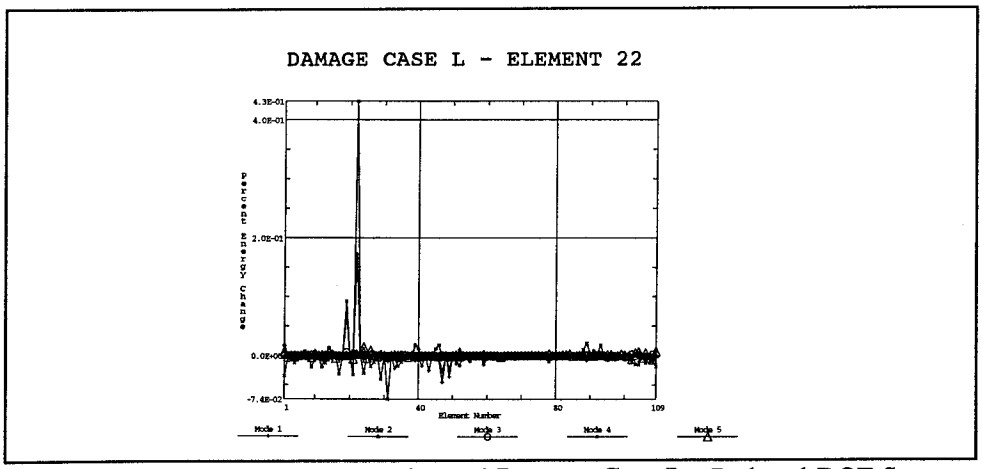


Figure 9: Eight-Bay Truss - Experimental **Damage Case L** - Reduced DOF Set

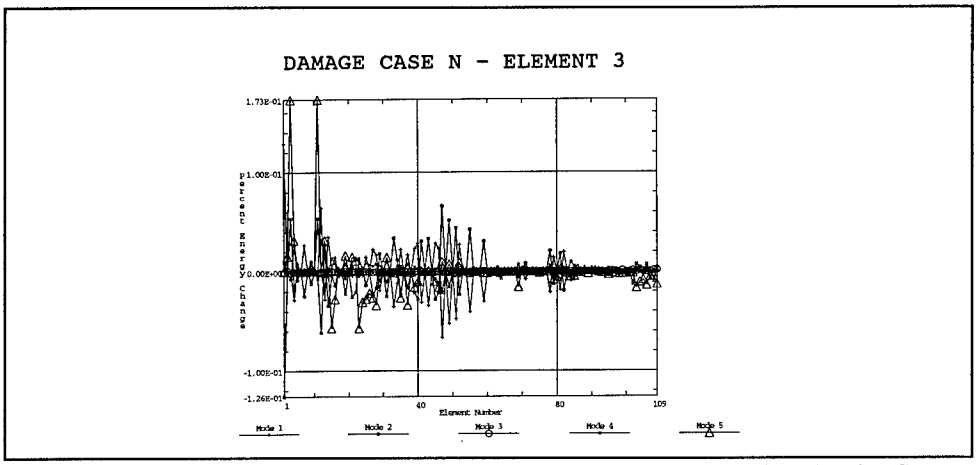


Figure 10: Eight-Bay Truss - Experimental **Damage Case N** - Reduced DOF Set

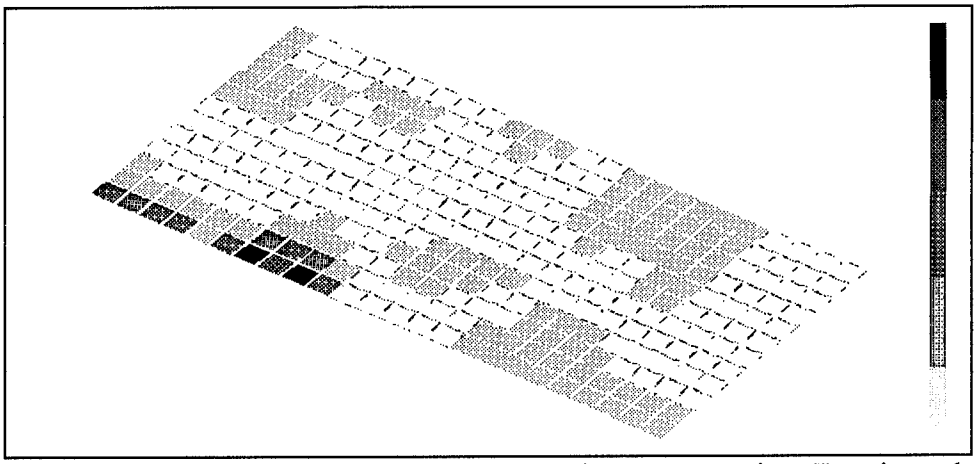


Figure 11: Difference in Normalized Element Strain Energy Density - Experimental

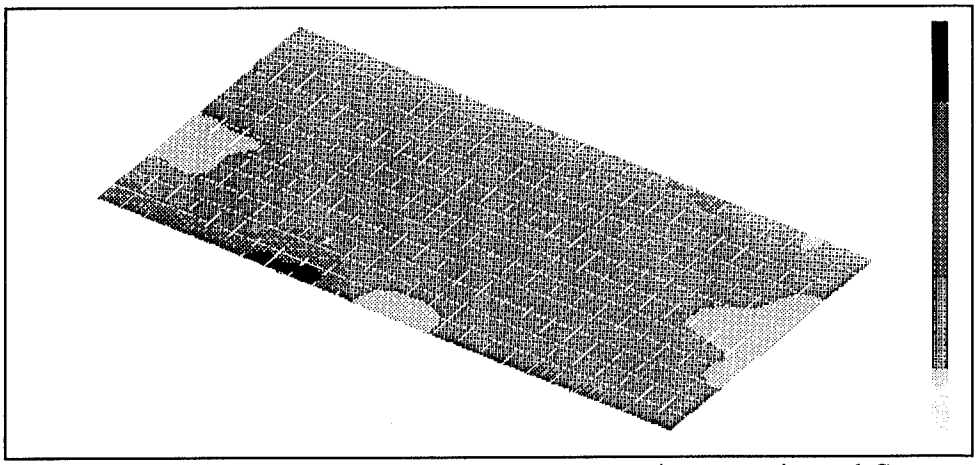


Figure 12: Difference in Normalized Strain Energy Density - Experimental Case

LIFE EXTENSION THROUGH HOLE COLD EXPANSION TECHNOLOGY

Len Reid, Vice President, Engineering
Fatigue Technology Inc.
100 Andover Park West
Seattle, WA, 98188 USA

Abstract: Safety, durability and structural integrity are basic design philosophies in the aerospace industry. Long life, safe operation and low life cycle costs are the ideals of the airline operators. Through necessity, the aerospace industry has exploited technology and design optimization to attain these goals. Hole cold expansion technology evolved to overcome fatigue problems by generating compressive residual stresses around the hole to effectively shield it from cyclic tensile loads. It has since become the most widely used method of preventing fatigue crack growth originating at holes in both new and aging aircraft structures. Derivative processes that expand bushings into holes at high interference levels have virtually overcome problems such as migration, fretting, rotation and corrosion associated with bushings under load. By simultaneously cold expanding the material around the hole, the fatigue life and durability of bushed installations is greatly increased compared to other bushing methods. This paper describes the basic cold expansion technology and processes. It discusses applications in the aerospace and some non-aerospace industries and how they could be applied to aging machinery and structures as a cost-effective method of failure prevention and service life extension.

Key Words: Bushing; Cold Expansion; Crack Growth; Fatigue; Residual Stress.

Introduction: Nearly all mechanical machinery and devices used in industrial, manufacturing, power generation and transportation industries are subject to dynamic and cyclic loading conditions that are conducive to fatigue damage or component wear and fretting. To overcome this, equipment designed to operate under high loads and stresses is invariably heavy or incorporates considerable reinforcement to reduce the magnitude and effects of cyclic loading, especially in the vicinity of holes or attachment lugs. Processes developed and used by the aerospace industry to minimize the effect of cyclic tensile loading could be effectively applied to mechanical systems to optimize design, reduce weight, increase structural reliability and prevent costly mechanical failure.

Background: With the advent of the jet age, the aerospace industry was challenged with optimizing structural design for long life and low weight. Structural fatigue became a serious issue and is still an on-going focus of investigation, research, and debate. This paper focuses on the Split Sleeve Cold Expansion$_{TM}$ (SsCx$_{TM}$) process developed over 25 years ago for the aerospace industry to improve the fatigue life and durability of fastener holes in metallic structure. The process, which has been exhaustively validated through test and in-service experience, provides fatigue and crack growth life improvement through the creation of residual compressive stresses in the metal surrounding the hole. The beneficial stresses work by effectively reducing the stress concentration effect at the hole.

The split sleeve process, conceived by The Boeing Company in the late 1960's and developed as an integrated system by Fatigue Technology Inc. (FTI) [1], has also been successfully applied to a number of non-aircraft applications over the years. Some of the areas this technology is used is in aircraft structures, gas turbine engines, railroad rails, the superstructure of maritime ferries, intermedullary medical nails, wind turbines and mining and drilling equipment. FTI has also

adapted the process to a variety of metallic materials from aluminum alloys to high and ultra high strength steels, nickel based alloys, cast iron and titanium.

The method of cold expanding holes has been further adapted for expanding bushings into mounting lugs, hinges and fittings and also for installing bushings to repair or resize holes which have been damaged by fatigue cracks, fretting or elongation. The two techniques known as ForceMate® ($FmCx_{TM}$) and BushLoc® ($BlCx_{TM}$), developed by FTI, involve radially expanding an initially clearance fit bushing or insert into the hole thereby imparting a high level of interference. In most cases, this process also cold expands the material around the hole to generate a residual compressive stress zone around the bushing. This combination enhances bushing retention, prevents fatigue cracking, and virtually eliminates migration and corrosion problems associated with bushing installations. These processes greatly enhance the integrity of a bushed hole or fitting.

This paper will discuss the principles of the cold expansion processes and the techniques used. Examples will also be presented of where the technology is currently applied, along with suggestions as to where it could be incorporated in other machinery and equipment to prevent failure, increase durability and inspection cycles, and improve production and repair techniques. Potential use of cold expansion to replace other mechanical or hydraulic techniques such as shot peening, roller burnishing and autofrettage will be discussed.

DEVELOPMENT OF LIFE ENHANCEMENT METHODS

Historical Perspective: Throughout modern industrial history, holes in metals and other materials have been the greatest source of fatigue problems. Holes, which are unavoidable in most structural designs and machinery, typically multiply the magnitude of the applied stresses by a factor of three or more and are inevitably the weakest part of the structure. It is also a well-established fact that holes contain manufacturing defects from drilling or reaming, or microscopic inclusions, which begin to grow as fatigue cracks soon after being subjected to service loads. The addition of cyclic tensile loading, such as that produced by a wing of an aircraft in flight, a railroad wheel passing over a bolted joint, or a bellcrank, frequently causes crack initiation, growth, and eventually fatigue failure.

Traditionally, manufacturers overcame fatigue problems by using thicker sections and better designed joints to reduce stress levels. In some cases, a complete redesign with higher strength material was required to correct the problem. Weight increase was the usual penalty for fatigue life improvement. In the aircraft industry in the 1950's, as aircraft range and payload requirements increased, fatigue cracking at holes became more prevalent because of higher operating stress levels. Mandrelizing, broaching and ballizing [2] were some of the techniques used at the time in an attempt to reduce the effect of manufacturing defects and the in-service cyclic stresses at fastener holes. These processes resulted in inconsistent, if any, fatigue life improvements. In the 1960's, the wider use of high-strength metals added to the fatigue and crack growth problems because of the higher crack sensitivity of these materials. In the 1970's, specialty high-interference fit fasteners were introduced as the preferred means of alleviating fatigue problems. These fasteners reduced the effective magnitude of the cyclically applied load by tensile pre-stressing the material around the hole. However, their effectiveness was very dependent on how well they were installed and the level of interference fit.

The creation of permanent compressive stresses near holes was recognized as a means of extending fatigue life of joints in fatigue critical components. The compressive stresses retard

crack initiation and growth by greatly reducing the stress ratio (ratio of the minimum to maximum stress levels) of the applied load. Methods commonly used to induce surface compressive residual stresses include shot peening, roller burnishing, coining, and extrude honing. However, these techniques produce only relatively shallow residual compression zones which are sensitive to manufacturing variables and, often, operator proficiency. Consequently, these surface treatment methods have only a limited ability to effectively prolong fatigue life, especially in holes.

High-interference hole cold expansion was developed for treating holes because unlike the previous processes, hole cold expansion: 1) generates a large, controllable residual compression zone with high compressive stresses; 2) is not significantly affected by hole quality and surface finish; and 3) has measurable parameters for process quality assurance. Over the past 25 years in the aerospace industry, hole cold expansion has become the most widely used and cost effective method of controlling initiation of fatigue cracks and retarding the growth of cracks from defects in holes.

The Cold Expansion Process: High-interference cold expansion of holes is accomplished by drawing an oversize tapered mandrel, which is pre-fitted with a lubricated split sleeve or bushing, through a hole (see Figure 1). The mandrel and sleeve are designed to generate a prescribed optimum amount of radial expansion and plastic deformation around the hole. The disposable sleeve protects the hole, reduces mandrel pull force, ensures radial expansion, and allows one-sided (blind) processing.

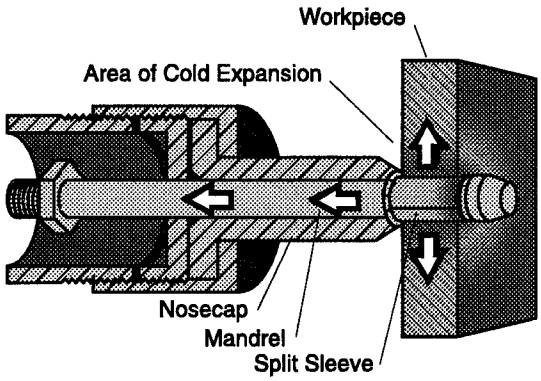


Figure 1. Split Sleeve Cold Expansion.

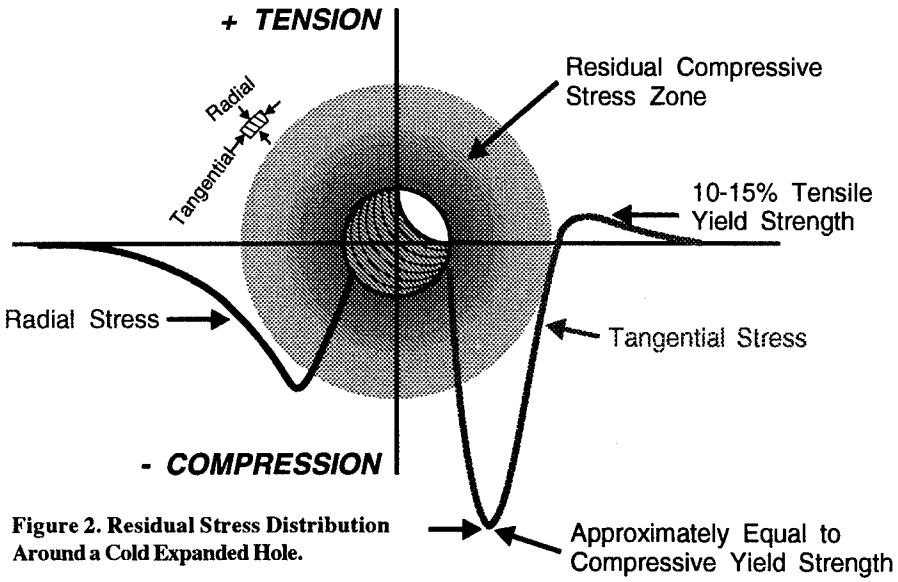


Figure 2. Residual Stress Distribution Around a Cold Expanded Hole.

FTI tests show the optimal fatigue performance is achieved when the hole is expanded by at least 3 percent for aluminum and mild steel, and at least 4.5 percent for titanium and high-strength steel, in typical hole diameters (up to 1-inch) and plate thicknesses. In a typical application of this process, the peak compressive stress near the edge of the hole is generally equal to or greater than the compressive yield strength of the material. The residual (tangential) compressive stress zone extends radially from the edge of the hole approximately one hole radius to one diameter, as shown in Figure 2. The compressive stresses are balanced by a zone of tensile stresses; approximately 10 to 15 percent of the material's tensile yield strength. The magnitude and location of the balancing tension zone are such that unforeseen surface cracking or stress corrosion cracking is precluded.

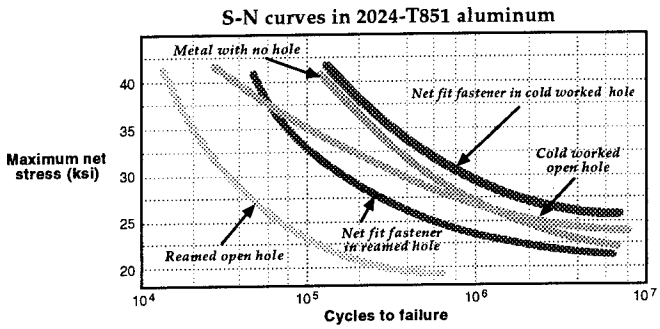


Figure 3. Fatigue Life Improvement - 2024-T851 Aluminum Alloy.

Cold expanded hole fatigue lives generally range from 2 to 10 times the fatigue life of similarly non-cold expanded holes is shown in the examples of Figure 3 for aluminum alloy [3] and Figure 4 for titanium [3].

Affect on Crack Growth: In most structures when cracks are found emanating from holes, usually the hole is enlarged to remove the damage and the hole is bushed or an oversized fastener is reinstalled. Hole oversizing will generally increase the localized stress levels which, when combined with the high probability of leaving undetected residual cracks in the holes, can lead to further accelerated cracking. Hole cold expansion has been shown by tests to arrest the growth of these small undetected cracks. Figure 5 shows that cracks of 0.040-inches in 7075-T6 aluminum alloy are totally arrested when subjected to the same applied cyclic loads. The residual compressive stress zone acts like a strong clamp on the material ahead of the crack. Crack growth is slowed by effectively reducing the stress intensity factor (K_1) associated with the crack [4].

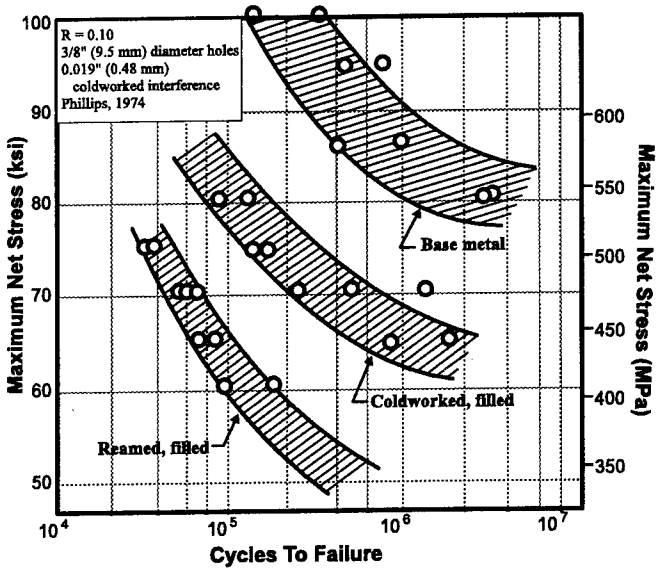


Figure 4. Fatigue Life Improvement - Titanium (6Al-4V)

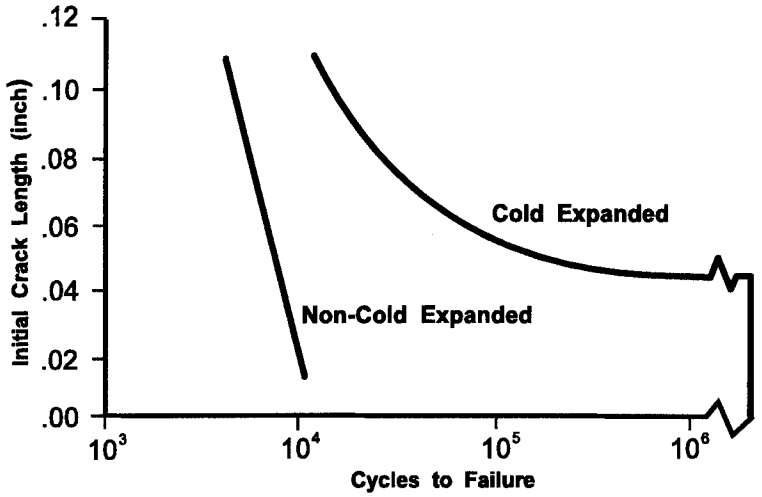


Figure 5. Effect of Cold Expansion on Stopping Crack Growth (7075-T6 Aluminum

To illustrate the effectiveness of this process in arresting, or significantly retarding, the growth of pre-existing cracks, the following actual aerospace coupon test programs in both titanium and aluminum alloy serve as good examples. The first test performed on titanium (Ti 6Al-4V) by Grumman Aerospace Corp. (now Northrop Grumman) shows the effect of cold expansion on crack growth life of a 0.200-inch diameter hole using a severe fighter spectrum load with a maximum stress of 82 ksi. Figure 6 [5] clearly shows the crack growth life (or measure of crack retardation) from a pre-existing 0.030-inch crack was increased by 14:1 using Split Sleeve Cold Expansion (Cx).

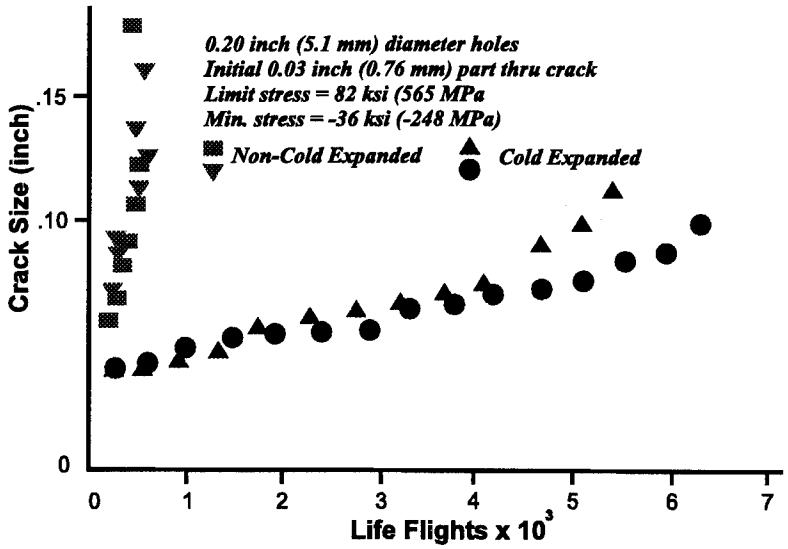


Figure 6. Crack Growth Life Improvement (Ti 6Al-6V-25n).

These modified crack growth characteristics also serve as the basis for determining structural maintenance planning for many military and some commercial aircraft operators. By determining the durability and damage tolerance characteristics of materials and the effectiveness of hole cold expansion in retarding crack growth, structural fatigue lives and inspection intervals can be

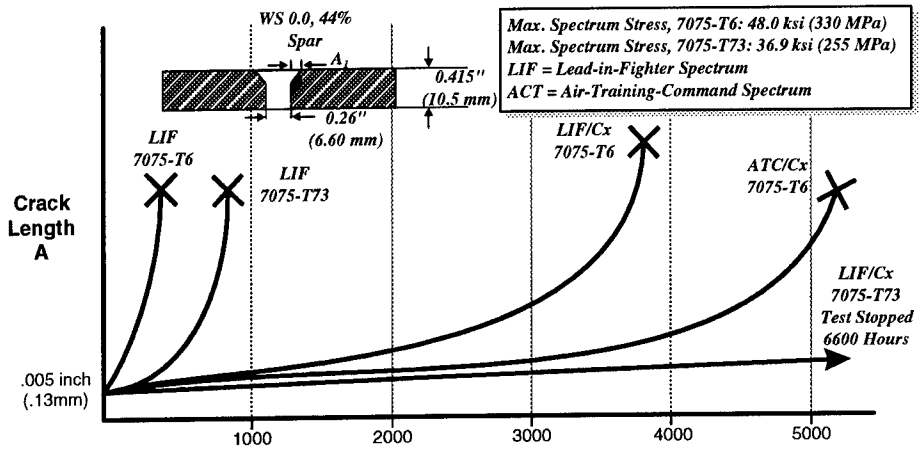


Figure 7. USAF T-38 Fatigue Crack Growth Data.

more accurately predicted. The second example shows how the retarded crack growth with cold expansion increased the inspection interval for a trainer aircraft.

As part of the USAF Durability and Damage Tolerance Analysis (DADTA) philosophy, each fatigue-critical hole is analyzed, assuming an initial rouge flaw (crack) size of 0.050-inch, to determine the safe residual operating life. When holes are cold expanded, a smaller initial flaw size (as small as 0.005-inch) can be assumed. Advantage can now be taken of the slow initial crack growth period to extend the life of the structure. Crack growth tests for the USAF T-38 trainer aircraft, in pre-countersunk holes, substantiated this decision after 0.005-inch flaws were virtually arrested in 7075-T73 aluminum alloy (see Figure 7 [6]). The results showed that inspection cycles could be greatly extended or, in fact, the fatigue life was extended beyond the projected service life of the aircraft when hole cold expansion was used in conjunction with a revised heat treated material. The same philosophy could be applied to determining inspection cycles and repairs on aging commercial machinery.

Stop Drill Repairs: One of the most effective uses of Split Sleeve Cold Expansion in maintaining or repairing aircraft structures is in the enhancement of stop drilled hole repairs. The FTI $StopCrack_{TM}$ ($ScCx_{TM}$) process involves drilling a hole ahead of a crack and then Split Sleeve Cold Expanding the hole by five or six percent of its initial hole diameter. The resulting residual compressive stresses around the hole effectively shield the hole from subsequent cyclic tensile stresses, thus retarding further crack growth. Although it is recommended as an interim repair action, StopCrack dramatically extends the fatigue life of the cracked structure. Testing is currently being conducted to apply the process to steel bridges and was successfully used to prevent further crack growth on the steel superstructure of Washington State ferries. It could be applied to cracked mechanical machinery as an interim repair or part of a reinforcing patch or doubler repair to enhance the durability of the final repair.

DERIVATIVE PROCESSES

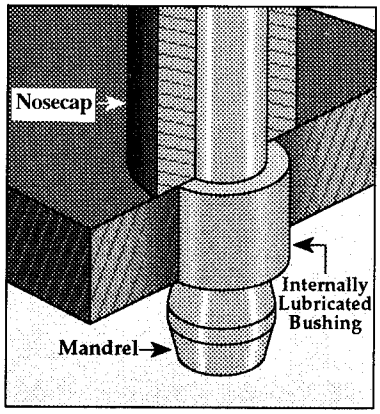


Figure 8. The ForceMate Bushing Installation Process.

Bushing Installations: A significant variation of the hole cold expansion process, known as FTI ForceMate (FmCx), involves permanent installation of bushings or thick-walled inserts at high interference levels [7]. This is accomplished by expanding a custom-made, pre-lubricated bushing into the hole as shown in Figure 8. The initially clearance-fit bushing (or insert) is locked in place by the mandrel-induced expansion and the resulting residual stresses in the surrounding material. The bushing is usually reamed afterwards to its final desired inside diameter, although it may also be expanded to size. The process can be performed with access to one side of the hole only. An important benefit of this technique is that the material surrounding the hole is simultaneously cold expanded as the bushing is installed due to the high expansion level. Bushings made from a variety of materials (steel, stainless steel, beryllium copper, aluminum nickel bronze, etc.) in diameters up to 4-1/2-inches have been successfully installed using ForceMate.

A derivative bushing installation known as BushLoc (BlCx), developed by FTI, is typically used as a hole repair procedure but may be applied to any bushing installation. Again, the initially clearance-fit bushing is radially expanded into the hole; however, the expansion mandrel is fitted with a pre-lubricated split sleeve, similar to that used for cold expansion of holes, instead of using a lubricated bushing (Figure 9). Although this process is not as tightly toleranced as ForceMate and does not result in as high an expansion or interference fit, the final interference level is much higher than can be achieved by thermal (shrink) fitted bushings. BushLoc also facilitates installation of locally manufactured bushings in a similar range of materials to ForceMate (FmCx) including some aluminum alloys.

In both methods, the resulting high-interference fit and degree of local cold expansion of

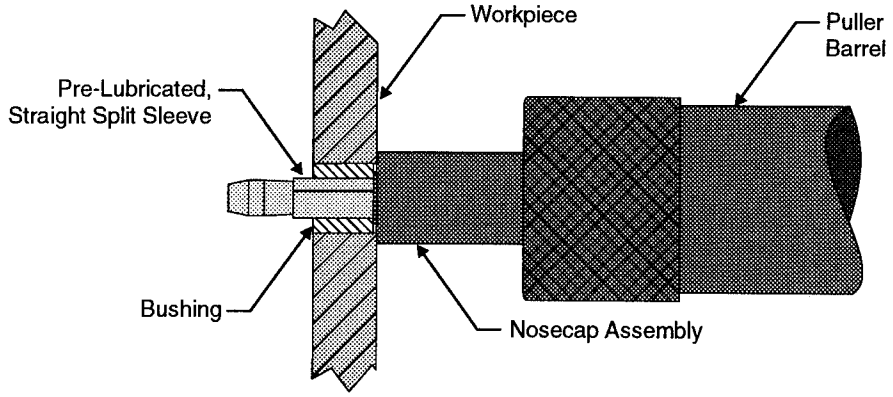


Figure 9. The BushLoc Process.

surrounding material produce a bushing installation that will resist bushing migration and movement (which induces fretting), resist corrosion by preventing installation damage (galling and scoring), and provide exceptional fatigue life of the bushed installation (see Figure 10).

REVIEW OF COLD EXPANSION APPLICATIONS

Although the hole cold expansion process was developed for the aerospace industries, the benefits of the process for life extension of aging machinery and structures is evident from both the aerospace and non-aerospace applications where the technology is currently used. The processes are applicable to a broad range of materials. FTI has also generated actual material test data and both analytical finite element and computer analysis algorithms to estimate and predict the behavior of the processes for different material characteristics and hole geometries. These analytical models have been verified by test and in-service experience.

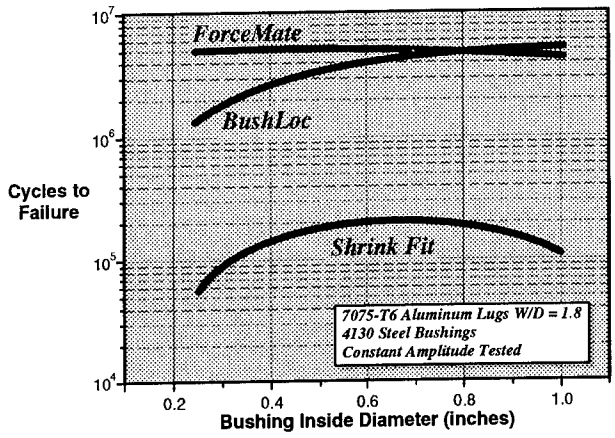


Figure 10. Fatigue Life Comparison of Bushing Installation Methods.

Most machinery and structures are made from various grades of steel. Cold expansion has been used by the railroad industry in steel rails to prevent growth of cracks in rail-end bolt holes. The process has virtually eliminated fatigue cracks thereby greatly extending inspection intervals of rail joints and eliminating a major source of derailments. Also cracks in brake attachment holes of railroad rolling stock were similarly prevented by cold expansion. Similarly, fatigue cracks originating at the corners of the upper steel deck structure of Washington State ferries have now been eliminated by cold expansion of the window attachment holes.

Attachment lugs and truss joints of cranes and similar structures or machinery that is subject to cyclic tensile loads or fluctuating load conditions could benefit from cold expansion if fatigue cracking is found or is a potential source of failure. The cost of inspecting joints and holes for cracks can be very expensive and time consuming. Inspection methods often require specialist inspectors to inspect these fatigue critical holes, which adds to the cost of the task. Besides possibly preventing catastrophic failure of the mechanism or system, cold expansion could also minimize or eliminate these costly inspections.

Hydraulic systems also experience cyclic load conditions that can eventually lead to fatigue cracks in housings, mountings for connectors, and component mounting holes. Cold expansion has been effectively used in aircraft hydraulic components and actuators, both in initial production of these items and field rework operations. The safety and liability consequences of failure of these components, and the downtime to repair them, can be expensive and may be eliminated by use of these processes.

Braking systems suffer similar failures to hydraulic components. The severe operating stresses can also cause loading conditions conducive to fatigue cracking. Even threaded holes can be cold expanded prior to threading to leave residual compressive stresses at the high stress concentration at the base of the thread form. This method is successfully used on aircraft braking systems.

Applications in the paper manufacturing industry are also under consideration. Pulp and paper rolls are very expensive items that operate under very high rotational dynamic loads. Catastrophic-fatigue failure from balance weight attachment holes and the drum are of enormous consequence, not only to the facility, but to the operators in the vicinity of it. Hole cold expansion could be incorporated into potential failure areas during production that would alleviate this potential problem. Retrofit incorporation may be of some value if fatigue cracking has not already exceeded safe operating specifications.

Wind turbines for power generation are already being reworked using cold expansion to correct major fatigue problems in blade joints. This machinery did not seem to be adequately designed to meet the extreme environmental and operating stresses. The operating stresses in the blades is very similar to aircraft wings or helicopter rotator blades. These rotating blades could well benefit from aerospace joint and fatigue design, both in new production units and field repairs.

Both hole cold expansion and ForceMate bushing installation have been successfully demonstrated and used in gas turbine engine components to meet stringent low cycle fatigue (LCF) requirements [8]. The use of these processes in such applications have been very effective even in some high temperature components made from ultra-high strength nickel based alloys. The high interference ForceMate bushings were also shown to be effective in repairing blade attachment holes in rotating compressor discs.

Highly loaded hinges and lugs are frequently bushed to facilitate repair and inspection. ForceMate bushings have proven to be very effective in eliminating fatigue cracks and bushing migration in aircraft engine pylon attachments, landing gear attachments, and wing fold fittings. An application on a mobile military bridge is under evaluation, and it is expected that ForceMate will be the answer to both a bushing migration and a lug fatigue problem. ForceMate or BushLoc could be very effective in reducing life cycle costs and improving the safety of bushed hole/lug applications.

Autofrettage is another method used to generate residual compressive stresses in gun barrels and in the walls of cylindrical pressure vessels or cylindrical tube sections used in fluid mining equipment. Very high hydraulic pressure is used to yield the wall of the cylinder, thereby generating the residual stresses. There are also safety concerns when applying this method. Autofrettage is only marginally effective when compared to mandrel cold expanding the same cylinders. Cold expansion is much safer and could be an effective and less costly alternative to autofrettage where this method is used.

CONCLUSIONS

Several cold expansion processes developed for the aerospace industry to prevent fatigue cracks in holes have been successfully applied to non-aerospace applications. Two processes discussed in this paper may be suitable for applications in aging machinery and structures:

1. Split Sleeve Cold Expansion of holes to prevent fatigue cracks and improve the durability and damage tolerance of holes and joints subjected to cyclic tensile loading conditions, and

2. ForceMate or BushLoc high interference bushing installation, which prevent bushing migration, fretting and corrosion and enhance the overall durability of bushing installations and assemblies.

These two processes are well proven in tests and in-service conditions in a wide range of materials and structural loading conditions. Their use in both new and aging machinery and

structures could provide life cycle cost benefits and enhanced fatigue life performance of holes and bushing installations through:

- enhanced structural integrity and durability,
- added safety and operational assurance through improved mechanical integrity of the structure or component,
- better integrity of attachment,
- reduced maintenance costs through elimination of cyclic fatigue failure,
- better component and structural reliability, and
- reduced inspection costs through extended inspection intervals.

REFERENCES

[1] FTI Split Sleeve Cold Expansion Process Specification FTI 8101C, September, 1994 (first issued June, 1981).

[2] Champoux, R.L., "Fatigue of Aircraft Structures", Fatigue Prevention and Design, Chameleon Press Ltd., London, 1986, pp. 41 et seq.

[3] Phillips, J.L., "Sleeve Coldworking Fastener Holes", Vol.I, 1974, Air Force Materials Laboratory report AFML-TR-74-10.

[4] Clark, G., "Modeling Residual Stresses and Fatigue Crack Growth at Cold Expanded Fastener Holes", Fatigue Fracture of Engineering Materials Structures, Vol. 14, No. 5.

[5] Armen, H., et al., "Elastic-Plastic Behaviour of Coldworked Holes", Proc. 24th AIAA/ ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conf., May, 1983 (AIAA Paper 83-0865-CF), pp. 1-11.

[6] "T-38 Damage Tolerance Assessment Program; NPN 3347 - Crack Growth Test Summary Report", Contract F41608-76-A010, Northrop Corp. Rpt. No. NOR 77-17, October, 1978.

[7] Champoux, R.L., and Landy, M.A., "Fatigue Life Enhancement and High Interference Bushing Installation Using the ForceMate Bushing Installation Technique", 1986 American Society for Testing and Materials (ASTM) Report STP 927, pp. 39-52.

[8] Rufin, A.C., "Extending the Fatigue Life of Aircraft Engine Components by Hole Cold Expansion Technology". International Gas Turbine and Aeroengine Congress and Exposition, Cologne, Germany, June 1-4, 1992, ASME 92-GT-77.

FIBER OPTIC BASED STRUCTURAL HEALTH MONITORING TECHNIQUES

A. D. Kersey and E. J. Friebele

Optical Sciences Division, Code 5600, Naval Research Laboratory
Washington, D.C. 20375

Abstract

Fiber optic sensing and multiplexing techniques provide a powerful enabling technology for the monitoring of the health and integrity of structural components and systems. Applications for this technology include analysis of load induced stresses, deterioration or damage on new or existing structures, such as aircraft, spacecraft, ships, bridges and buildings.

Key Words: Bragg grating; Distributed sensing; Fiber optics; Multiplexing; Structural monitoring; Strain sensor.

Introduction

Optical fiber based sensors offer a number of significant advantages and unique features which make them attractive for use in structural monitoring. These include the capability to sense a variety of physical parameters, the ability profile a measurand (e.g. strain or temperature) over the length of the fiber, to multiplex several measurands onto a single sensing fiber, and high compatibility for embedding in a range of structural materials. In the case of strain monitoring, a fiber sensor can be configured to have a long gage length for path integrating measurements, or with a short gage length for quasi-point monitoring. In either case, the sensors can be multiplexed using a single or multiple fibers to give distributed measurements. In addition to having the capability of being embedded directly within the structural material, fiber sensors can also be surface-mounted on new or existing structures. The ability of fiber sensors to 'profile' the strain at a number of locations distributed throughout a structure or structural component in real-time with very high resolution and wide dynamic range represents a powerful enabling technology for the monitoring of the health and integrity of structural components and systems. This technology will find application in the analysis of load induced stresses, deterioration or damage on new or existing structures, such as aircraft, spacecraft, ships and offshore structures, bridges, buildings, pipelines and storage tanks for example.

Although a wide variety of fiber sensor types are possible, recently, the use of short gage-length quasi-point 'intrinsic' fiber optic sensors have attracted most research interest. One example of such a device is an 'in-line' fiber etalon sensor, which acts as a miniature resonant cavity within a fiber, and the interferometric response of the device can be used to monitor localized strain. Another example of an intrinsic quasi-point sensor is the fiber Bragg grating. These devices can be optically written into photosensitive Ge doped fibers by exposure to a periodic UV interference pattern. Gratings can be fabricated by side exposure using holographic based means or via a diffractive mask, and represent one of the most attractive forms of fiber sensor for embedded sensing within structural materials.

In-Line Interferometric Sensors

Interferometric 'phase-encoded' sensors [1] have been developed for a range of high sensitivity applications, such as acoustic sensors (hydrophones) and magnetic field sensors . Although highly sensitive, for most applications these devices do not represent viable options for embedded sensing. A type of interferometric 'phase' encoded sensor more suitable for embedded sensing is based on interference between light reflected from two closely spaced surfaces which form a short gage-length Fabry-Perot (FP) type sensor. A number of variations of this design have been devised, both intrinsic and extrinsic in nature, as shown in Figure 1. In the intrinsic forms, the Fabry-Perot (IFP) sensor is formed between two reflective interfaces within the fiber. As the cavity is a fiber region, the IFP sensor configuration has the flexibility of supporting either a long [2] or short [3] gage lengths. The Extrinsic Fabry Perot Interferometer (EFPI) is formed by inserting the two fibers into a larger diameter tube [4]. Reflections from the two fiber end-faces form the cavity, which now must be short to achieve reasonable optical efficiencies. This configuration has the obvious disadvantage that the sensor is a physical discontinuity along the fiber, which can lead to stress concentrations when the device is embedded into a structural material. Furthermore, as the fibers are bonded into the tube using an epoxy, the gage length is not accurately determined and a calibration process is required for each sensor fabricated. One advantage of the EFPI configuration, however, is it exhibits a low susceptibility to thermal offset due to the fact that the cavity is in air, and thermal apparent strain arises due only to the thermal expansion coefficient of the hollow silica tube used to align the fibers ($\approx 10^{-7}$/°C).

A more recent development has been the In-Line Fiber Etalon (ILFE) [5], shown in Figure 1.c, which combines the advantages of both the intrinsic and extrinsic Fabry-Perot forms. In the ILFE, an cavity is formed between two fiber endfaces by splicing them to a short length of 'hollow-core' fiber (micro-tube). Figure 2 shows a photograph of such a sensor element. This sensor retains the low thermal sensitivity, has a continuous profile, well defined gage length, and high strength.

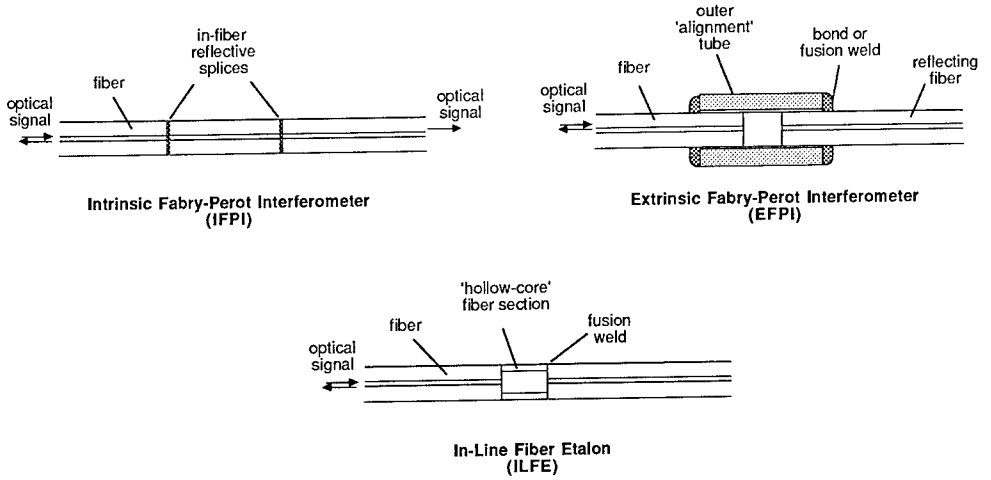


Figure 1. In-line interferometric sensor configurations: a) Intrinsic and b) Extrinsic Fabry-Perots, and c) the In-Line Fiber Etalon Sensor.

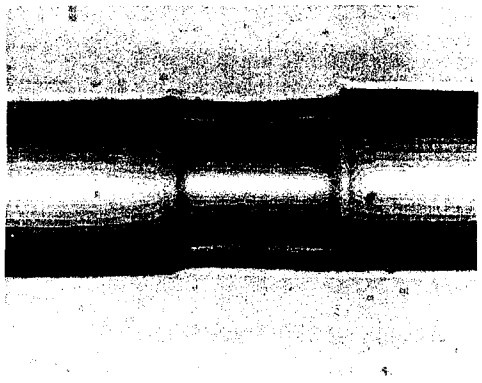


Figure 2. Photograph of the ILFE sensor.

Fiber Bragg Gratings

Fiber Bragg gratings (FBG) represent one of the most exciting developments in the area of fiber optic sensing in recent years, and are currently receiving considerable research interest. The optical transmission characteristics of a fiber grating is a simple narrow band notch filter, the wavelength of which is determined by the Bragg resonance of the device [6]. For light within this resonance condition, a strong reflected optical signal is produced, as illustrated in Figure 3. The Bragg condition is set by the pitch, Λ, of the grating and the refractive index, n, of the glass. When a fiber Bragg grating is strained, the pitch of the grating is altered, and the wavelength at which the Bragg resonance is satisfied shifts. The measurement performed with grating sensors thus reduces to a determination of the Bragg resonance of the device, usually achieved by illuminating the sensor with broadband light and detecting the reflected wavelength. This inherently wavelength-encoded mode of operation of these devices [7] has a number of distinct advantages over other sensing schemes. One of the most important of these is that as the sensed information is encoded directly into wavelength, which is an absolute parameter, the output does not depend directly on the total light levels, losses in the connecting fibers and couplers or source power. In addition, the sensor can easily be operated intermittently without the need for recalibration or re-initialization of the system. The wavelength encoded nature of the output also facilitates wavelength division multiplexing by assigning each sensor to a different slice of the available source spectrum.

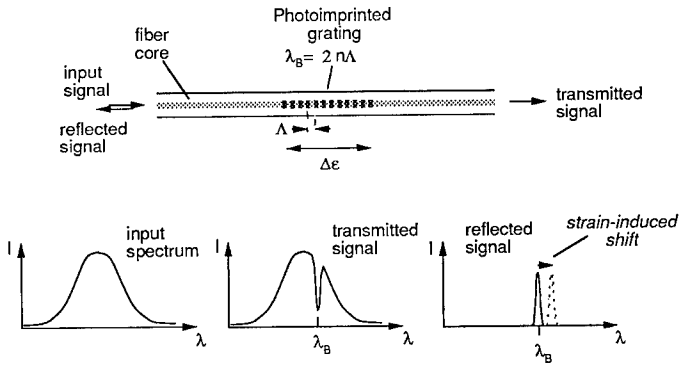


Figure 3. Fiber Bragg Grating (FBG) sensor

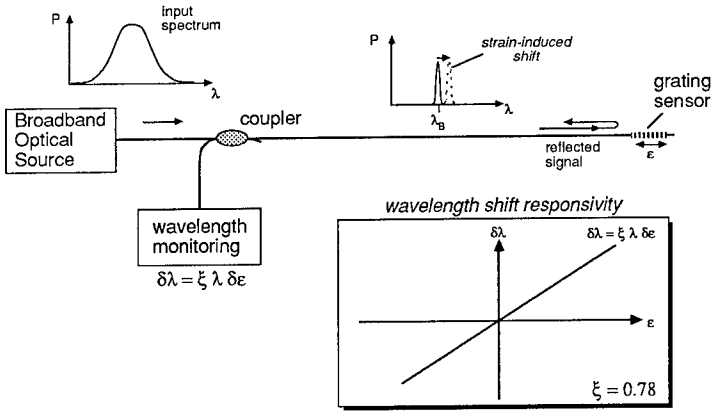


Figure 4. FBG sensor configuration.

The key to a practical sensor system based on FBGs lies in the development low cost fabrication techniques [8,9], and instrumentation capable of determining the relatively small shifts in Bragg wavelength of FBG elements induced by strain or temperature changes in these sensor elements. This area has received significant attention lately [10-15], with a variety of approaches demonstrated. The most straightforward means for the optical interrogation of a FBG sensor element is based on passive broadband illumination of the device, as is depicted in Figure 4. Here, light with a broadband spectrum which covers that of the FBG sensor is input to the system, and the narrowband component reflected by the FBG is directed to a wavelength detection system. Several options exist for measuring the wavelength of the optical signal reflected from a FBG sensor element, including the use of a miniature optical spectrometer, broadband optical "edge" filters, and tunable bandpass filters. Simple filtering techniques based on the use of broadband filters allow the shift in the FBG wavelength of the sensor element to be assessed by comparing the transmittance through the filter to that from a direct 'reference' path [10]. A relatively limited sensitivity is obtained using this approach due to problems associated with the use of bulk-optic components and alignment stability. This can be improved using a fiber device with a wavelength dependent transfer function, such as for example, a fiber optic wavelength-division coupler which splits the light into one of two paths depending on the exact wavelength of the light [11]. One of the most attractive filter based techniques for interrogating FBG sensors is based on the use of a tunable passband filter for tracking the FBG signal. Examples of the type of filter include Fabry Perot filters [12], acousto-optic filters [13], interferometric [14], and FBG based filters [15]. This approach allows a number of gratings sensors located along a fiber path to be 'addressed' using a single instrumentation system. This concept is illustrated in Figure 5. Arrays of 8 or more FBG sensors have been addressed using such interrogation approaches.

By combining time and wavelength division multiplexing [16], where different 'series' of gratings ($\lambda_1 - \lambda_N$) are addressed sequentially by the instrumentation system, the number of grating sensors which could potentially be addressed increases to the tens to possibly one hundred devices (depending on the strain resolution and measurement bandwidths required). This concept is depicted in Figure 6. Such a sensing capability will allow the strain at a large number of locations over a structure to be 'mapped' in real time, providing data regarding the loading, deformation and potential failure of a structure or component. This represents a powerful new enabling technology which will

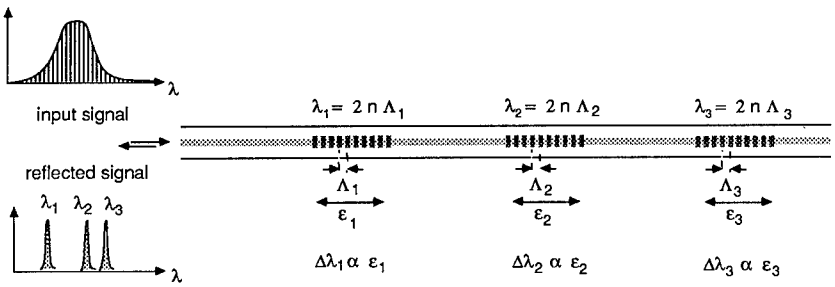


Figure 5. Multiplexed FBG sensor system utilizing wavelength division addressing

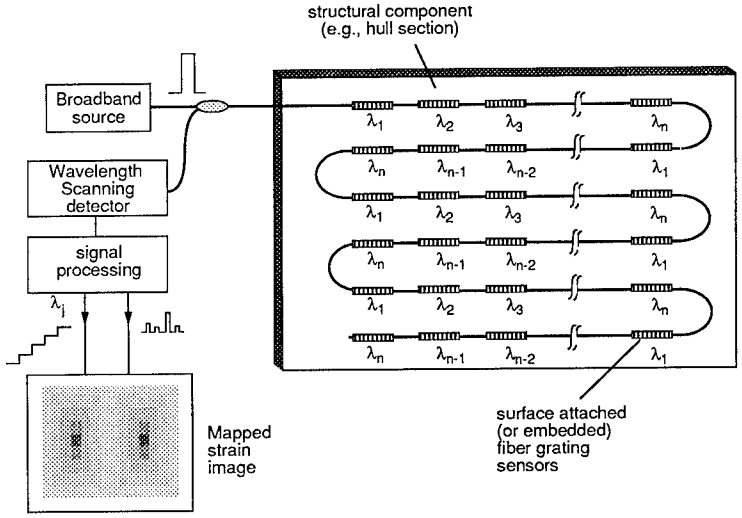


Figure 6. Multiplexed FBG sensor system utilizing wavelength division addressing

impact sensing and structural monitoring in a wide range of applications. Recently developed fiber draw-tower fabrication techniques will allow the production of such FBG arrays at low cost [9], facilitating more widespread usage of these sensors in structural monitoring applications.

Summary

This paper has briefly reviewed two types of fiber optic based sensors for use in structural monitoring of strain. The key issues which will determine the utility of these devices for widespread application in structural monitoring include the size, intrinsic strength, and gage length of the sensor, the ease of interrogation and multiplexing of the device, and the ability to intermittently read the sensor without recalibration or re-initialization of the interrogation system.

Acknowledgements

This work is supported by the Office of Naval Research, the Advanced Research Projects Agency, and the Federal Highway Administration.

References

1. A. Dandridge and A. D. Kersey, *Proc SPIE* , **985** (1988) 34
2. A. D. Kersey et al., *Opt. Comm.,* **45** (1983) 71
3. C. E. Lee and H. F. Taylor, *Electron. Lett.,* **24** (1988) 193
4. K. A. Murphy et al., *Optics Lett.*, **16** (1991) 273
5 J. Sirkis et al., *Optics Lett.*, **18** (1993) 1973
6. G. Meltz, W.W. Morey, and W. H. Glenn, *Optics Lett.*, **14** (1989) 823
7. W.W. Morey, J. R. Dunphy and G. Meltz, *Proc. SPIE*, **1586** (1991) 216
8. K. O Hill et al., *Appl. Phy. Lett.*, **62** (1993) 1035
9. C. G. Askins et al., *Optics Lett.*, **19** (1994) 147
10. S. M. Melle et al., *Photonics Technol. Lett.*, **5** (1993) 263
11. M. A. Davis and A. D. Kersey, *Electron. Lett.*, **30** (1994) 75
12. A. D. Kersey, T. A. Berkoff and W. W. Morey, *Optics Letters,* **18** (1993) 1370
13. J. Dunphy et al., *Proc. SPIE* , **2071** (1993) paper 01; also, J. L. Archambault et al., *Proc. SPIE,* **2071** (1993) paper #04
14. A. D. Kersey, T. A. Berkoff and W. W. Morey, *Electronics Lett.*, **28** (1992) 236
15. D. A. Jackson and A. B. Lobo Ribeiro, *Proc OFS-9* (1993) 43

EXTENDING THE WEAR LIFE OF DRY SLIDERS USING MICRO VIBRATIONS FROM SURFACE WAVINESS

Michael D. Bryant and Atul Tewari
Department of Mechanical Engineering
University of Texas
Austin, Texas 78712-1063

Abstract: Sliding wear tests will be presented in which a carbon graphite block slides against "smooth" (flatness tolerance < 0.051 mm) and "wavy" rotor surfaces (possessing surface waves of tens of microns). The slider was spring-loaded against the rotor. Wear rates (μgm/sec) versus rotor speed for the smooth and wavy rotors were identical at some speeds, but at other speeds, wear rates on the wavy rotor were nearly half.

Slider vibrations were larger (ten to hundreds of microns) at speeds where the wear rates for the wavy rotor were most reduced. Resonance calculations suggest that at these speeds certain waviness harmonics on the rotor passed beneath the slider and excited natural frequencies of the slider-spring-rotor dynamic system. Friction was similar on the wavy and smooth rotors. Measurements of contact voltage drop (caused by an applied voltage) implied that the slider never separated from the wavy rotor. These reduced wear trends were also observed when the slider conducted up to ± 40 amperes DC.

This study suggests 50% reductions in wear rates for sliders (brakes, electrical brushes, clutches, etc.) can be engineered by applying/creating micro-vibrations of proper amplitude and type. Thus maintenance and part costs could be halved.

Key Words: Resonance; Sliding wear; Surface waviness; Vibrations; Wear reduction

Introduction: Under normal operation, lifetimes of sliding contact devices such as brakes, electrical brushes, and clutches are limited by wear. Following are schemes that utilize micro-vibrations to reduce sliding wear rates and double device lifetimes.

Friction and wear depend on interfacial conditions [1],[2]: loads, geometry, relative surface motion, and environment. Adhesive wear exhibits low to moderate wear rates and usually predominates at lower sliding speeds and forces. If conditions intensify (large loads at high speeds), more severe wear can be activated. With abrasive wear, hard particles entrapped within the sliding interface cut or plow material from surfaces. With thermal mounds [3],[4] found on brushes [5]-[7], brakes [8],[9], and seals [10],[11], heat generated by friction and electric currents (if present) elevate temperatures and expand material near the contact spots, causing material at these spots to bulge. Faster growing bulges promote loss of large particles: carbon graphites harden, grow, and break off like tiny pencil points, and metals soften and particles shear off. Formation and detachment of a thermal mound occurs in about 4 [5] to 100 msec. [7].

Surface undulations in the form of narrow grooves [12] or small pits cut into a rotor [13] can appreciably reduce friction, contact resistance, and wear [13]-[15]. Hard particles normally separate and abrade surfaces, but with undulated surfaces, particles fall into grooves/pits (reducing abrasive wear) and contacts shift to other sites (disrupting the thermal process that forms a thermal mound).

Contact vibrations can increase or decrease friction [16]-[20] and wear [21]-[31], depending on conditions. Contact vibrations are generally perceived to increase brush wear [21] and encourage arcing [22] and impact wear [23]. In machining, tool wear is accelerated [24],[25] and surface finish marred. Contact vibrations can also reduce wear. Ultrasonic vibrations applied normal to a slider drastically reduced steel pin on steel disk wear rates [26]. Others [27] found that vibrations sometimes increased and sometimes decreased wear rates, depending on the materials. In single point diamond turning of steel [28], ultrasonic vibrations reduced tool wear without harming surface finish; ultrasonic vibrations also extended carbide tool life ten to twenty-fold in the machining of glass [29]. Santini and Kennedy [9] realized 30% and 44% reductions in surface temperature and wear rate without a significant change in friction, after they cut a slot into a copper based brake pad and ran it against a brake disk. They produced similar results by pressing the pad against the disk with a more compliant spring.

For carbon samples sliding against wavy copper surfaces with [30] and without [31] current, Bryant et al reported reductions in wear rate (in μgm/sec) compared to wear rates for sliding against smooth surfaces. Wear rates for the smooth and wavy rotors were identical at some speeds, but at certain other speeds which were correlated to resonant frequencies of the brush mass--load stiffness system, the wear rates for the wavy rotors were nearly half those of the smooth rotors. Bryant and Lin [7], [31] also noticed thermal mounds on the smooth rotor but none on the wavy rotor. They hypothesized [30], [31] that resonance induced micro-vibrations may have altered the interfacial physics.

In this article, we review data in references [30] and [31] in which carbon slider wear rates were reduced 50%. Potential applications to devices such as brakes, electrical brushes, and clutches will then be discussed.

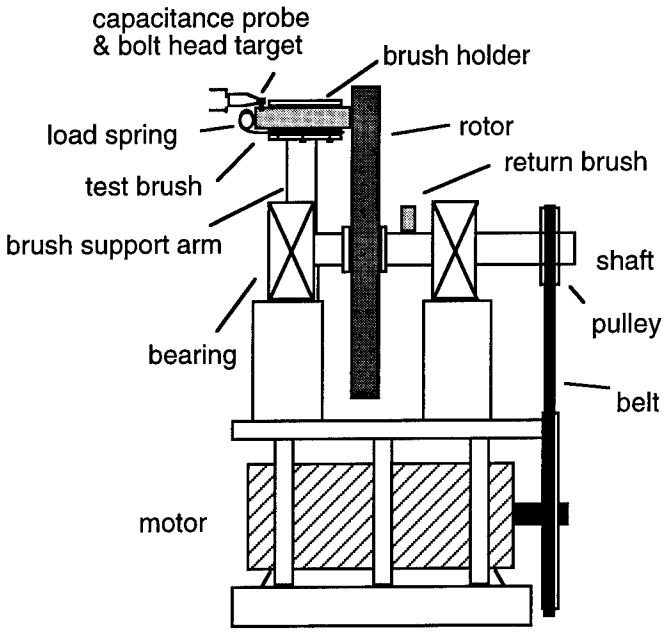


Figure 1: The test rig with motor, drive train, rotor, and brush holder/spring loading system. The return brush on the shaft completes a series electrical circuit consisting of power supply (not shown) test brush, slip ring, rotor frame, shaft, and return brush. A capacitance gauge probe measured slider displacements normal to the rotor surface.

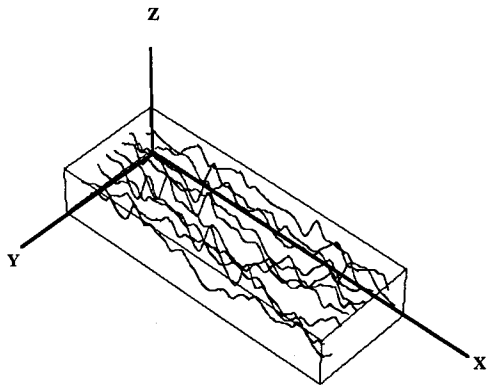


Figure 2: Surface profile of wavy rotor "a" measured along ten circular rings. Spacing between circles was 2.54 mm. The surface height Z ≤ 100 μm. The circumferential distance X ≤ 88 cm, and the radial distance from the inner ring Y ≤ 2.54 cm.

Wear Testing Machine: Details of the wear testing machine (shown in figure 1) are given in Bryant and Lin [7]. A stationary carbon slider slid against specially designed rotor tracks (radius 13.97 cm) at speeds from 10 to 2,500 rpm with normal loads of up to 111.2 N. Rotors possessed smooth and wavy surfaces: Sliders ran against a 250 μm thick soft gage copper-nickel alloy sheet bonded atop underlying layers of either steel or hard polycarbonate. When bonded against steel, the surface was smooth, with a flatness tolerance less than 0.08 mm. When bonded against polycarbonate, some rotors had wavy surface profiles (figure 2) with peaks and valleys forming ridges across the sliding track.

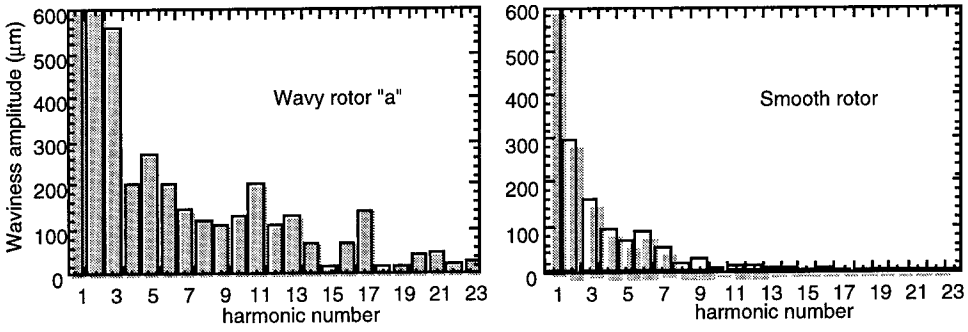
Probe measurements of axial (perpendicular) motions targeted a metal bolt attached to the slider (figure 1). At slow speeds axial motions produced a Z versus X trace roughly related to the maximum heights across the track width Y. Fourier analysis produced the spectra of displacement amplitudes (in μm) versus harmonic number shown in figure 3. Only higher harmonics represent surface features: the first harmonic is due to rotor tilt while the second harmonic is a flatness error.

Wavy polycarbonate backed surface "a" (figure 3a), shows harmonics 5 through 17 enhanced compared to figure 3b, a smooth steel backed surface. In figure 3a compared to figure 3b, harmonic 3 is very large, harmonic 5 large, with slightly smaller harmonics 4, 6, and 11 about the same size; harmonics 7-10 and 12-13 are yet smaller; and after 13 only harmonic 17 is appreciable. Another similar wavy polycarbonate rotor "b" (not shown) had surface harmonics 9 through 16 enhanced with harmonics 10, 11 and 16 most appreciable.

Conforce springs with constant load 24.4 N (or 20 N) pressed the slider against the rotor along the top of the trailing edge (see figure 1). To maintain a constant force over a large displacement range, spring stiffnesses were weak.

For experiments with current, sliders were carefully insulated from the holder [30]. Resistance between the brush and machine frame measured about 30 mega-ohms. Clearances between the slider and inner surfaces of the holder were less than a millimeter.

The holder support arm was a beam 37.5 cm long (2.54 cm x 2.54 cm cross section) cantilevered to the machine frame. By altering masses (slider and spring) and spring stiffnesses, the lowest resonant beam frequency could be changed from about 110 Hz to about 170 Hz.



(a) (b)

Figure 3: Spectral density (μm) versus harmonic number, for a.) wavy rotor "a" featuring enhanced harmonics 7 through 17 and b.) the smooth steel backed rotor.

Measurements: Measured were wear rate WR = $\frac{\Delta W}{\Delta t}$: the weight ΔW lost from the slider while sliding during time Δt); rotor speed; slider displacements perpendicular to the rotor; contact forces (via strain gauges mounted on the holder arm); slider current (when used as a brush); and when sliding without current, contact connect/disconnect voltage drop: the voltage drop across the contact due to a 4 volt DC source with a 220 ohm series resistor in series with the figure 1 series circuit consisting of test brush--slipring--rotor--shaft--return brush. When the slider contacted the rotor, the contact connect/disconnect voltage drop was of ten milli-volt order (low). When disconnected, the voltage drop approached the 4 volt source (high). Finally, the coefficient of friction was calculated as the ratio of the tangential force at the beam root (measured by strain gauges) to the spring force, which was assumed constant and equal to the spring load.

Table 1: Properties of carbon sample material.

grade	resistivity (μW cm)	hardness (scleroscope)	flexural strength (kg/cm^2)	density (gm/cm^3)
634	1780	20	53	1.28

Table 2: Slider geometries used in the experiments.

slider geometry	width (cm)	length (cm)
B	0.406	2.286
C	0.381	2.159

Table 3: Rotor geometries and loading conditions used in the experiments.

expt	prominent waviness harmonics	spring force (N)	beam natural frequency (Hz)	friction coefficient
a	11, 17	24.4	110 - 160	0.37
b	9, 16	20	150 - 175	0.21

Results: Sliders were NECC brush grade 634, with material properties listed in table 1 and geometries in table 2. All brushes weighed about 0.085 to 0.092 kg and slid against the smooth and wavy rotors at 500 rpm to 1,500 rpm.

Two sets of experiments, "a" and "b" were conducted, each with different conditions and rotors; table 3 summarizes. Both sets compared wear rates for wavy rotors to wear rates for smooth rotors. Each test ran for at least 20 to 24 hours. All data points were measured at least twice (some several times) to verify curve shape and repeatability. The coefficient of friction was measured over a short interval for every test. Experiments were run with DC currents of 0, ± 10, ± 20, and ± 40 amperes.

Wear rates WR_W for sliding against the wavy polycarbonate rotors (fig. 3a) and WR_S for sliding against the smooth steel (fig. 3b) backed copper surfaces versus speed are plotted in figure 4. No currents were conducted and conditions "b" in table 3 apply. Both WR_W and WR_S tend to increase with speed. Below 1000 rpm, WR_S is nearly linear with moderate slope; above 1000 rpm, the slope steepens. In general $WR_W \leq WR_S$: $WR_W \approx 0.5\ WR_S$ near 600 and 1000 rpm and $WR_W \approx WR_S$ near 500 and 750 rpm. The fractional difference $FD = \frac{WR_S - WR_W}{WR_S}$ is 0.4 to 0.5 near 600 and 1000 rpm, about zero near 500 and 750 rpm, and asymptotically approaches about 0.4 beyond 1200 rpm.

Figures 5 and 6 show wear rate versus speed for geometry B sliders conducting 10, 20, and ± 40 amperes and sliding against the smooth (curves with solid symbols) and wavy rotors (curves with hollow symbols). Figure legends indicate experimental conditions (see table 3). For almost all data, $WR_W \leq WR_S$. Figure 6 shows wear rates for cathode brushes considerably higher than those for anode brushes; for cathode brushes, as for anode brushes, $WR_W \leq WR_S$, and in fact WR_W is nearly half WR_S for most of the speed range.

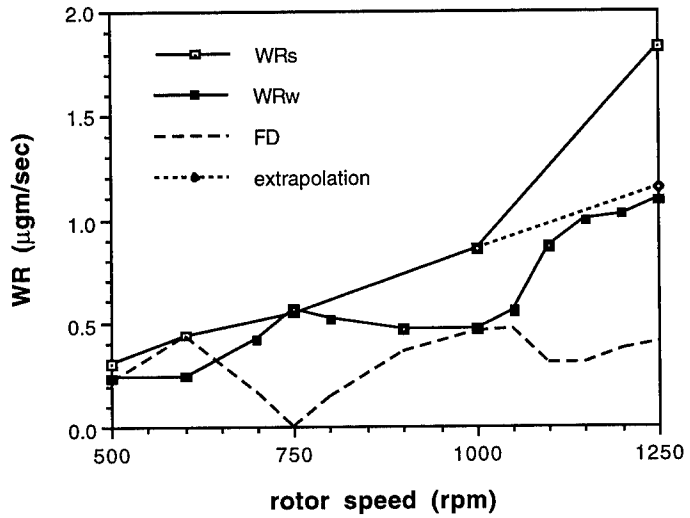
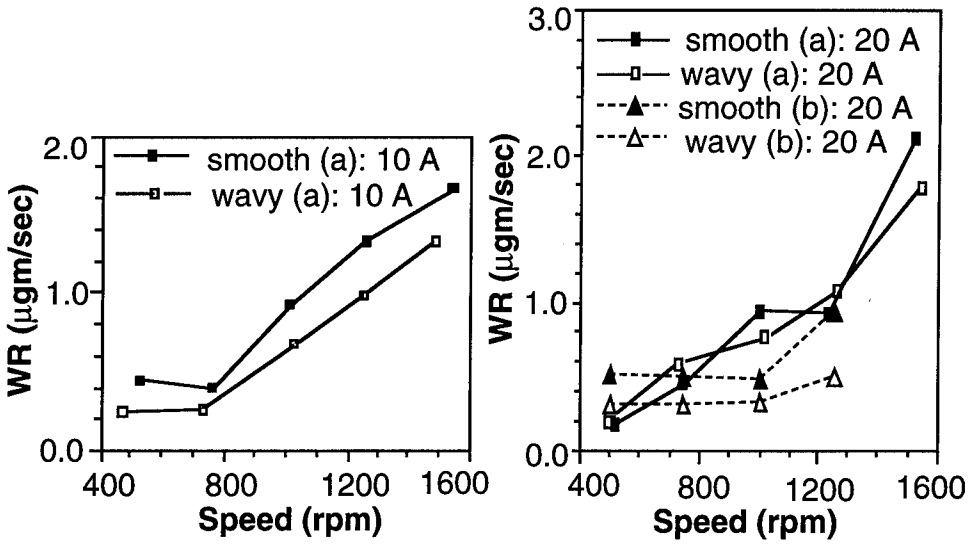


Figure 4: WR_W, WR_S, and $FD = \frac{WR_S - WR_W}{WR_S}$ vs. speed; the dotted curve extrapolates the low speed WR_S to high speeds. Generally $WR_W \leq WR_S$. Near 600 and 1,000 rpm $WR_W \approx 1/2\ WR_S$; near 500 and 750 rpm $WR_W \approx WR_S$. At high speeds WR_W approaches the extrapolation of WR_S.



(a) (b)

Figure 5: Wear rates (μgm/sec) for geometry B carbon sliders (brushes) conducting 10 (5a) and 20 (5b) amperes and sliding against wavy and smooth surfaces "a" and "b" versus rotor speed (rpm). Generally $WR_W \leq WR_S$.

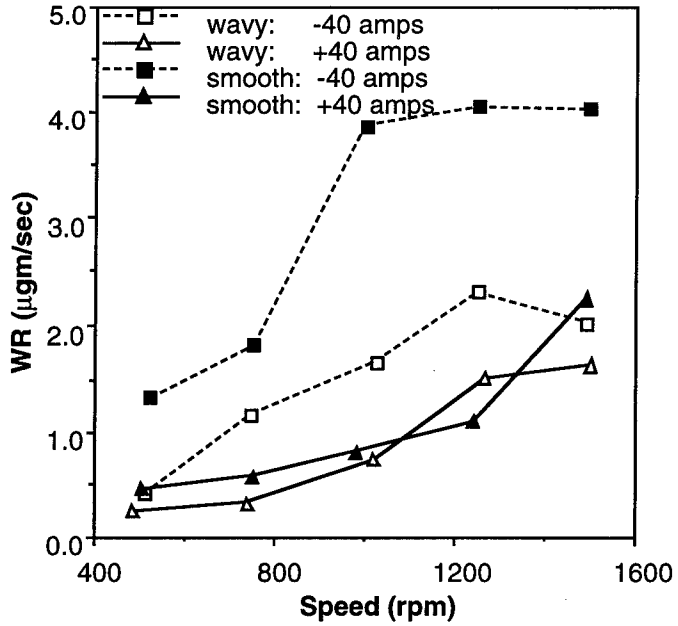


Figure 6: Wear rates (μgm/sec) for geometry B carbon sliders (brushes) conducting ± 40 amperes and sliding against wavy and smooth surfaces "a" versus rotor speed (rpm). Slider geometry B was used. Although wear rates are larger for cathode brushes (- 40 A) than for anode brushes (+ 40 A), generally $WR_W \leq WR_S$.

Figure 7 is an amplitude (μm) spectra versus frequency (Hz) of displacements normal to the sliding surface, derived from sliding against the wavy rotor of figures 2 and 3a at 1,000 rpm. In the figure the abscissa ranges from 0 to 400 Hz, and the first twenty or so peaks---equally spaced on the abscissa axis---are prevalent. Peaks represent harmonics of vibrations activated by sliding. The fundamental near 17 Hz is consistent with the rotor speed of 1000 rpm activating the first waviness harmonic.

Peaks at higher frequency should be induced in part by surface waves of higher harmonics passing beneath the slider. Figure 7 exhibits a relative ordering of peak heights qualitatively similar to figure 3a, except that in figure 7 peaks 8 and 9 have grown relative to their neighbors (compare these peaks to their respective counterparts in figure 3a). Peaks 7 through 10 fall within a band extending from about 110 to 160 Hz, which overlaps the first resonant frequency of the brush mass-spring-beam holder arm system "a" given in table 3.

Relative amplification of peaks about harmonic 8 in figure 7 suggests a resonant frequency near this range (100 to 120 Hz). Peak 8 has grown relative to peak 9 in figure 7 (compared to the same peaks in figure 3a).

Measured normal and tangential forces yielded frictional coefficients independent of speed with rapid fluctuations about 0.37 for the "a" experiments, and 0.21 for the "b" experiments. Figure 8 shows traces taken at 1,000 rpm while running against wavy rotor "a" of figures 2 and 3a. Fluctuations range from about 0.29 to 0.57. Traces on the smooth rotors were similar to traces on the wavy rotors.

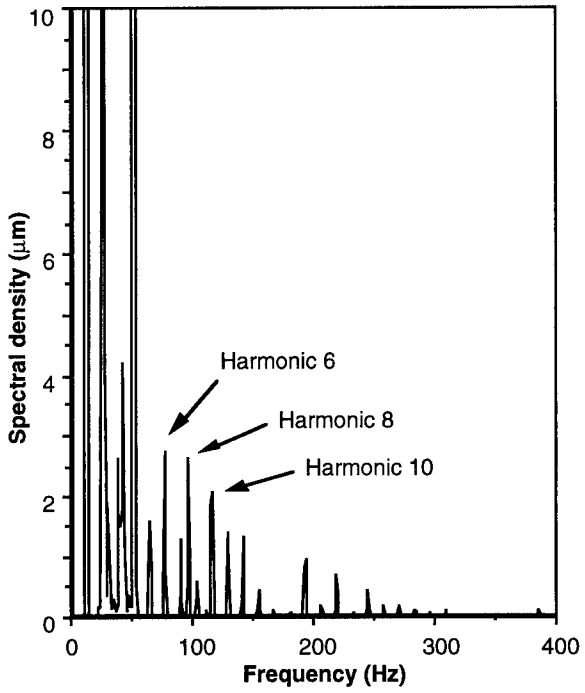


Figure 7: Spectral density (μm) of the brush deflection against the wavy rotor of experiment "a" at 1000 rpm. The horizontal axis is the harmonic frequency in Hz.

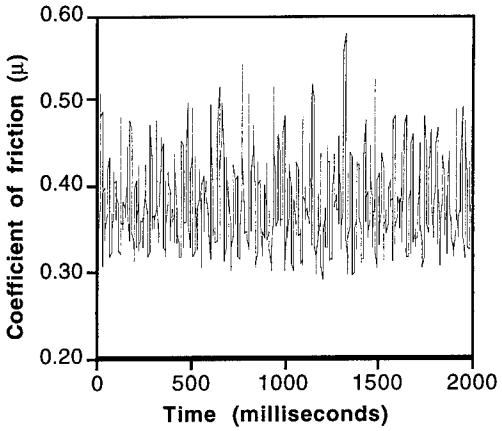


Figure 8: Time history diagram of the coefficient of friction (μ) signal at 1000 rpm. The average value is 0.37.

Oscilloscope traces of the contact connect/disconnect voltage drop for runs without current taken at 750, 1000, and 1525 rpm revealed small spikes at intervals of about 800 pico seconds. A typical spike was about 25 mV, the highest was less than 50 mV. After every run with current, the copper sliding track of the wavy rotor was inspected. No evidence of arcing or micro-arcing (melt zones, pits, craters, etc.) was observed.

Discussion: Figures 4 to 6 imply that a tiny amount of surface waviness (usually 10 to 100, and up to 300 μm) on a rotor may appreciably reduce slider wear. This was observed at most speeds and currents.

Without current, Bryant and Lin [31] reported reductions in wear rate and hypothesized that resonance induced micro-vibrations changed interfacial physics. In figure 4, a local maximum in the fractional difference FD corresponds to a local minimum in WR_W relative to WR_S. At 600 and 1000 rpm where FD maximizes, products of rotor speed and prevalent harmonic number on wavy rotor "b" are related: 600 rpm x 16 ≈ 1000 rpm x 10. These products in harmonic number-revolutions per second (160 Hz and 167 Hz, respectively) are close to the resonant frequency of the brush holder support arm (150 to 175 Hz) for setup "b" (see table 3).

The micro vibrations created by the rotor surface waves running beneath the slider seem to reduce wear, but not by disconnecting the slider from the rotor. The resonant frequency of the holder arm beam for the "a" set of experiments was approximately 110 to 130 Hz. If the brush/rotor contact disconnect small voltage spikes were caused by rotor waviness induced beam resonance, disconnects would have appeared at time periods dictated by the inverse of the beam resonant frequency, roughly every 10 milli-seconds. The observed small disconnects appeared about every 800 pico-seconds, suggesting no relationship to resonant beam vibrations. In addition, a complete disconnect would demand a 4 volt disconnect voltage: 200 times the largest of the observed 50 mV contact connect/disconnect voltage drop. Finally, for runs with current, arcing----the result of contact separation----- was not observed.

Figure 4 suggests that the waviness retarded severe wear. At about 1000 rpm the slope of the smooth rotor curve WR_S abruptly steepens (more than triples), indicating a more severe wear mechanism such as thermal mounding or abrasive wear [1]-[8]. In general, $WR_W < WR_S$. At higher speeds, wear rates WR_W on the wavy rotor approaches the dotted line,

which extrapolates the low slope portion of WR_S to higher speeds. Since WR_W follows the low speed slope, low speed wear mechanisms are likely still operative on the wavy rotor at higher speeds, whereas on the smooth rotor, more severe wear mechanisms reign.

A small amount of surface waviness that synthesizes micro-vibrations seems to discourage severe wear and limit moderate wear. Without disconnecting the contact, perpendicular vibrations could limit wear only by modulating the contact forces. Rocking vibration modes coupled to the perpendicular modes could disrupt interfacial physics without disconnecting the contact by rapidly transferring the contact to new locations within the slider-rotor interface, as the rocking causes the interface to touch at new sites. This could disrupt thermal processes needed to form thermal mounds and/or open clearances in the contact interface, allowing abrasive particles to escape.

Application to Slider Designs: Vibrations could regulate friction, wear, and contact resistance with only minor changes to the sliding system. Micron level surface waves machined onto the mating surface to generate micro-vibrations of a slider could lower wear rates by as much as 50% and double slider lifetimes: for example, on a slip ring a small amplitude surface waviness with wavelength correlated to the slip ring speeds could achieve low wear brushes in motors without exotic brush materials, sealed and/or pure atmospheres, or other esoteric and expensive techniques. The technique could double operating lifetimes of friction brakes and clutches, reducing (by half) maintenance labor and part costs.

Acknowledgements: Support from the National Science Foundation Tribology program (grant MSS-9302210) is gratefully acknowledged.

References:

[1] J.F. Archard, "Wear theory and mechanisms," Wear Control Handbook, M.B. Peterson and W.O. Winer, editors, New York, NY: ASME, 1980, pp. 35-80.

[2] D. Tabor, "Friction and wear---developments over the last 50 years," Keynote Address, Proc. International Conf. Tribology---Friction, Lubrication, and Wear, 50 Years On, London, Inst. Mech. Eng., pp. 157-172, 1987.

[3] J.R. Barber, "The influence of thermal expansion on the friction and wear process," Wear, vol. 10, pp 155-159, 1967.

[4] R.A. Burton, "Thermal deformation in frictionally heated contact," Wear, vol. 10, pp 1-20, 1980.

[5] T. A. Dow and J. W. Kannel, "Thermomechanical effects in high current density electrical slip rings," Wear, vol. 79, pp 93-105, 1982.

[6] D. K. Lawson and T. A. Dow, "The sparking and wear of high current density electrical brushes," Wear, vol. 102, pp. 105-125, 1985.

[7] M.D. Bryant and J.W. Lin, "Photoelastic visualization of contact phenomena between real tribological surfaces, with and without sliding," Wear, vol. 170, pp. 267-279, 1993.

[8] F.F. Ling and M.B.Peterson, "A reassessment of the wear of copper based brake material at elevated temperatures," Wear, vol. 102, pp. 43-50, 1985.

[9] J.J. Santini and F.E. Kennedy, "An experimental investigation of surface temperatures and wear in disk brakes," Lubrication Engineering, vol. 31, pp. 402-417, 1975.

[10] J. P. Netzel, "Observations of thermoelastic instability in mechanical face seals," Wear, vol. 59, pp 135-148, 1980.

[11] F.E. Kennedy, C.K. Chuah, and F.O. Brote, "Thermomechanical contact phenomena in face seals," Wear, vol. 102, pp. 127-140, 1985.

[12] Morgan Electrical Carbon, Carbon Brushes and Electrical Machines, Morgan Electrical Carbon Handbook, Quadrant Press Limited, Swansea, Great Britain, pp. 165, 1978.

[13] S.T. Oktay and N.P. Suh, "Wear debris formation and agglomeration," ASME J. of Tribology, vol. 114, pp. 379-393, 1992.
[14] N. Saka, M.J. Liou, and N.P. Suh, "The role of tribology in electrical contact phenomena," Wear, vol. 100, pp. 77-105, 1984.
[15] D. M. Tolstoi, G. A. Borisova, and S. R. Grigorva, "Friction reduction by perpendicular oscillation," Soviet Physics-Doklady, vol. 17, no. 9, pp. 907-909, 1973.
[16] S. R. Grigorva, D. M. Tolstoi, and A. V. Chichinadze, "Eliminating self-induced vibrations due to friction," Soviet Physics-Doklady, vol. 17, no. 1, pp. 60-62, 1972.
[17] Z. Broniec and W. Lenkiewicz, "Static friction processes under dynamic loads and vibration," Wear, vol. 80, pp. 261-271, 1982.
[18] A. Lehtovaara, " Influence of vibration on the kinetic friction between plastics and ice," Wear, vol. 115, pp. 131-138, 1987.
[19] T. Skare and J. Stahl, "Static and dynamic friction processes under the influence of external vibrations," Wear, vol. 154, pp. 177-192, 1992.
[20] D. Godfrey, "Vibration reduces metal to metal contact and causes an apparent reduction in friction, " ASLE Trans., vol. 10, pp. 183-192, 1967.
[21] I.R. McNab and J.L. Johnson, "Brush wear," Wear Control Handbook, M.B. Peterson and W.O. Winer, editors, New York, NY: ASME, pp. 1053-1102, 1980.
[22] E. I. Shobert, Carbon Brushes, the Physics and Chemistry of Sliding Contacts, New York, NY: Chemical Publishing Company, Inc., pp. 80, 178, 1965.
[23] P.L. Ko, "Metallic wear---a review with special reference to vibration-induced wear in power plant components," Tribology International, vol. 20, no. 2, pp. 66-76, 1987.
[24] N.K. Mehta, P.C. Pandey, and G. Chakravarti, "An investigation of tool wear and the vibration spectrum in milling," Wear, vol. 91, pp. 219-234, 1983.
[25] M.S. Selvam, "Tool vibration and its influence on surface roughness in turning," Wear, vol. 35, pp. 149-157, 1975.
[26] H. Goto and M. Ashida, "Effects of ultrasonic vibration on the wear characteristics of a carbon steel: analysis of the wear mechanism," Wear, vol. 94, pp. 13-27, 1984.
[27] K. Kato, A. Iwabuchi, and T. Kayaba, "The effects of friction-induced vibration on friction and wear," Wear, vol. 80, pp. 307-320, 1982.
[28] T. Moriwaki, and E. Shamoto,"Ultraprecision diamond turning of stainless steel by applying ultrasonic vibration," Annals of the CIRP, vol. 40, no. 1, pp. 559-582, 1991.
[29] H. Weber, J. Herberger, R. Pilz, and T.H. Karl-Marx-Stadt, "Turning of machinable glass ceramics with an ultrasonically vibrated tool," Annals of the CIRP, vol. 33, no. 1, pp. 85-87, 1984.
[30] M.D. Bryant, Atul Tewari, and J.W. Lin, "Wear Rate Reductions in Carbon Brushes Conducting Current and Sliding Against Wavy Copper Surfaces," Proceedings of the 40th IEEE Holm Conference on Electrical Contacts, pp. 229-237, Chicago, 1994.
[31] M.D. Bryant and J.W. Lin, "Reductions of Wear Rate and Contact Interface Observations for Carbon Samples Sliding Against Wavy and Smooth Copper Surfaces," Effects of Mechanical Stiffness and Vibration on Wear in Test Devices and Applications, ASTM STP 1247, Raymond G. Bayer, Ed. American Society for Testing and Materials, Philadelphia, 1995.

OPERATION OF MULTIPLE HEARTH FURNACES WITH DEFICIENT CENTER SHAFTS

By: Donald O. Dusenberry, Atis A. Liepins, and Hamid Vossoughi
Simpson Gumpertz & Heger Inc.
297 Broadway
Arlington, Massachusetts 02174

Abstract

In 1989, the city of Columbus, Ohio constructed two new multiple hearth furnaces for incineration of sludge at a waste water treatment plant. Each furnace contained a 32-in.-diameter cast steel center shaft that turned rabble arms to stir and distribute sludge. Soon after the furnaces were placed in operation, both experienced failures of the sockets that connected rabble arms to the center shafts. Since these furnaces are essential components of the waste treatment facility for this major city, rapid reactivation was a priority.

The authors were retained to investigate the failure and to develop procedures that would allow these damaged and deficient furnaces to be placed back into service on an interim basis while plans were made for replacement of the center shafts. The work included detailed inspections of the failed components, metallurgical tests of the center shaft materials, stress analyses for the sockets, analyses of maintenance and usage records, center shaft torque monitoring, analyses of the operational forces in the sockets, and development of torque limiter setting procedures for the existing torque limiters.

The study described herein resulted in reactivation of the subject furnaces at reduced, but temporarily acceptable, production rates.

Key Words: Failure; multiple hearth furnace; rabble arm; torque monitoring

Description of Furnaces

Measuring approximately 25 ft in diameter and 41 ft tall, each furnace had eight hearths, positioned one above the other. A 32-in. diameter cast steel center shaft extended from the bottom to the top of each furnace, turning two to four rabble arms at each hearth.

Sewage sludge is delivered to the top hearth of each furnace. In the top three hearths, excess water (the sludge is usually between 18% and 35% solids) is boiled off in preparation for burning. The temperature in the top hearths ranges from approximately 600°F to 1,000°F. The sludge is incinerated in the fourth and fifth hearths where the temperature typically is near 1,800°F. Grease and oil skimmings are burned from time to time in the fourth and fifth hearths. The sixth through eighth hearths are the cooling hearths in which the ash temperature is reduced to 200°F to 400°F. The furnaces were designed to incinerate sludge at a rate of 10 tons/hour.

The hollow center shaft for each furnace is 32-in. in diameter and has 2-in. wall thickness. The center shaft has interior baffles at each hearth to direct cooling air, supplied at the base of the furnace, to the hollow rabble arms. The rabble arms attach to the center shaft at sockets which have 1-in.-thick rims that project outward 5 in. from the center shaft. The sides and tops of each socket are braced by 1-in.-thick stiffeners. Rabble arms are secured by a single 1/2-in. stainless steel pin that passes through a 9/16-in. hole in the top of each socket. The center shaft and the rabble arms are coated with castable refractory.

The center shaft has four sockets at each hearth, although only the top three hearths have four rabble arms. Hearths below the third have only two rabble arms; two sockets at each hearth are not used. (Photo 1 shows the center shaft, a rabble arm, an unused socket, and a severed rabble arm in the hearth above.) The rabble arms are 123-1/2 in. long. Each is fitted with a series of rabble teeth that are 18 in. long in plan and extend downward 6 in. from the underside of the rabble arm. The teeth are canted at a 22-1/2° angle, in plan, to the axis of the arm. The arms in successive hearths are canted in opposite directions so that sludge is stirred and plowed radially inward in the odd-numbered hearths and radially outward in the even-numbered hearths to "drop holes" at the extremes of the hearth floors.

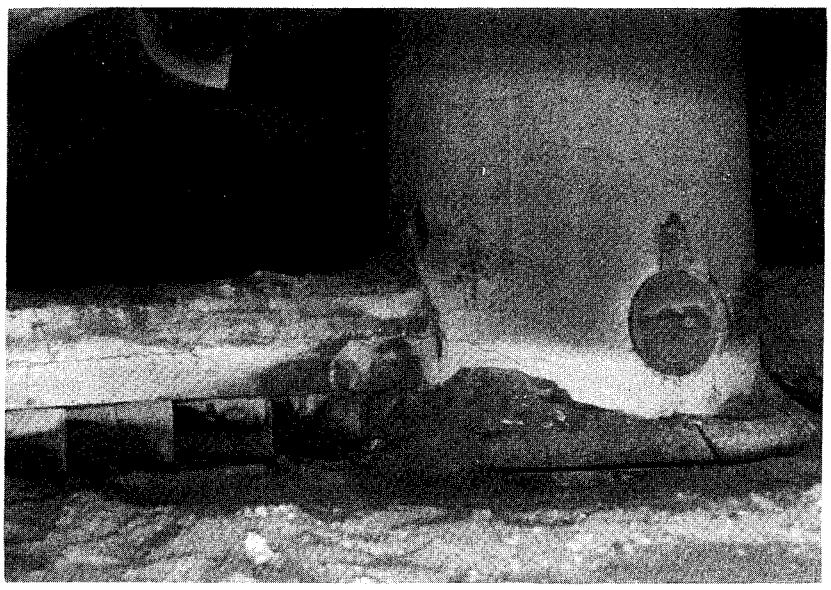


Photo 1. Center Shaft and Rabble Arms

In the top three hearths, rabble teeth are spaced 15-1/2 in. apart along the length of the rabble arms. In the other hearths, tooth spacing is 10 in. In all hearths, the positions of the teeth on the various arms are offset so that furrows created by each rotating arm are caught by the teeth of the trailing arm. In all hearths, each furrow is advanced radially a distance equal to the tooth spacing on a rabble arm during one revolution of the shaft. Shaft rotation speed typically is 0.5 rpm.

The center shaft of each furnace is driven by a 30 HP electric motor which runs at a synchronous speed of 1,800 rpm, delivering a torque of 89.1 ft-lb and drawing a

current of 36.5 amps at full load. The clutch, which also acts as a gear reducer, is an eddy current drive that is capable of controlling output speeds at constant rates of between 50 rpm and 1,710 rpm. The eddy current drive is equipped with a torque limiting device which monitors the voltage across the drive coil. When voltage raises to the level of the torque limiter setting, the torque in the center shaft ceases to increase and the center shaft stalls. The torque limiter has no specific indication of the torque limit; the torque limit is set by trial.

Beyond the eddy current drive in the drive train is a gear reducer with a speed reduction ratio of 83:1. The gear reducer operates at an efficiency of 80%. The drive train engages a bull gear on the center shaft through a pinion gear on the gear reducer. The pinion/bull gear set has a speed reduction ratio of 6.85:1. Thus, the combined speed reduction of the gear reducer and pinion/bull gear set is approximately 568:1. The combined torque multiplication, using 80% efficiency of the reducer, is approximately 455:1.

The Failure

Both furnaces were incapacitated by failures of the sockets in the fifth hearths (this is one of the hearths where oil and skimmings are burned). Both arms had separated from the center shafts when the sockets failed. The insides of these hearths were coated with clinker. In one of the furnaces, the coating of clinker was so thick that the rabble teeth had cut grooves into the coating on hearth floor. Rabble teeth were bent. Some witnesses claimed that the clinker build-up was caused by malfunction of the oil-burning system; other witnesses claimed that the furnaces were not properly maintained.

Apparently, the sockets in the fifth hearth of each furnace became overloaded when the rabble arm struck an obstruction created by the build-up of clinker. The second socket failed when the center shaft rotated 180° and the second arm struck the severed first arm. The nature of the failure suggested that the torque limiter setting for the eddy current drive must have been too high for the strength of the center shaft sockets.

The two unused sockets in each of the lower hearths allowed reconfiguration of the rabble arm pattern in the furnace after the failure. Therefore, it was possible to restore the correct number of functioning arms to each hearth. With proper control of torque and modified operation of the furnace to minimize formation of clinker, the potential existed for interim operation of the furnaces while new center shaft segments were fabricated.

The Investigation

Simpson Gumpertz & Heger Inc. was charged with evaluating the strength of the center shafts and developing controls that would allow the furnaces to be operated until permanent repairs could be developed. This required substantial analyses and testing to evaluate stresses in the elements of the drive shaft system and the performance of the torque limiter.

Analyses

Metallurgical Analyses: Pieces of the fractured rabble arm sockets were collected for laboratory analyses and testing to determine whether the cast iron used in the

center shafts conformed to the project specifications. This part of the investigation provided the following:

1. Polished and etched microstructure specimens showed a structure typical of pearlite gray cast iron.

2. No evidence was found of decomposition of carbides to form graphite. This indicated that strength was not degraded by operating temperatures.

3. Chemical analysis indicated a cast iron roughly equivalent to ASTM A319 Grade I-B or II-B. The project specifications required ASTM A319 Grade III.

4. Hardness tests indicated a tensile strength in the range of 27,000 psi to 30,000 psi. The tensile strength of ASTM A319 Grade III can be expected to be 40,000 psi. Therefore, the center shaft castings had much lower strength than that required by the project specifications.

Visual examination of the failed surfaces of the rabble arm sockets showed no evidence of casting flaws, fatigue cracks, or excessive heat damage.

These analyses, tests, and visual examinations supported the conclusion that the rabble arm sockets failed from overload. At the retaining pin hole with a diameter of 9/16 in., the fracture toughness was estimated to be in the range of 11,000 psi to 12,000 psi.

Structural Analyses of Center Shaft Components: Structural analyses of center shaft components were performed to determine torques in the center shaft that fracture the components and to provide guidance in setting the maximum permissible torque in the center shaft.

In the calculations of center shaft torques, we considered loadings which absorb the shaft torque and the load at the gear teeth which impresses the torque on the shaft. The loadings that absorb the torque are due to no-load operation, primarily due to friction in the bearings; normal stirring of sludge; and abnormal operation, primarily due to obstructions caused by clinkers.

Quantitative information about any of these loadings did not exist. Torques due to no-load operation and normal stirring of sludge were best determined by test. The permissible torque at abnormal operation was estimated from the load capacity of the components in the system together with an appropriate factor of safety. The components considered were rabble arm sockets in the center shaft, retaining pin that secures the rabble arm in the socket, rabble arm teeth, rabble arm, and gear teeth.

Rabble Arm Sockets: The rabble arm sockets have complicated geometry and loading and, therefore, the load capacity of the socket was best computed with a finite element model. A portion of the center shaft approximately one diameter in length above the socket and one diameter below it was modeled. Above and below this portion, the center shaft was considered to act as a simple shaft carrying torque, moments, shears, and axial load. These effects were included in the model as loads applied to the top and bottom extremities of the model. A torque of 1,000,000 in-lbs was applied at the lower extremity and reacted at the socket by a system of loads that were consistent with an obstruction at the end of the rabble arm. The finite element model gave a maximum tensile stress of 40.9 ksi on the inside surface of the socket at the pin hole.

In addition to torque, the temperature gradient through the thickness of the wall of the socket was computed as follows. The insulation of the rabble arm and the center shaft was interrupted at the end of the socket. This exposed the end surface of the socket to the hot gasses in the hearth and possibly to direct impingement by flames. The temperature gradient in the wall of the socket based on the one-dimensional heat flow equation, a thermal coefficient of 0.112 cal/cm/sec/^{o}C, and available heat from the flame of 1.3 Btu/hr/sq ft was calculated to be 40^{o}F/in. Approximating the socket by a long, thin cylinder, the thermal stress due to this gradient, with a thermal coefficient of expansion of 5.9 x 10^{-6} in/in/^{o}F, an elastic modulus of 14,000 ksi, and Poisson's ratio of 0.3, was calculated to be +2.4 ksi on the inside surface of the socket and -2.4 ksi on the outside surface.

Using an allowable stress of 8 ksi (corresponding to a factor of safety on fracture toughness of approximately 4/3), the thermal stress, and the finite element result, the permissible increase in the center shaft torque above the no-load operating torque is 11,500 ft-lbs.

The breaking strengths of a rabble arm tooth, rabble arm, retaining pin, and the pinion/bull gear set were estimated using strength of materials calculations and the nominal ultimate tensile strengths of the steels specified by the project specifications. The corresponding center shaft torques were calculated to be at least twice as high as that for the rabble arm socket.

Conclusion of Analyses: These calculations show that the weakest component of the center shaft is the rabble arm socket. This is consistent with the observed failures. The calculated permissible increase in the center shaft torque of 11,500 ft-lbs above the no-load operating torque is approximately 1/4 of the full load torque that the motor can deliver to the center shaft.

Torque and Torque Limiter Testing

Field investigations of forces in the components of the center shaft were performed after the rabble arm patterns were reconfigured. The purposes of the test program were:

1. to develop recommendations for center shaft torque limiter settings and

2. to operate the furnace at a production rate that was acceptable to the owner at the lowest possible risk of center shaft failures.

In normal operations, the furnaces are heated to temperatures up to 1,800^{o}F. At this temperature, direct instrumentation of rabble arms and sockets was impractical. The only convenient location to monitor loads in the center shaft was at its base, directly above the bull gear. At this location, there is a smooth section of the center shaft that is below the bottom hearth of the furnace.

In consultation with the manufacturer of the eddy current drive and the torque limiter, the authors determined that the setting of the torque limiter for the eddy current drive could be monitored by noting the voltage between two pins in the control panel for the torque limiter (at the time of the failure, the specific relationship between this voltage and the stall torque had never been determined; one goal of the subject investigation was to attempt to establish this relationship).

The instrumentation installed to measure the center shaft torque of each furnace is listed below.

1. Strain gages on the drive section of the center shaft below the base of the furnace (above the bull gear). These strain gages monitored torque in the center shaft.

2. An induction powered transmitter mounted on the center shaft adjacent to the strain gages. This transmitter sent strain gage data to a receiver mounted on fixed supports adjacent to the rotating center shaft.

3. A multiple channel strip chart recorder in the control room to continuously record the following data:

    - current drawn by the drive motor (this data was collected from the current indicator on the control panel in the control room),
    - center shaft speed (this data was collected from the shaft speed indicator on the control panel in the control room),
    - excitation voltage in the eddy current drive coil,
    - voltage across relevant terminals in the torque limiter circuit board, and
    - torque in the center shaft (as monitored by the strain gages).

These instruments were used to collect data during restart of the furnaces. Using an incremental procedure, the sludge feed rate was slowly increased in the furnace with the torque limiter at low settings. This induced shaft disengagements that allowed correlations among motor current, eddy current drive coil excitation voltage, torque limiter voltage, stall torque, feed rate, and sludge percent solids. It also permitted setting the torque limiter at levels that were just above the operational threshold; this allowed operation of the furnaces with minimum risk of failure.

The data permitted the development of the relationship between torque limiter setting and stall torque. For each stall that occurred during beddown of the furnace, initial start-up testing (some stalls were intentionally induced at specific settings), and during the early production phase, the maximum center shaft torque and the torque limiter setting were noted. The data and the best-fit curve are shown in Figure 1. The best-fit relationship was used subsequently to determine theoretical stall torques for various torque limiter settings.

Concurrently, risk to the furnace was evaluated (as discussed later in this paper) and the permissible incremental increase in feed rate was determined. The feed rate was increased slowly over a period of several days for each furnace until the production rate became acceptable to the owner or until the risk of damage to the center shaft reached the limit of acceptability.

The torque limiter setting corresponding to the highest achieved value acceptable to the owner based on production and risk became the interim setting for operation of the furnace until it could be taken out of service for replacement of the damaged and deficient center shaft.

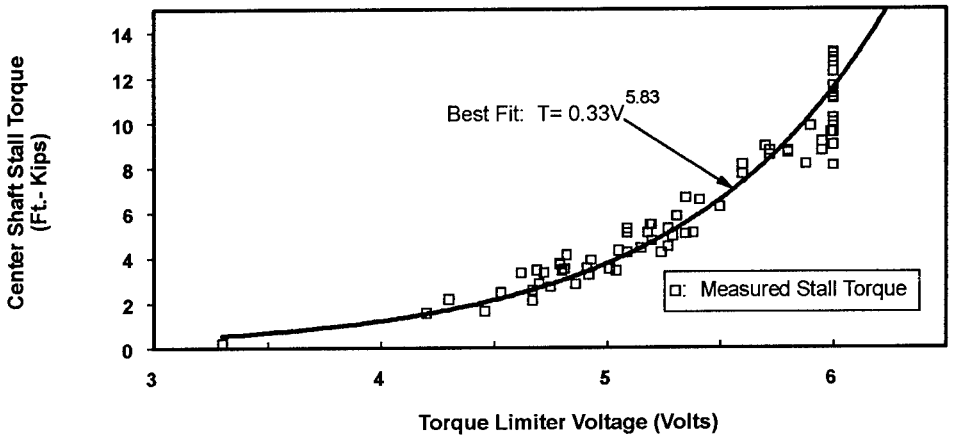


Figure 1. Stall Torques

Evaluation of Torque

Metallurgical analyses revealed that the material used for the center shafts did not conform to the project specifications or specified standards. Interviews with designers revealed that the ranges of expected forces for elements of the drive and center shaft systems were not known at the time of our investigation. These circumstances required that the conditions that caused overstresses be assessed accurately and the proper setting of the torque limiter be determined.

It was noted from the strip chart trace of the motor current, eddy current drive coil voltage, and center shaft torque that each of these traces has a cyclic pattern with a frequency that is four times the shaft revolution speed. Data also revealed that center shaft torque drops to very low values within approximately two revolutions of the center shaft after sludge feed is ceased. These observations suggest that the torque in the center shaft is heavily influenced by the four rabble arms in the first hearth, where each arm successively strikes an accumulation of sludge that is introduced continuously at a single point in this hearth.

To evaluate the forces in the rabble arm sockets, the authors first postulated the pattern of loads on the center shaft. In theory, the forces in arms of hearths below the first hearth are relatively constant and the forces in the arms of the first hearth consist of constant and varying components. Therefore, the torque in the center shaft is the sum of a constant (time invariant components of the first hearth arm forces plus arm forces from the other seven hearths) and four varying forces (variable components of first hearth arm forces).

The form of the variable component of rabble arm forces can be assessed from the data. Forces rise rapidly as an arm strikes the accumulation of sludge at the source. Then there is a period of rapid oscillation of torque (due to drive system dynamics) superimposed on essentially linear decay of torque until the next rabble arm reaches the accumulation of sludge. At that point, the cycle repeats.

The magnitude of the excursion of force in each arm varies from arm to arm, suggesting perhaps that the height of each arm above the floor of the first hearth is not constant (some arms dig deeper into the accumulation of sludge than others).

To evaluate the torque in each rabble arm, data were extracted at 20-minute intervals from the record of center shaft torque at various constant feed rates. Values of peak torque, minimum torque, and the maximum and minimum torque excursion during a revolution of the center shaft were plotted as a function of feed rate. Best-fit straight lines were evaluated for each of these relationships. Figure 2 illustrates the relationship between peak torque and feed rate.

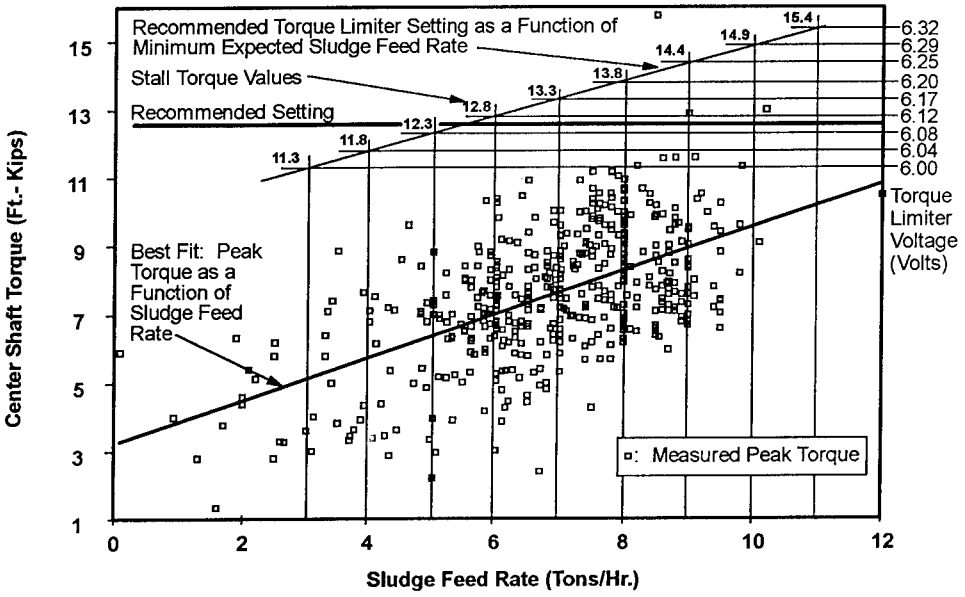


Figure 2. Peak Torques and Torque Settings

Assuming a linear decay of the varying forces in the rabble arms during the period of each revolution of the center shaft, an analytical model of the variation in center shaft torque was developed using rabble arm excursions selected from the data. For the excursion of force in one arm, the value was selected from the best fit to the data for maximum excursion at selected feed rates. For the other three arms, the value from the best fit for the minimum excursion was selected.

By solving simultaneous equations for the peak and minimum torques using this analytical model, the magnitude of the constant component of the center shaft torque was calculated. The constant component tended toward zero for peak torques less than approximately 8,000 ft-lbs, and ranged up to approximately 28% of the peak torque for peak torques of approximately 14,000 ft-lbs. The relatively small magnitude of constant torque and the observed rapid reduction of center shaft torque when sludge feed is terminated suggest that most of the constant component of torque is generated in the first hearth.

Torque Limiter Settings

The torque limiter controls the total torque in the center shaft. For it to function properly, it must be set high enough to allow the furnace to operate at suitably high feed rates (which implies high center shaft torque), but also be sensitive enough to detect increased load in a single rabble arm that strikes an obstruction. The torque limiter setting problem is exacerbated by the need to operate the furnaces over ranges of feed rates.

For high feed rates, the torque limiter setting must be high enough that nuisance stalls are minimized during normal operation. However, the setting can not be arbitrarily high because rabble arm sockets are normally relatively highly stressed for high feed rates and, therefore, the margin to overload is relatively low. For low feed rates, the total torque on the center shaft is relatively low, causing the margin between shaft operating torque and the torque limiter setting to be large. For this condition, an arbitrarily high setting might not protect even lightly loaded rabble arms from overload.

There are two probable causes for obstructions when the furnace is operating at high feed rates. One is the formation of clinker, which usually occurs in the fourth or fifth hearths where the temperatures are the highest and the rabble arms are relatively lightly loaded. The other is overfeeding the first hearth where the rabble arms are relatively heavily loaded. Since the lightly loaded rabble arms have larger reserve against overload than do the heavily loaded rabble arms, conditions in the first hearth control for high feed rates.

To evaluate the torque limiter setting for the first hearth, the constant component of the torque was assumed to be generated equally by the four arms in this hearth. To establish the force in the most heavily loaded rabble arm in the first hearth for each feed rate, one-quarter of the constant component of the center shaft torque was summed with the corresponding maximum torque excursion. Then, the reserve capacity of the rabble arm at each feed rate is the difference between its capacity to generate torque in the center shaft (calculated previously to be 11,500 ft-lbs) and the calculated maximum force at that feed rate. Finally, the maximum acceptable torque setting is the sum of the center shaft peak torque and the torque reserve at that feed rate.

At relatively low feed rates, it is unlikely that the furnace will be stalled due to overfeeding in the first hearth. The probable cause of obstructions at low feed rates is clinker in the fourth or fifth hearth. Under these circumstances, lightly loaded rabble arms in these hearths control.

To evaluate the torque limiter setting for low feed rates, a 500 ft-lb minimum value of rabble arm contribution to center shaft torque was assigned to each critical arm. Then, the reserve capacity of the rabble arm at each feed rate was taken as 11,000 ft-lbs (rabble arm capacity to generate center shaft torque minus this assigned minimum value). Finally, the maximum acceptable torque setting is the sum of the minimum center shaft torque at each feed rate and this torque reserve.

For all feed rates studied (10 tons/hour or less), the calculated maximum permissible torque limiter settings based on calculations for the first hearth were higher than those for the other hearths. Therefore, conditions that dictate the setting of the torque limiter are formation of obstructions in lower hearths at relatively low feed rates. Values of maximum permissible torque limiter stall torques and corresponding voltage

settings for the torque limiter based on these calculations are shown in Figure 2 for feed rates between 3 and 10 tons/hour.

In consultation with the owner, it was concluded that the minimum feed rate that was likely to result in formation of obstructions was approximately 5 tons/hour. At this feed rate, the center shaft torque that may result in overstresses is approximately 12,300 ft-lbs (see Figure 2).

We recommended a setting of 6.1 volts, which was sufficient to protect against overloads at feed rates of approximately 5 tons/hour, while allowing the furnace to be operated at feed rates up to approximately 8 tons/hour with only occasional nuisance center shaft stalls. While this feed rate was below the intended design capacity of the furnace, it allowed the owner to address current needs for incineration of sludge while planning for replacement of the damaged and deficient center shaft.

Conclusions

Based on this investigation of the performance of two multiple hearth furnaces, the authors concluded that, even with deficient materials and damage from prior failures, it was possible to set the torque limiters to specific maximum torques, calculated on the basis of structural analyses of the center shafts and analyses of torque data, to allow the furnaces to be operated at acceptable levels while planning for replacement of the center shafts.

Recommendations

The authors recommend procedures such as those described herein to evaluate strength and performance whenever it is crucial to operate damaged or deficient machinery.

Acknowledgments

The authors wish to acknowledge the assistance provided by Mr. Lester Smiley and Mr. Arthur Gibson of Columbus Southerly Wastewater Treatment Plant, Mr. Andrew Hruska and Mr. Timothy Jasina of U.S. Motors, Dr. Thomas Eagar (metallurgist), and Mr. Lissimachos Vassilopoulos and Mr. Ivan Laxa of Maritech, Inc. for their assistance during our investigation.

RECIPROCATING MACHINERY LIFE CYCLE MAXIMIZATION

Grant John Gorton
NKF Engineering, Inc.
4200 Wilson Blvd., Suite 900
Arlington, VA 22203-1800

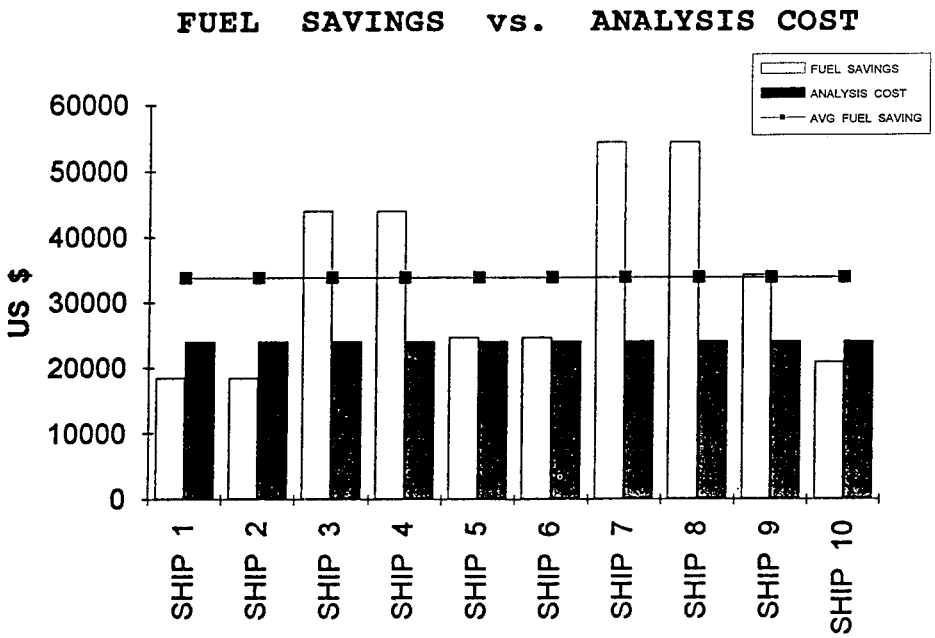
Proper operation of a diesel engine, air compressor, or other reciprocating compressor can have a major impact on equipment life and cost efficiency. Condition monitoring of reciprocating machinery can identify major equipment problems which, if left uncorrected, could result in catastrophic failure or material condition degradation with attendant life reduction. For example, improper diesel fuel injection timing can create abnormal peak pressures in engine cylinders which will place stress on pistons, rods, bearings, and crankshafts. Correction of such a problem will increase engine efficiency, reduce stress on internal components, reduce engine degradation, and increase engine life. Sticking or broken valves on a gas compressor can result in reduced efficiency and add significant operating cost to the equipment. This presentation will include several case studies of actual condition monitoring tests which revealed equipment problems. The presentation will also include a case study of cost savings documented for a fleet of diesel powered ships using a condition monitoring system. One method for condition monitoring compressors and diesel engines is the Compressor and Reciprocating Machinery Analyzer (CARMA). CARMA was developed initially to support condition monitoring for compressors used in natural gas pumping stations for Canadian oil and gas companies. The equipment monitors cylinder pressures, vibrations, and ultrasonic signals. Another method is the use of the MAILIN system which monitors cylinder pressure only. A third method is the use of the ERNASKO on line monitoring system which has 63 sensors for engine temperatures, auxiliary temperatures, and pressures.

Keywords: Analysis; compressors; condition monitoring; diesel engines; life cycle costs; maintenance; performance testing; pressure; reciprocating machinery; ultrasonic; vibration

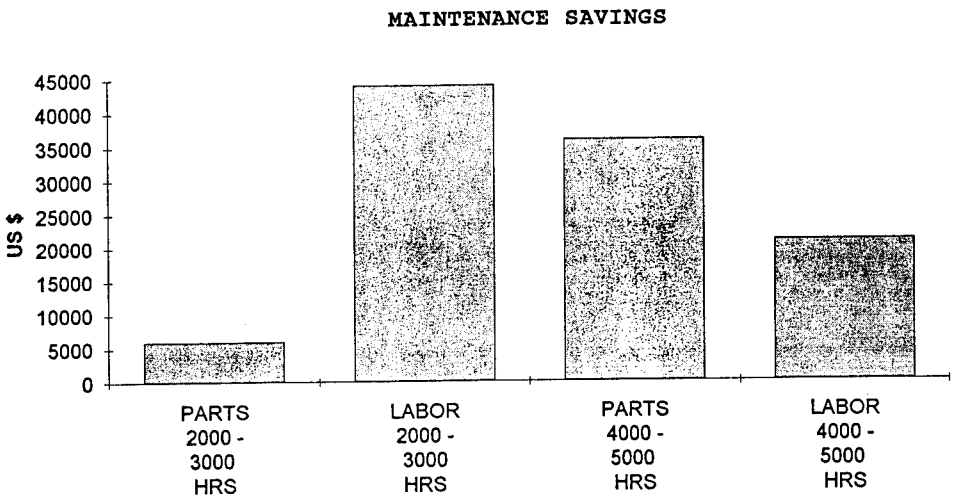
The CARMA system monitors actual cylinder pressure, vibration and ultrasound for a diesel or compressor one cylinder at a time. Data collection is keyed to shaft rotating position by a mechanically attached encoder so the analysis software can create a "virtual machine" from the data. This allows analysis of the machine as whole including the identification of cross talk signals or interference between cylinders during data acquisition. The encoder device monitors shaft position every degree of rotation so data is accurately keyed to shaft position. The cylinder pressure analysis combined with the ability to identify "events" occurring in the machine give an

engineer a very effective condition monitoring tool. The MAILIN system monitors only cylinder pressure and uses an optical pickup to determine shaft position. The optical pickup method is easier to use, but shaft position is only monitored once per revolution. For a very slow speed engines with only 5 or 6 cylinders, shaft speed changes during a single revolution can introduce position timing errors. The analysis is limited to power balancing of the engine or observing cylinder pressure problems. The ERNASKO system is an online system which monitors diesel engine auxiliaries as well as the engine itself. The system does not monitor individual cylinder pressures but uses cylinder temperatures to analyze engine balance and power. The sensor suite is less expensive and easier to install and maintain, however, balancing an engine on temperature does not always result in a true power balance.

CASE 1: A Mideast tanker company has been using routine (quarterly) diesel engine monitoring for over a year now on a fleet of ten tankers. The company has documented changes in fuel consumption and maintenance over the monitoring period. This particular company chose to contract the anlaysis service and compared contract costs to fuel savings. The following table shows the estimated fuel savings based on calculations of measured fuel consumption for the main engines before the testing and after testing-recommended adjustments were completed. The total annual fuel savings after contract testing expenses were approximately $120,000.00 for the ten ship fleet.



Based on the analysis results, this company has extended the manufacturers recommended 2000 hr time based maintenance to 3000 hours and extended the 4000 hr time based maintenance to 5000 hours. The following table shows projected maintenance savings from the reduction in parts and manhours required to support the main engines. Further reduction in parts and labor is expected as the accuracy and effectiveness of determining engine material condition increase with continued analysis. The total maintenance cost savings were approximately $107,000.00 for the ten ship fleet.



CASE 2: An electric power generation company was using three ALCO 251 2600 HP / 2000 kW diesel generators with a major overhaul scheduled every 30,000 hours, head overhauls at 20,000 hours, ring overhauls at 10,000 hours, and maintenance checks at 4000 and 2000 hours. The company had experienced three turbo chargers damaged over a one year period due to broken exhaust valve parts and one unit experienced such high vibration (shaking) that operators were reluctant to run the machine and opposed to more than a 70% load when it would run. Condition monitoring with a CARMA system showed wide variations in valve opening and closing events. Valves clearances ranged from 0.040" to 0.007" instead of the specified 0.034 +/- 0.004 inches. Valves were either closing very hard or candidates for hot gas damage with both conditions resulting in potential valve breakage. Valve clearance adjustment and

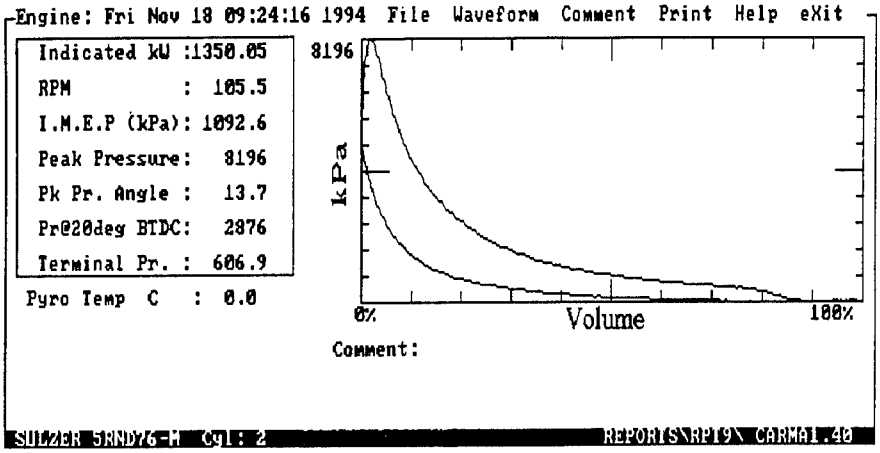
continued valve timing monitoring with CARMA has eliminated turbo charger damage. The unit with excessive vibration had a severe power imbalance in addition to the valve problems. The analysis revealed two defective injectors, a defective fuel pump, and indicated fuel rock adjustments were needed. Correcting these problems resulted in smoother engine operation and made a noticeable change the ratio of indicated engine horsepower to generated electric power due to efficiency increase.

CASE 3: The recent analysis of a main engine on a tanker operating out of Florida illustrates material condition assessment capability. The engine is a Sulzer 5 cylinder 14,000 HP diesel with over 70,000 hours of operation. The analysis detailed was based on data gathered while underway from Port Everglades to Tampa with the main engine operating at a steady load.

General Remarks on the Main Engine: The engine power balance was fairly good with less than 13% spread between cylinders. There were, however, indications of mechanical wear on many of the engine components. Cylinder No. 4 had recently been fitted with a new liner and piston rings. This was reflected dramatically in the data collection. Cylinder No. 2 and cylinder No. 3 both had indications of vibration occurring in the BDC area. Considering the age of components of this engine, these BDC events should be investigated at the next possible convenient occasion. A thorough inspection of the piston assembly bearings including bottom end bearings, all piston rings and cylinder liner surfaces should be performed at the next convenient maintenance period. The data traces for cylinders 2 and 3 are shown as figures 1 and 2.

Cylinder 1: Vibration event just before exhaust blowdown - *The possible sources of this vibration are skirt contact with the port area or lower ring clipping the exhaust port.*
Abnormal vibration at the beginning of the compression stroke - *This could be an indication of a liner mark. A corresponding vibration event would have to be witnessed at the same location on the power stroke to confirm this. Unfortunately the vibration caused by blowdown masks any indication. Another possible culprit for this type of event is piston ring vibration as they expand with rising cylinder pressure of the compression stroke. Also, because of a corresponding low pressure measured at 20 deg. before TDC, this could be an indication of a slight compression leak.*
Vibration and ultrasonic events on either side of TDC - *This is usually a characteristic of a liner mark. It may well be a piston ring clipping a lubricator groove.*
Vibration and ultrasonic events at 56 deg. after TDC - *This could be indicative of a problematic piston ring.*
Cylinder 2: Vibration event just before exhaust blowdown - *Same comments as Cylinder 1.*
Bottom Dead Center vibration and ultrasonic event. Both events carry on after BDC - *This could be an indication of excessive clearances in the power transmission assembly (i.e. bottom end bearing piston pin bushing...etc.)*
Vibration and ultrasonic events apparent at 12 deg before TDC - *This could be an*

FIGURE 1 CYLINDER 2 TRACE



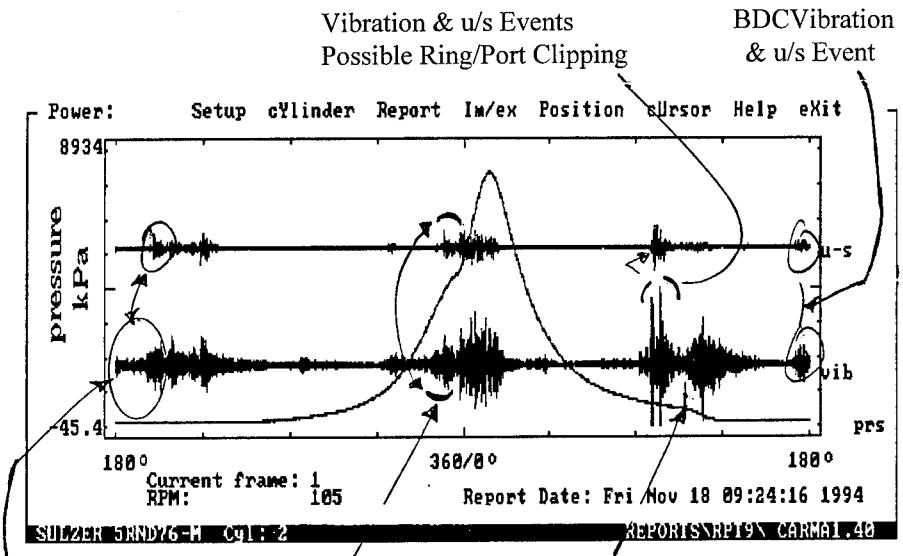
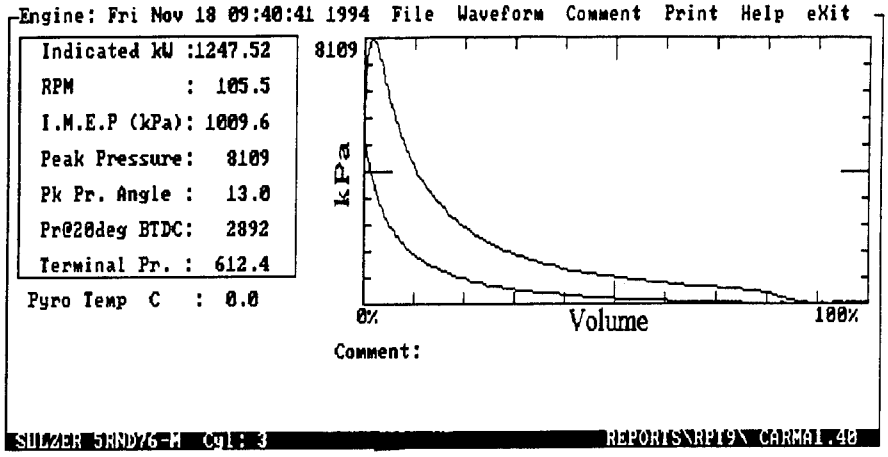
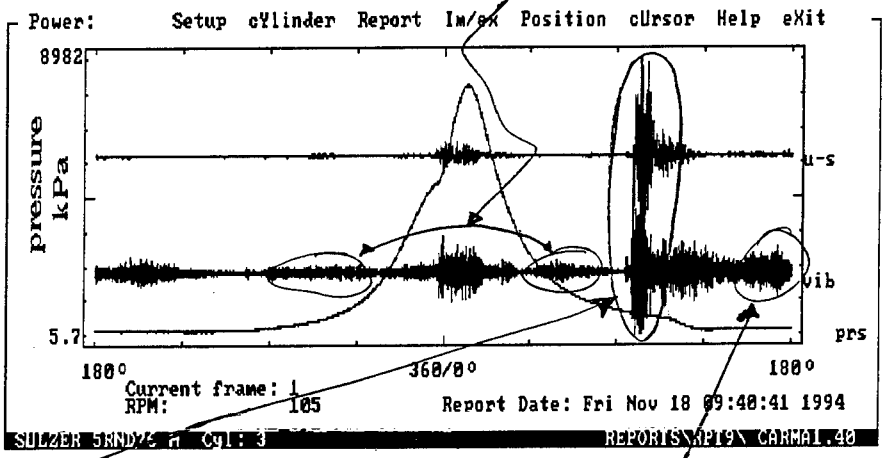


FIGURE 2 CYLINDER 3 TRACE



Vibration Equi-distanced on Either Side of TDC.
Possible Liner Roughness



Heavy Ring/Port Vibration
& u/s Events just before Blowdown

BDC Vibration Event.
Begins @ 20° BBDC and Carries Through to 17° ABDC

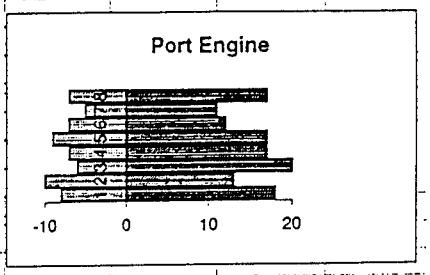
indication of sticking piston rings. As pressure rises above the restrictive force holding the ring back (deposits in ring groove) the ring has a tendency to snap out against the liner wall.
Cylinder 3: Extended vibration traces equi-distanced on either side of TDC - *Again, this can be an indication of cylinder liner condition. This type of trace would be characteristics of a rough liner surface.*
Vibration and ultrasonic events just before exhaust blowdown - *Same comments as Cylinder 1.*
Extended vibration event beginning at 20 deg before BDC and carrying on through 17 deg after BDC - *Again this could be an indication of excessive clearances or problems in the power transmission assembly.*
Cylinder 4: Very heavy vibration and ultrasonic traces throughout the cylinder - *This liner had been in operation for 97 hours when data was collected. It is still in its wear in period and should be monitored closely. It is good practice to increase lubrication to a cylinder during the break in period. The heavy traces are a result of the new piston rings mating with the new liner surface.*
Cylinder 5: Vibration just before exhaust blowdown - *The possible sources of this vibration are lower rings clipping the exhaust port, or skirt contact with the port area.*
Extended vibration event at the beginning of the compression stroke - *This could be an indication of piston ring vibration as they are forced out to the liner wall with rising cylinder pressure.*

CASE 4: The recent analysis of a main engine on a Gulf Coast ship illustrates material condition assessment on a very different type of diesel. This engine is a Hitachi - B & W model 2x8K45GTC diesel which has two banks of eight cylinders each on separate crank shafts. The analysis was performed as if the unit were two engines.
General Remarks on the Main Engine: There is a considerable difference in engine performance between the port and stbd banks. The port bank produced on average 3687 IHP while the cylinder to cylinder power spread was 79%. The reason for such a large cylinder to cylinder power spread on this bank is due to the fuel being limited to cylinder 7 (Fuel Rack set to 10mm). Ignoring cylinder 7, the power spread between the other cylinders is considered fair at 18%. The stbd bank produced on average 2480 IHP, while the cylinder to cylinder power spread was 26%. It should be noted that the engine speed varied by as much as 7 RPM during the analysis. This speed variation will have an effect on the power produced. The injector timing diagram (figure 3) shows that there is a substantial difference in the timing of the stbd bank as compared to the port bank. All injectors were changed before the test. The ultrasonic traces taken on the fuel injectors allow us to determine the injector opening and closing. To improve power balance on the port bank, items such as fuel injector performance, fuel rack settings, etc... should be checked and verified. On the stbd bank, several cylinders display early peak pressure angles. The danger of having early peak pressure angles is that a great portion of the combustion pressure is exerted on the piston while the piston is still travelling in its upward compression stroke. This results in increased stresses on the piston, connecting rod and associated bearings. This will ultimately result in premature failures if not dealt with. Factory timing of

FIGURE 3 INJECTOR TIMING

Cylinder injector timing

Cyl No		Inj start	inj end	duration
1	352	-8	18	26
2	350	-10	13	23
3	354	-6	20	26
4	353	-7	17	24
5	351	-9	17	26
6	353	-7	12	19
7	355	-5	11	16
8	353	-7	17	24



Cyl No		Inj start	inj end	duration
1	345	-15	7	22
2	345	-15	4	19
3	343	-17	6	23
4	349	-11	10	21
5	341	-19	3	22
6	346	-14	6	20
7	352	-8	14	22
8	348	-12	12	24

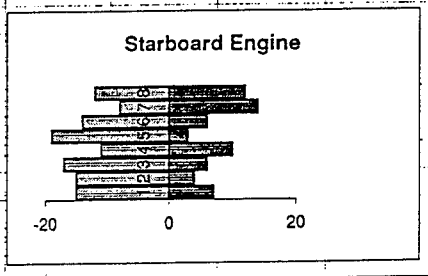
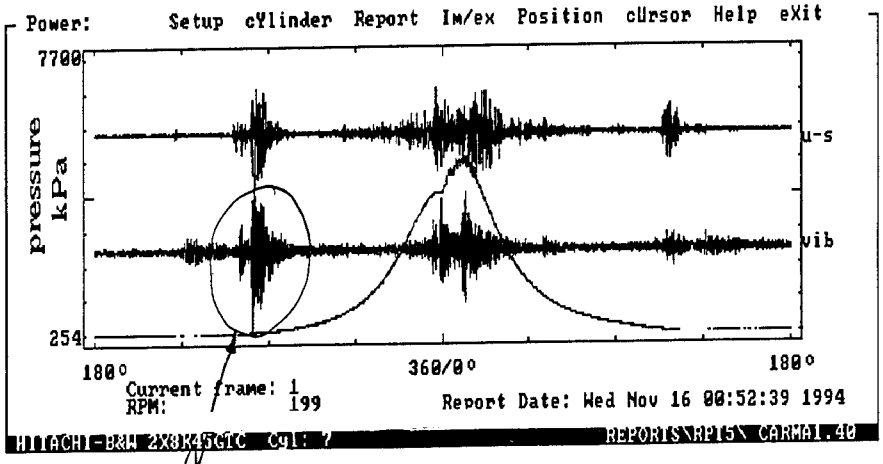


FIGURE 4 HARD VALVE CLOSURE TRACE



Hard, Multiple Impacts on Exhaust Valve Closures

this bank should be verified to identify the source of the early fuel injection and resulting early Peak Pressure Angles. The CARMA unit can acquire statistical data for 25 engine cycles and display the variations in Peak Pressure and Peak Pressure Angles for each cylinder. Statistical data acquired on both banks show that some cylinders show substantial cycle to cycle variation in Peak Pressure Angle. The usual cause of this is deteriorating performance of the fuel injectors. Cylinders displaying wide variations in Peak Pressure Angle have their fuel injectors should be tested. Vibration traces taken on the cylinder head and the exhaust valve housing of all cylinders indicate that several cylinders have exhaust valves that close very hard with some of them showing double impacts. Exhaust valve operating mechanisms on these cylinders should be checked and adjusted to lessen the severity of the impact on exhaust valve closures. Several cylinders display signs of increased vibration levels when the piston rings pass the scavenge ports, especially on the power stroke. The piston rings and the scavenge ports on these cylinders should be inspected. The vibration traces taken on the cylinder head show that some cylinders show increased baseline vibration or excess vibration at certain locations of the cycle. The normal vibration is either a rough liner and/or damaged piston rings. The liners and piston rings of the cylinders displaying these symptoms should be inspected.

Individual Cylinder Comments on the STBD Bank: Cylinder 1, Vibration evidence of rings clipping the scavenge ports on the power/exhaust stroke. *Examine rings and ports at next maintenance frequency.*
Hard, double impacts on exhaust valve closure as shown in figure 4. *Check/adjust valve operating mechanisms.*
Cylinder 2, Vibration evidence of the rings clipping the scavenge ports on the power/exhaust stroke. There is also evidence of increased baseline vibration which is usually attributed to a rough cylinder liner. *Examine rings and ports at next maintenance frequency. Also examine the liner for roughness.*
Hard, double impacts on exhaust valve closure. *Check/adjust valve operating mechanisms.*
This cylinder produces the lowest IHP. The Peak Pressure Angle as shown on the performance summary is very early at an average of 4.3 degrees ATDC while the compression pressure is much higher than any of the other cylinders. This cylinder has the symptoms of early combustion. *Test/replace the fuel injectors as required. Check/adjust fuel rack setting as required.*
Cylinder 3, Vibration traces indicate that the rings are periodically clipping the scavenge ports on the power/exhaust stroke. The vibration traces also show that there is a slight increase in vibration as the rings pass the scavenge port on the compression stroke. *Examine rings and, scavenge ports. Also examine the liner for roughness.*
The Peak Pressure Angle as shown on the performance summary is very early at an average of 4.4 degrees ATDC while the terminal pressure is much higher than any of the other cylinders. *Test/replace the fuel injectors as required. Check/adjust fuel rack setting as required.*
Hard impacts on exhaust valve closure. *Check/adjust valve operating mechanisms.*
Cylinder 4, Vibration evidence of the rings clipping the scavenge ports on the power/exhaust stroke. *Examine rings and ports at next maintenance frequency.* Hard,

double impacts on exhaust valve closure. *Check/adjust valve operating mechanisms.* The Peak Pressure as shown on the performance summary is low compared to the other cylinders. Statistical data acquired on this cylinder indicates a wide spread on Peak Pressure Angle from cylinder to cylinder. *Test/replace the fuel injectors as required.*

Cylinder 5, Vibration evidence of the rings clipping the scavenge ports on the power/exhaust stroke. *Examine rings and ports at next maintenance frequency.* Hard, double impacts on exhaust valve closure. *Check/adjust valve operating mechanisms.* The Peak Pressure Angle as shown on the performance summary is very early at an average of 5.0 degrees ATDC. *Test/replace the fuel injectors as required. Check/adjust fuel rack setting as required.*

Vibration traces taken on the cylinder head show a vibration event 45 degrees ATDC. However, vibration readings on the main bearings show no abnormal vibration. This situation should be monitored during the next analysis.

Cylinder 6, Hard, double impacts on exhaust valve closure. *Check/adjust valve operating mechanisms.*

Cylinder 7, Very slight evidence of the rings clipping the scavenge ports on the power/exhaust stroke. Monitor next analysis to determine if the vibration is increasing. Hard, double impacts on exhaust valve closure. *Check/adjust valve operating mechanisms.*

Cylinder 8, The Peak Pressure Angle as shown on the performance summary is very early at an average of 4.5 degrees ATDC. *Test/replace the fuel injectors as required. Check/adjust fuel rack setting as required.*

The author would like to recognize the analysis work of Dr. Howard Malm and John Solmon of Resigat Development, Inc., 715-50 O'Conner Street, Ottawa, Ontario, Canada K1P 6L2.

DIAGNOSTICS II

Cochairmen: **Patrick J. Sincebaugh**
U.S. Army Research Laboratory

James Taylor
HSB Reliability Technology

INCIPIENT FAULT DETECTION IN LOCOMOTIVE DC TRACTION MOTORS

By

Michael P. Treanor
Martin Marietta,
KAPL Inc., PO Box 1072
Schenectady, NY 12301

&

Gerald B. Kliman
GE Corporate Research
& Development, PO Box 8
Schenectady, NY 12301

Abstract: The problem of machine condition monitoring of dc traction motors on diesel-electric locomotives is addressed. Techniques for detecting poor commutation to allow an operator to take action to prevent flashover were studied. The focus of the study was frequency-domain analysis of the motor current at the bar-passing frequency, although other techniques were found to be successful. A fast-fourier-transform of the motor current indicated a peak at the bar-passing frequency which increased in magnitude by a factor of five when the commutation quality was degraded. A relation is provided between the visual sparking level and the magnitude of the bar-passing peak. A radio-frequency spectrum of the armature voltage was also obtained and found to increase in magnitude, for frequencies below 5 Mhz., with the spark level. A method of analyzing the spectrum to account for different load and operating conditions is described. O_3 and NO_x were monitored by sensing tabs, which indicated that the presence of O_3 may be used to detect poor commutation.

Keywords: brushes, commutation, dc, diesel-electric, locomotive, motor-current-signature analysis, sparking, traction.

Introduction: The ability to monitor commutation quality degradation as a precursor to flashover in dc machines would provide a significant payoff in steel mills, paper mills, and, in particular, the locomotive industry. One of the largest sources of maintenance effort on diesel electric locomotives is the traction motor, which has warranted some investigations [1, 2]. In part this is due to problems inherent in all electric motors but exacerbated by the very difficult thermal, humidity, contaminant and shock environment of the traction motor which far exceeds that of industrial motors in general. A further complication and added cost is that, when something is seriously wrong with the motor, it must be removed from the truck which must first be removed from the locomotive by means of a drop table. It is often the case that locomotives may be backed up for many days waiting for a turn at the drop table adding to the already high cost of removal. Advance

knowledge of developing problems in the motors would be a major advantage in scheduling scarce repair facilities. It is also very important to note that some motor problems can be corrected without removal of the motor if they are caught early enough.

Diesel-electric locomotives have dominated the railroad industry for almost 70 years. In that time the diesel engines have grown in power from around 1500 HP to 4500 HP on dc drive locomotives. At the same time, the 4, 6, or 8 dc traction motors on a locomotive have had a corresponding growth in rated power from around 300 HP to 750 HP with much of that growth not accompanied by an increase in size. The motors are always axle-hung, hence they are subjected to a high level of road vibration and shocks in excess of 50 g and more. All of the motors are cooled with ambient air that may range from -40 deg F to + 130 deg F. This air, taken from the right of way with little or no filtering, may contain particulates, chemical contaminants and moisture ranging from zero to 100% relative humidity. To complete the stress picture the motors are normally run at very low speeds at high loads for extended periods as well as at high speeds over rough track.

The conditions described above may quickly degrade the commutators in the dc motors. A degraded commutator will show excessive sparking, as the brushes bounce over the rough surface and the conduction period for a particular segment ends prematurely, or a winding remains shorted for too long by a spark region which extends between segments. Ultimately a short circuit will develop, through an extended spark region over the commutator bars, between opposite-polarity brushes. This "flashover" is often severe enough to destroy the motor and can cause the axle to lock.

It has been found that flashovers are usually preceded by poor commutation. This condition can be recognized by excessive sparking while the motor is running, or by careful examination of the commutator surface for an improper film development, out of round, scoring, and other signs. These examination methods are not practical for a running locomotive traction motor, nor is it practical to train the operators or to perform the level of inspection necessary to certify each motor periodically. It is clear that there is a need for automated monitoring of the quality of the commutation for locomotive traction motors.

Several methods were investigated for gaging the quality of commutation in dc traction motors. The first method, motor current signature analysis, is similar to methods which have been applied to other motor types [3]. The motor current was examined in three frequency bands: below 100 Hz, below 10 kHz., and 10 kHz. - 10 MHz. Spectra were obtained as the commutation quality was degraded by moving the brushes or by adjusting the current in the commutating poles. Spectral analysis of the armature current is described in detail. Other techniques were also considered;

Antenna pickup of sparking, photodiode measurement of sparking, acoustic analysis, and O_3 and NO_x detectors. These other techniques, which were either less successful or were ruled out for practical reasons, are described briefly.

Preliminary tests and discussions with locomotive maintenance personnel resulted in a focus on motor current signatures as the best route to a useful instrument for actual locomotive road testing. If successful it could then be incorporated, at low cost, into the locomotive monitoring systems already in place.

General Description of Tests: Two laboratory tests with motors on the test stand in the steady state were carried out. One was a 10 HP industrial motor on which the brush ring was movable. The other was a standard 750 HP locomotive traction motor on which the brush position was fixed at neutral but where the commutation poles were accessible for buck and boost. Results were similar in a general way but of course the numerical results were quite different.

Figure 1 indicates the layout of the test apparatus for the 10 HP motor test. Measurements were made on the 10 HP machine acting as a generator in order to decrease the switching noise from the rectifier. Figure 2 indicates the test setup for the GE752-AH1 traction motor. The series connection of the motor and generator allowed a practical amount of power to be used to conduct the tests.

Motor Current Signature Analysis

10-HP Motor Measurements: Measurements were made on a 10 hp shunt-wound dc machine. The test machine had 99 bars on its commutator. This was run as a generator, being driven by another shunt-wound motor, in order to isolate the measurements from converter noise. It is recognized that, for a practical application of this technique, power supply noise will need to be considered. The brush ring was moveable on the test machine, which provided a means to cause poor commutation. By moving the position of the brushes with respect to the commutating poles, large voltages were induced on the segments undergoing commutation.

Figure 3 shows the power spectra for the armature current, between 0 and 312.5 Hz., for two conditions of commutation for the 10 HP machine. Figure 3a is the power spectrum with the brushes properly adjusted for minimum sparking, while Figure 3b shows the power spectrum with approximately "C"-level (or moderately bad) sparking. In both cases, a peak appears at the frequency of rotation and at every harmonic of this frequency. Peaks also appear at 60 Hz. and harmonics of 60 Hz., due to the power supplies, as indicated by the large 120 Hz. peak.

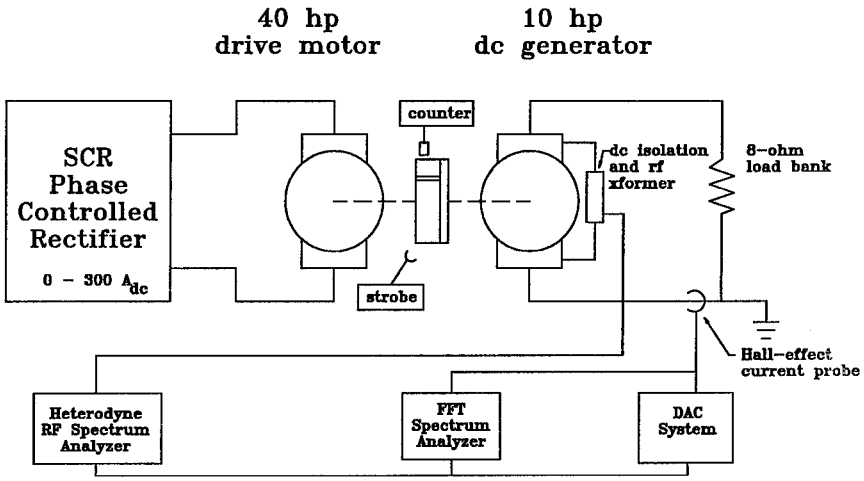


Fig. 1. 10-HP Commutation Test Setup at GE CR&D.

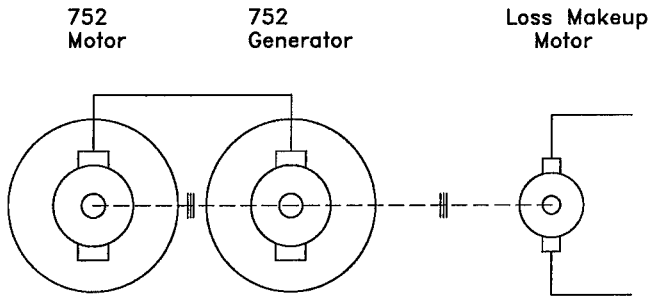


Fig. 2. Locomotive Traction Motor Test Setup at GE Transportation Systems

The magnitude of the peak at the rotation frequency was slightly affected by the quality of the commutation. The average value of the power spectrum was also affected in this frequency range. By averaging over this frequency range, however, the measurement becomes sensitive to unrelated events, such as traveling over sections of rough track. The most notable feature of these spectra was the clear indication of the rotation frequency.

Figure 4 shows the power spectra for the same motor over a wider frequency range. It can be seen that several peaks appear with poor commutation. One peak which is easily identified is at the bar-passing frequency. The bar-passing frequency can be determined by multiplying the number of commutator bars by the speed of the motor. These peaks were measured at 3.35 kHz. and 3.33 kHz. with the brushes at the neutral position and with the brushes rotated to cause sparking, respectively. The magnitude of the peak at the bar-passing frequency increased by approximately five times when the brushes were rotated. By observing such a drastic change at a frequency which is distinctly related to the

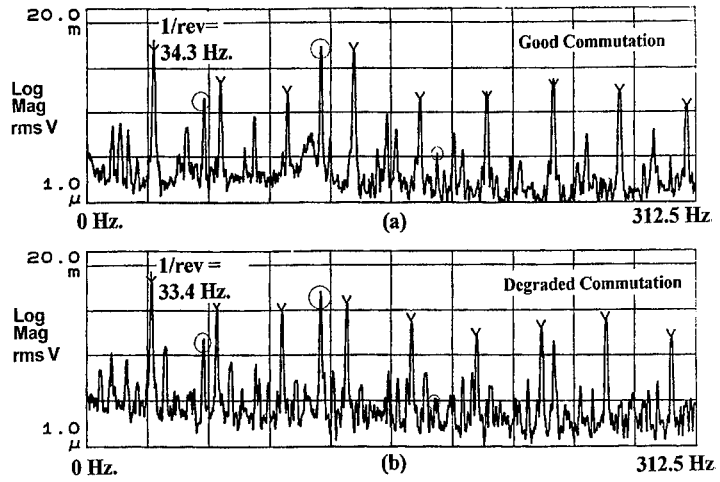


Figure 3. 0 - 312.5 Hz Current Spectrum for (a) Good and (b) Poor Commutation. 60 Hz. and once-per-rev harmonics are marked.

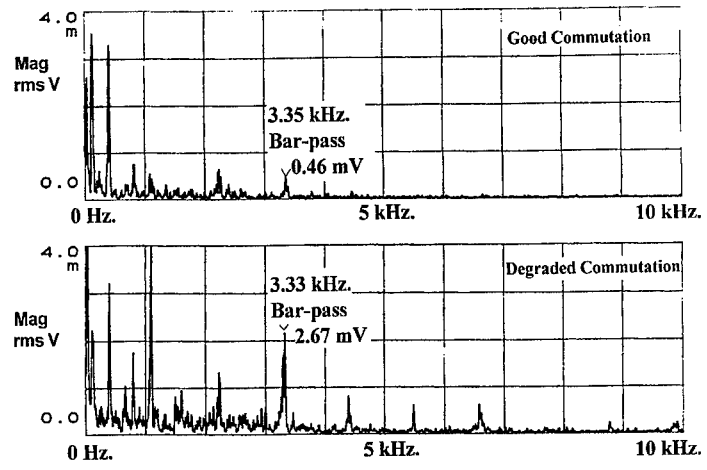


Figure 4. 0 - 10 kHz Current Spectrum for (a) Good and (b) Poor Commutation, showing a large increase in the bar-passing signal.

commutation process, it becomes clear that this measurement provides an effective gage of the commutation quality.

750-HP Traction Motor Measurements: Measurements were made on a 750 HP GE752-AH1 dc traction motor connected in the arrangement indicated in Figure 2. The traction motor had 184 commutator bars and the windings for the commutating poles were available for bucking or boosting.

The brush ring could not be moved. Results which were similar to the small motor in that the frequency of rotation was easily discernible and the current magnitude at the bar-passing frequency was found to be a strong function of the level of sparking. An interesting distinction between these results and results with the small motor is that 12 Hz., and its harmonics, appear, which were determined to be related to resonant frequencies in the test stand. The magnitude of the bar-passing harmonic appears to be linearly dependent on the magnitude of the armature current.

Figure 5 indicates the bar-passing current at three levels of sparking at 1220 RPM and 2440 RPM while boosting the commutating poles. The level of sparking is visually observed at the brushes. It was found that the magnitude of the bar-passing current was independent of speed, so that, in Figure 5, straight lines were drawn through the buck and boost data regardless of speed. The data appear to fit to these straight lines. For each spark level, there was a larger bar passing current while boosting than while bucking the commutating pole current. Lines of the same slope as in Figure 5, but displaced downward, fitted well to the buck data. This non-symmetrical behavior is typical and can be seen clearly in Figure 6, which summarizes the variation of bar-passing current with spark level at one speed and armature current level.

It may be noted that the identical motor and generator were connected in series, so that the "A" sparking is the output of two machines, both assumed to be commutating well. It is reasonable, therefore, to subtract 1/2 of the "A"-sparking value at all points to obtain the contribution from one machine.

Since a plot of the current at the bar-passing frequency as a function of the dc armature current produces a straight line, a plot can be made of the slope of this line as a function of the spark level. This is done in Figure 7, where 1/2 the "A"-spark level has been subtracted as described above. The figure illustrates that the bar-passing harmonic can be related to the spark level, as discussed above.

Radio-Frequency Analysis of Armature Voltage: A number of monitoring methods have been used based on the intensity of the sparking at the brushes. Visual observation of the brightness of the spark was first used [4]. This technique defined a "black band", which is a range of commutating-pole current which produces no spark at the brushes. The method requires no instrumentation and does provide a means of initially adjusting the commutating poles. However, it suffers from several disadvantages as a condition monitoring method and is not practical on a moving locomotive. Sensing the spark with photodetectors has been described [5-6]. The rugged environment and contamination of the lens surfaces are concerns with this technique as well as added wiring.

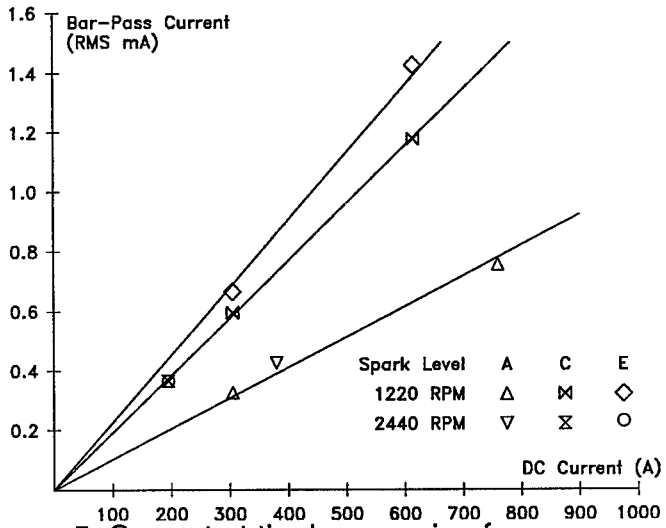


Figure 5. Current at the bar-passing frequency as dc current and spark level are varied.

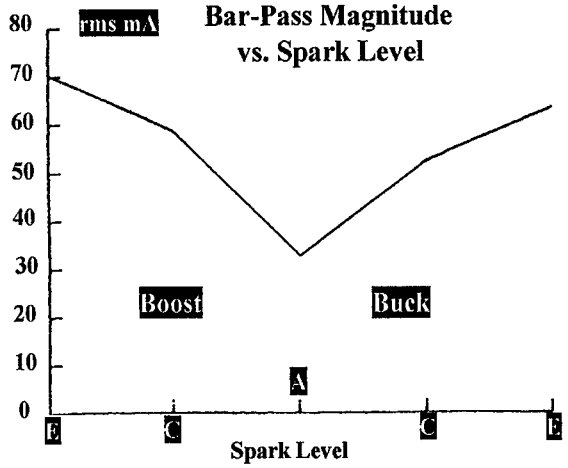


Figure 6. Bar-passing current magnitude at 1220 RPM, 615 A_{dc} as commutating pole current is varied.

Sparking is known to be associated with wide-band radio-frequency (rf) signal generation, although the exact nature of the spark, the machine characteristics and the ambient conditions will control the power spectrum of the signal. Previous work has described circuits which monitor the brush voltage, isolate the rf portion of the voltage using passive filters, rectify the rf signal and display the result on a milliammeter [4,7]. These circuits can not account for the variations in the spectrum which are possible with a varying environment and load. Thus, the use of a radio-frequency spectrum analyzer allows the detailed capture of the rf activity needed to

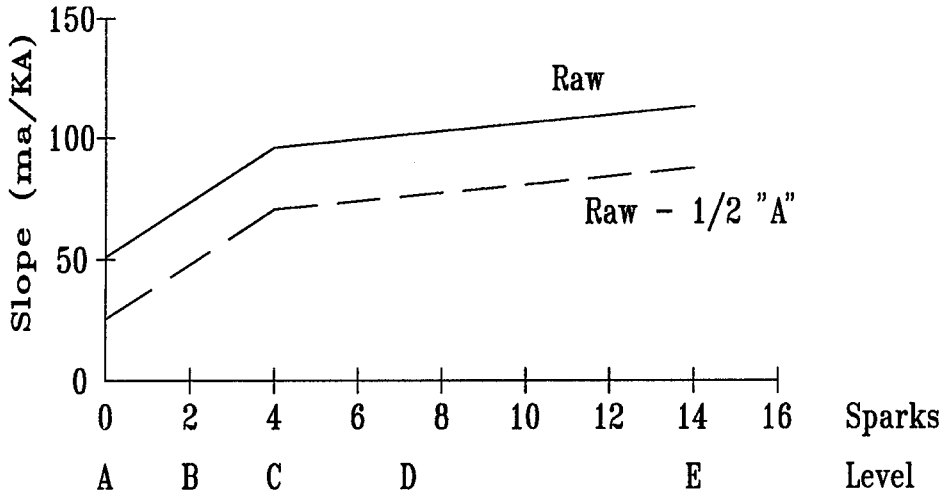


Figure 7. Slope of Bar-passing current vs. Dc current as a function of spark level.

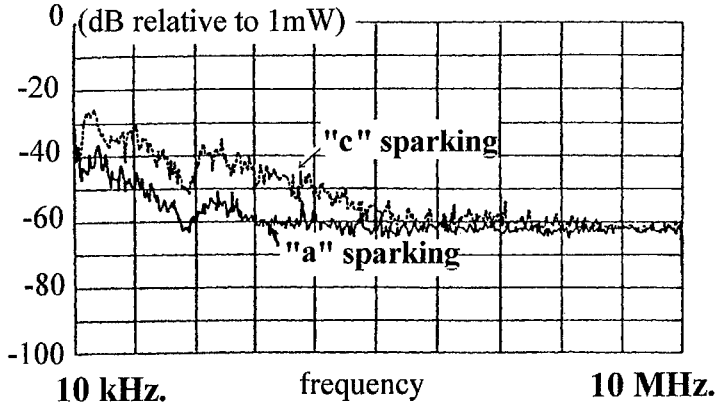


Figure 8. RF spectrum, brush-to-brush voltage, for two levels of sparking on GE752 motor. Scales are logarithmic.

account for these variations. Figure 1 shows the placement of an appropriate probe and the spectrum analyzer to accomplish detection. The RF filter grounded low-frequency signals and provided dc isolation between the motor leads and the instrumentation. The use of a spectrum analyzer permits some flexibility in determining what frequency ranges are most indicative of sparking.

The rf spectrum analysis for two levels of sparking in the 750 hp traction motor is illustrated in Figure 8. It can be seen that "C"-level sparking shifts the rf spectrum upward for frequencies below approximately 5 MHz. It is known that these spectra will be affected by ambient conditions and loading, so that a logic scheme which interprets the spectra would be needed. Ambient conditions may include inverter noise and the effect of the cable routing between the rf pickup and the instrumentation.

Alternate Methods: Other techniques of detecting poor commutation were investigated or considered. It is known that sparking is associated with the production of ozone and nitrous oxides. Ozone and nitrous oxide may be detected using paper tabs chemically impregnated with materials which change color upon exposure to these gases. Chemically impregnated paper has been used in safety devices intended to alert the wearer of the device to excess exposure to these gases. Commercial devices are sensitive to, for example, 0.1 ppm (parts per million) exposure to ozone for a period of 15 minutes. The paper is fairly inexpensive. A test was performed by placing this paper near the brushes. It was found that after a 15-minute exposure to sparking brushes the ozone-sensitive paper changed color. Intermittent exposure, over several days, to properly commutating brushes did not affect the paper. The change in the color of these tabs is more easily and quickly recognized than subtle and inaccessible damage to the commutator surface.

Antenna pick-up of the radio signals created by sparks has been investigated [8]. In some cases this method has the advantage that the antenna need not be mounted at the brushes. The needed signal can propagate through the air to an antenna. This was considered for the dc traction motor, but was considered to offer little advantage. The motor casing provides an effective shield which contains the rf signal. The degree to which the signal escapes the casing and the effect on the power spectrum of the signal depend on the cable routing and other variables which would be expensive to control. Thus, the antenna mounting must be near the brushes in this case, so that little advantage exists over monitoring the radio-frequency component of the brush voltage.

Conclusion: Several methods have been discussed by which the condition of the commutator of a dc traction motor can be monitored. The results of frequency analysis of the motor current, and the effect of poor commutation on the current magnitude at the bar-passing frequency were explained in detail. It was found that the motor speed can be determined by recognizing the frequency of rotation within a low-frequency spectrum. From this information, the bar-passing frequency can be estimated. The magnitude of the motor current at the bar-passing frequency was found to be a strong function of the commutation quality, and could be related to the well-known visual spark levels. It should be noted that the speed of a dc motor is often known by other means, so that the frequency-analysis routines can be made simpler. Acquisition times should be kept as short as possible. This method was found to provide a significant remote indicator of commutation quality.

Radio-frequency analysis of the brush voltage was also found to indicate the presence of sparking. "A" and "C" sparking levels were easily distinguished in a spectrum taken from 10 kHz. to 10 MHz., Because the

power spectrum for noise generated by sparking is known to be sensitive to ambient conditions, an algorithm for interpreting the spectrum would be needed.

Alternate methods of detecting poor commutation were described briefly. Paper impregnated with ozone-sensitive chemicals was found to successfully distinguish good and poor commutation. Methods of calibrating this technique and a practical application method were not discussed. Antenna pickup of the radio-frequency noise was also discussed.

References:

[1] D.D. Steeples, "Electrical Characteristic Changes from Artificially Induced Defects on an EMD D57 Traction Motor," Submitted to IAS (1994)

[2] D.D. Steeples, G.B. Kliman, "On-Line Commutation Monitoring of Locomotive DC Traction Motors In Service," in preparation.

[3] M.E. Steele, R.A. Ashen and L.G. Knight, "An Electrical Method for Condition Monitoring of Motors", IEE Conference on Electrical Machines - Design and Applications, Pub. #213, July 1982, pp. 231-35.

[4] E.I. Shobert, Carbon Brushes - The Physics and Chemistry of Sliding Contacts, Chemical Publishing Co., Inc., New York

[5] Yu. E. Nebolyubov, "Photo-electric Method of Testing and Adjusting the Commutation of Electrical Machines", Elektrichestvo, No. 11, Nov. 1956, pp. 34-37

[6] V.N. Wheaton, "Use of Photomultiplier in Commutation Analysis Measurements", G.E. TIS Paper No. DF61LC992, 2/24/61

[7] S.J. Roumanis, "A New Method of Commutation Analysis", G.E. Report No. GIZ545 IE-1007

[8] Y. Michiguchi, S. Tanisaka and S. Izumi, T. Watanabe and I. Miyashita, "Development of a Collector Ring Monitor for Sparking Detection on Generators, IEEE Trans. on Power Appa-ratus and Systems, Vol. PAS-102, No 4, April 1983, pp 928-33.

MOTOR CURRENT ANALYSIS FOR THE DIAGNOSIS OF FAULT CONDITIONS IN A MOTOR DRIVEN PUMP

Ens. Jonathan A. Siegler, USN
Antal A. Sarkady
Department of Electrical Engineering
U.S. Naval Academy
Annapolis, MD 21402

Christopher P. Nemarich
Naval Surface Warfare Center, Carderock Division
Annapolis, MD 21402

Abstract: The detection and diagnosis of faults in Navy shipboard equipment requires a large investment in man-hours and money. The diagnosis of machinery problems often requires that equipment be shut down and taken apart. In addition, the machinery may be located in an area of the hull that is very difficult to reach. Motor current signature analysis (MCSA) can provide a solution to some of these problems. It is a non-invasive technique for the monitoring and diagnosing of mechanical problems in motor driven equipment. This paper presents the development of signal processing and classification techniques used to distinguish between the normal and eroded impeller conditions of a submarine sea water pump. Demodulation techniques are presented for the extraction of relevant signal features used to create pattern vectors for classification. The ultimate goal of this research is to develop MCSA techniques which may be applied to a broad variety of shipboard motor driven machineries.

Key Words: Analytic signals; condition based maintenance; demodulation; digital signal processing; electrical current monitoring; fault detection; machinery monitoring; neural networks; pattern recognition; shipboard machinery

Introduction: The U.S. Navy is presently developing condition based maintenance technologies for the detection and diagnosis of shipboard machinery faults. In concert with this effort, the Navy is pursuing motor current signature analysis as a diagnostic tool for motor driven machineries.

Tests were conducted on a submarine service motor driven sea water pump. The goal of this project was to find rugged signal features in the electrical current signal that could be used to distinguish between a pump working with a good impeller and a pump working with an eroded one. Signal processing, demodulation and classification techniques were then developed and applied in an attempt to distinguish normal pump operation from eroded impeller operation.

The Problem: Over a sea water pump's operating life, the original smooth finish of its impeller becomes worn and pitted. This results in inefficient operation and may ultimately result in the structural failure of the impeller. Since the pitted or eroded impeller condition is not easily detected with conventional vibration analysis, motor current signature analysis was applied as a possible solution. It was postulated that changes in the motor torque load due to the eroded impeller condition would cause changes in the input current signal driving the induction motor. In other words, changes in the reflected load from the pump would be detectable in the electrical current signal.

Pump Test and Data Acquisition: A submarine service, centrifugal sea water pump driven by a three phase, 60 Hz, 440 VAC induction motor was used for these tests. Tests were conducted on a pressurized circulation loop located at NSWC Carderock Division's Annapolis Detachment. The tests used for analysis were performed at the pump's high speed and at a high suction pressure. Tests were run with an impeller in good condition and with a badly eroded impeller which was pulled from submarine service.

The pump was instrumented with current and potential transformers, accelerometers and a tachometer. The current transformers were attached to each phase lead and the potential transformers were used to measure the individual phase voltages. The analog phase voltages were multiplied by their corresponding phase currents with analog circuitry to obtain the real power signal, P. Each phase current and voltage, the real power, the accelerometer signals and the tachometer signal were digitized and recorded using a 16 channel, digital audio tape (DAT)recorder.

Signal Analysis: The analysis of the pump motor current is complicated by the low frequency (60 Hz) and large amplitude of the power source. It was determined in previous efforts that one solution is to sum the squares of the phase currents and voltages to produce the power signal, P.[1,2] P retains the information available in the phase current signal, I, but the large, 60 Hz carrier is removed. This method appears to be most sensitive to small

signals superimposed on the source current but it is not the only means of removing the 60Hz carrier. Others have demonstrated demodulation techniques which use only a single phase current signal. [3,4]

The first step in the analysis was to examine the frequency spectra of the power signals corresponding to both a good impeller and an eroded one. Twelve, 2 minute time series samples taken using the eroded impeller and eight, 2 minute time series samples recorded using the good impeller were selected for comparison. The signal-to-noise ratios of the original power signals were poor. Although there seemed to be signal peaks in the good impeller spectrum that did not exist in the eroded impeller spectrum, they were not statistically significant and could have been due to the variance of the background signal.

Ensemble Averaging: To improve the signal to noise ratio of the power signal, a technique of data segmentation and ensemble averaging known as the Bartlett smoothing procedure was applied.[5] This technique can be used to detect spectral lines buried in additive Gaussian noise even with extremely poor signal-to-noise ratios (SNR).

Ensemble averaging requires the original signal be divided into smaller, mutually exclusive time records or partitions. After the original series is partitioned, the magnitude of the Fast Fourier Transform for each record is computed and stored. Since the records are shorter than the original time series, the resolution in the frequency domain will diminish by a factor of $N_{records}$.

Each of the individual short, frequency domain estimates, or spectral records, have the same SNR as the original long time series. Taken separately, each spectral record is of no more use than the original series, but if the records are ensemble averaged together, the SNR improves approximately as the square root of the number of averaged records. If the original series is longer than necessary to resolve any two required frequencies, then segmentation is possible and averaging is an excellent method of improving the SNR. It is important to note that the magnitude of the frequency spectrum must be used in the averaging process. It is impossible to average in the time domain because both the period and phase of the original, information carrying signal are unknown. Taking the magnitude prevents phase cancellations from occurring in the frequency spectrum.

For a fixed total data length, segmentation and the ensemble averaging process have their trade-offs. An increase in the SNR can only be achieved at the expense of reduced frequency resolution. A compromise had to be made

between both the frequency resolution and the size of the FFT. In order to increase the frequency resolution at the lower frequencies, both the sampling frequency and record lengths were modified. The digital data were low-pass filtered at the Nyquist frequency to prevent aliasing from occurring and resampled. A tenth order digital Butterworth filter was designed and implemented to accomplish this.[6]

Figure 1 shows the effect of the averaging process on the power signal spectra. After ensemble averaging 24 records together, the SNR increased by approximately a factor of 5. The variance in the noise was significantly reduced and spectral peaks were considered reliable. As hypothesized, there seemed to be some type of modulating process in the signal. This was evidenced by the sidebands associated with the spectral lines at 60 Hz, 120 Hz, 180 Hz, 240 Hz, and 360 Hz. At this point, it was decided to digitally demodulate the power signal in order to extract

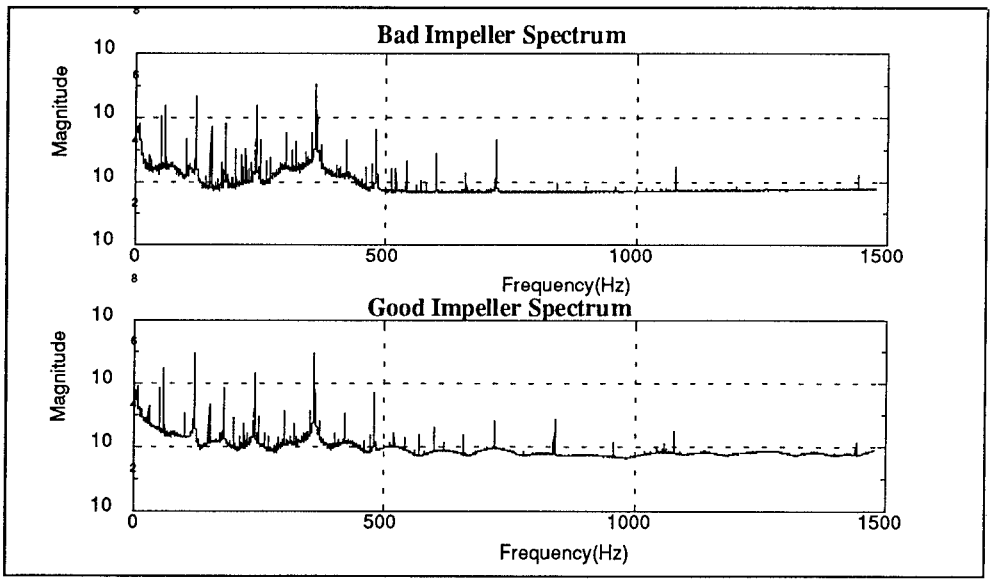


Figure 1 - Power signal spectra ensemble averaged 24 times.

fault related signal features.

Digital Demodultaion: As mentioned before, the forces acting on the pump impeller modulate the motion of the motor shaft and therefore, these forces show up as modulations of the power signal. In order to separate the desired signal information, the power signal was demodulated with the Hilbert Transform. If s(t) is a real signal, then the Hilbert transform [7] is defined as the convolution of s(t) by $(1/\pi*t)$:

$$s_h(t) = s(t) * \left(\frac{1}{\pi \cdot t}\right) = \frac{1}{\pi} \cdot \int_{-\infty}^{\infty} s(\lambda) \cdot \frac{1}{t - \lambda} \, d\lambda$$

After taking the Fourier Transform of $s_h(t)$, and shifting to the frequency domain, we obtain:

$$F\left(s_h(t)\right) = S_h(\omega) = S(\omega) \cdot F\left(\frac{1}{\pi \cdot t}\right)$$

and

$$F\left(\frac{1}{\pi \cdot t}\right) = \frac{1}{\pi} \cdot \int_{-\infty}^{\infty} \frac{1}{t} \cdot e^{-j \cdot \omega \cdot t} \, dt = -j \cdot \mathrm{sgn}(\omega)$$

where F is the Fourier Transform operator, and sgn(ω) is defined as: sgn(ω) = 1 for ω>0, 0 for ω=0 and -1 for ω<0

The Hilbert Transform is known as a quadrature function because each component of $S_h(\omega)$ is in phase quadrature with $S(\omega)$. In other words, the frequency components are 90 degrees out of phase with the original spectrum. In order to demodulate s(t), the analytic signal, z(t) is created from the original signal, s(t) and the Hilbert transform signal, $s_h(t)$, as:

$$z(t) = s(t) + j \cdot s_h(t)$$

It is important to note that, while s(t) is a signal that exists in the real world, z(t) is a contrived, complex signal. After taking the Fourier Transform of the analytic signal, and moving to the frequency domain we obtain:

$$F(z(t)) = Z(\omega) = S(\omega) + j \cdot S_h(\omega) = S(\omega) + j \cdot (-j \cdot \mathrm{sgn}(\omega) \cdot S(\omega))$$

and

$$Z(\omega) = S(\omega) \cdot (1 + \mathrm{sgn}(\omega)) = \begin{cases} 2 \cdot S(\omega) & \text{for } \omega > 0 \\ S(\omega) & \text{for } \omega = 0 \\ 0 & \text{for } \omega < 0 \end{cases}$$

Z(ω) is an upper single-sidebanded signal in the baseband. In other words, it is created by zeroing the negative side of the original frequency spectrum, and doubling the positive side, excluding the DC component. This same technique is used to produce single-sideband modulation.

The Hilbert transform was applied as follows. The original

power signal , s(t) contains amplitude, phase and frequency modulation signals and can be written as:

$$s(t)=a(t)\cdot\cos\left(\omega_c\cdot t+\phi(t)\right)$$

Where a(t) is the amplitude modulating signal, $\phi(t)$ is the angle modulating signal, and ω_c is the carrier frequency. It can be shown that z(t) is of the form:

$$z(t)=s(t)+j\cdot s_h(t)=a(t)\cdot\cos\left(\omega_c\cdot t+\phi(t)\right)+j\cdot a(t)\cdot\sin\left(\omega_c\cdot t+\phi(t)\right)$$

$$z(t)=a(t)\cdot e^{j\cdot\left(\omega_c\cdot t+\phi(t)\right)}$$

Z(t) is now a complex representation of the original signal, s(t) and can be used to AM, PM, and FM demodulate the original signal.

$$|z(t)|=|a(t)|\cdot\left|e^{j\cdot\left(\omega_c\cdot t+\phi(t)\right)}\right|=|a(t)|\cdot 1=|a(t)|$$

$$Ph(z(t))=\theta(t)=\arctan\left(\frac{Im(z(t))}{Re(z(t))}\right)=\omega_c\cdot t+\phi(t)$$

Taking the absolute value and phase of z(t) produces the amplitude modulation a(t). The phase of z(t) leads to the sum of the carrier signal and the angle modulation $\phi(t)$. The carrier signal can be subtracted, leaving only the phase modulation. The FM modulation is found by taking the derivative of the angle modulation.

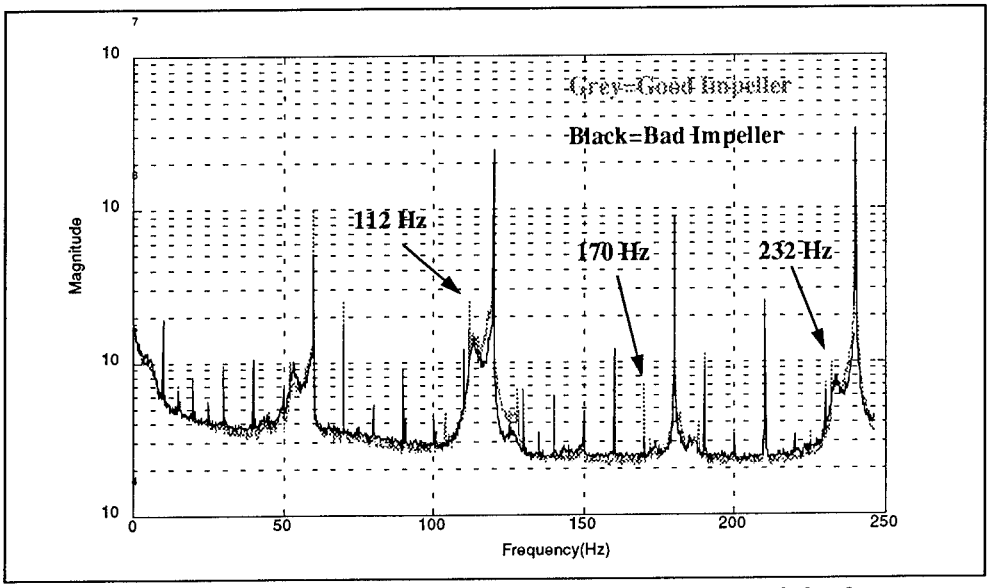


Figure 2- AM demodulated templates from good and bad impeller signals.

This scheme was applied to all of the original power data. Figure 2 depicts the results of the AM demodulation process. Not only did the processed data have a reliable SNR, there seemed to be spectral lines that were enhanced by the demodulation process. At certain frequencies there were significant energy differences between the two spectra. From an initial inspection of the different samples, the disparities seemed to be statistically significant. All the spectra of a certain condition exhibited the same characteristics. Similar results occurred in both the PM and FM demodulated spectra.

Signal Classification: Data from tests conducted with the same impeller under the same conditions but at different times were examined. Although the background noise level shifted slightly, the shapes of the demodulated spectra remained identical. This suggested that the pump process itself was stationary and that the selection of rugged signal features could be performed.

To identify features that were common to a certain condition and that could be used to differentiate it from the other condition a good and a bad set of templates were formed. A template was created by taking the ensemble averaged spectra of a given condition and ensemble averaging them together again. Templates were formed for the original spectrum, and the AM, PM, and FM demodulated spectra. These eight templates, four good and four bad, provided a visual representation of the average signal produced from a pump working with a good impeller, and the average signal produced from a pump working with an eroded impeller.

The four sets of corresponding templates were compared, and examined for significant differences in energy. A 1% test was developed to determine whether or not a signal feature was rugged. To pass the 1% test, there had to be less than a 1% chance that the difference in energy at a certain frequency was due to the variance of the noise, or, there had to be more than a 99% chance that the energy difference was due to the information and not due to the noise. Since, each of the spectra was ensemble averaged 24 times, the distribution of the noise was assumed to be Gaussian. The marked frequencies shown in Figure 2 easily passed the 1% test. A total of 29 rugged features were found throughout the four sets of templates as follows: 2 from the original spectrum template; 4 from the AM demodulated template; 8 from the PM demodulated template; and 15 from the FM demodulated template.

Although the 1% test provided a criterion to determine if a signal feature is rugged, it did not provide a method by

which to quantify it. To accomplish this, a low-pass moving average filter was developed. The moving average filter compares the area beneath a signal peak to the area beneath the noise surrounding it, producing a signal-to-noise ratio figure. The moving average filter was used to measure the SNR throughout the four spectra of each sample. Consequently, each power data sample had 29 SNRs associated with it, each serving as a component of a 29 dimensional pattern vector. It was this pattern vector or 'fingerprint' that was used to classify the pump condition that created it.

In order to determine the condition of the pump, two classification techniques were employed; the nearest neighborhood technique and a neural net known as the perceptron.

The first scheme implemented was the nearest neighborhood technique.[8] The first step was to create the average good and bad pattern vectors. Two sets of templates were formulated using a training set. The training set consisted of samples whose classification were previously known by the computer. Using the training set, the four distinct templates were created for both the good impeller and the eroded impeller. The moving average filter was implemented to quantify the previously determined rugged signal features. Then, both the average good and the average bad template vectors were created using these SNRs.

After creating the template pattern vectors, the test set was applied to the classification scheme. The test set consisted of pattern vectors whose condition was unknown to the computer, and that were not involved in the formation of the templates. The mean squared distance between the test pattern vector and each of the template vectors was calculated. The test vector was then classified as the template vector that produced the smaller distance value.

Using this technique, 90% of the test set was classified correctly. Even in the worst case where the system broke down, 19 of the 29 individual components were classified correctly.

It is important to note that the nearest neighborhood technique weights each of the individual components of the pattern vector equally. This led to the hypothesis that weighting the components differently would lead to a higher classification efficiency. In order to weight the individual components differently, a simple neural net known as a perceptron was implemented.[9] The perceptron multiplies each of the individual components of the pattern vector by a specific weight. These weighted components are added together and then added to an offset. The function

hardlimit is equal to 1 when the input is greater than 0 and equal to 0 when the input is less than or equal to 0. Consequently, when the total sum is greater than 0, the neuron fires producing a value of A=1. If the total sum is less than 0, then the neuron does not fire producing a value of A=0. The perceptron was designed to produce a value of A=1 for a good impeller, and a value of A=0 for an eroded impeller.

The weights and offset are calculated in a recursive training process following an established learning rule. Again, the training set consists of samples whose classifications are known to the computer. The weights and offset are adjusted until all members of the training set are classified correctly. The following rule is used: Case (1) If after the presentation of a training vector, the output of the neuron is correct, the weights and offset remain unchanged. Case (2) If the output of the neuron was a 0 and should have been a 1, the weights are increased by the value of the individual components of the training vector, and the offset is increased by 1. Case (3) If the output of the neuron was a 1 and should have been 0, the weights are decreased by the value of the individual components of the training vector, and the offset is decreased by 1. Following this rule it took approximately 100K recursions to train the perceptron.

Using the trained perceptron, 100% of the test set was classified correctly. By simply shifting the weights of the individual components it was possible to raise the classification efficiency from 90% to 100%. It is important to note that 100% classification of the training set can only be achieved if the vectors are linearly separable, otherwise, a more complex neural network must be employed.

Conclusion: By using both the digital demodulation and the process of ensemble averaging, a signal processing scheme was developed and implemented for classifying the impeller condition of a sea water pump. The algorithm involves breaking the original two minute time series into 24, 5 sec time records. Both the frequency spectrum and analytic signal were then formed from each of the individual records. The analytic signal was further used to AM, PM, and FM demodulate the power data. Finally, each of the different types of spectra were ensemble averaged 24 times to improve the SNR. Rugged signal features were selected from these signals and pattern vectors were formed. Using the nearest neighborhood technique, the pump condition was classified correctly 90% of the time. Using the trained perceptron, 100% of the test sets were classified correctly.

Although the basic algorithms have been created, further testing is required to create a technique which can be used at different pump speeds and suction pressures. To potentially reduce the amount of required instrumentation, the sensitivity of single phase current analysis will be compared to the three phase analysis performed here. Finally, it is hoped this MCSA effort will provide the Navy with an efficient and inexpensive method for the monitoring of motor driven pumps.

References:

1. Tate, Ronald C., Christopher Nemarich, and Diane Porter, "Current Analysis for the Condition Assessment of Shipboard Motor Driven Machinery", Proceedings of the 46th Meeting of the MFPG, Vibration Institute, Willowbrook, IL, April 1992, pgs 277-289.

2. Tinston, Michael and Antal A. Sarkady, "Motor Current Analysis for the Diagnosis of Air Compressor Defects", Proceedings of the 46th Meeting of the MFPG, Vibration Institute, Willowbrook, IL, April 1992, pgs 291-300.

3. Smith, S.F., K.N. Castleberry, and C.H. Nowlin, "Machine Monitoring Via Motor Current Demodulation Techniques", Proceedings of the 44th Meeting of the MFPG, Vibration Institute, Willowbrook, IL, April 1990, pgs 87-96.

4. Smith, S.F., and K.N. Castleberry, "Advanced Techniques in Current Signature Analysis", Proceedings of the 46th Meeting of the MFPG, Vibration Institute, Willowbrook, IL, April 1990, pgs 263-275.

5. Jenkins, Gwilyn and Donald Watts, Spectral Analysis and Its Applications, Holden-Day, CA, 1968.

6. Hush, Don and Samuel Sterns, Digital Signal Analysis. Prentice Hall, NJ, 1990.

7. Stremler, Ferrel, Introduction to Communication Systems, Addison-Wesley Publishing Company, MA, 1990.

8. Kapouleas, Ioannis and Sholom Weiss, "An Empirical Comparison of Pattern Recognition, Neural Nets and Machine Learning Classification Methods" in Readings in Machine Learning, Ed. Jude W. Shavlik and Thomas G. Dietterich, Morgan Kaufmann Publishers, CA, 1990, pgs 177-183.

9. Kosko, Bart, Neural Networks and Fuzzy Systems, Prentice Hall, NJ, 1992.

A DEMONSTRATION OF SENSORLESS TORQUE MEASUREMENT

Gerald B. Kliman, Rudolph A. Koegl & Rik W. DeDoncker*

General Electric, Corporate R&D
Schenectady, NY 12301-0008

ABSTRACT: It has been demonstrated that, in theory, the electromagnetic torque of an induction motor may be precisely calculated, in a fairly straightforward manner, using only the terminal voltages and currents without knowing any of the motor dimensions or parameters except for stator resistance. The torque theorem has been implemented on a low performance PC-AT. Steady state torque estimation via the theorem was verified to a precision and accuracy of better than 1% in either polarity using a standard, low cost, 3/4 HP induction motor and a laboratory grade, high precision electromechanical torque meter. Dynamic data acquisition, processing and display was performed to demonstrate that it could be done on a low performance platform.

KEY WORDS: AC motor; induction motor; sensorless; torque; torque measurement; torque estimation.

INTRODUCTION: In many industrial applications it is necessary to measure the torque applied to the load of an induction motor. This is typically done by fitting the installation with an electromechanical torque sensor. The installation is expensive and can be impossible in many situations hence unsatisfactory estimators, such as current magnitude, are used as a low cost expedient [1].

The usual means of measuring torque is to apply strain gages to the shaft or a special section of shafting which is part of an instrument. Signals are transferred between the rotating shaft and the stationary system by slip rings, transformers, infra-red optical couplers or RF depending on the manufacturer. In the past special cradled motors were often used. These days a common technique is to place force sensors under each foot of the motor and combine the outputs. Any such mechanical modifications will be expensive and, in many cases, not feasible. Finally the torque may be computed via on-line motor models if the motor parameters are known. Such methods usually need a speed input in addition to the voltages and currents. If sufficient computing power is available and the load and/or line conditions vary sufficiently, it may be possible to continuously adapt an on-line motor model to calculate torque.

---

* Now with Silicon Power Corporation, Malvern, PA

Power electronic drive designers now use control systems known as "field oriented" or "vector" controls in which the flux and current of an induction motor may be separately controlled much as in a dc motor [3],[4]. Part of the background of this technique provides an algorithm by which, in principle, the electromagnetic torque can be calculated from a knowledge of the line currents and voltages only, without knowledge of the motor parameters.

THE TORQUE ALGORITHM: The dynamic behavior of all induction machines may be analyzed by means of an equivalent two phase machine model because the rotating field can be realized regardless of the number of phases actually used. The equivalent two phase machine has sets of windings that are perpendicular making the self inductances decoupled which greatly simplifies the model equations. To transform a three phase (a, b, c) machine into a two phase (d, q) machine the following transformation rule, in complex notation, is used:

$$\underline{x}(t) = x_d(t) + jx_q(t) = x(t)e^{j\varphi(t)} = \frac{2}{3}\left[x_a(t) + ax_b(t) + a^2x_c(t)\right]$$

with: $x_a(t), x_b(t), x_c(t)$: instantaneous (a,b,c) phase quantities

$x_d(t), x_q(t)$: instantaneous (d,q) phase quantities

$\underline{x}(t)$: space vector

$x(t) = \sqrt{x_d^2(t) + x_q^2(t)}$: space vector amplitude

$\varphi(t) = \arctan\left[x_q(t) / x_d(t)\right]$: space vector phase

$a = e^{j\frac{2\pi}{3}}$: 120° rotation operator

Dynamic models for induction machines typically neglect winding asymmetries, winding slot harmonics, magnetic core loses, winding skin effects and air gap unbalance (causing homopolar fluxes). With these assumptions in place, the state equations of the two phase induction machine become:

stator voltage equations:

$$v_{sd} = R_s * i_{sd} + d\psi_{sd}/dt \qquad (1)$$

$$v_{sq} = R_s * i_{sq} + d\psi_{sq}/dt \qquad (2)$$

rotor voltage equations:

$$0 = R_r i_{rd} + \frac{d}{dt}\psi_{rd} + p\omega_m\psi_{rq} \qquad (3)$$

$$0 = R_r i_{rq} + \frac{d}{dt}\psi_{rq} - p\omega_m \psi_{rd} \qquad (4)$$

where: ω_m represents the mechanical speed of the rotor.

The flux linkages, ψ , may be expressed as functions of the stator and rotor currents. Usually each flux component consists of a leakage inductance term and a main (or mutual) inductance term:

$$\underline{\psi_s} = L_{sl}\underline{i_s} + L_h(\underline{i_s} + \underline{i_r}) = L_{sl}\underline{i_s} + \underline{\psi_h} \text{ : stator flux} \qquad (5)$$

$$\underline{\psi_r} = L_{rl}\underline{i_r} + L_h(\underline{i_s} + \underline{i_r}) = L_{rl}\underline{i_r} + \underline{\psi_h} \text{ : rotor flux} \qquad (6)$$

$$\underline{\psi_h} = L_h(\underline{i_s} + \underline{i_r}) \text{ : air gap flux} \qquad (7)$$

The machine parameters in the equations are:

R_s stator resistance

R_r rotor resistance

L_{sl} stator leakage inductance

L_{rl} rotor leakage inductance

L_h main (self or magnetizing) inductance

p number of pole pairs

Normally the resistances are temperature dependent while the inductances are dependent on the saturation level. In practice, their dependencies have little effect when operating at medium to high frequencies (20 to 80 Hz).

The electromagnetic torque produced by the induction machine may be calculated from:

$$T_{em} = \frac{3}{2} pI_m\left[\underline{\psi_n} \bullet \underline{i_r^*}\right] = \frac{3}{2} p(\psi_{hd} i_{rq} - \psi_{hq} i_{rd}) \qquad (8)$$

Equation (8) states that torque results from the interaction between air gap flux ψ_h and the rotor current i_r . However both the air gap flux ψ_h and the rotor current are difficult to measure. Fortunately, the flux linkage expressions (5) to (7) allow equation (8) to be rewritten as a function of stator quantities that can be measured directly hence:

$$T_{em} = \frac{3}{2} pI_m\left[\underline{i_s} \bullet \underline{\psi_s^*}\right] = \frac{3}{2} p(i_{sq}\psi_{sd} - i_{sd}\psi_{sq}) \qquad (9)$$

In equation (9) the current components i_{sd} and i_{sq} may be measured using conventional current sensors while ψ_{sd} and ψ_{sq} are derived from the stator voltages v_{sd} and v_{sq} according to equations (1) and (2) or in complex format:

$$\underline{\psi_s} = \psi_{sd} + j\psi_{sq} = \int_o^t (\underline{v_s} - R_s \underline{i_s})dt \qquad (10)$$

As a result, an algorithm that depends only on the estimated (or measured) value of the stator resistance R_s may be derived to calculate the instantaneous value of electromagnetic torque. First, the d, q components of the stator voltages and currents are determined using (1):

$$v_{sd} = v_a \qquad (11)$$

$$v_{sq} = \sqrt{3}/3(v_b - v_c) = \sqrt{3}/3\, v_{bc} \qquad (12)$$

$$i_{sd} = i_a \qquad (13)$$

$$i_{sq} = \sqrt{3}/3(i_b - i_c) = \sqrt{3}/3(i_a + 2i_b) \qquad (14)$$

Next, the d, q components of the flux vector, $\underline{\psi_s}$, are computed using the integration formula (10). To avoid start-up effects (when the integration is applied to a machine that is already excited) a high pass filter is required. This high pass filter has to be carefully selected to prevent phase shift errors in the d, q flux components. A simple, first order, high pass filter (sometimes called a lag filter or a forgetting filter) is adequate when frequencies are typically above 10 to 15 Hz:

$$\underline{\psi_s} = \int_o^t (\underline{v_s} - R_s \underline{i_s} - \omega_f \underline{\psi_s})dt. \qquad (15)$$

where: ω_f represents the lag filter bandwidth,
$\underline{\psi_s}$ stator flux,
R_s stator resistance,
$\underline{v_s}$ stator voltage,
and $\underline{i_s}$ stator current.

CHALLENGES TO THE ALGORITHM: Objections to this method of calculating torque focus on the assumptions of the model set out in the previous section. When the algebra of the derivations is carried out in detail a number of cancellations of leakage inductance terms in the perpendicular axes occur. The identity of these terms has been questioned in view of cross couplings that could occur when the induction machine magnetic core is saturated [2]. Other issues involve the neglect of core losses and stray load losses. As may be seen from the test data these effects are either non-existent or negligible in real induction machines.

OBJECTIVES OF THE TEST: While the algorithm has been known to a narrow group of electronic control specialists for some time it has been used only as a rough indication of system performance or as a feedback signal. Its usefulness as a monitoring and diagnostic tool in general industry applications has not been explored. Even the accuracy of the method has been called into question on theoretical grounds [2]. This is a serious issue if the method is expected to compete with laboratory grade instruments having accuracy and precision of better than +/- 0.5%. Furthermore, the feasibility of implementing the algorithm, at low cost, in a free standing instrument was in question.

The investigation reported here was undertaken to first verify that a practical system could be implemented on low performance, low cost platforms. In the second place it was necessary to verify that the method, when implemented on such platforms, will deliver high accuracy and precision even when applied to low cost, highly saturated, inefficient and unbalanced "commodity" motors. Furthermore, this performance had to be consistent and drift free for extended time periods.

TEST SYSTEMS: A test dynamometer was set up with a low cost, 3/4 HP, NEMA frame 56, TEFC, three phase, 230/460 volt, single circuit, concentric wound, Y connected induction motor as the test article. The test motor was loaded by a permanent magnet DC motor, of the same rating, running as a generator with a controllable resistor load. The load on the generator consisted of fixed resistors and an electronically variable resistor for maximum flexibility. A conventional, laboratory grade, in-line, torque meter was installed between the test motor and the load. The data acquisition system was implemented using an old IBM 286 AT and "off the shelf" amplifiers, filters and A/D converter boards for laboratory convenience, which could be plugged into the expansion slots of the computer. The data acquisition, torque computation and display software were compiled and run on the AT. A depiction of the dynamometer and data acquisition systems organization is shown in figure 1.

The arrangement of the motor circuit is shown in figure 2. Conventional, switchboard grade current and potential transformers were used to simulate, as closely as possible, the situation that would be encountered in the field. Three phase power was supplied to the motor through a variac. This was done because, in early tests, the electromechanical torque meter experienced a set in its measuring element manifested as a zero shift every time the motor was started on the line. Eventually the instrument was severely enough damaged that it had to be replaced. The variac was used in these tests only to soft-start the motor and avoid damage to the torque sensor. Precision (0.1%) resistors were used for the burdens and voltage dividers.

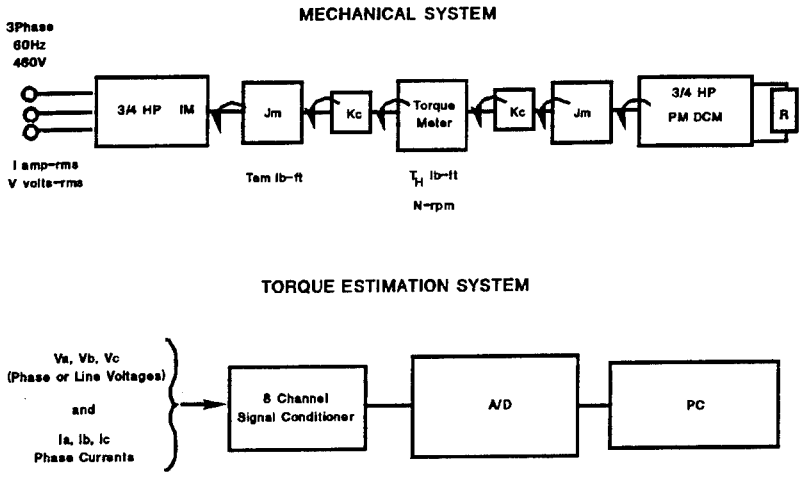


Figure 1 - Dynamometer and Data Acquisition Systems

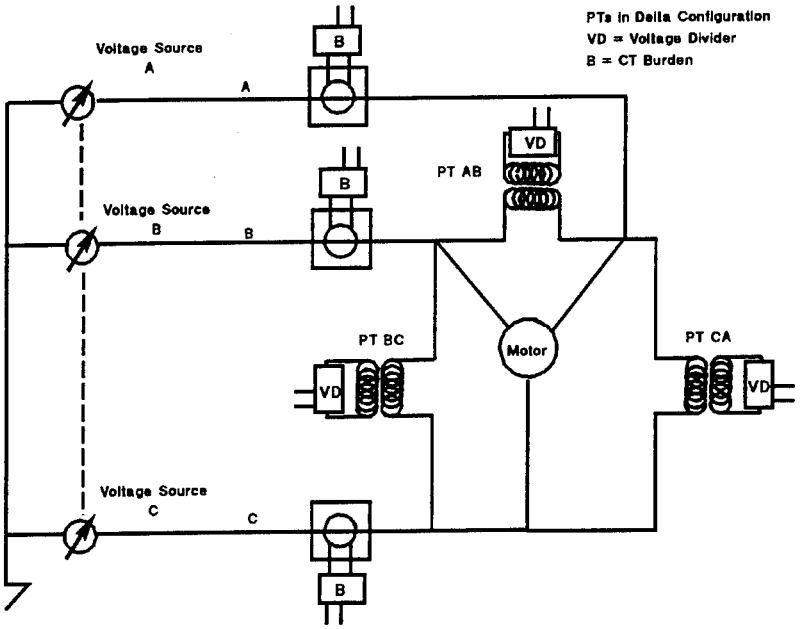


Figure 2 - Motor and Instrument Circuit

To establish the calibration as closely as possible the conventional torque meter was first recalibrated using certified weights and a precision torque arm. The transformer's calibrations were verified using ordinary electronic laboratory instruments. No other adjustments or calibration factors were applied to the algorithm. To obtain the best accuracy in estimating motor flux the motor resistance was measured with an ohm meter so that the IR drop could be subtracted from the line voltage in the computations.

Although not a precision measurement in itself the resistance drop is such a small part of the voltage that this approach is quite adequate.

The data acquisition system utilized high quality, digitally controlled anti-aliasing filters and amplifiers. A 16 bit, sample and hold, A/D converter board with eight differential channels was plugged into a spare slot in the computer. Three voltages and three currents were acquired continuously at a 1 kHz sampling rate. The remaining two channels were used to record the analog speed and torque outputs of the electromechanical torque meter.

TEST RESULTS AND DISCUSSION: The test results showed that the electromagnetic torque calculation was at least as precise and accurate as the electromechanical torque meter. Figure 3 is a plot of steady state electromagnetic torque, as calculated from the voltages and currents via the algorithm, versus the torque measured by the electromechanical instrument. To assure simultaneity and precision the analog output of the torque meter was fed to one channel of the A/D converter that was used to acquire the voltages and currents. The agreement of the torques is illustrated by the straightness of the line connecting the data points. The slight jog near the origin turned out to be due to a small but persistent offset in the torque meter analog output that could not be adjusted out although it did not appear to be present in the meter's digital display.

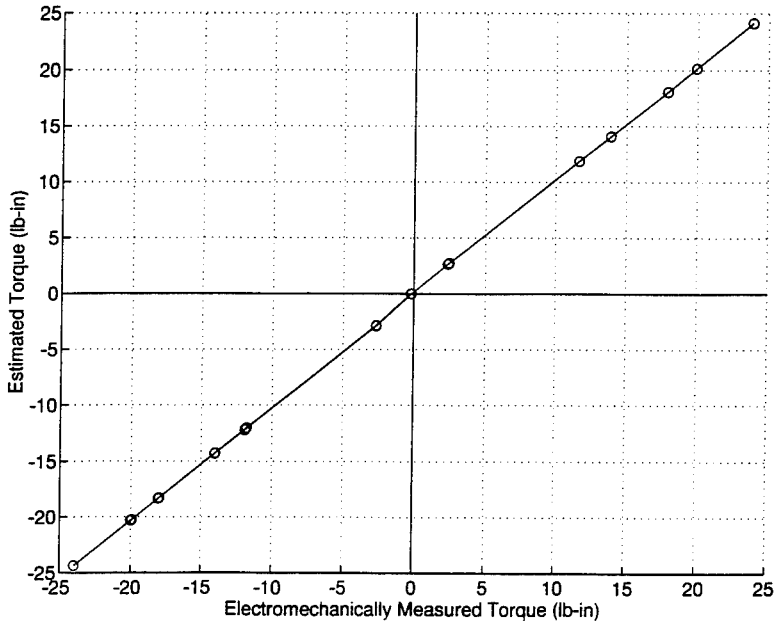


Figure 3 - Steady State Torque Comparison

The only systematic errors in the data, that could be traced, were in the conventional instrument leading us to believe that the method was, in fact, more precise than the electromechanical instrument. Figure 4 utilizes the same data as figure 3 but is

processed to display the torque error, defined as the electromagnetically calculated torque minus the electromechanically measured torque, versus the measured torque. The effect of the torque meter analog output offset is clearly seen. Even with the offset, the two torques track within 1.5 %. It is clear that, if the offset could have been removed, the torques would have tracked to within about 0.3 % (+/- 0.15 %). The mechanical hysteresis of the torsion shaft, on which the meter's strain gages are mounted, is apparent and accounts for about 0.1 % of the error.

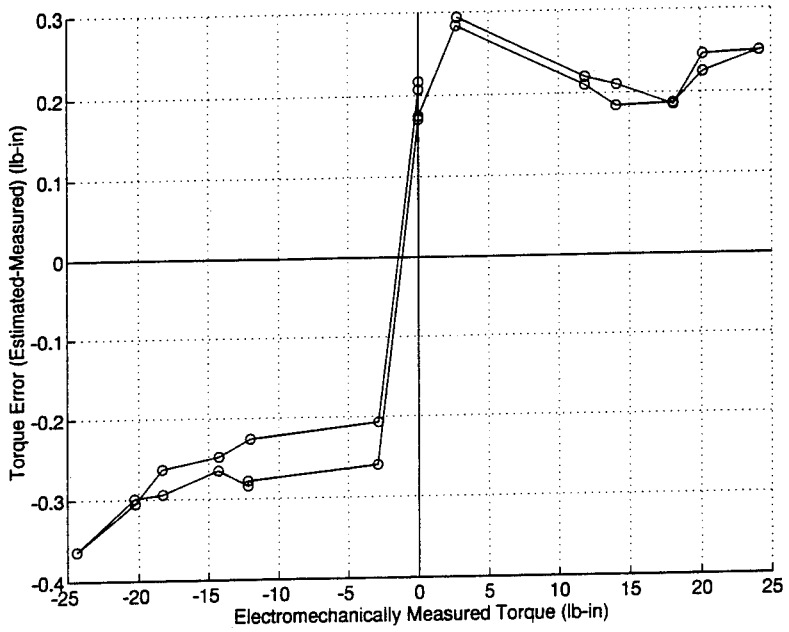


Figure 4 - Torque Error vs Electromechanically Measured Torque

As the motor was stepped through a series of loads in both directions the data was processed to display current magnitude or real power and both torques so that all of them could be compared in real time. The result of one such demonstration is shown in figure 5 where average real power is displayed on a scale adjusted to coincide with the maximum torque. It is easily seen that the electromagnetically calculated and electromechanically measured torques coincide over the entire range of positive and negative torque. The only exceptions are brief periods during the startup of the motor when there is sufficient acceleration to yield significant inertia torque. It is also clear that core losses and magnetizing current losses cause considerable error in using real power as an indicator of load, especially at low loads. Of course, the real power is not sensitive to the sign of the torque. A similar, but greater, disparity exists when attempting to use the current as an indicator of load.

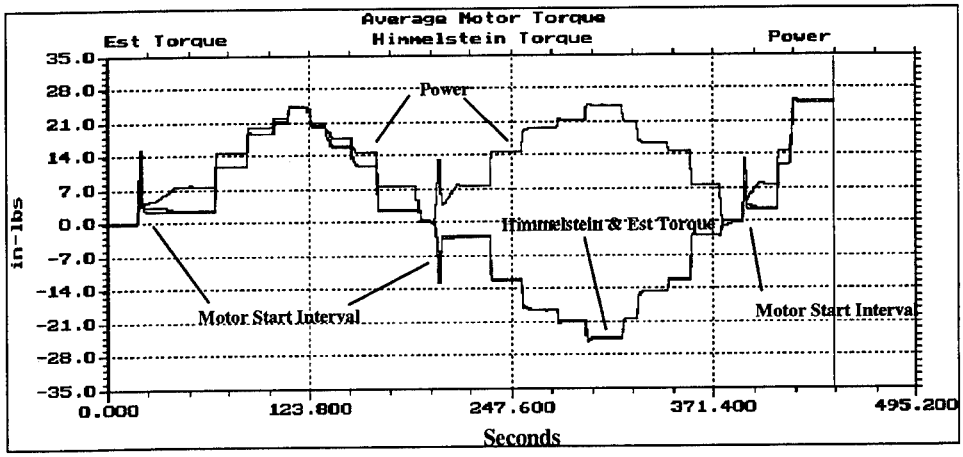


Figure 5 - Torque Demonstration Screen

CONCLUSION: The steady state torque of induction (and synchronous) motors may be straightforwardly calculated from the terminal voltages and currents without knowledge of the motor parameters or dimensions. A simple measurement of stator resistance will improve the accuracy. Furthermore the calculation can be implemented, in real time, on low performance platforms. The resulting accuracy and precision is dependent on the calibration of the current and voltage sensors but can be better than laboratory grade electromechanical instruments and on the order or +/- 0.3 % or better.

REFERENCES:

[1] Kliman, G.B., Srinivason, K., & Ritscher, D.E, "The limits of Motor Current Signature Analysis", Economic Implications of Mechanical Failure Prevention, The Vibration Institute, 1992

[2] Vas, P., "Parameter Estimation, Condition Monitoring and Diagnosis of Electrical Machines", Clarendon Press, Oxford 1993

[3] Blaschke, F., "Das Verfahren de Feldorientierung zur Regelung der Asynchronmaschine", Siemens Forsch.-u.Entwickl-Bei,Springer Verlag, Bol 1 Nr.1, 1972, p184-193.

[4] DeDonker, R., "Synthesis and Digital Implementation of Adaptive Field Oriented Controllers for Induction Machines.", Diss. Thesis, University Leuven, Belgium, Dec. 1986.

DETECTION OF ROTOR DEFECTS BY ENHANCED TORSIONAL VIBRATION MONITORING

W.C. Hernandez
R.E. Sutermeister
W.D. Tang

Monitoring Technology Corporation
2731 Prosperity Avenue Suite A
Fairfax, VA 22031-4308

Abstract: Traditional accelerometer techniques for measuring translational vibration of rotating machinery are ineffective in detecting rotor related problems like cracked shafts and cracked blades. Torsional vibration is much more sensitive to these types of problems, but practical methods for accurately measuring torsional vibration have not been available. With this in mind, MTC is developing shaft encoder techniques and new software analysis methods which are easy to employ and have the necessary reliability. In particular, traditional FFT techniques tend to be limited in their application to this problem because of spectral component leakage and RPM variation. MTC is developing methods to eliminate these limitations and achieve new levels of sensitivity. Examples of applications to fan blade crack detection and turbine blade/shaft resonance detection are discussed.

Key Words: Blades; defects; detection; faults; resonances; rotating machinery; rotors; shafts; signal processing; torsional vibration

Introduction: This paper discusses the use of torsional vibrations methods for detecting certain types of faults in rotating machinery. MTC attempted to solve this problem several years ago because of the magnitude of turbine failure problems in the power industry, and related problems in military surface ships, aircraft turbine engines and, in general, critical turbomachinery.

The most common methods of machinery vibration monitoring basically rely on the use of accelerometers and/or proximity sensors for measuring translation vibration, i.e., vibration in the x, y and z directions. These methods are generally very effective for detecting problems like rolling element bearing defects, and balance and alignment problems. However, they are fairly ineffective in detecting certain rotor related defects such as cracked shafts, blade problems and rotor head

cracks. These types of defects tend to reveal themselves primarily as changes in the torsional (or angular) vibration of the machine, however, instrumentation for reliable and accurate measurement of torsional vibration is generally not available. The technology that is available tends to be either invasive, too expensive, difficult to use or lack sufficient sensitivity.

Shaft encoder based techniques are non-invasive and relatively simple to employ on machines, but generally lack the reliability to detect the problems of interest. MTC has substantially increased the reliability of their techniques through the development of new innovative software for processing the signals. We shall discuss this further with two examples. The first is a laboratory test we conducted on cracked fan blades to verify the basic physics of the technique. The second example consists of data collected on a steam power turbine to illustrate the sensitivities required for turbine blade/shaft crack detection.

Example 1, Fan Blade: This example verifies the basic physics for detecting blade defects via a torsional vibration approach where a shaft encoder is employed as the rotational sensor.

Figure 1 shows the basic experimental setup. A 4 bladed propeller type fan is driven by a one horsepower 4-pole induction motor which, without slip, would rotate at 30 Hz. A 60 tooth sprocket with a passive magnetic sensor serves as the shaft encoder. The signal processor calculates a zoomed power spectrum which shows a carrier of 30 x 60=1800 Hz and shaft frequency sidebands located at 1800 $\pm$ 30 x n Hz where n is an integer. We shall see that blade vibration FM sidebands will appear as indicated.

Figure 2 shows, in more detail, that the fan blades are 24" in diameter and the approximate position of a cut introduced into one of the blades to simulate a cracked blade.

Figure 3 shows the instantaneous RPM (in Hz) as a function of time. The average RPM is 30 Hz, but a strong once per shaft turn FM component is also indicated. Each blade also resonates at its fundamental modes which results in a modulation of the rotation frequency. An example of such a blade FM contribution is also indicated.

Figure 4 shows a theoretical spectrum with a carrier located at 1784 Hz (allowing for motor slip) and multiple shaft sidebands. The skirts about the carrier and sideband components result from FFT leakage, caused by the fact that there are generally not an integral number of cycles of each of these components in each data record used for calculating each FFT. Moreover, there is always present some slow variation in the RPM (caused by slow variation in line frequency and voltage) which tends to spread the components out even further. This

leakage/RPM variation problem is the main factor limiting the sensitivity of this technique for detecting small FM components.

Figure 5 shows a segment of the actual right hand sideband for fan data (from 147 Hz to 183 Hz above the carrier) as the crack is introduced. The crack ranges in size up to $1^1/_8$". We clearly see the moveout of the cracked blade from the 3 good blades so that there is a final separation of 10 Hz in these spectral components.

Figure 6 shows a linear regression of crack site versus frequency shift for the five (5) data points. A very good $R^2 = .994$ was obtained as expected.

Figure 7 shows the results of a similar power spectral analysis where translation vibration is detected using an accelerometer mounted on the shaft bearing nearest the blades. Even though the largest crack ($1^1/_8$" length) is present, the translational vibration fails to show any sign of the blade frequencies at 158 Hz and 168 Hz.

Example 2, Steam Turbine: This example shows the results of using this technology for detecting blade and shaft resonances of an operating steam turbine.

Figure 8 shows the basic data collection setup. A bull gear (with 78 teeth) and proximity probe were used as the shaft encoder. The system shaft and the blades of the LP section are the main components of interest. The data was actually digitally recorded and then analyzed off line.

Chart 1 shows the resonances calculated and identified by G.E. engineers for certain torsional modes of this turbine. L-0 and L-1 stage blade resonances of the LP sections are indicated. The other frequencies of 19.1, 124.3 and 144.9 Hz were calculated but the origins were not indicated.

Figure 9 shows a portion of a high resolution zoomed spectrum in the vicinity of the first harmonic of the carrier (30x78=2340 Hz). FM components with modulation frequencies of 19 Hz, 103 Hz, 124 Hz and 145 Hz are readily identifiable. Again we see the strong spreading of the carrier and shaft components which tend to hide the presence of other smaller FM components caused by blade and shaft FM resonances.

Figure 10 shows a smaller portion of the spectrum upper sideband with modulating frequencies between 122 Hz and 148 Hz. In addition to the 124 Hz and 145 Hz components which coincide with the G.E. measurements, we also see components at 128 Hz and 134.5 Hz which G.E. says probably corresponds to torsional modes which they did not calculate.

Figure 11 shows an example of analysis on synthetic data where two small FM signals are present. The higher level spectrum was calculated using a standard

FFT zoom technique. Because of limited resolution, the FFT leakage skirts are very large and they mask the presence of the FM signals of interest. The lower level spectral signal was calculated by a new technique under development at MTC for eliminating leakage/RPM variation effects. The skirts are entirely removed and the FM signals are easily detected. The tremendous increase in dynamic range and sensitivity are obvious. Details of the method will be presented in the near future, as soon as permitted by our foreign patent applications.

Summary: The examples presented here show the value of torsional vibration monitoring for detecting rotor defects like cracked shafts and blades. Ordinary translational vibration methods are ineffective for these types of problems, and the currently available torsional monitoring systems are usually impractical either because they are ineffective or too difficult and expensive to use.

Shaft encoders are a convenient method for measuring torsional vibration, but they tend to be limited in sensitivity because ordinary FFT processing, which is the analysis method generally used, is not adequate for many applications. Several factors contribute to this lack of sensitivity, but the main factor is FFT leakage and the RPM variation of the machine. Thus, in order to measure torsional vibration reliably and with large dynamic range this leakage problem must be solved. MTC has developed software which eliminates FFT leakage and RPM variation effects. As a result, small defects which produce FM sidebands as much as 150 dB below the carrier will be detectable.

Acknowledgment: These developments were supported, in part, by the Electric Power Research Institute (EPRI) and ABB, Combustion Engineering, Inc.

BASIC EXPERIMENTAL SETUP

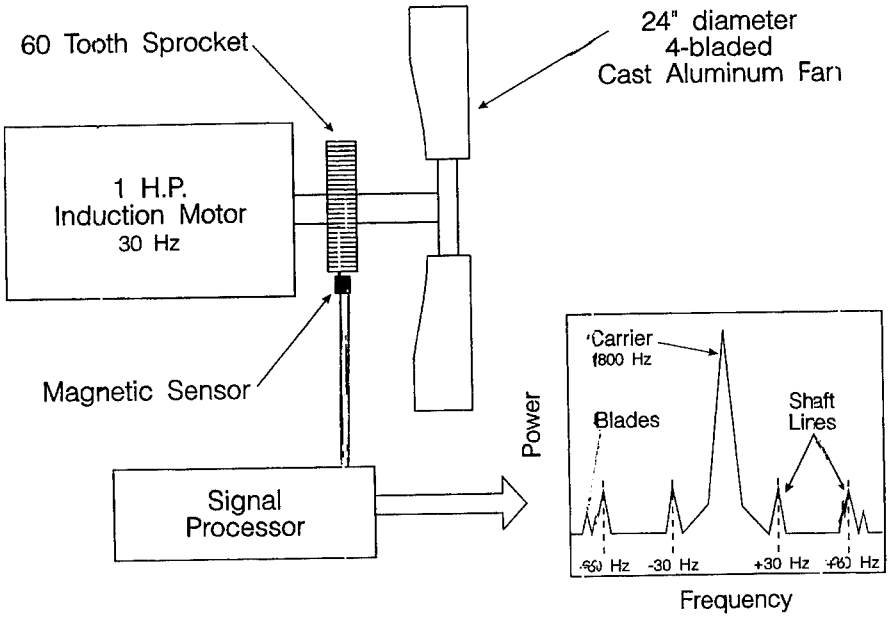


Figure 1

EXPERIMENTAL CAST ALUMINUM FAN WITH 1 CRACKED BLADE AND 3 GOOD BLADES

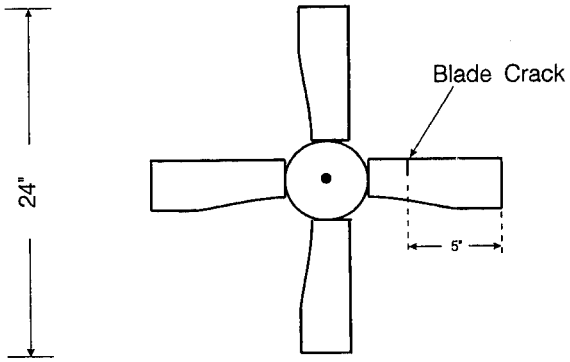


Figure 2

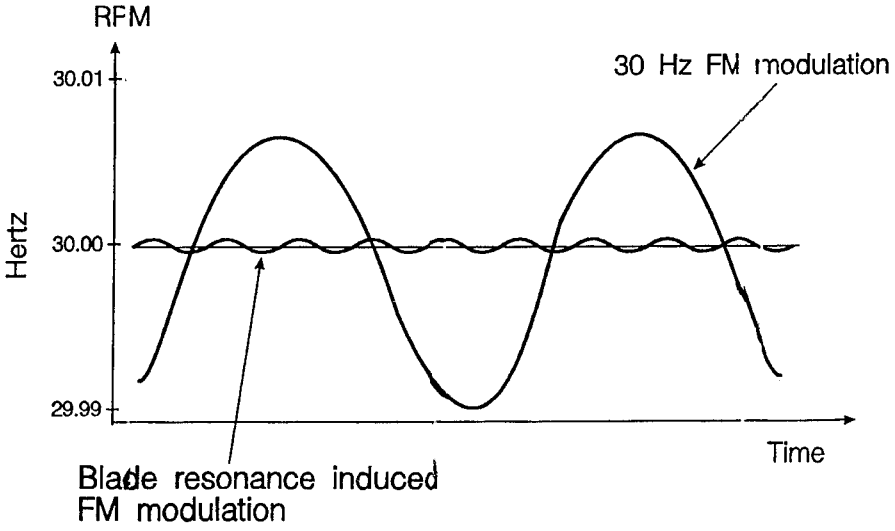


Figure 3

LIMITED SIGNAL PROCESSING

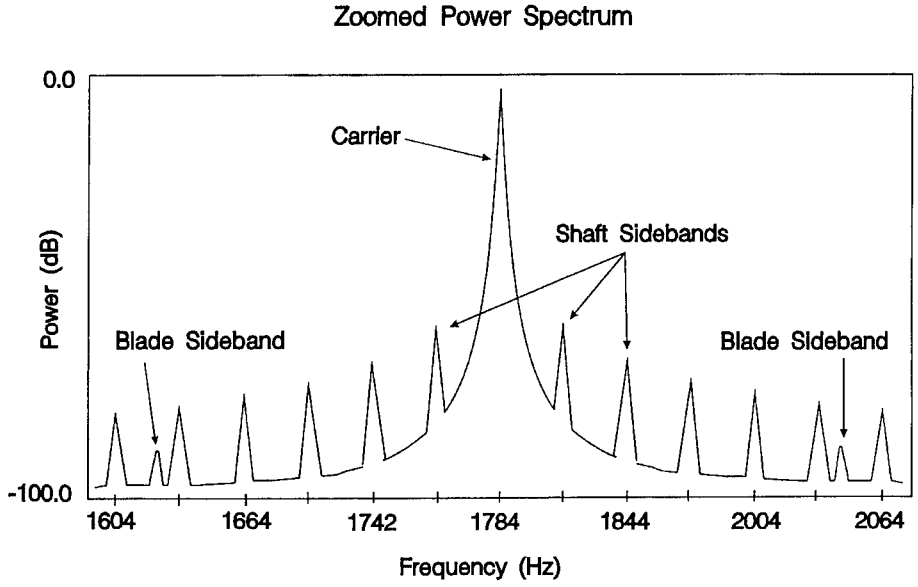


Figure 4

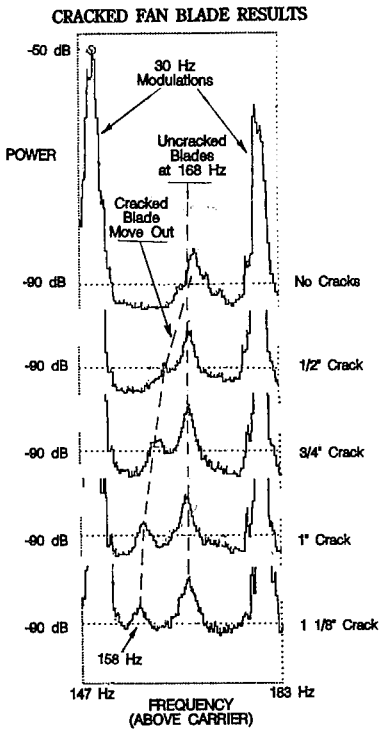


Figure 5

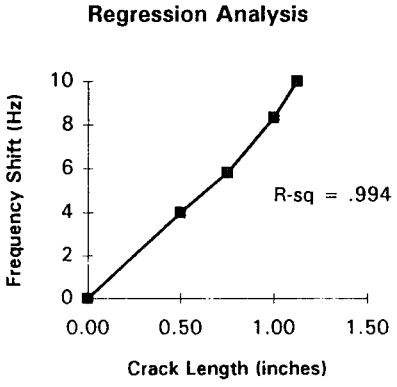


Figure 6

CONVENTIONAL VIBRATION SPECTRUM FOR SINGLE BLADE CRACK OF 1 1/8" LENGTH
NO BLADE RESONANCES ARE PRESENT

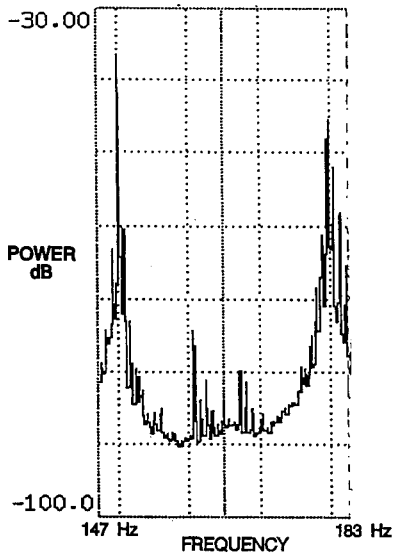


Figure 7

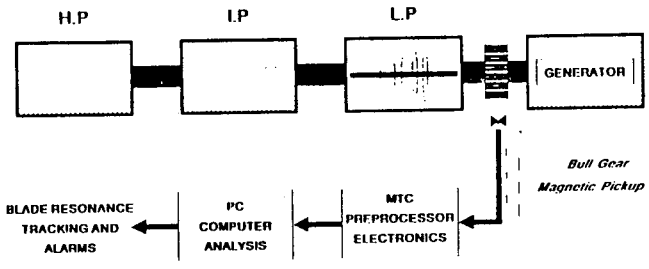


Figure 8

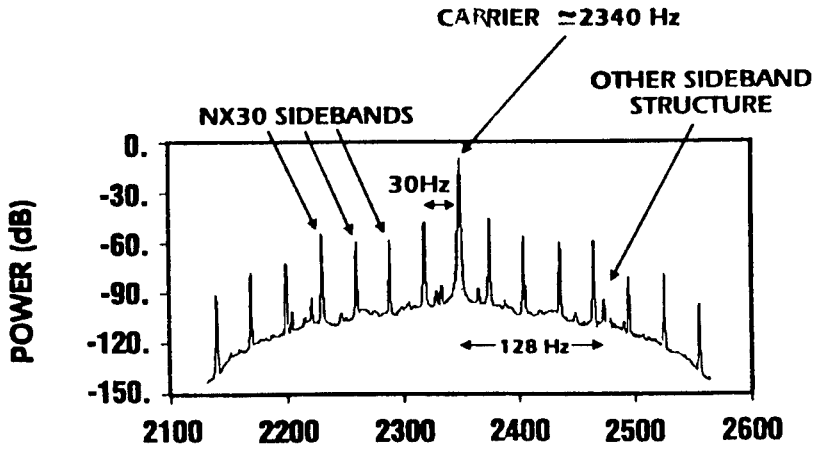


Figure 9 HighResolution Power Spectrum
(all components present)

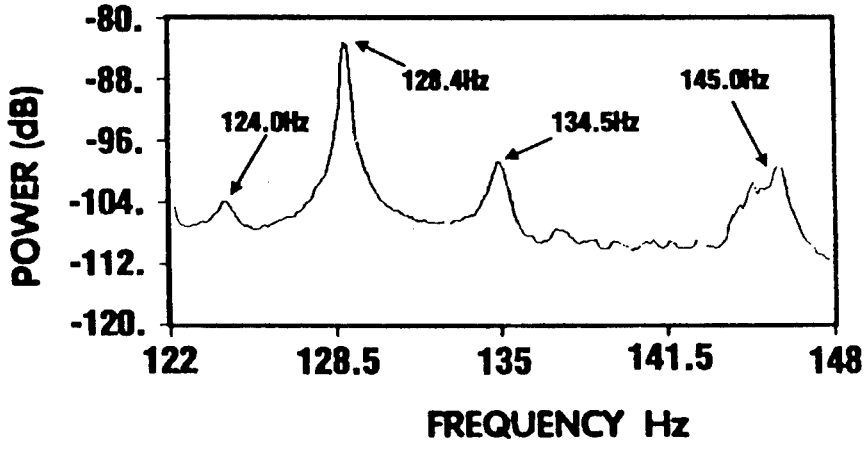


Figure 10 Power Spectrum in Region 122-128 Hz

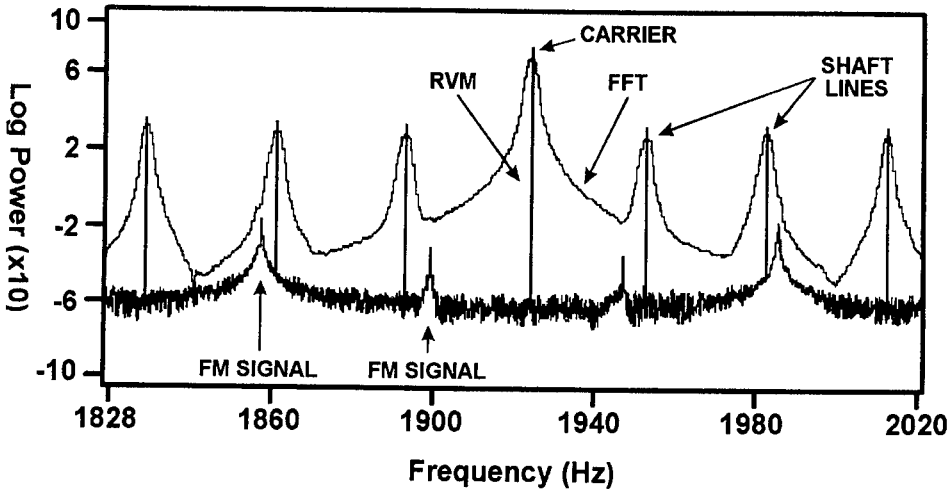


Figure 11

- 19.1 Hz
- 78.4 Hz (L-0)
- 82.6 Hz (L-1)
- 102.6 Hz (L-0)
- 105.9 Hz (L-1)
- 124.3 Hz
- 144.9 Hz

Chart 1 General Electric Turbine / Potomac #2 Torsional Resonances (calculated for certain modes)

SENSORS TECHNOLOGY

Cochairmen: Henry R. Hegner
ECO, Inc.

Christopher Nemarich
Naval Surface Warfare Center

Optical Debris Monitoring

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****

Laser Physics Branch, US Naval Research Laboratory,
Washington, DC 20375

Abstract: We describe an on-line real-time oil debris monitor capable of detecting ferrous, non ferrous and non metallic debris in lubricating oil, hydraulic fluid and fuel. The monitor is based on optical imaging and image processing and can be used to identify the severity and source of the fault through analysis of the morphology of the debris particles.

Key Words: Bearings; early warning; catastrophic failure; fuel; gears; hydraulic fluid; lifetime; neural nets; oil debris; on-line; optical; particulate contamination; real-time; shape classification; wear debris

Introduction: We describe an optical monitor for detecting and characterizing particles in lubricating oil, hydraulic fluid and fuel of engines, gearboxes and other rotating machinery. The optical technique is based on illumination of the fluid column, forming images in transmission of objects in the fluid and analyzing the images with a suitable image processor. The monitor is capable of on-line real-time measurements of the size, shape, and rate of production of particles. It can detect ferrous and non ferrous metallic particles as well as sand and debris from ceramic bearings. It will allow identification of the type of fault from the size and shape of the debris particles and the severity of faults from the size and rate of production of the debris particles.[1-3] It can provide a cumulative monitor of the health and condition of engines and gearboxes, as well as provide warnings for crew action as needed for avoiding catastrophic failure.

The optical imaging technology can be scaled to detect particles in various size ranges from a few micrometers to millimeters or more. The size range detected in a particular application will in general be determined by a combination of the required viewing area, the desired resolution and the practical size of photodetector arrays. In our work we have found it convenient to operate the optical debris

monitor for detection of particles in two size ranges: 1). larger than 50 micrometers and 2). smaller than 50 micrometers. Operation in the larger size range is suitable for on-line, real time detection of failure related debris in lubricating oil. Operation in the smaller size range allows detection of debris produced by normal or early wear in lubricating oil, as well as particulate matter in hydraulic lines and fuels. We will describe our work in both of these areas.

Detection of failure-related debris: A schematic of the optical oil debris monitor in a configuration suitable for real-time on-line detection of failure related debris in a helicopter or other aircraft is shown in Fig. 1. Here the oil column is illuminated by light from a single mode diode laser and is imaged in transmission onto a photodiode array. The resulting image is analyzed in an onboard image processor to identify any debris particles that are present and to identify their size, shape, edge irregularity and time of appearance. This data is stored in on board memory and allows size, shape and rate of production of debris particles to be determined. This information provides a cumulative record of the condition of the equipment, and can be used either to provide information for interrogation by ground crew or to provide recommended actions or warnings for the flight crew as the situation warrants.

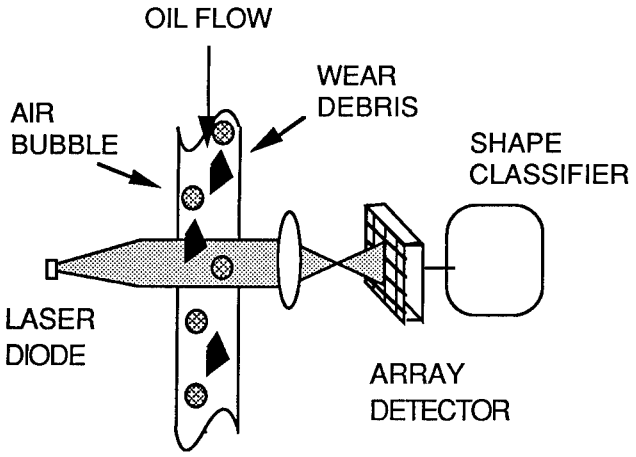


Fig. 1. A schematic diagram of the optical oil debris monitor, showing the illuminating laser diode, the array photodetector and the shape classifier. Objects suspended in the oil flow are imaged in transmission onto the array detector, appearing as dark shadows against a bright background.

The size range of particles detected in this configuration is driven primarily by the viewing area and the size of available photodetector arrays. Typical oil scavenge lines for most platforms are in the range from #12 to #20 (0.62 in to 1.25 in). Because the particles are relatively large, and therefore will be trapped in the filter, the oil debris monitor must detect each particle as it passes the viewing port. As a result it is necessary to illuminate all of the oil flow for possible particles. A photoarray with 1000 x 1000 elements will allow imaging through lines of this size with resolution of the order of 10 - 15 μm, allowing detection of particles of the order of 50 μm, and size and shape determination for particles about 100 μm and larger.

The location of the real time failure debris monitor in a lube flow system is illustrated in Fig. 2. Because the larger particles are trapped in the filter, it must be located between the scavenge pump and the scavenge filter.

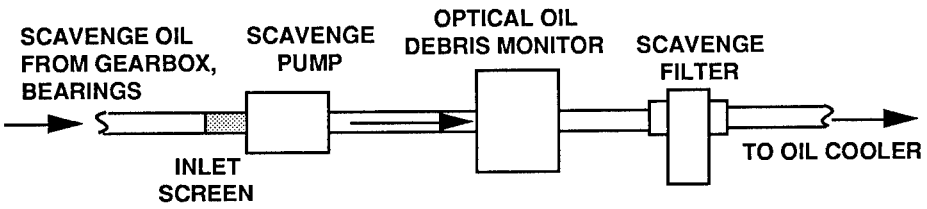


Fig. 2. Schematic diagram of a typical location of the optical oil debris monitor for failure related debris detection.

An example of the parameters suitable for an on-board system are shown in Table 1. The illumination wavelength is chosen to be compatible with efficient, uncooled photodetector arrays and laser diodes, and to give a relatively high optical transmission through used oil. Measurements of optical transmission on samples of engine and gearbox oil with about 790 hours of use have indicated that although the oil is black in the visible, it will transmit greater than 10% through a 3/4 inch path for wavelengths longer than about 800 nm. Optical transmission typically decreases for wavelengths longer than 1 μm, with transmission windows near 1.3 and 1.5 μm, and again becomes black for wavelengths longer than about 2 μm. The use of a laser diode with a wavelength of about 850 nm is a reasonable choice for both optical transmission and sensitivity of silicon photodetectors. The laser pulse duration is determined by the requirement to freeze the motion of the particles within a pixel to allow accurate shape identification. For typical applications lube oil transport rates of the order of 15-30 gpm in lines of the size considered above result in

linear flow speeds of the order of 10 m/sec. A pulse duration of 1 μsec therefore will freeze the motion to within 10 μm, comparable to the resolution provided by the 1000 x 1000 element array. The pulse repetition rate is determined by the need to illuminate all of the oil column. With a flow rate of 10 m/sec and a viewing area of 1 cm in the flow direction this requirement is satisfied by a repetition rate of 1 kHz. If a smaller array is used (e. g. 100 pixels in the flow direction) a higher rate of 10 kHz is required. Both of these rates are well within the capability of diode lasers. A peak laser power of 100 mW is estimated to be required to provide enough photons per pixel to be detected by semiconductor arrays, while the average power will be of the order of 100 μwatts to 1 mwatt depending on the repetition rate.

TABLE 1. SAMPLE SYSTEM PARAMETERS FOR AN OPTICAL OIL DEBRIS MONITOR

COMPONENT	VALUE	AFFECTED BY
LASER WAVELENGTH	850 - 950 nm	OIL TRANSMISSION DETECTOR SENSITIVITY
PULSE DURATION	1 μsec	SPATIAL RESOLUTION FLOW SPEED
REPETITION RATE	1 kHz	ILLUMINATION AREA FLOW SPEED
DETECTOR ARRAY SIZE	1000 x 1000 pixels	ILLUMINATION AREA PARTICLE SIZE RESOLUTION
OIL TRANSMISSION (2 cm path)	0.1 (790 hour gearbox sample)	OIL CONDITION
LASER POWER	100 mW peak 0.1 mW average	OIL TRANSMISSION DETECTOR SENSITIVITY
SIGNAL LEVEL	4×10^4 photons/pixel	LASER POWER ILLUMINATION AREA OIL TRANSMISSION

Once the image of the objects has been detected, it must be analyzed for inclusion of debris particles. The read and analysis time of a frame must be compatible with a rate of 1000 frames/sec. This requires a high speed photoarray, a wide i-o bus and a high speed processor.

Conventional ccd's and HDTV systems are too slow for the photodetector requirements. One of the methods we are looking at involves an active pixel matrix array developed by NASA's Jet propulsion Laboratory for high altitude imaging. This array reads an entire row of pixels in parallel and digitally encodes the resulting information.

Following the high speed photo array is a parallel digital processor. One of the major obstacles to successful operation of the optical oil debris monitor is the need to distinguish debris particles and sand from air bubbles. Air bubbles scatter light very effectively and have been a major difficulty in other optical monitoring systems. In our system air bubbles are distinguished from sand and debris particles by the image processor. In the size range being considered here, debris particles are generally irregular in shape, while air bubbles will be round or at worst slightly elliptical. The digital processor screens each frame to eliminate any air bubbles that may be present. The screening algorithm calculates the variance of the length of radial spokes of objects in the image to identify any that are round.

Images obtained with the optical oil debris monitor on flowing test stands have confirmed that it possible to distinguish metal debris particles from air bubbles on the basis of their shape. Sand (Arizona road dirt) is further distinguishable because it appears gray in the images.[2,3] The bubble screening algorithm has been demonstrated using sample pictures taken on flow stands and in the laboratory. When implemented on a serial processor it has been shown to screen out air bubbles and identify irregular debris particles successfully. An example of the operation of the screening algorithm is given in Fig. 3. In Fig. 3a only air bubbles are present and the screening algorithm has identified all objects as round. In Fig. 3b there is one metal particle present, and the screening algorithm has identified it on the basis of its non-round shape. When implemented on a parallel processor, this algorithm is estimated to be capable of processing 1500 frames per second.

Any particles that are identified as not round, as in Fig. 3b, are passed to a post processor for calculation of size and shape. The post processor can operate at lower speeds since the rate of production of failure related debris particles is expected to be low compared to the framing rate.

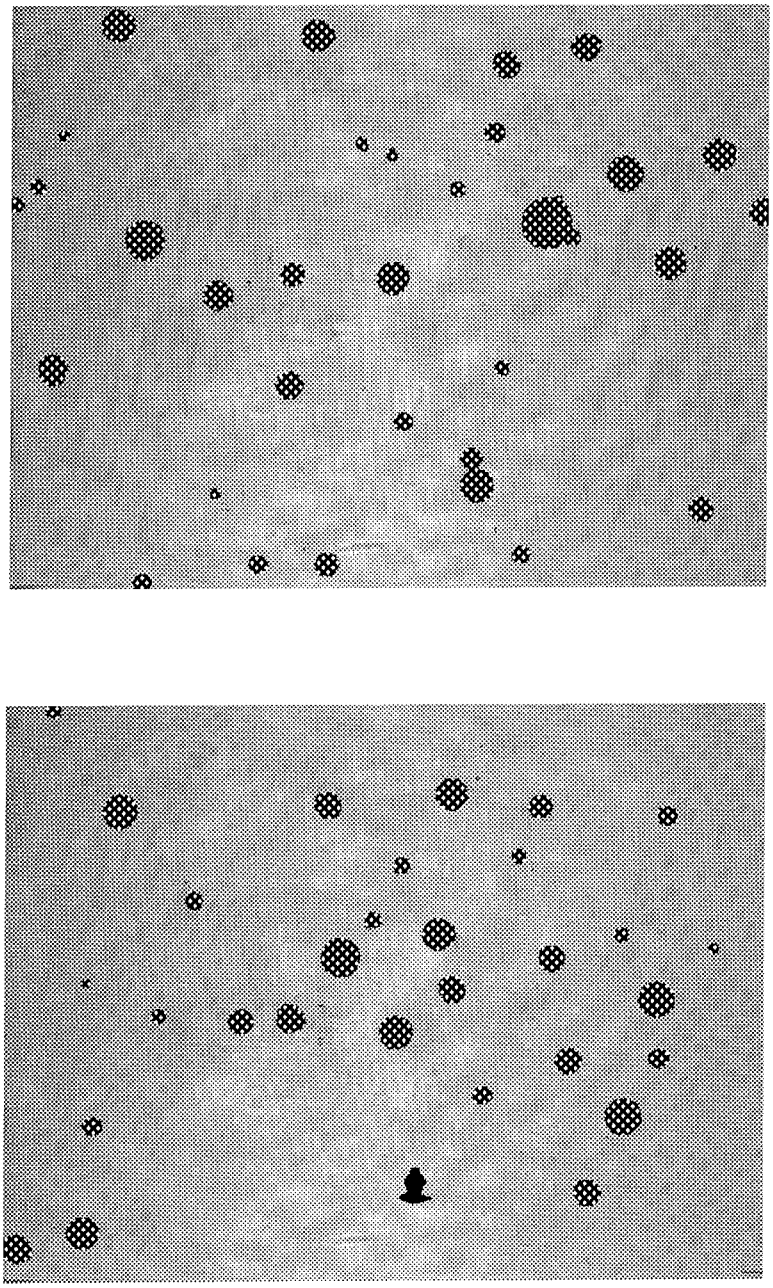


Fig. 3. a(top) Image containing air bubbles and no metal flakes. The image processor identified all particles as round. Particles smaller than 5 pixels across were ignored. b(bottom). Image containing air bubbles and one metal chip. The image processor correctly identified the one particle that was not round.

We are also examining an alternative approach in which the image is detected and tested for debris particles with a focal plane array. In this approach all the processing is done on the photodetector chip, and only information on debris particles is transferred to the post processor.

Detection of fines: In addition to detection of failure related debris there is a need for detection of normal and early wear debris in lubricating oil as well as particulate contaminants in hydraulic fluid and fuel. The optical imaging technique described above is scalable to detection of particles in the size range of 5-50 μm that is appropriate for these applications. The fine particle detector can be implemented either as an on-line continuous monitor, or a portable unit for periodic monitoring. Successful implementation of this type of detector can reduce or eliminate the need for sending samples to laboratories for analysis with its attendant long turn around time. In addition, the optical imaging detector will allow the morphology of particles to be used to identify the source of the wear, augmenting and enhancing the capabilities of other techniques such as SOAP.

In changing the size of particles that are detected to the new range, some changes in the implementation of the detector must be made. First, the magnification must be increased to the order 10x instead of 1x, giving a minimum resolution size of the order of 1 μm, and a viewing area of 1 mm instead of 1 cm. In addition the entire oil flow does not have to be examined since the particles that are under detection here are the same ones that are currently taken from static oil samples for laboratory analysis and remain suspended in the oil. As a result spatial and temporal sampling are acceptable for this application. In addition multiple color illumination using red or green laser diodes can be used to obtain information as to the identity of the individual particles.

We have made initial examinations of calibration samples of hydraulic fluid provided by Pratt&Whitney using the setup shown in Fig. 4. The sample cell is thinner than the one used for the failure debris monitor. This allows the potential use of illuminators of different wavelengths, which can aid in the identification of various particles. In our setup the overall magnification of the image was about 25x, providing a full viewing area of about 400 μm. Larger viewing areas can be obtained with smaller magnification.

Examples of images obtained with this technique are shown in Fig. 5. Here a distribution of particles of different size are evident, with the smallest particles ranging down to below 5 μm.

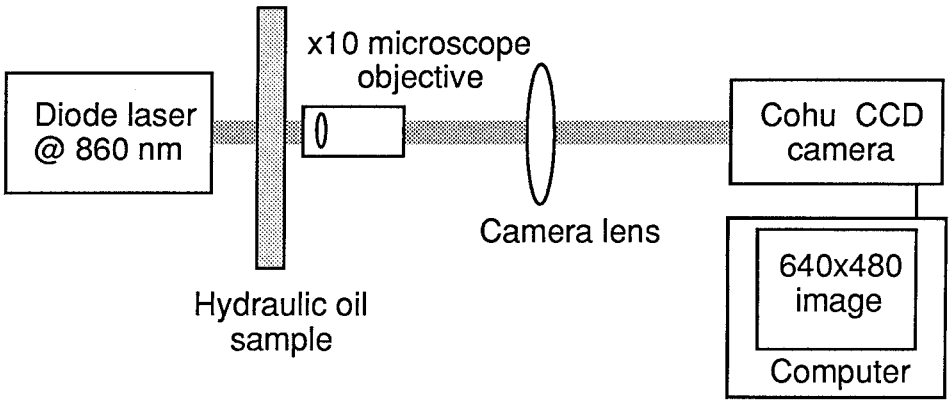
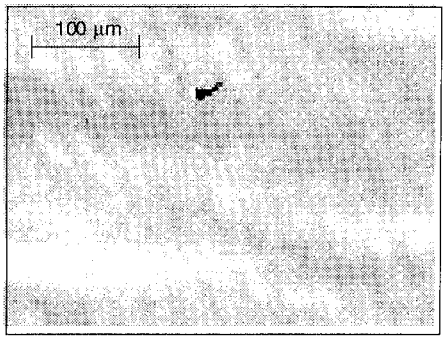


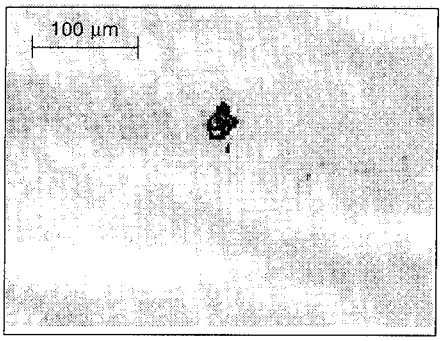
Fig. 4. Schematic diagram of setup used to image and analyze Fines in a hydraulic fluid sample.

We have started development of an image processor that will give size and shape of the particles. Although the processing speed requirements are not as high as for the failure debris monitor, the need for making a number of measurements to obtain a statistically significant sample produces a requirement for reasonable speed in the calculation. In addition, since some wear debris particles from processes such as fatigue cracks can be round, it is not possible to reject round objects in this application. We have shown previously, however, that air bubbles can be distinguished from debris particles on the basis of characteristic highlights under transverse illumination.[2,3]

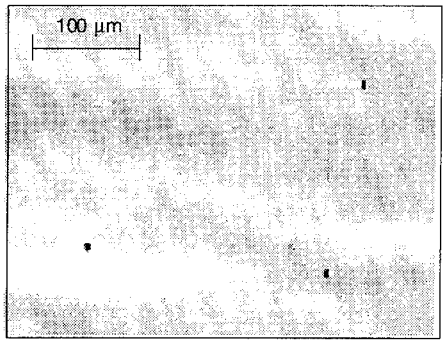
An example of the results obtained from the samples studied so far is given in Table 2. For the measurements presented here no attempt was made to obtain statistically significant measures of particle size distribution. Rather the sample was scanned manually for interesting distributions of particles. In its present form, the algorithm can process a single 640 x 480 pixel frame containing 40 particles in 1 sec. If the algorithm is implemented in a parallel processor to allow it to operate at video frame rates, we estimate that a sampling of 70,000 particles could be analyzed in 1 minute, giving a statistically significant reading. Extension to measurements of the order of 200,000 particles could be done in 3 minutes, or in 1 minute with an increase in the photodetector speed to 100 frames/sec, with pixel transfer rates comparable to those in HDTV.



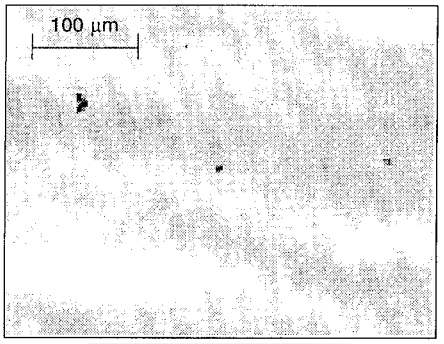
Size (µm)	Number
3 - 5	1
5 - 15	3
15 - 25	0
25 - 50	0
50 - 100	0
> 100	0



Size (µm)	Number
3 - 5	1
5 - 15	1
15 - 25	0
25 - 50	1
50 - 100	0
> 100	0



Size (µm)	Number
3 - 5	3
5 - 15	2
15 - 25	0
25 - 50	0
50 - 100	0
> 100	0



Size (µm)	Number
3 - 5	2
5 - 15	1
15 - 25	0
25 - 50	0
50 - 100	0
> 100	0

Fig. 5. Examples of images of Fine particles in a hydraulic fluid sample obtained with the setup of Fig. 4. The tables show the results of the image analysis for the sizes of particles in each picture.

TABLE 2. COMPOSITE FINES ANALYSIS FOR 9 IMAGES OF HYDRAULIC FLUID SAMPLE

SIZE RANGE (μm)	NUMBER*
3-5	28
5-15	19
15-25	0
25-50	2
50-100	1
>100	0

* No attempt was made to obtain statistically significant sampling for this data

Summary: We have demonstrated an optical oil debris monitor that is capable of real time on line detection of wear debris in lubricating oil, as well as particulate matter in hydraulic fluid and fuel. The monitor can detect ferrous and non ferrous metals, ingested sand and debris from ceramic and hybrid bearings. It is based on optical imaging and image analysis, allowing it to distinguish debris particles and sand from air bubbles on the basis of the shape of the various objects. The ability to identify the shape of individual debris particles provides information on the source and severity of the fault.

References:

1. 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

2. J. Reintjes, R. Mahon, M. D. Duncan, L. L. Tankersley, A. Schultz, V. C. Chen, D. J. Kover and P. Howard, "Optical Oil Debris Monitor," in "Condition Monitoring '94", M. Jones, ed., pp 335-343 , 1994

3. J. Reintjes, R. Mahon, M. D. Duncan, L. L. Tankersley, A. Schultz, V. C. Chen, D. J. Kover and P. Howard, "Optical Oil Debris Monitor", in "Advanced Materials and Process Technology for Mechanical Failure Prevention", H. C. Pusey and S. Pusey, eds. pp. 57-66 , 1994

* Jaycor, Vienna, VA, USA
† Department of Physics, US Naval Academy, Annapolis, MD USA
** Radar Division, Naval Research Laboratory, Washington, DC USA
†† Naval Surface Warfare Center, Annapolis Detachment, Annapolis, MD USA
*** P. L. Howard Enterprises, Westchester, PA, USA
††† JASON Associates, ONR Code 342, Ballston Center Tower 1, Arlington, VA.
****LNK Corp., 6811 Kenilworth Ave., Suite 306, Riverdale, MD 20737

HIGH SPEED IMAGE PROCESSING FOR WEAR DEBRIS MONITORING

Naresh Gupta, Sridhar Srinivasan, and Srini Raghavan
LNK Corporation Inc.
Riverdale, MD.
Tel: (301) 927 3223

John Reintjes, Victor Chen, and Abe Shultz
Naval Research Laboratory
Washington D.C.
Tel: (202) 767 2175

Abstract: This paper presents a real-time image processing algorithm for differentiating debris particles from air bubbles in laser generated images. The laser imaging system developed by the NRL scientists delivers 1000 binary images (1024 x 1024 pixels) per second. The algorithm differentiates between circular air bubbles and irregularly shaped debris particles. The high speed computational requirements and the voluminous nature of this data pose significant challenges to solving the discrimination problem in real-time. The problem is further complicated by the presence of overlapping air bubbles. The primary difficulty with shape discrimination algorithms is the inherently serial nature of the connected component extraction. The algorithm described here decomposes the connected component extraction algorithm into two phases; a computationally expensive but inherently parallel stage followed by a few serial operations with minimal computational requirements. A noise-robust and extremely efficient discrimination function is used to differentiate between debris particles and air bubbles to achieve the desired detection rate without false alarms. In preliminary testing, serial implementation of the algorithm takes approximately 30 msec on a Sparc-10 workstation. For the full system the processing time will be reduced significantly by implementing the parallel segments of the algorithm on CNAPS-512 parallel hardware. The combined configuration of CNAPS-512 processor and Sparc-10 system gives a throughput of 100 frames a second. A faster serial chip with multiple processors (e.g., DEC alpha with 3 cpu's) combined with the CNAPS-512 processor parallel hardware will achieve the desired throughput of 1 Giga pixels per second, with 1000 frames per second and 1024 x 1024 pixels per frame.

Key Words: Connected components; image processing; parallel computation; real time on-line oil debris monitoring; wear debris monitoring.

Introduction: Monitoring debris in engine and gearbox oil is one of the principal mechanical diagnostic systems for on-line health monitoring. On-line health monitoring of machinery allows condition-based maintenance to be implemented with early fault detection, with resultant savings in parts and maintenance operations.

The NRL Lasernet automatic oil debris monitor identifies the shape and size of debris particles suspended in oil in real time using advanced image analysis techniques. the knowledge of particle morphology allows identification of the source of failure, differentiation of failure related debris from benign debris and detection of any particulate matter, ferrous, metallic and non-metallic, On-line monitoring provides a cumulative

record of the engine health as well as advance warning of deteriorating and potentially catastrophic conditions.

This paper describes an algorithm for automated detection of debris particles in the lubricant flow of engines, gearboxes and other rotating machinery. The algorithm operates on optical images generated as the fluid flows across an illuminated area. A sequence of frames is analyzed by the algorithm, which identifies debris particles and distinguishes them from air bubbles. In the following discussion, the image acquisition process, the steps in the algorithm for automated debris detection, and the proposed hardware scheme for achieving real-time operation are discussed. A summary of preliminary results of our analysis of sample data will is also given.

The Lasernet Image Acquisition Process: The fluid is illuminated by a laser diode that is pulsed with a duration designed to freeze the motion of the oil stream. Images of opaque items in the cross section of the oil flow are acquired by a 1k x 1k element photodetector array as shown in Figure 1. The scan resolution is 10 micrometers per pixel. One thousand such images are acquired every second. The data rate is of the order of 10^9 bit/sec. The design considerations that determine these numbers are the fluid velocity, the optical transmission characteristics and the achievable camera resolution.

Air bubbles appear as round objects in the image. Debris particles measuring 10 or more pixels across are identified by their irregular shape. A single frame may contain many bubbles, while a debris particle may occur once every 20-30 minutes (i. e. every 1-2 million frames). Figure 2 is an illustrative example of a frame having bubbles, overlapped bubbles, bubbles on the periphery of the image, debris particles and debris on the periphery. The algorithm detects debris particles in such images automatically in near-real time. The algorithm needs to have a very small failure rate since detection of every debris particle is vital. Additionally, the false alarm rate must be minimal.

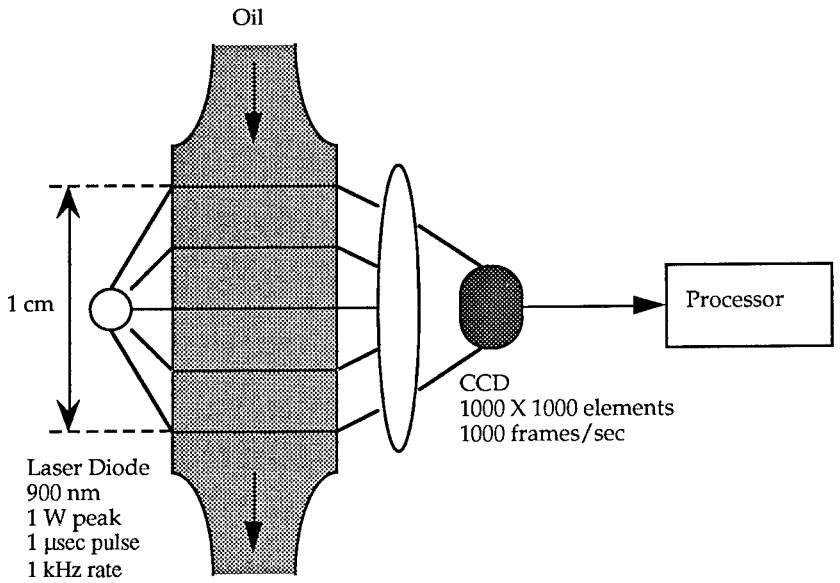


Figure 1. Lasernet functional schematic

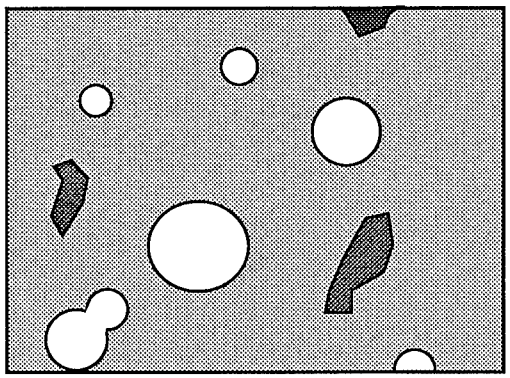


Figure 2. Illustration of an image frame

Description of the algorithm: The algorithm for real-time debris detection consists of the following five steps:

1. Data form the photo array is scanned horizontally to detect object edge points.

2. Connected regions are listed in terms of their edge points.

3. These regions are tested for circularity.

4. Regions failing the circularity test are analyzed for overlapped circular structure.

5. Regions failing tests described in Steps 3 and 4 are then identified as debris particles.

The sub-images of the suspect debris particles are then sent to the post processor for further analysis. We will now look at each step in detail.

Step 1:

The first step of the algorithm is an edge detection process that marks the transition points along each scan row as a boundary point. Although computation of a single transition point is a trivial operation, repeating this operation for all pixels in the image accounts for a large part of the total computation.

Significant speedup can be achieved by implementing this step of the algorithm in parallel. Edge detection is inherently parallel as it involves performing the same primitive operation of comparing adjacent pixels of the image. The parallelism can be exploited by a simple SIMD (single instruction multiple data) machine. A SIMD machine as its name suggests, performs the same operation on all its processors.

In preliminary analyses, we used the CNAPS^{TM} parallel computer with a configuration of 512 processors in a linear array. Two rows of the image were stored on each processor, with the edge detection step executed in parallel in all the processors.

Along with marking the transition points, a concise array of runs of 1's is generated from the data. The position and length of runs sufficiently characterize the entire image. This step reduces the data size by over an order of magnitude.

Step 2:

Once the data has been reduced from a raster scan description to a set of run lengths, a *partial connected component* analysis is performed to group runs that are connected in the vertical direction. This differs from conventional connected-component analysis where each pixel is labeled with the component identifier. A data structure is formed for each component or region, populated by the set of all connected runs. For the class of data encountered in the current application this algorithm performs two orders of magnitude faster than conventional connected components.

In addition, the region data structure contains information on whether or not the region lies at the periphery of the image. Steps 1 and 2 together implement a partial connected component identification which ensures that all convex objects are connected together. The succeeding steps operate on this structure.

Step 3:

This eliminates from consideration all regions that are round, including those on the periphery of the image that are segments of circles.

A circularity test is performed by fitting a complete circle to the set of transition points. The center of the candidate circle is estimated to be the centroid of the region bounded by the set of points. This estimate is robust to minor perturbations in the boundary. The circle fit is accomplished by exploiting the constraint of minimal deviation of the boundary points from a well-formed circle with the centroid of the region as the center.

Step 4:

The regions failing Step 3 could potentially correspond to overlapped bubbles or to debris particles. To address this, the region is modeled as two overlapping circles. On the basis of quality of fit, the region is identified as overlapping bubbles or a debris particle. A closed form solution to this problem has been identified.

Hardware Design: The algorithm has been implemented on off-the-shelf hardware in order to validate its performance. For operation at the speed necessary for near-real-time oil debris detection some customization is still required. In this section, we describe a hardware design for achieving the required speed. The system is based on the CNAPS^{TM}

1016 parallel computer chip followed by DEC ALPHA 21164 processors for the sequential stage.

The photo detector array downloads the captured image as 64 bit words at 16 MHz, delivering data at the rate of 128 MByte/s or 10^9 b/s (one billion bits per second). Each row of the image is composed of 16 such words, and rows are sent sequentially. Each image spans 1024 x 16 = 16 k words. The hardware buffers each image before it is processed further, as shown in Figure 3.

The data from the array is read into the CNAPS^{TM} processors via a dual high-speed buffer. The CNAPS^{TM} is a true SIMD processor. Multiple processors (16, 32, 64) running between 16 and 25 MHz are connected as a linear array on a single chip. Several such chips can in turn be connected together to provide a massively parallel structure.

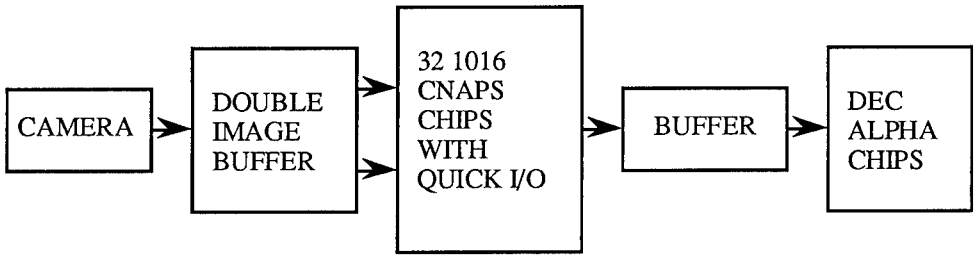


Figure 3. A top level architecture of the proposed hardware

For I/O operations, each CNAPS^{TM} chip allows access from the external world separately, a feature known as Quick I/O. The required I/O and processing rates of 1000 frames/s are achieved by 32 CNAPS^{TM} 1016 16-processors that are equipped with Quick I/O which allows the processor memory to read and write directly from and to the input and output buffers (Fig. 4). This configuration of CNAPS chips offers an I/O rate of 512 MByte/s or 8 billion multiply-accumulate operations per second. The CNAPS^{TM} board running at 16 MHz performs Step 1, which is the most computationally intensive step in our algorithm, in parallel.

Once the image is stored, all 512 processors execute the transition detection algorithm in parallel on their local data. The input buffer is a composite of two image buffers to facilitate simultaneous reading and writing. When one buffer is being accessed by the camera for input data, the other is simultaneously by the processing board.

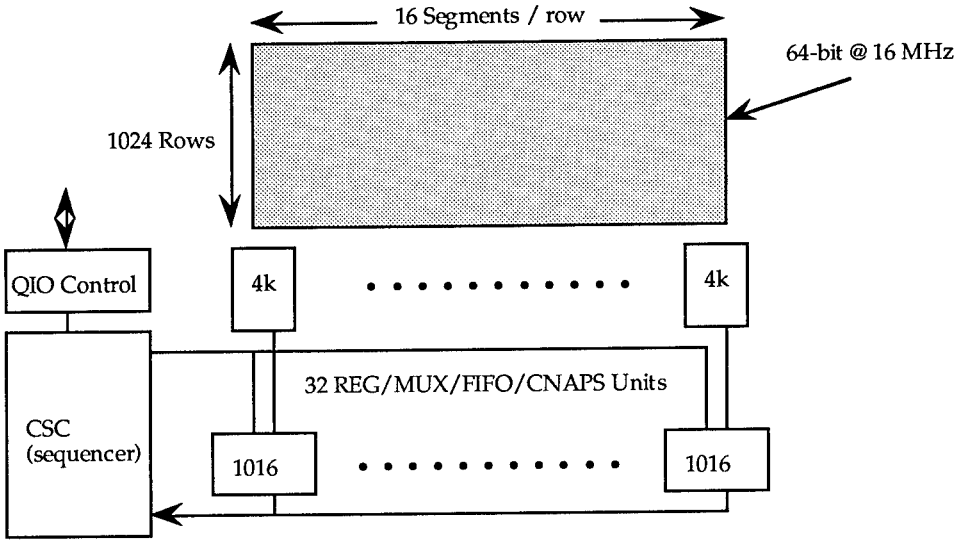


Figure 4. Configuration of the CNAPS^{TM} section of the hardware

On completion of processing, the CNAPS^{TM} board stores the result on an output buffer which is accessible for further processing by the DEC ALPHA (21164) chips. The output operation is anticipated to require around 1 k clock cycles.

Steps 2, 3 and 4 are executed on the sequential DEC ALPHA (21164) processor running at 300 MHz. The unique design of the 21164 chip allows a total throughput of 1.2 billion instructions per second. The output of Step 1, which amounts to around 10 kBytes/image, is stored into objects which are then discriminated by the algorithm in steps 2, 3, and 4. These steps are anticipated to require 2 ms/image on one DEC ALPHA chip. Three such chips running in synchronous mode provide sufficient computational power to process the full rate of 1000 images/second.

Results: The algorithm has been implemented and tested on a hardware setup consisting of a CNAPS^{TM} 512-processor parallel computer followed by a Sun SPARC 10 workstation. A test data set of several hundred real images of the cross section of oil flow was provided by NRL. In addition to this set, we extracted around 150 bubbles including 20 cases of overlapped bubbles and 40 debris particles from real images. Using these, we generated a test set of over 1000 frames. The algorithm distinguished air bubbles from debris on all frames of the test data. Some of the results are shown in Figures 5, 6, 7, and 8. Debris particles are marked black and bubbles white.

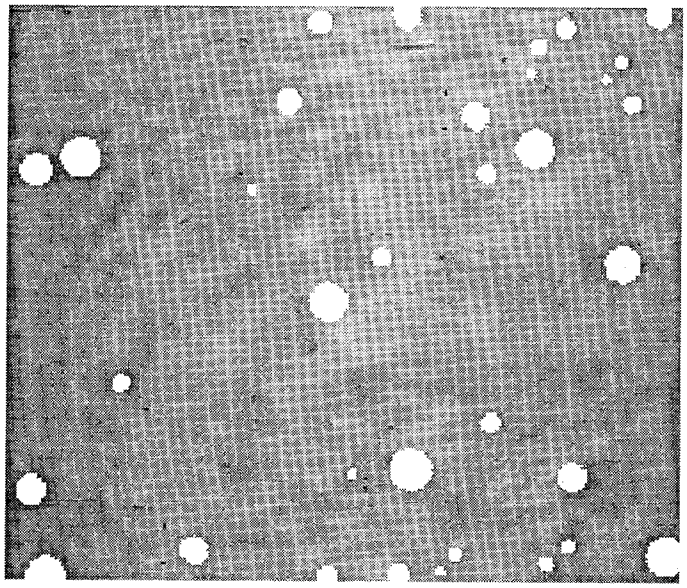


Figure 5. Results of the Debris Detection algorithm. This frame contains neither debris nor overlapping bubbles

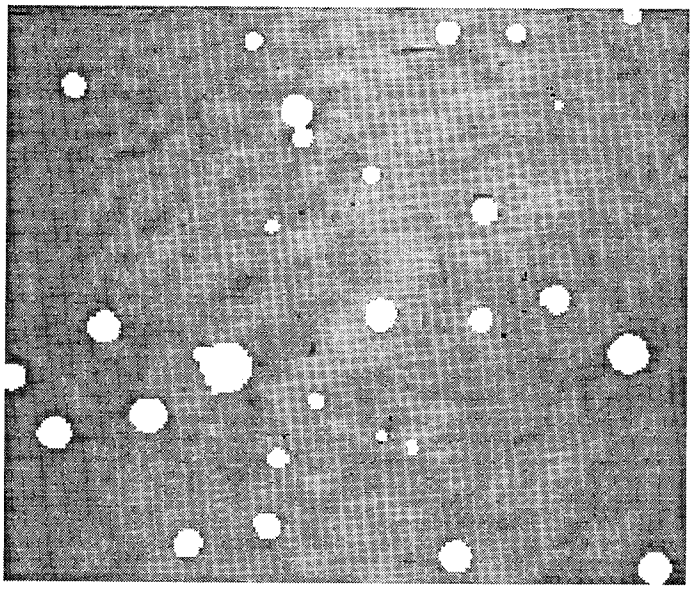


Figure 6. Results of the Debris Detection algorithm. The frame does not contain any debris particles, but contains overlapping bubbles.

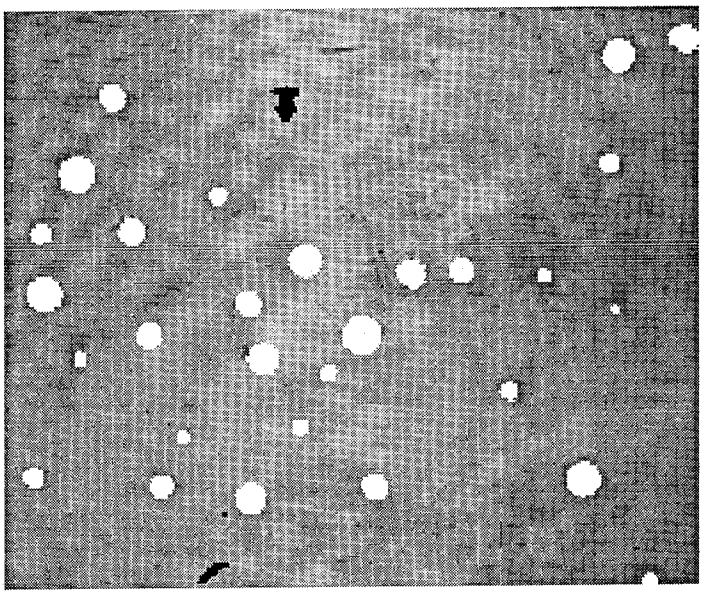


Figure 7. Results for the Debris Detection algorithm. The frame contains debris particles, but no overlapping bubbles.

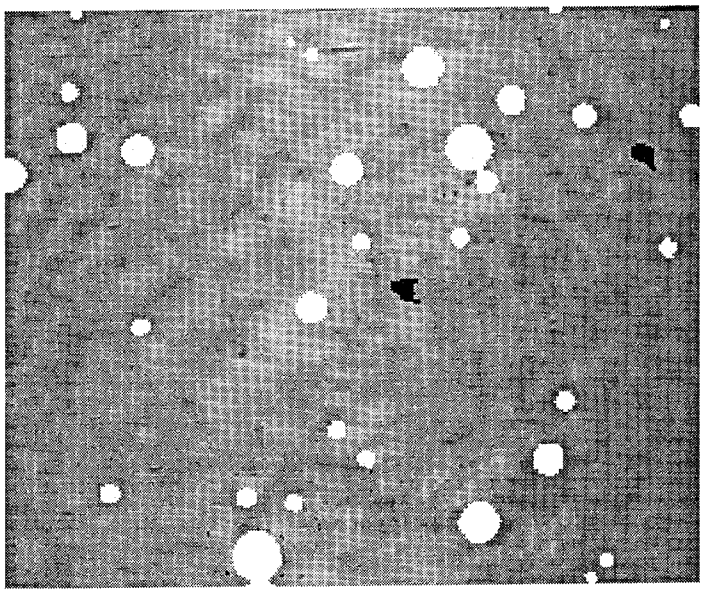


Figure 8. Results of the Debris Detection algorithm. The frame contains debris particles and overlapping bubbles.

TEMPERATURE SENSORS FOR DETECTING FAILURE OF TRIBOLOGICAL COMPONENTS

F.E. Kennedy, A.K. Henning, D. Frusescu, L.M. Caballero, X. Tian and T.M. Cook

Thayer School of Engineering
Dartmouth College
Hanover, NH 03755

ABSTRACT

Frictional heating and the resulting increase in surface temperatures can have an important influence on the tribological behavior and failure of sliding or rolling/sliding components, whether they be lubricated or unlubricated. To avoid failure of those components, surface temperatures must be kept below critical values. Similarly, incipient failure of the components can be detected if the surface temperatures can be monitored during service. This monitoring task requires sensors which can respond very quickly and reliably to changes in temperatures, but which do not adversely change the operating characteristics of the components. This paper describes the development of single thin film thermocouples (TFTC) and arrays of TFTC devices, and the use of these devices in monitoring actual surface temperatures during operation of sliding mechanical components. TFTC sensors may be used to warn of impending surface failure of sliding components, such as thermoplastic bearings, and would aid in avoidance of component failure.

Key Words: Frictional heating; sliding bearings; surface temperatures; thin film thermocouples; wear

INTRODUCTION and BACKGROUND:

Sliding friction results in a loss of mechanical energy, and it has long been known that the vast majority of frictional energy is transformed into heat. This frictional energy dissipation takes place in the immediate vicinity of the real area of contact, where frictional interactions occur. The transformation of frictional energy to heat, called frictional heating, is responsible for increases in the temperatures in the contact region of the sliding bodies. Frictional heating and the resulting increase in surface temperatures can have an important influence on the tribological behavior and failure of sliding or rolling/sliding components. It has been known for many years that scoring (or scuffing) of gears is related to the maximum surface temperature rise (or flash temperature) in the contact zone, but work goes on to elucidate that relationship [1,2]. Mechanical face seals often fail by the phenomenon known as thermocracking or heat checking; this phenomenon has been shown to be directly related to frictional heating [3]. For dry sliding bearings, one of the primary design parameters, the PV (= pressure x velocity) factor, is really an indication of the amount of frictional heating, and therefore the surface temperature [4]. Many dry bearings are made of polymers, and it is well known that polymer wear is very dependent on surface temperature [5]. Much wear of sliding metallic or ceramic surfaces has been found to be thermomechanical in nature, with surface temperatures playing a critical role in its

occurrence [6]. In fact, the recently developed wear-mechanism maps which enable the mode of wear of metallic surfaces to be determined are in actuality based on maps of surface temperature [7].

From these examples, it should be clear that surface temperatures play a major role in surface failure of sliding or sliding/rolling mechanical components, whether they be made of metals, ceramics or polymers. Because of the importance of surface temperatures, attempts have been made for many years to calculate those temperatures and use them in the design of critical components. Analytical models have been developed for calculating surface temperatures in sliding contacts [8-10], and a recently-published handbook contains a chapter devoted to surface temperature prediction [11]. All of these methods, however, require knowledge of a critical parameter which is seldom, if ever, known in actual sliding contacts - the actual area of contact over which frictional heat is being generated. Therefore, the accurate prediction of actual surface temperatures in sliding or rolling/sliding mechanical components is seldom possible. As an alternative, experimental measurements of surface temperature can be made.

Many different surface temperature measurement techniques have been used with some degree of success, and those methods are described in a recent handbook article [12]. Although each of the techniques is a useful engineering or research tool, each is also subject to some limitations, so there is no single surface temperature method which can be used for all applications. Some techniques, such as thermocouples, give a measure of temperature at a single point, while others, such as infrared imaging, give a field measurement. Some, such as thermocouples or thermistors, can be mounted on actual machine components, while others, particularly optical techniques, require a special counterface material (eg., a transparent sapphire). Response time varies considerably, from <50 ns for photon detection, to 1 μs or less for thin film thermocouples, to several seconds for infrared photography, and nearly a minute for metallographic techniques. The work described in this paper has aimed to develop fast response temperature sensors for use in measuring surface temperatures on machine components.

METHODS:

Temperature Sensor Selection:

Because of the peculiar environment encountered in sliding situations, the sensors developed for contact surface temperature measurement in this work had to satisfy the following requirements:

(1) The highest surface temperatures occur only at localized real contact areas and have very short duration. The sensor must enable measurement of the temperatures within the actual contact area. It must have small measuring junction dimensions and small thermal mass in order to achieve rapid thermal response and local area thermal sensitivities.

(2) The sensor must cause minimal disruption to the normal operation of the sliding mechanical component. It should be able to be installed in a wide variety of actual components and should pose no significant restriction on the choice of materials for the components.

(3) The sensor must have good wear resistance and reasonable durability during operation which includes friction and wear at the sliding interface.

To meet these requirements, the sensors developed in this work were thin film thermocouples deposited directly on the sliding surfaces of actual mechanical components. The thermocouples were designed to have miniature measuring junctions, sandwiched between wear-resistant, electrically insulating protective layers. A cross-section of a typical thin film thermocouple (TFTC) is shown in Figure 1.

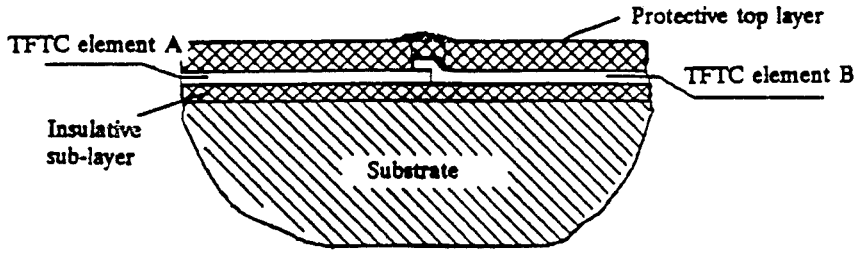


Figure 1. Schematic cross-section of thin film thermocouple.

Thin Film Thermocouple Fabrication:
The thermocouple devices used in this study were made from thin films of copper and nickel, vapor deposited on glass or metallic substrates. The TFTC measuring junctions have been from 0.5 to 2 μm thick, and are sandwiched between thin films of a hard, non-conducting ceramic (Al_2O_3 in our work) to insulate the thermocouple electrically from the substrate and protect it during sliding. Tests have shown that the TFTC devices have extremely rapid (< 1 μs) response to a sudden temperature change and do not significantly disturb the heat flow from the sliding contact [13].

Previously, our TFTC devices were deposited using a shadow mask technique, which enabled the production of devices with measuring junctions as small as $(80\ \mu m)^2$ [13]. In this work, we used microelectronic fabrication techniques to achieve junctions as small as $(10\ \mu m)^2$. Substrate surfaces for TFTC deposition were first polished smooth, then washed in water detergent, followed by ultrasonic degreasing in acetone and subsequent wiping with an methanol-soaked low particulate fabric. The substrates were precoated with Al_2O_3 evaporated from an electron gun source through a reactive oxygen vacuum chamber backfill of 1×10^{-4} Torr. This layer served to improve metal adhesion to nonconductive substrates (0.15 μm thick Al_2O_3), or to isolate the TFTC layers from metallic substrates electrically (> 1000 MΩ for 5 μm thick Al_2O_3). After the surface was cleaned, nickel from a 99.95+% pure source was vapor deposited on the surface to a thickness of about 0.5 μm. The substrate was then removed from vacuum, and photoresist was spin coated on the surface. The photoresist was then exposed lithographically and developed, to define the desired nickel TFTC contacts and wires (the darker portion of Figure 2). After exposure, the nickel was etched in a solution of H_3PO_4:HNO_3:HAc:DI water, until the nickel lines were defined visually. The remaining photoresist was then stripped, and the surface was cleaned using solvents. Copper was then vapor deposited using a 99.99% pure source in an alumina-coated, tungsten resistor boat. The copper was of the same thicknesses as the nickel, up to 1 μm. A second layer of photoresist was applied, exposed and developed, thus defining the copper TFTC contacts and wires (the lighter portion of Figure 2). After etching to remove the unwanted copper, the photoresist was stripped and the surface was cleaned. A protective layer of 2 μm Al_2O_3 was deposited over the completed TFTC in a manner similar to the precoat, except that a shadow mask was used to prevent Al_2O_3 coverage of the bond pads. Extension wires were connected to the finished device using rosin core tin solder.

Since a thermocouple gives a measure of temperature only at one point, it is often desirable to have many thermocouples on a sliding surface so that a distribution of surface temperatures can be obtained. We have produced arrays of up to four individual TFTC devices, each with its own output wiring, and have used those arrays successfully to determine how surface temperature distributions vary with time at a sliding interface. The techniques used to produce the thermocouple arrays were similar to those described above for a single thermocouple. A diagram of a TFTC array with three measuring junctions is shown in Figure 2. The junctions in that particular array were each (80 μm)^2 and were located in an area approximately (0.5 mm)^2.

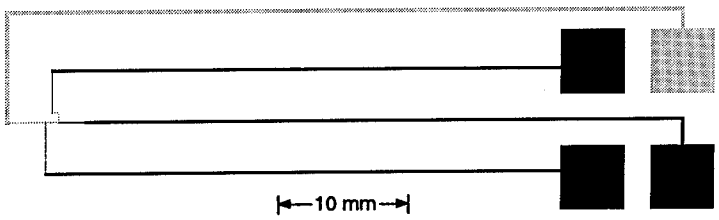


Figure 2. Top view of array of three thin film thermocouples.
Thermocouple junctions (40μm)^2 are at left. Bonding pads (5 mm)^2 are at right.
The lighter line on top is the common copper line.
The three darker lines are the nickel lines from each of the three thermocouples.

After testing the multiple TFTC arrays, we found that array design and data acquisition problems increase significantly as the number of individual thermocouples increases. Each thermocouple junction requires an additional bonding pad and an additional line running from the junction to the bonding pad. These lines cannot contact any other lines except at the active junction, posing severe geometrical constraints on the design of the array. In addition, each of the thermocouples must be attached to a separate channel of the data acquisition system, thus slowing down the data acquisition rate for each channel. The combination of these problems essentially prevents having TFTC arrays with more than four or five active junctions. To overcome these problems, a novel device has been developed to measure spatial surface temperature profiles. This device, a thin film thermocouple MOSFET array, enables measurement of a surface temperature distribution by switching through the measuring junctions of a TFTC array [14]. A cross-section of such a switchable thermocouple is shown in Figure 3. The MOSFET in the semi-conducting substrate acts as an electrically controlled switch connecting the two metals of a thermocouple pair. Because the switching device is fabricated using microelectronic techniques, the size of the junction can be kept very small, validating the assumption of constant temperature across the junction. Arrays have been designed and fabricated using a conventional bit-line/word-line scheme, allowing for increased packing density and ease of data acquisition. We have successfully fabricated several different arrays of sixteen Ni-Cu thermocouples, each with a junction size 40 μm square and less than 1 μm thick [15, 16]. A diagram of one such array is shown in Figure 4. These are by no means the limits of the

size of TFTC thermocouple arrays; we foresee arrays of 100 or more thermocouples, with sizes of individual thermocouple junctions being less than 10 μm square. The arrays were produced on silicon wafers using a microelectronic fabrication process described in detail elsewhere [15]. The final stages of the process, i.e., the fabrication of the nickel and copper lines, are essentially the same as described above.

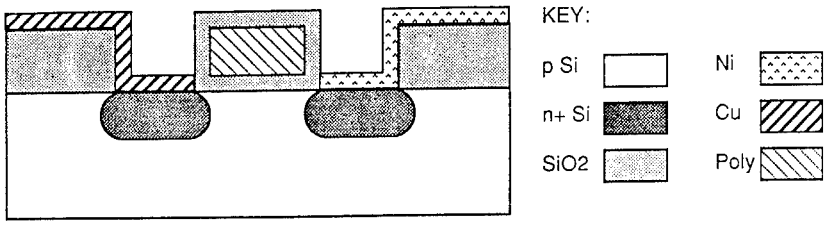


Figure 3. Cross-section of thermocouple MOSFET.

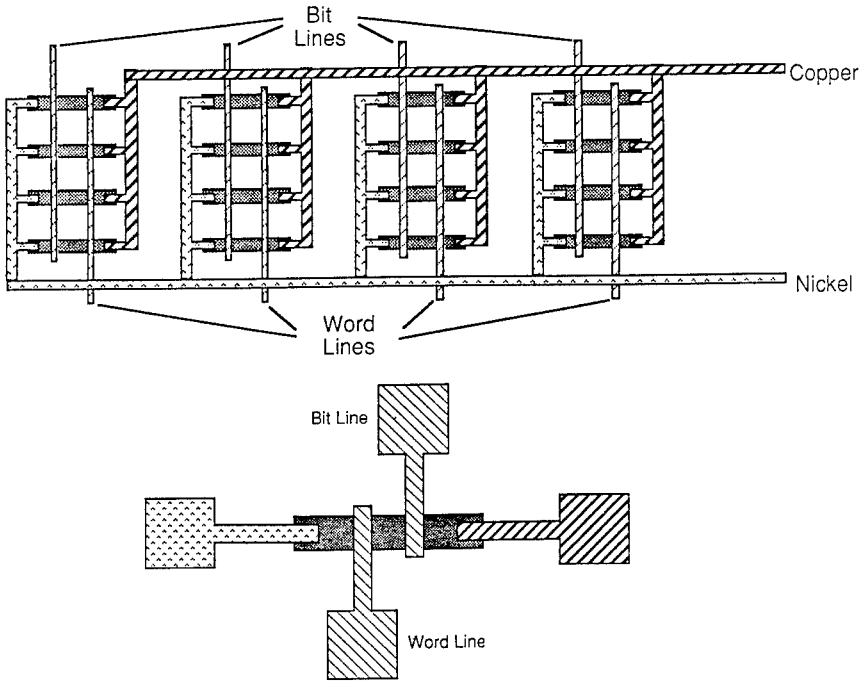


Figure 4. Layout of thin film MOSFET thermocouple array.

Testing:

Many tests were performed to confirm the efficacy of the TFTC devices as temperature sensors for sliding components. The particular application studied was dry sliding bearings made from a thermoplastic polymer. Such bearings are widely used in situations where lubrication during operation is difficult or impossible, and failure of the bearings is a common occurrence. In the tests, a polymer specimen was rubbed against a flat counterface containing TFTC devices on its contact surface. Two thermoplastic polymers were tested, Ultrahigh Molecular Weight Polyethylene (UHMWPE) and Polymethylmethacrylate (PMMA). Pins of dimensions 2mm x 4mm (4 mm in the sliding

direction) were machined from the polymer materials and were held in a stationary pin holder on a test apparatus which had been used in earlier studies of polymer wear [17]. A weight was applied to the top of the specimen holder to provide the desired normal load. The flat specimens were made from either glass or stainless steel, and Ni-Cu thermocouples were deposited on their contact surface as described above. The specimen holders for both pin and flat specimens were fitted with thermoelectric heaters which enabled their background (or bulk) temperature to be controlled at temperatures ranging from 15°C to 50°C. Tests were conducted at a wide range of normal loads, sliding speeds and background temperatures. All tests were run in air. During a test, a chosen normal force was applied to the top of the polymer pin, background temperatures were set to their desired values, and the flat specimen was set in motion at the chosen oscillatory frequency. The friction force was measured by a piezolectric force transducer and the linear wear of the pin specimen was monitored using a displacement transducer (LVDT). Contact temperature, friction force, and linear wear were all monitored continuously with the aid of a computer-based data acquisition system.

RESULTS AND DISCUSSION:

A typical plot of measured surface temperature rise (above background temperature) as a function of time during a test is shown in Figure 5. In this test a PMMA pin was in contact with an alumina-coated glass flat upon which a three-element thermocouple array (similar to that shown in Figure 2) had been deposited. Figure 5 shows the variation of each of the three thermocouples in the array during a 5 second period of the test. As can be seen, all of the thermocouples show two temperature peaks during each oscillation cycle, one as the pin slides over the thermocouple junction in one direction and the other as the pin passes the junction during the return trip. These temperature peaks (called flash temperature rises) are superimposed upon a nominal temperature rise which is due to the repeated frictional heating of the entire surface. In this case, the nominal temperature rise (above the 27°C background temperature) was between 20°C and 24°C, while the flash temperature rise added an additional 32-53°C, depending on the thermocouple. Thus, the peak surface temperature ranged from about 78-87°C for thermocouple #3 to over 100°C for thermocouple #1. The difference in flash temperatures between the three thermocouples is presumably due to differences in contact pressure at the three points, since contact temperature rise is proportional to the rate of frictional heat generation [11]. The frictional heat generation at a point is determined by μPV, where μ is the friction coefficient, P is the contact pressure, and V is the sliding velocity. V is the same for all three points, and differences in friction coefficient across the surface are probably small. The relative magnitude of contact temperature for the different thermocouple locations changed over time because of variations in contact pressure resulting from wear of the polymer pin. Whereas thermocouple #1 gave the highest temperatures during this particular time period, its temperature could be lower than that of the other thermocouples at other times during the sliding test. It is difficult to tell beforehand where the highest contact pressure, and therefore the highest contact temperature, will occur at any particular time. For that reason, a single thermocouple on the contact surface will be unlikely to register the highest surface temperature at any given instant. An array of thermocouples would be much better in detecting the actual maximum surface temperature.

The testing apparatus allowed the variation of background temperature as well as sliding velocity (or oscillation frequency), and in some tests both parameters were varied. Partial results for one such test are shown in Figure 6. During that test the data aquisition system gathered temperature data from the thin film thermocouple in 5 second bursts every 30 seconds. [Temperature data were not gathered continuously because the 125 sample/sec data aquisition rate would rapidly result in too many temperature data for the computer to handle.] Friction data were gathered continuously at a slower acquisition rate. Oscillation frequency was set at 2 Hz for most of the test and at 4 Hz for two intervals. The background temperature was held constant at 25°C for the first 380 seconds and then

increased to 45°C for a period before ramping back to 25°C. It can be seen that both higher oscillation frequency (higher sliding velocity) and higher background temperature led to a higher peak contact temperature, although the higher background temperature did not result in a higher flash temperature rise, or temperature change per oscillation cycle. This is in agreement with theoretical calculations, which show that flash temperature rise is a function of heat generation rate, but is independent of background temperature [18]. In fact, the contact temperatures measured here were in very good agreement with those predicted analytically [18].

An examination of the friction data in Figure 6 shows that an increase in friction is brought about by either an increase in velocity (frequency) or an increase in background temperature. In fact, it appears that the friction is quite temperature dependent and an increase in contact temperature brings about an increase in friction, whether the temperature increase is caused by an increase in velocity or by a change in background temperature.

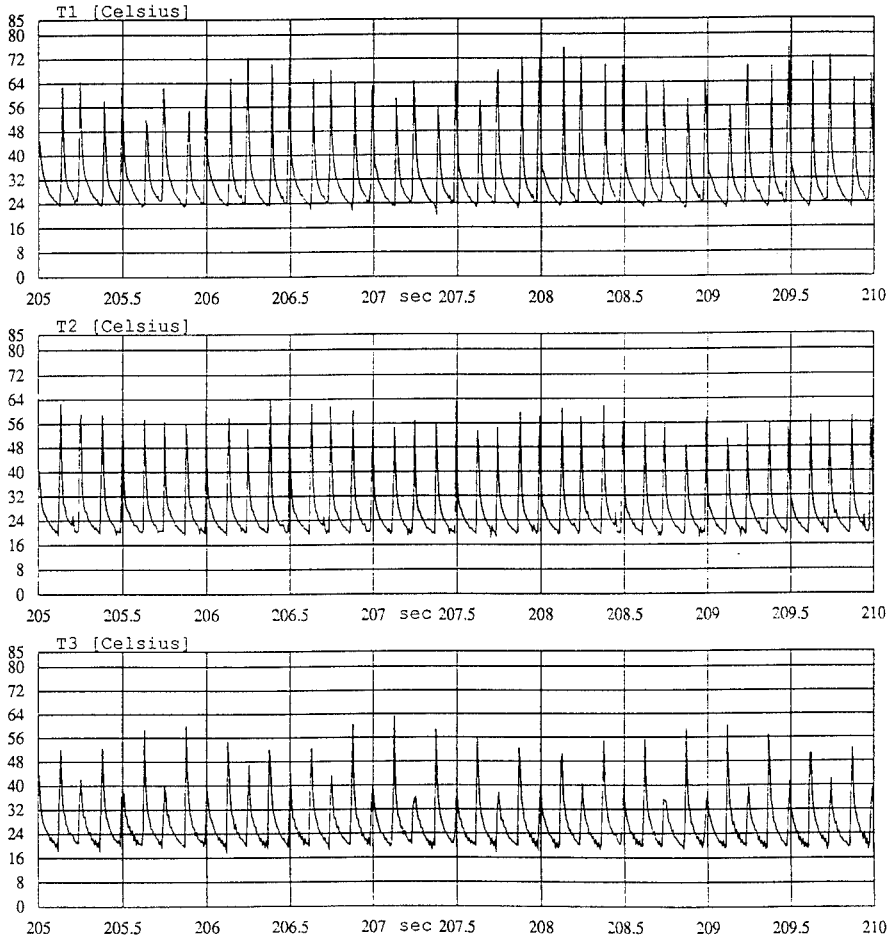


Figure 5. Temperature output from 3-element thin film thermocouple array during 5 sec. period for sliding test of PMMA pin against oscillating alumina-covered glass specimen. Normal load = 9.8 N. Oscillating frequency = 4 Hz. Background temperature = 27°C.

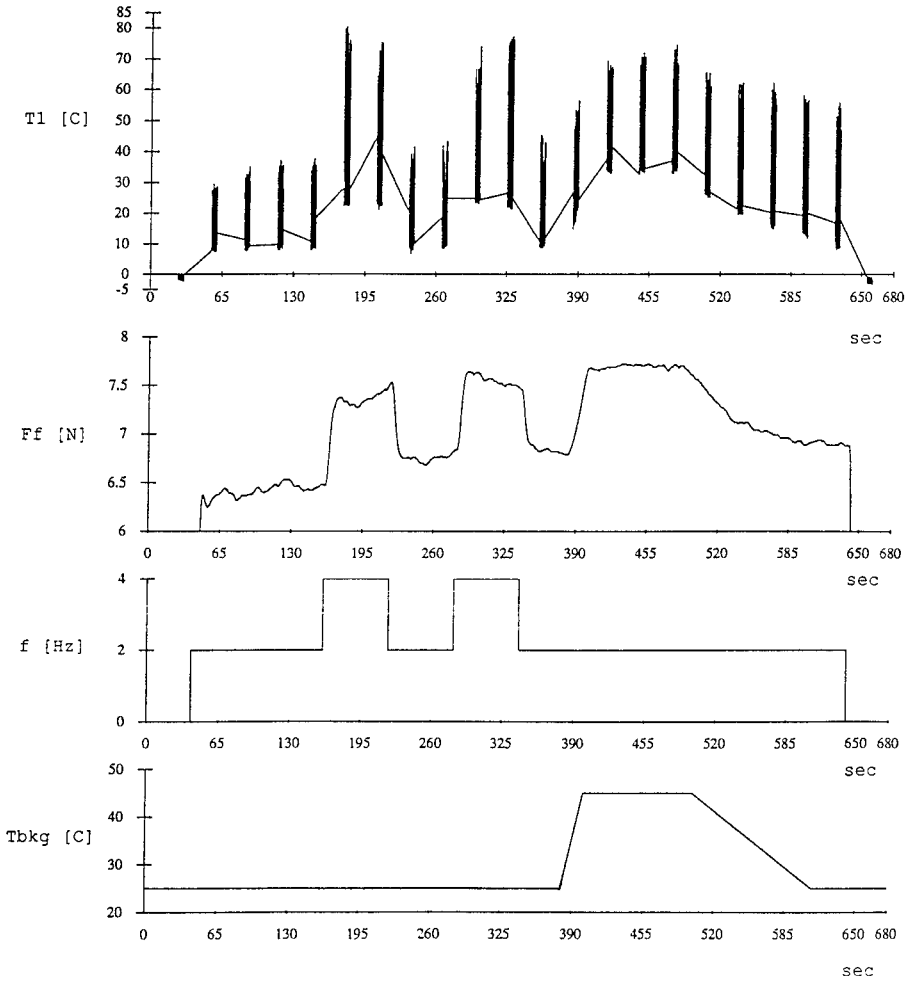


Figure 6. Contact temperature (T1) and friction (Ff) measurements during sliding test of PMMA pin against oscillating alumina-covered glass specimen. Normal load = 9.8 N. Frequency (f) varied from 2 to 4 Hz. Background temperature (Tbkg) varied from 25°C to 45°C.

Tests of both PMMA and UHMWPE were run at a large number of operating conditions leading to a wide range of surface temperatures. The resulting wear rates are shown in Figures 7 and 8 as a function of the measured total surface temperature. The wear coefficient used in the figures is defined as the volume lost per unit sliding distance per unit normal load, and it has units of m^2/N. It is evident from the figures that both materials experienced relatively low wear until the peak surface temperature reached a critical value. For the PMMA material, the critical temperature was about 162°C, whereas for the UHMWPE material the critical temperature was about 137°C. The critical temperature for UHMWPE is approximately equal to the melting temperature of that material. For PMMA, however, the critical temperature is about 30-35°C below its melting temperature and is determined by the temperature at which the PMMA material softens while under

compressive stress. It might also be noted that the wear coefficient of UHMWPE was much lower than that for PMMA, even in the severe wear regime at $T_{critical}$.

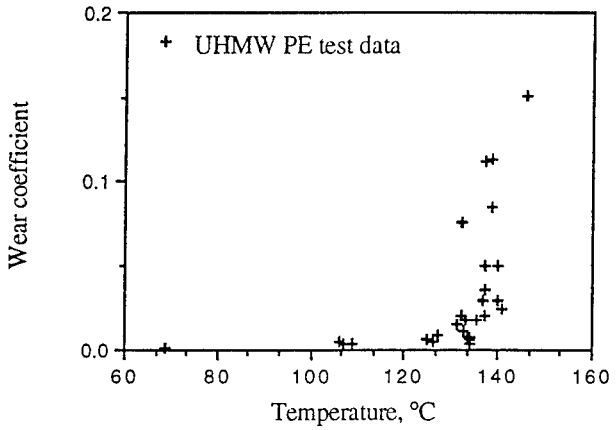


Figure 7. Wear data for UHMWPE pins in oscillatory sliding against glass flats at different contact tamperatures. The wear coefficient is defined as volume lost per unit sliding distance per unit normal load.

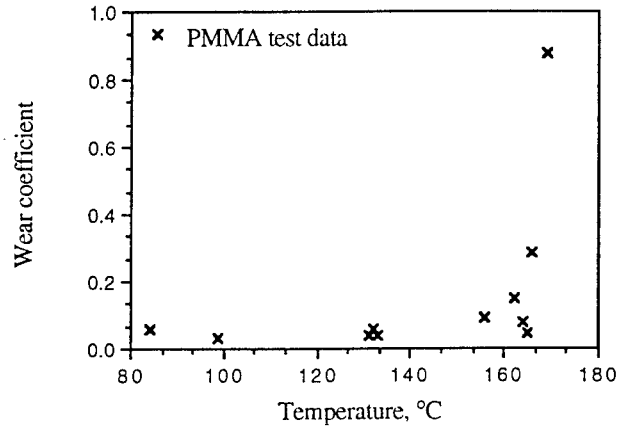


Figure 8. Wear data for PMMA pins in oscillatory sliding against glass flats at different contact tamperatures. The wear coefficient is defined as volume lost per unit sliding distance per unit normal load.

Tests have shown that the drastic increase in wear rate of these thermoplastic materials as their contact temperature approaches $T_{critical}$ is relatively unaffected by the method used to raise the surface temperature [19]. Thus, the temperature increase could be caused by an increase in either sliding velocity, background temperature, or normal load, or a combination of the three factors. It was also found that this result holds true whether the material is amorphous, like PMMA, or semi-crystalline, like UHMWPE. When the critical temperature is reached, changes occur in real area of contact and/or friction coefficient to keep the maximum surface temperature from exceeding the critical value [19].

Use of the TFTC device enabled accurate measurement of the sliding contact temperature responsible for failure of the polymer. Measured temperatures correlated well with a

recently-developed analytical model for prediction of surface temperature rise at the contact interface for cases of oscillatory sliding [18].

It is apparent from the evidence presented above that significant changes in wear of thermoplastic materials can occur if the sliding surface temperature reaches the critical temperature for the material. It is important that such temperature excursions not occur in bearings employing thermoplastic bearings. The peak surface temperatures are dependent on the actual friction coefficient and the real area of contact during sliding, and those quantities are seldom known with any certainty, especially since they vary during operation. One way to overcome the possibility of temperature-induced wear failure of the components is to use in-situ temperature sensors to determine the actual surface temperatures in service. An array of thin film thermocouples would be a very appropriate sensor for such a measurement. The multi-thermocouple array would have a better chance of detecting the highest contact temperature than would a single thermocouple or any other single temperature sensor. The sensors could provide a warning when the peak surface temperature approached the critical temperature, and they can therefore enable failure of the components to be avoided. This could be especially helpful when sliding conditions occur which are more severe than had been anticipated in the design stage.

An array of TFTC devices would be very useful in monitoring potentially damaging contact temperatures over a large nominal contact area in a machine component. The design of such thermocouple arrays and the aquisition of data from them would be rather difficult if the arrays were of the form shown in Figure 2. Incorporation of a switch in each thermocouple, such as that shown in Figure 3, would greatly ease the data acquisition task. Switchable thermocouple arrays similar to the one shown in Figure 4 could make it possible to monitor surface temperature at a large number of points within a contact region without very much data acquisition burden (a single temperature output channel is all that would need to be monitored). Our tests of these arrays have been limited to devices on silicon [15,16], and more development work is needed before similar switchable arrays are possible on other substrates.

Although the work reported here concentrated on only one type of mechanical component (thermoplastic sliding bearings), the methodology used here could be applied to a wide variety of tribological components which encounter friction and frictional heating during operation. Although the failure mechanism for the component might be different, as long as that failure mechanism is temperature-dependent, the use of thin film thermocouple contact temperature sensors could aid in detecting the presence of surface temeratures which could lead to failure of the components.

CONCLUSIONS:

Thin film thermocouples and arrays of such devices have proven to be very effective in measuring surface temperatures in sliding contacts. The TFTC devices have a very small measuring junction which enables rapid response to changes in contact temperature resulting from changes in sliding velocity, contact pressure, or background temperature.

It was found that the wear rate of thermoplastic bearing materials increases dramatically when the contacting surface of the material reaches a critical temperature which is related to its melting or softening temperature. This holds true whether the material is amorphous, like PMMA, or semi-crystalline, like UHMWPE.

An in-situ temperature sensor, such as the thin film thermocouple employed in this study, is one of the most effective ways to accurately determine actual sliding surface temperature during operation. Such a sensor could be used to warn of impending surface failure of

thermoplastic components, and could be of similar benefit in avoiding surface failure of other tribological components.

ACKNOWLEDGEMENT:

The work reported here was supported by the U.S. Office of Naval Research under contract number N00014-93-I-0542. Dr. Peter Schmidt is the ONR contract monitor. The authors are grateful to James J. Deacutis and Christopher G. Levey for assistance in thin film thermocouple fabrication.

REFERENCES:

1. H. Blok, "The Postulate About the Constancy of Scoring Temperature", in *Interdisciplinary Approach to the Lubrication of Concentrated Contacts*, NASA SP-237, 1970, pp. 153-248.
2. S.C. Lee, "Scuffing Modelling and Experiments for Heavily Loaded Elastohydrodynamic Lubrication Contacts", Ph.D. dissertation, Northwestern Univ., 1989
3. F.E. Kennedy and S.A. Karpe, "Thermocracking of a Mechanical Face Seal", *Wear*, v.79 (1982), pp. 21-36.
4. R. Pike and J.M. Conway-Jones, "Friction and Wear of Sliding Bearings", in *Friction, Lubrication and Wear Technology*, Metals Handbook, v. 18, 10th ed., P.J. Blau, ed., ASM International, 1992, pp. 515-521.
5. V.R. Evans and F.E. Kennedy, "The Effects of Temperature on Friction and Wear in Oscillatory Motion of Polyethylene Against Stainless Steel," *Wear of Materials 1987*, ASME, Houston (1987) pp. 427-433.
6. B.Y. Ting, "Thermomechanical Wear Theory", Ph.D. dissertation, Georgia Institute of Technology, 1988.
7. S.A. Lim and M.F. Ashby, "Wear Mechanism Maps", *Acta Metallurgica*, v. 35 (1987), pp. 1-24.
8. X. Tian and F.E. Kennedy, "Temperature Rise at the Sliding Contact Interface for a Coated Semi-Infinite Body", *ASME. J. of Tribology*, v.115 (1993), pp. 1-9.
9. X. Tian and F.E. Kennedy, "Contact Surface Temperature Models for Finite Bodies in Dry and Boundary Lubricated Sliding", *ASME J. of Tribology*, v. 115 (1993), pp.411-418.
10. X. Tian and F.E. Kennedy, "Maximum and Average Flash Temperatures in Sliding Contacts", *ASME. J. of Tribology*, v.116 (1994), pp.167-174.
11. R.S. Cowan and W.O. Winer, "Frictional Heating Calculations", in *Friction, Lubrication and Wear Technology*, Metals Handbook, v. 18, 10th ed., P.J. Blau, ed., ASM International, 1992, pp. 39-44.
12. F.E. Kennedy, "Surface Temperature Measurement", in *Friction, Lubrication and Wear Technology*, Metals Handbook, v.18, 10th ed., P.J. Blau, ed., ASM International, 1992, pp. 438-444.
13. X. Tian, F.E. Kennedy, J.J. Deacutis and A.K. Henning, "The Development and Use of Thin Film Thermocouples for Contact Temperature Measurement", *Tribology Transactions*, v.35 (1992), pp. 491-499.
14. J.J. Deacutis and A.K. Henning, "Switchable Thermoelectric Element and Array", U.S. Patent #5,261,747, (1993).
15. L. Caballero, "Development of a Thin Film Thermocouple Array", Master of Science Thesis, Dartmouth College, June 1994
16. T.M. Cook, "Fabrication of Switchable Thin Film Thermocouples Using MOSFETs", Senior Honors Thesis, Dartmouth College, June 1993.
17. F.E. Kennedy, S.C. Cullen and J.M. Leroy, "Contact Temperature and Its Effects in an Oscillatory Sliding Contact", *ASME J of Tribology*, v.111 (1989), pp. 63-69.

18. X. Tian and F.E. Kennedy, "Prediction and measurement of surface temperature rise at the contact interface for oscillatory sliding", *J. of Engineering Tribology*, in press, 1995.
19. F.E. Kennedy and X. Tian, "The Effect of Interfacial Temperature on Friction and Wear of Thermoplastics in the Thermal Control Regime", *Dissipative Processes in Tribology*, D. Dowson and C.M. Taylor, eds., Elsevier, Amsterdam, (1994), pp. 235-244.

CONDITION MONITORING USING THE TIME STRESS MEASUREMENT DEVICE (TSMD)

Leonard Popyack

RL/ERSR
525 Brooks Road
Griffiss AFB, NY 13441-4505

Joseph Kubler

Reliability Analysis Center
201 Mill Street
Rome, NY 13440-6916

Abstract: New technologies facilitate the accurate and non-intrusive measurement of actual shock and vibration stresses in operational environments. These devices often referred to as Time Stress Measurement Devices (TSMD), integrate sensors with processing and data storage technologies to provide a record of stress frequencies and magnitudes over an operating period. These devices were originally designed to help isolate and resolve reliability design issues and to assist in the development of system design requirements. To date, TSMDs have been used by the U.S. Army, Navy, and the Air Force on the B1-B, B52, EF-111, and F-22. Through a number of development efforts at the United States Air Force's Rome Laboratory (RL), several different versions of TSMDs have been developed by various contractors.

TSMD devices were designed to provide a small self contained data collection system for deployment on military platforms. These devices are also well suited for industrial or commercial applications having mobility requirements or size restrictions, such as those applications found in the transportation industry. TSMDs can monitor and record shock and vibration as well as other parameters such as temperature, pressure, relative humidity, and voltage transients or on/off cycling.

Results of an on-going study aimed at identifying commercially available TSMD-like devices will also be presented.

Key Words: Condition monitoring; data collection; TSMD

Introduction: A Time Stress Measurement Device (TSMD) is a device that integrates sensors with processing and data storage technologies to provide an

operational stress history over an operating period. A TSMD consists of a suite of physical parameter sensors connected to a microcomputer and non-volatile memory for storage of stress data through signal conditioning circuitry. The data is recorded with time tag information so an accurate reconstruction of the operating parameters can be reviewed at a later time. The TSMD was designed to collect data in the actual usage environment while remaining transparent to the host system's operation. Originally designed to resolve reliability related issues in military avionics, the TSMD lends itself to collecting data needed for a predictive maintenance program where system mobility is a requirement. Currently, there are several TSMDs deployed by the USAF in airborne applications. The term TSMD is a term used by RL which applies to a technology rather than a specific product or manufacturer. Over the last 8 years, there have been several RL TSMD development programs involving different manufacturers. During this same time period, new technologies have allowed for the development of commercial data collection systems similar in capability to the RL developed TSMDs.

RL TSMDs are well suited for data collection applications in mobile or harsh environments that will not permit large bulky data collection systems or are not easily accessible to personnel during operation. In addition to stress data storage, TSMDs can be configured to provide an indication of when sensed parameters have exceeded user determined tolerances or thresholds.

Rome Laboratory TSMD Device Family: RL has developed several types of Time Stress Measurement Devices. The intended function of all RL developed TSMDs is to determine what impact the environment has on a particular system of interest. The TSMDs provide the capability to record data on *operational* aircraft without the need for special dedicated test aircraft. The first TSMD produced was the TSMD module (see Figure 1).

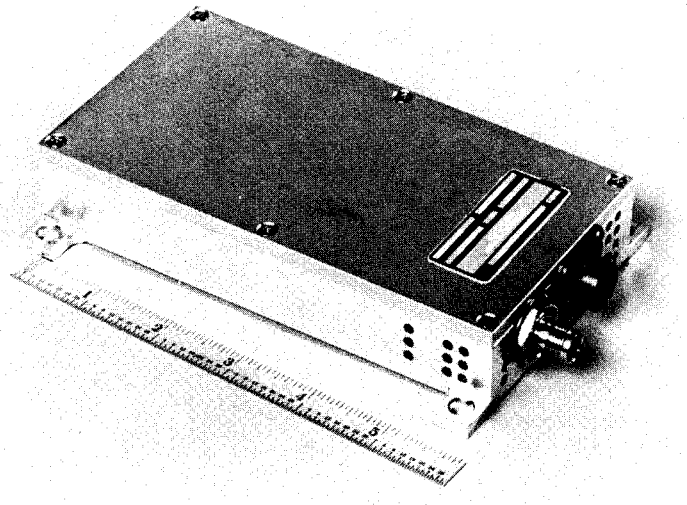


Figure 1: TSMD Module

Fifteen modules were produced under a contract to Honeywell. The module size is 6"x4"x1" and it measures and records temperature, vibration, humidity, shock, corrosion and power transients. The TSMD module operates independently of the host through the use of a rechargeable sealed lead acid battery. The module records its parameters and stores the data in memory for later retrieval through an RS-232 interface. Excursions of the avionics beyond user defined limits for humidity, aircraft power, and vibration/shock are recorded in memory bins. Each time one of the limits is exceeded, the appropriate bin is incremented. Each bin contains the time above a limit in 0.1 second increments. At the end of each day, the bins contain the number of times the limit was exceeded for that day. The bin counts are saved in memory and the bins are cleared for the next days use. Environmental parameters and power cycles are time-stamped with their values. The module collects vibration data by recording the accumulated exposure time to vibration above a field adjustable threshold from two gs' to five gs' in the four frequency bands that follow: 20 Hz to 80 Hz, 80 Hz to 200 Hz, 200 Hz to 2000 Hz, and 2000 Hz to 4000 Hz. The outputs of the rectifier filters and signal conditioners for vibration and shock are fed respectively into comparators that have adjustable reference inputs. Each reference is adjustable through software. The output of the comparators produce an interrupt input to the microprocessor. When the envelope of the vibration signal persists for an extended time, the comparator reference is raised so that an interrupt is no longer generated. The signal is then monitored on a periodic basis. When the vibration level drops back below the original threshold level, periodic sampling is discontinued, and the comparator is reset. The vibration and shock sensor used is the Vibrometer CE501 accelerometer. This module uses 32K battery-backed SRAM for data storage.

The second TSMD developed is the Micro TSMD (see Figure 2). Twenty devices were delivered under contract with Honeywell. The Micro TSMD is physically a 1"x2" hybrid flatpack with leads on 50 mil centers. The Micro TSMD measures and records temperature, vibration, shock, voltage and voltage transients and has memory reserved specifically for storage of textual data. The typical text data stored is information such as serial numbers of the equipment and repair history. A real time clock is used to time stamp the significant events. The Micro TSMD is a circuit card mountable device that requires 5 volts dc from the host system. The Micro TSMD user interface and data are accessible through an RS-232 interface. All sensors are internal to the small 1"x2" package, however the Micro TSMD can also accommodate remote sensors for vibration and temperature. Additionally, there are four spare channels that can be used by any sensor that has an output voltage in the range of two to ten volts. The vibration data is collected in a much better fashion than the older modules. The internal accelerometer senses vibration and shock in the plane perpendicular to the hybrid. A 64 point Fast Fourier Transform is performed on the vibration data immediately. The Micro TSMD has a fourth order Butterworth filter to eliminate anti-aliasing of the vibration signal. The vibration signal is sampled at 8 KHz. Records of time at overall RMS level and time at g^2/Hz are retained in tables. The magnitude thresholds are user definable in software. Shocks above 3g are time stamped with their value and stored in nonvolatile EEPROM. A shock counter can also be used when the system is powered down. This requires a small button cell battery to wake-up the shock circuitry when an event occurs. The

measurable range for shocks is ±64g with a 0.2g resolution. The vibration and shock sensor used in the Micro TSMD is the Endevco 22M5. The Micro TSMD has 256K of EEPROM memory for data storage of all measured data. Thirty six kilobytes of assembly language program is built-in for the operating firmware.

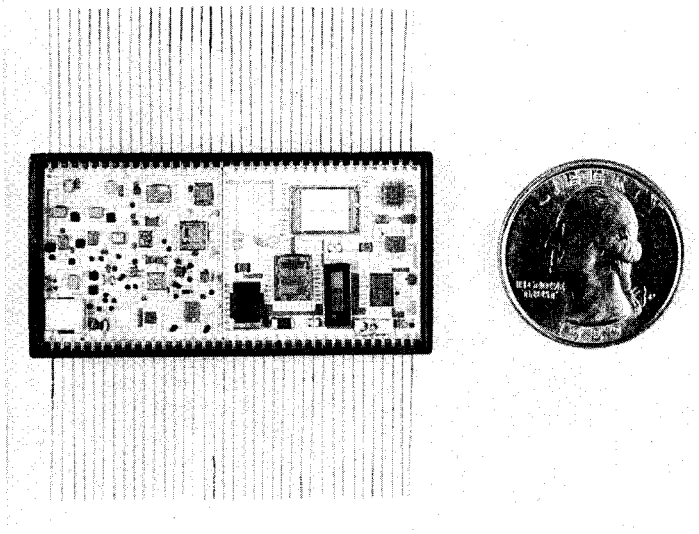


Figure 2: Micro TSMD

The latest TSMD device is referred to as the Advanced TSMD. It is based on the MC68333 Motorola 32-bit microprocessor. The sensors are external to the hybrid circuit. This device is built specifically to monitor host system faults and record the operational environment before, during, and after a built-in test (BIT) detected event.

A graphical representation of actual vibration data is shown in Figure 3.

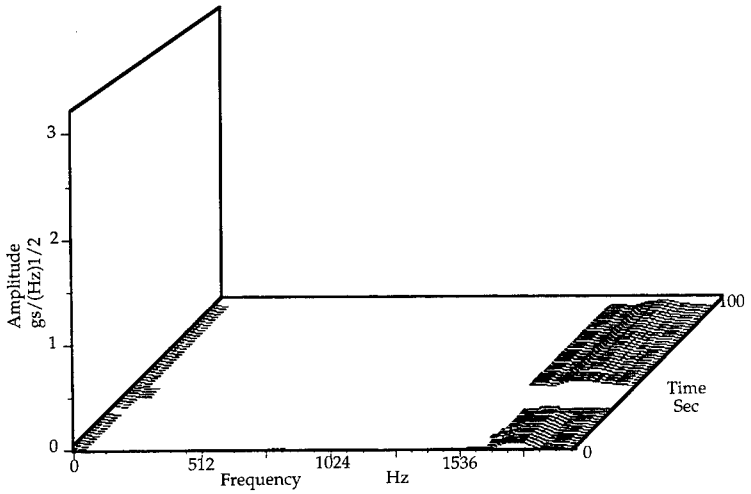


Figure 3: Graphical Representation of Vibration Data

Market Survey: The Reliability Analysis Center (RAC) has been involved with several of the RL TSMD programs. The RAC is a DoD information analysis center (IAC) that disseminates information and performs special studies pertaining to reliability and maintainability issues. RAC is brokering RL TSMD technology and commercial data recording devices to organizations expressing a need for environmental characterization to resolve reliability issues. At this point in time, Rome Laboratory developed TSMDs are not "off the shelf" equipment. The lack of commercial availability has deterred small scale programs from using the devices due to fiscal and time constraints. This has prompted RAC to perform a market study to assess the commercial availability and applicability of data collection systems similar to RL TSMDs.

RAC's ongoing market survey has encompassed over three hundred companies involved with data acquisition of environmental parameters. Product information from solicited venders is summarized by capability, technical specifications, and target applications. The limiting factors for most commercially available TSMD type devices for measuring shock and vibration are data storage capability and data sampling rates. The survey has identified manufacturers whose products can be deployed with a host system to record shock and vibration data unobtrusively. The prices of these commercially available devices range from several hundred dollars to several thousand dollars, depending on size and capability. Additionally, a number of companies offer customization of their commercially available products or build custom devices for specific applications. Example companies that offer stand-alone shock and vibration measuring and recording equipment include:

Dallas Instruments, Inc. offers the "Peak Acceleration Logger (PAL)" and the "Saver". The "PAL" records tri-axial peak accelerations using three internal accelerometers with the option of using externally mounted accelerometers. The "Saver" offers eight high speed channels and six low speed channels with an internal tri-axial accelerometer and temperature and humidity transducer as standard equipment. User programmable options include: selection of measurement sensitivity, low pass filter selection, sampling intervals, and time and date clock. Programming these devices and downloading the data is accomplished via an RS-232 serial interface.

Instrumented Sensor Technology, Inc. offers several devices that measure shock and vibration. Models include; "EDR-1,2,3", "MSR-1", and the "Snapshock". The "EDR" product line records acceleration measurements in three axes through the use of internal or external accelerometers, The "EDR" can also record temperature and humidity data. The "MSR-1" has the capability of recording nine acceleration channels as well as three channels dedicated for temperature and humidity measurements. The "Snapshock" records peak accelerations in one axis and time and date stamps the information. Instrumented Sensor Technology will provide some factory modifications allowing the "Snapshock" to interface with other types of sensors. Programming the device and downloading the data is accomplished via an RS-232 serial interface.

Electronic Development Corporation offers the Model 650 acceleration recorder as well as custom device development. The Model 650 is a battery powered, multiple event, triaxial acceleration recorder. The trigger source can be acceleration, rms vibration, velocity change, or a threshold set on one of the three external channels. The unit records the date, time, battery voltage, and internal temperature for each data frame. Programming the device and downloading the data is accomplished via an RS-232 serial interface.

Out of the three hundred companies contacted, RAC has identified 40 plus manufacturers that produce self-contained data collection systems with various sensing capabilities. All of these identified manufacturers have products that can record one or more of the following; vibration, shock, temperature, humidity, voltage, current, and pressure and are applicable to a predictive maintenance scenario.

The capabilities of earlier versions of the RL's TSMDs can be found readily in the commercial market place. However, most commercial systems are not sufficiently packaged for certain military applications. The newer models of RL TSMD's offer minimal size and weight the packaging which cannot be found commercially. The majority of the commercial devices including the Rome Laboratory TSMDs use an RS-232 interface for data downloads and programming.

Device selection is based solely on user requirements and the specific application constraints. The market survey is an ongoing process that will be continuously updated as new sources of TSMD technology are identified. Rome Laboratory TSMD devices are not available as off-the-shelf hardware at this time and must be manufactured to order. For the purpose of this paper, only a few of the commercial systems were highlighted. The information presented is based on received literature and vendor claims and not actual experience. RAC does not endorse any particular manufacturer's product.

Maintenance Parameter Prediction Monitoring: As with most predictive maintenance programs, initially the user must determine the normal operating parameters for the system. Once the user has determined the normal operating parameters of a system, the data collection system can be programmed to record only exceeded tolerances to conserve data storage space. The data collection system thresholds must also account for different operating modes and conditions, such as; RPMs, ambient temperature, loading, etc. Depending on the application, the measured parameters may be frequency response or amplitude, temperature, pressure, or any other sensed parameter. Once installed, the data collection system will collect data limited only by battery life and data storage capacity. However, most systems will typically overwrite the oldest data as new data is received so that the most current information is not lost. The data collection period is solely user dependent and data is typically retrieved via a serial port with a laptop computer. As monitored equipment degrades, the user may want to shorten the data collection period and change the thresholds to reflect changing operating parameters.

A method for predicting equipment degradation and possible failure with on-line monitoring of equipment environmental stress was described in [1]. By

monitoring and computing stiffness parameters in real time, a method exists which can be used to flag faulty equipment (or soon to be faulty equipment).

The method uses multiple accelerometers to measure and record the frequency spectrum of vibrations normally present within the equipment (see Figure 4). The TSMD acquires the data and processes the Fast Fourier Transform (FFT) information. Next, extraction of the stiffness model equation co-efficients are stored. By monitoring the time-lap sequence of the stiffness parameter coefficient drift, one can determine through experience when a material will fail (see Figure 5). This technique can be thought of analogous to bending a wire back and forth many times. Eventually it will break.

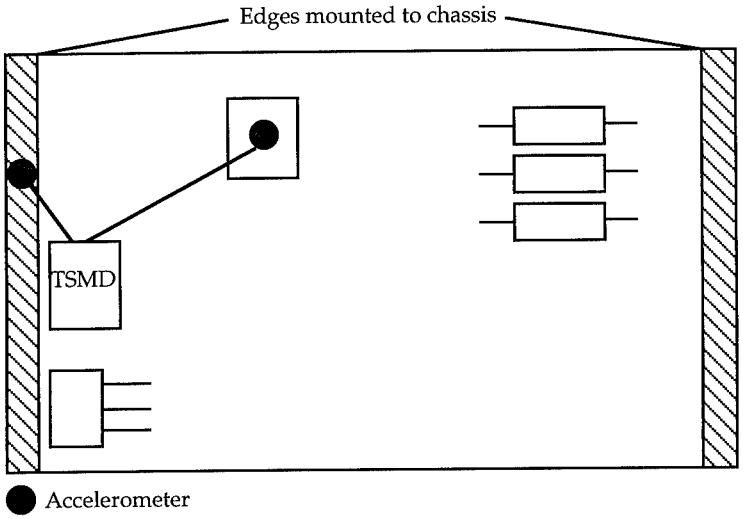


Figure 4: Micro TSMD Installation on Circuit Card

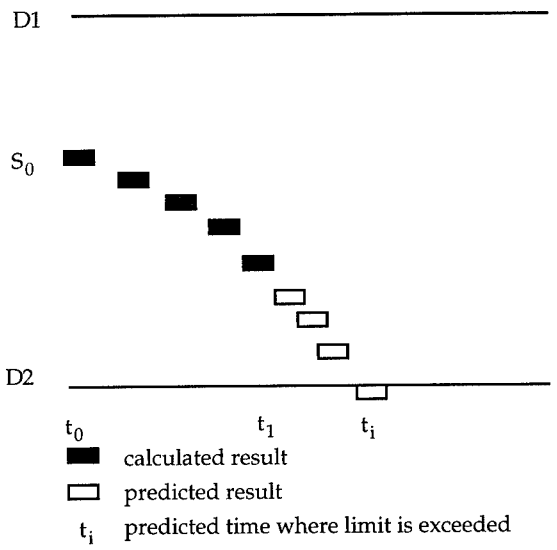


Figure 5: Stiffness Parameter Coefficient Drift

TSMD Module: Currently, RAC is managing a RL TSMD program that was designed to characterize system parameter's leading to and at the time of failure. The TSMD is monitoring an oxygen generating system aboard a USAF B1-B aircraft. Measured parameters are inlet and outlet pressure, relative humidity and temperature. The TSMD is programmed to collect data only when system inlet pressure reaches a user set threshold. Collected flight data is downloaded approximately once every week using a laptop PC at the aircraft. This particular TSMD system employs rechargeable lead acid batteries as a prime power source. Typically the battery pack is swapped when the technician downloads the data. Ideally, data from this program will be used to isolate an unknown recurring failure mechanism in order to implement corrective design changes.

Advanced TSMD: The Advanced TSMD is being flown on operational B-1B aircraft (see Figure 6). It monitors the environment inside the plane's forward looking terrain follow radar system. When the radar detects a fault within itself (BIT event), the TSMD records the past 40 seconds and the next 20 seconds of environment data. This "fault window" shows what conditions were present when the fault occurred. The Advanced TSMD monitors vibration, mechanical shock, internal avionics temperature, cooling air pressure, cooling air temperature, and power supply conditions.

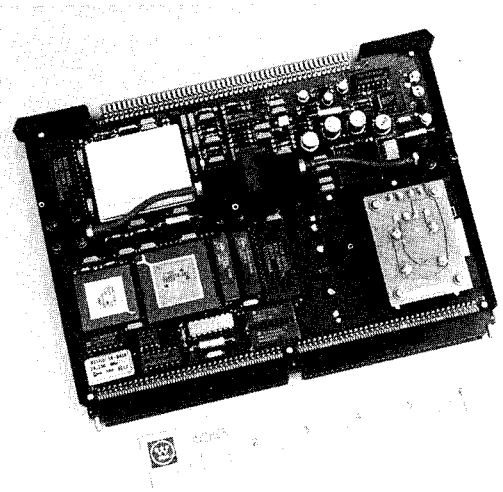


Figure 6: Advanced TSMD

In a flight test program, up to 40 B-1B aircraft are fitted with the Advanced TSMD. The program has measured the environment and correlated it to BIT faults covering the past 13 months (ending December 1994). Over 600 Megabytes of environmental data have been collected from aircraft in Texas and California. The data is currently being analyzed for correlation of single and multiple environment parameters affecting the host equipment. A

statistical clustering model is being constructed to look for highly correlated data.

Summary: Incorporating condition monitoring equipment to systems with size restrictions and mobility requirements can be difficult unless provisions are made by the manufacturer. Technology has allowed for the miniaturization of stand alone data collection devices which can be affixed or incorporated into existing equipment. As shown with the B1B application, TSMDs can be successfully embedded in or attached to systems to offer a permanent condition monitoring capability. The commercial market place offers a variety of off the shelf data collection systems that match capabilities of earlier versions of RL TSMDs. The size, packaging and capabilities of the Micro and the Advanced TSMDs are not readily available in off the shelf products at this time.

References: (1) Popyack, L.J., Skormin, V.A., and Plaskon, S.L., "Transfer Function Method for Diagnostics of Electronic Circuit Boards Exposed to Mechanical Vibrations", May 1993 NAECON Conference, Dayton, OH.

TIME FREQUENCY ANALYSIS

Cochairmen: **Kam W. Ng**
Office of Naval Research

Leon Cohen
Hunter College

APPLICATION OF TIME-FREQUENCY ANALYSIS TO THE MONITORING OF MACHINING PROCESSES

Pat Loughlin[1], Les Atlas[2], Gary Bernard[3] and Jim Pitton[4]

[1]Dept. of Electrical Engineering
University of Pittsburgh
Pittsburgh, PA 15261

[2]Dept. of Electrical Engineering
University of Washington
Seattle, WA 98125

[3]Boeing Commercial Airplane Group
P.O. Box 3703, MS 5K-14
Seattle, WA 98124-2207

[4]AT&T Bell Laboratories
600 Mountain Ave.
Murray Hill, NJ 07974

ABSTRACT: Many manufacturing operations require close monitoring of the health of the machines in operation and the integrity of the parts being manufactured. The mechanical and acoustic vibrations produced by this process are a source of cues for such monitoring. Spectral analysis, and more generally time-varying spectral analysis, or time-frequency analysis, provides a means for identifying changes in the vibrations produced by machines as they degrade. In this paper, we report on our applications of time-frequency analysis to drilling and grinding operations. We show that, while the short-time spectrum, or spectrogram, often gives a good indication of the gross evolution of degradation, newer methods of time-frequency analysis can reveal far more of the detail and may allow earlier detection of problems.

Key Words: machine monitoring; manufacturing; spectral analysis; time-frequency analysis.

INTRODUCTION: In many manufacturing operations involving controlled removal of material with sharp tools, it is important to monitor in real time the quality of the cut surface and the health of the tool doing the cutting. As has long been recognized by skilled machine operators, mechanical and acoustic vibrations are a source of cues for such monitoring.

The physics of rotating machines leads naturally to spectrally analyzing the vibrations produced by rotation as a means for monitoring the condition of the manufacturing process. This analysis is typically achieved by Fourier transforming a machine signal $s(t)$ (e.g., a sound recording of the machine during operation), and taking the squared-magnitude to obtain the spectral density $P(f)$ of the signal:

$$P(f) = |S(f)|^2 = \left|\int s(t)\, e^{-j2\pi ft} dt\right|^2. \quad (1)$$

The spectral density reveals the strengths of the frequencies in the signal. The frequencies, in turn, reflect the condition of the machine. For example, the spectrum of

a 21-tooth gear rotating at a speed of three revolutions per second will exhibit a peak at the gearmesh frequency $f = 63$ Hz, which is the number of teeth multiplied by the rotation speed. Defects generate other frequencies. For example, imbalance generates a frequency at rotor speed (3 Hz in this example); a loose rotor will generate harmonics of the gearmesh frequency; and loose gears can also generate harmonics, as well as sidebands about the gearmesh frequency due to amplitude modulations (AM) [18].

While informative and useful for diagnostic analysis of rotating machines, the spectrum provides an incomplete picture of the degradation process. Specifically, the spectrum can not describe the *evolution* of the degradation process, since time is not a parameter of the spectral density. What the spectrum provides is a global picture of the entire process. The spectrum does not indicate when the different frequencies occurred, only that they did.

Acknowledging this limitation of the spectrum, current practice involves the computation of spectra over successive short-time intervals of data. A comparison of the short-time spectrum at one time interval to the short-time spectrum at some later time interval can provide a gross indication of the evolution of degradation. This simple but often effective technique falls under the category of "time-frequency analysis." Although more sophisticated and accurate methods will be presented here, in essence, time-frequency analysis allows one to determine at what time a particular frequency was present in the signal. For processes where changes in the frequency content are physically important, as in machine degradation, time-frequency analysis holds great promise as a means for allowing early detection of problems.

The early identification of problems in manufacturing processes is important for it enables immediate corrections to be applied, thereby maintaining process quality and ensuring safe and efficient operations. Safety and process integrity are currently maintained through operator experience developed over years of machining. However, there are more manufacturing needs than there are skilled experts, and machinists are still prone to human error. Thus, a conservative approach is taken, and manufacturing tools are typically serviced long before they are expected to become a problem (based on statistical data of similar tools). In a globally competitive market, such a maintenance schedule is inefficient and costly. What is needed are methods to assist machinists in assessing the state of a particular machining operation so that corrective maintenance may be applied *as needed.* Time-frequency analysis provides the basis for such a method. We have investigated the utility of time-frequency methods for monitoring drilling and grinding processes. In this paper, we report on our work in this regard.

BACKGROUND: Rohrbaugh [14] has applied new methods of time-frequency analysis [21], [12] to the assessment of submarine rotors and other marine machines. He has demonstrated that time-frequency analysis can reveal degradation problems long before they appear in the spectrum or short-time spectrum. For example, Rohrbaugh presents the case of a fan running at constant speed, wherein a phenomenon of "spectral splitting," which indicates a potential bearing problem, begins to appear in the time-frequency plot three months before any problem is indicated by conventional spectral analysis (figures 1b and c in [14]).

Forrester has applied time-frequency methods to detect gear damage in helicopters [7]. By examining gears with and without damage, he has shown that there are distinct

differences between the time-frequency patterns for these two states, and furthermore that time-frequency analysis can reveal the type and extent of damage. Wang and McFadden have also investigated the utility of time-frequency analysis for detecting gear damage [19]. They give an example where this analysis clearly reveals an anomaly in the time-frequency pattern of the vibration signal of a gear. Upon visual inspection, the gear was found to have a small spall on one tooth face. Other contributors include Williams and colleagues for Ford (car door slams and windshield wiper noises) [20], Rizzoni (engine knock) [15], [16], and Shin, Jeon and Spooner (machine monitoring) [17].

THEORY OF TIME-FREQUENCY ANALYSIS: Perhaps the most straightforward and readily understood method of time-frequency analysis is the short-time Fourier transform, the magnitude-squared of which is called a "spectrogram" or short-time spectrum [9]. Developed in the 1940s, the basic idea behind the spectrogram is simple yet powerful: to measure the changing spectral content of a signal, successive short intervals of the signal are isolated via multiplication with a window, and then Fourier transformed. The squared-magnitude is taken to obtain an energy density. Mathematically, we have

$$P_{sp}(t,f) = \left| \int s(\tau)\, w^*(\tau - t)\, e^{-j2\pi f\tau} d\tau \right|^2 \qquad (2)$$

where $P_{sp}(t,f)$ is the spectrogram of signal $s(t)$, $w(t)$ is the window, and * denotes complex conjugation. While simple and effective for many applications, the spectrogram has limitations. Principally, the window distorts and sometimes obscures the time-frequency structure of the signal [2], [4]. Figure 4 shows a good example of this limitation.

New methods of time-frequency analysis that overcome this limitation of the spectrogram have been developed in recent years. These new "time-frequency distributions," or TFDs, are obtained by specifying a particular "kernel" $\phi(\theta,\tau)$ in Cohen's formulation [3]-[5]:

$$P(t,f) = \iiint s(u+\tau/2)\, s^*(u-\tau/2)\, \phi(\theta,\tau)\, e^{-j2\pi\theta(t-u) - j2\pi f\tau} du\, d\tau\, d\theta\,. \qquad (3)$$

The kernel dictates properties of the distribution. For example, to ensure the correct answer for measurements from the TFD such as duration and bandwidth (calculated as the standard deviation in time and frequency, respectively), the kernel should be chosen such that [3]-[5]

$$\phi(\theta,0) = \phi(0,\tau) = 1\,. \qquad (4)$$

The spectrogram does not satisfy this condition, and consequently overestimates the duration and bandwidth of the signal [5]. Other measurements are considered in detail in [2]-[5] and references therein.

Significant progress has been made in designing kernels to obtain accurate and informative TFDs [1], [2], [4], [6], [8], [11], [12], [21]. Examples include the Wigner, Choi-Williams, RID, Page, Rihaczek, Bessel, sinc and cone kernel TFDs. Different distributions are appropriate for different signals; hence, the choice of the "best" kernel to use depends on the signal at hand. Recently, a general method for automatically designing a signal-dependent kernel to use in constructing the TFD has been developed

by the authors [10], [12], [13]. We apply this general method and others to the analysis of manufacturing signals.

ANALYSIS OF DRILLING: One conventional approach to fault prevention in machining is to replace a tool well before expected failure, based on statistical analyses of the lifetime of similar tools. In large-scale drilling operations, routine practice is to replace the drill after completing a specified number of holes, independent of the actual dullness of the drill. This technique can not prevent all failures, because some drills become dull well before the statistical lifetime is reached. A method for automatically assessing the quality of the tool is desirable and more efficient than the statistical-based approach. Our efforts to develop measures of drill dullness have focussed on the break-in and break-out regions of the drill time series. Figure 1 shows a spectrogram and the time series of a dull drill over the entire drilling of one hole.

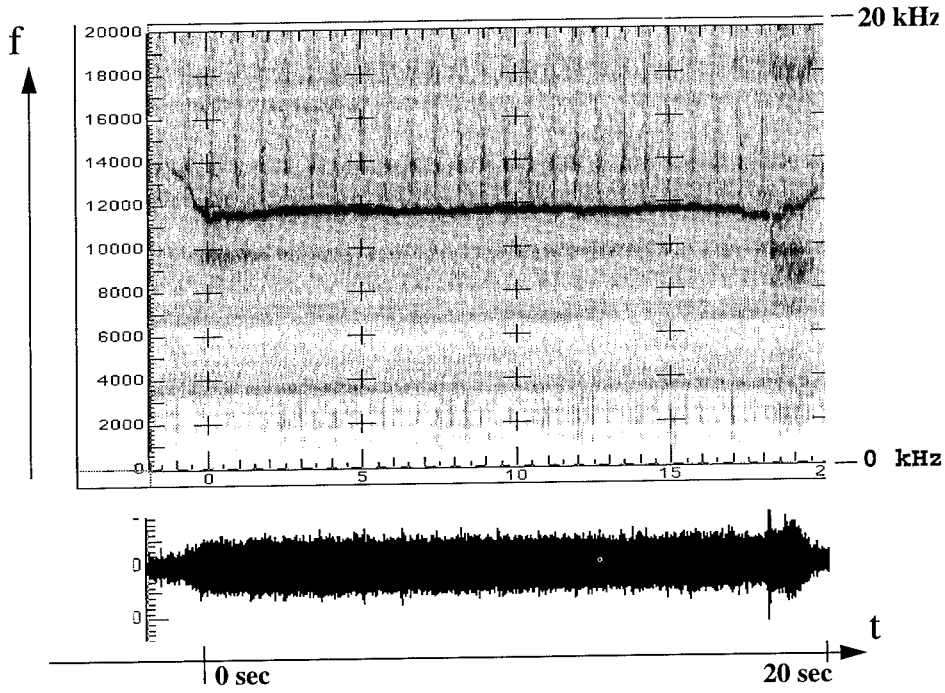


Figure 1: Spectrogram (top) and time series (bottom) of a dull drill (u39).

Drill dullness causes strong vibrations that can be detected by analyzing the sounds generated during drilling. A comparison of time waveforms, spectrograms, and cone kernel TFD [21] of drill signals is shown in Figure 2. In the time signals, a visible difference is apparent between the sharp drill, the medium drill, and the dull drill. The TFD of the drill data exhibits marked time-varying spectral differences among the drills. For sharp drills there is almost no energy in the vicinity of 500 Hz, unlike the situation for dull drills. The duller the drill, the stronger the power distribution at and around 500 Hz and the longer the duration of this component. The plots also show that the frequency localization around 500 Hz tends to happen mainly during break-out. Thus,

time-frequency analysis is useful here for determining not only the drill's dullness but also the time location of break-out events.

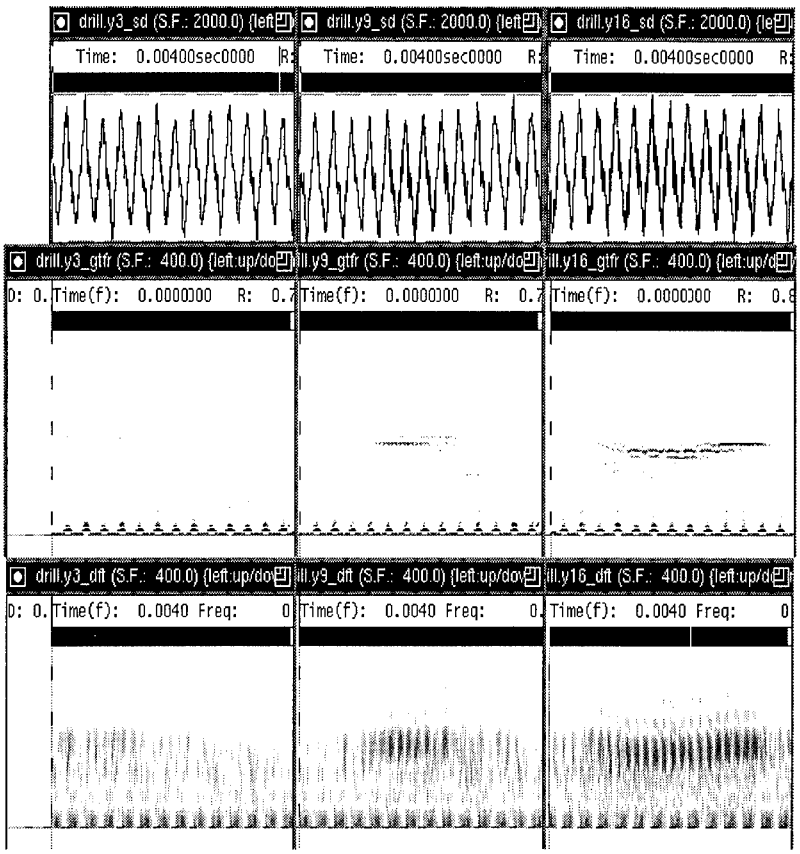


Figure 2: Comparison of spectrogram and cone-kernel TFD on sharp (left), medium and dull (right) drill time series. As the drill becomes dull, energy around 500 Hz increases. Note the clarity and sharpness of the cone-kernel technique, and the time-varying spectral changes not only from one time-series to the next, but also within a single time-series. Frequency range is 1000 Hz, time range is 0.8 seconds per time-series.

ANALYSIS OF GRINDING: In surface grinding of steel and aluminum, an abrasive wheel is used to remove material from a flat surface of a piece mounted on a longitudinally-translating bed. After each pass of the bed, the grinding wheel jogs a lateral increment across the bed, so that eventually the entire surface of the work-piece is covered. Then the wheel is advanced closer to the surface and the process is repeated. Catastrophic failures can occur. For example, aluminum is soft and has a lower melting point than steel. If a wheel is advanced too much on aluminum, the grinding can generate enough heat to cause local melting of the aluminum. This local melting severely plugs the grinding surface of the wheel. The cutting ability of the wheel is quickly reduced, and the wheel rubs against the part, rather than removing material. The friction from rubbing produces more heat, causing more melting and hence, more plugging of the

wheel. This unstable cycle accelerates until the wheel and part seize, damaging the part, the wheel, or the machine.

Expert operators rely on several cues to aid them in machining, and thereby avoid such catastrophes. A significant cue is acoustic; an experienced operator can make judgements on the state of the process based on how it sounds. Given the importance of sound in assessing manufacturing operations, acoustic recordings of grinding operations were made with a hand-held microphone onto a Sony DAT recorder.[1] These data were then analyzed via time-frequency methods.

A spectrogram from a recording of a steel-grinding operation (Figure 3) revealed that the signal contains, among other spectral characteristics, spectral bands of energy located at approximately 1500 Hz and 3100 Hz. These bands are present over intervals of 2.2 seconds, corresponding to the duration of a single pass of the grinding wheel. There is also significant background noise below 1000 Hz, corresponding to the sound of the motor, coolant, and other ambient noise sources. This noise is always present; it does not fall off with the completion of a pass of the wheel, as the grinding components do.

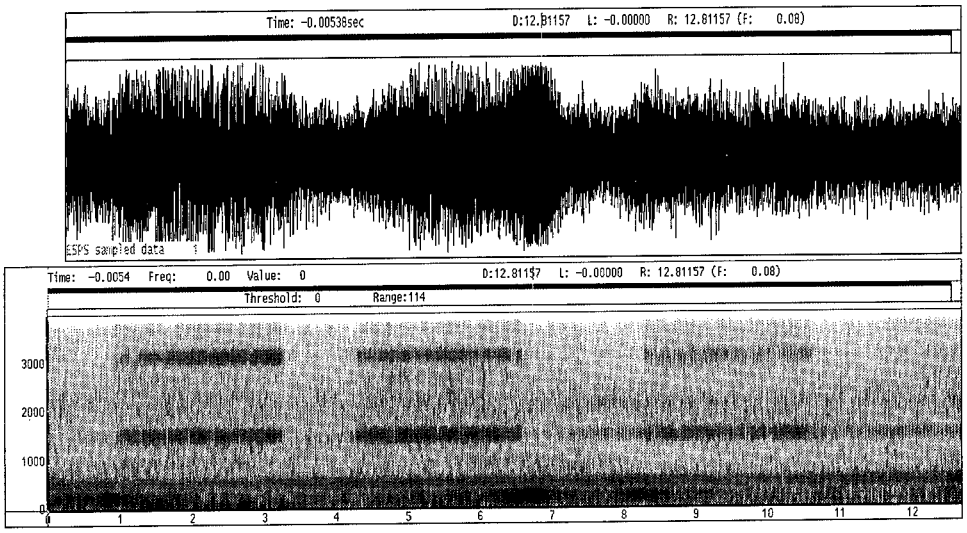


Figure 3: Time series and spectrogram of a segment of steel grinding data.

Figure 4 illustrates the application of a new technique [12] to the analysis of the same acoustic recording of the grinding operation. The figure shows detail over a small region of the time-frequency plane. Note that the new technique shows far more detail than does the spectrogram (top plot). In particular, this technique revealed that significant amplitude modulations (AM) were present in the grinding-related frequency components of the signal. Strong harmonic structure was observed in the AM from the fresh wheel, periodic with a 20 Hz fundamental frequency. This frequency corresponds

1. The method of recording was purposely simple, so that the recorded data would be of the same quality expected in a working system.

to the frequency of rotation of the grinding wheel. As with the case presented by Rohrbaugh [14], the transient AM in grinding appears in the time-frequency plot before it appears in the Fourier spectrum. Additionally, it is clearly apparent in the plot of the new technique, but not in the plot of the commonly used short-time spectrum, or spectrogram.

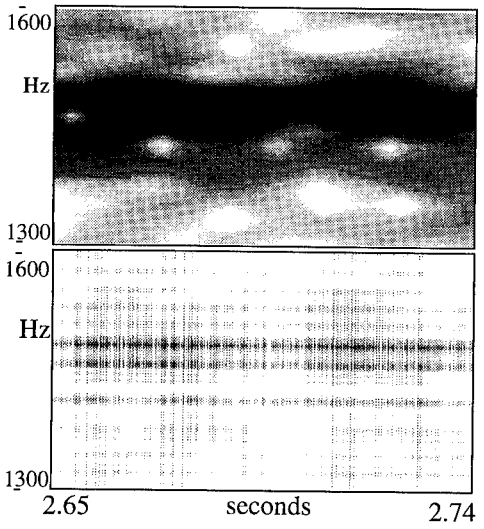


Figure 4: Spectrogram (top), and a TFD generated via a signal-adaptive method developed by the authors [12] (bottom) of an acoustic recording of a steel grinding operation. Note that the TFD reveals far more of the detail of the time-frequency structure of the signal (e.g., the time-varying harmonic structure). This harmonic content was found to relate to the state of the grinding wheel (e.g., sharp *vs.* dull).

As grinding continued and the wheel became plugged, the harmonic structure appeared to weaken. However, absolute energy changes are not a reliable measure of the state of the grinding wheel, as many factors other than a plugged wheel could change the energy level of the signal. For example, movement of the plate and wheel closer to or further from the microphone would change the energy level of the signal.

Given that the strength of the AM is dependent on the overall strength of the signal, correlation between individual AM components was measured. This measurement is unaffected by global level changes. Correlation is likely, too, since the frequency components observed were produced by the same periodic source, namely, the rotating grinding wheel. The cross-spectrum of the AM extracted from two frequency components (1500 Hz and 3100 Hz — see figure 5) of the grinding data was computed over 19 minutes of data (figure 6). This analysis revealed that the AM signals on the 1500 Hz and 3100 Hz components were indeed strongly correlated at the same fundamental frequency (20 Hz), and the amount of correlation increased as the grinding process progressed. After dressing, the correlation was noticeably reduced.

A very important concept in machine monitoring is generalization. Tools developed for manufacturing sensor signal analysis should be applicable to a broad range of processes for cost-effectiveness of the signal processing tools. In addition to analyzing the

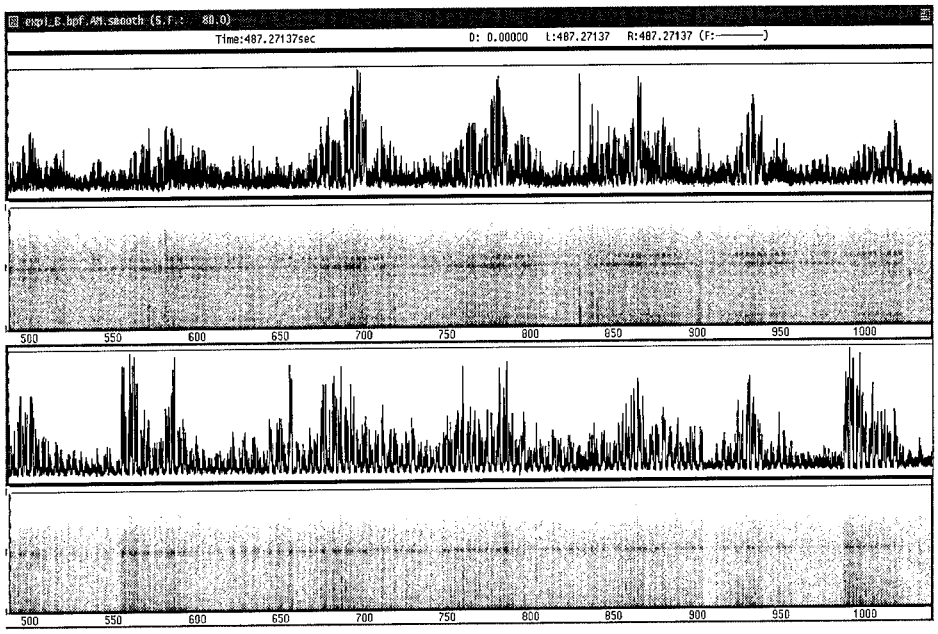


Figure 5: Time-series and spectrogram of the AM on the 1500 Hz component (top two plots) and the 3100 Hz component (bottom two plots). There is correlation between the AM in these bands - see Figure 6. Frequency range is 0-40 Hz.

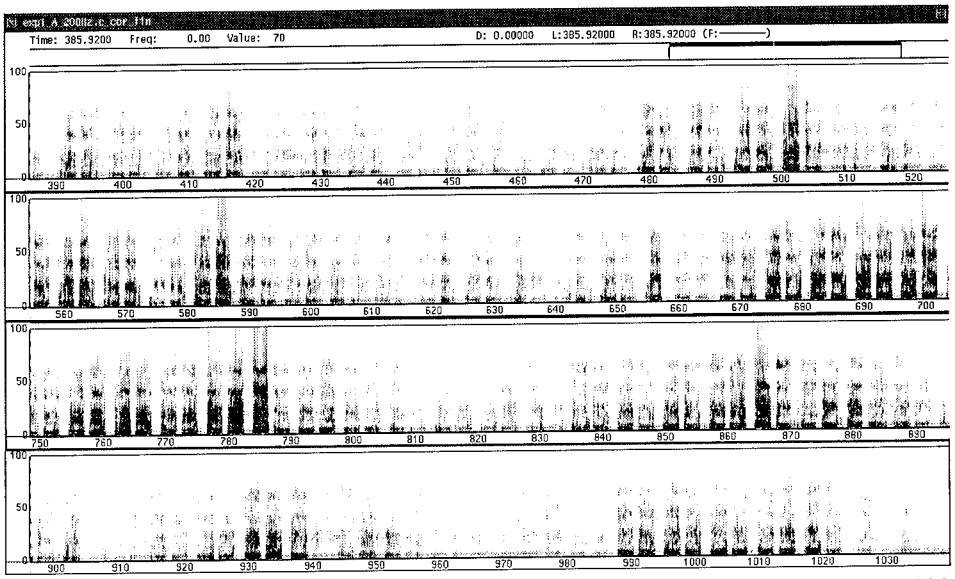


Figure 6: Time-varying cross-spectrum between the AM on the 1500 Hz and 3100 Hz bands of the grinding data. The four panels represent progression from sharp (top), to medium to dull to sharp again after dressing at 910 seconds (bottom). Note the intensity of the cross-spectrum increases as the wheel becomes dull, and then drops again after dressing.

grinding of steel, we collected data from an aluminum grinding operation, and performed the same time-frequency analysis as on the steel data to determine whether time-varying AM was a factor here as well.

As with the steel data, narrow spectral bands coincident with the actual grinding were observed, though at different frequencies than with steel. Also, time-frequency analysis indeed revealed AM present on these frequency components (Figure 7). The fundamental frequency of the AM corresponds to the rotational frequency of the grinding wheel, and, as before, changes as the state of the wheel changes. Thus, it appears that these changes in AM as the wheel dulls are ubiquitous features in grinding processes.

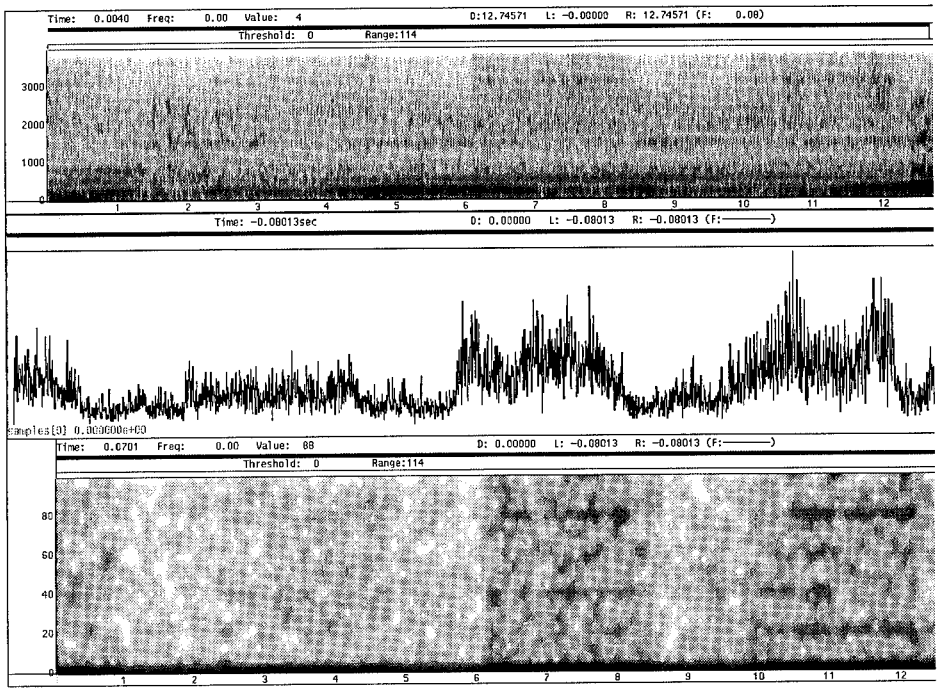


Figure 7: Aluminum grinding data: spectrogram of raw data (top), time series of AM (middle), and spectrogram of AM (bottom).

CONCLUSION: While the short-time spectrum, which is an efficient method for obtaining a gross characterization of the time-varying spectrum of a process, is often effective in revealing features indicative of degradation, the previous examples demonstrate that new techniques can sometimes reveal features that don't appear in the short-time spectrum. For some operations, the early detection of these features may be important for maintaining process quality.

Time-frequency analysis provides a means to accurately identify the changing frequencies that occur with degradation; these spectral changes in turn reflect the state of the process. We have developed new methods for time-frequency analysis [12], [13], [21], and applied these along with conventional methods to the analysis of manufacturing processes. We have found that the new methods can reveal features of degradation (e.g., AM) well before they appear in the spectrum or short-time spectrum.

ACKNOWLEDGMENT: The authors thank Jing Fang for her assistance in generating some of the figures, and for her contributions to the analysis of the drill data.

REFERENCES:

[1] H. Choi and W. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. ASSP*, vol. 37., no. 6, pp. 862-871, 1989.

[2] T. Claasen and W. Mecklenbrauker, "The Wigner distribution - A tool for time-frequency analysis" *Phil. J. Res.*, vol. 35, pp. 217-250, 276-300, 372-389, 1980.

[3] L. Cohen, "Generalized phase-space distribution functions," *J. Math. Phys.*, vol. 7, no. 5, pp. 781-786, 1966.

[4] L. Cohen, "Time-frequency distributions - A review," *Proc. IEEE*, vol. 77, no. 7, pp. 941-981, 1989.

[5] L. Cohen, Time-Frequency Analysis, Prentice Hall, 1995.

[6] L. Cohen and T. Posch, "Positive time-frequency distribution functions," *IEEE Trans. ASSP*, vol. 33., no. 1, pp. 31-38, 1985.

[7] B. Forrester, "Analysis of gear vibration in the time-frequency domain," *Proc. 44th Meeting of the Mechanical Failures Prevention Group*, pp. 225-234, 1990.

[8] J. Jeong and W. Williams, "Kernel design for reduced interference distributions," *IEEE Trans. Sig. Proc.*, vol. 40, no. 2, pp. 402-412, 1992.

[9] R. Koenig, H. Dunn and L. Lacy, "The sound spectrograph," *J. Acoust. Soc. Am.*, vol. 18, pp. 19-49, 1946.

[10] P. Loughlin, J. Pitton and L. Atlas, "An information-theoretic approach to positive time-frequency distributions," *IEEE Proc. Int. Conf. Acous., Speech and Sig. Proc.'92*, vol. V, pp. 125-128, 1992.

[11] P. Loughlin, J. Pitton and L. Atlas, "Bilinear time-frequency representations: new insights and properties," *IEEE Trans. Sig. Proc.*, vol. 41, no. 2, pp. 750-767, 1993.

[12] P. Loughlin, J. Pitton and L. Atlas, "Construction of positive time-frequency distributions," *IEEE Trans. Sig. Proc.*, vol. 42, no. 10, pp. 2697-2705, 1994.

[13] J. Pitton, L. Atlas and P. Loughlin, "Applications of positive time-frequency distributions to speech processing," *IEEE Trans. Sp. Aud. Proc.*, vol. 2, pp. 554-566, 1994.

[14] R. Rohrbaugh, "Application of time-frequency analysis to machinery-condition assessment," *Proc. 27th Asilomar Conf. on Sigs., Syst. and Comps.*, vol. 2, pp. 1455-1458, 1993.

[15] R. Rizzoni and X. C. Chen, "Detection of internal combustion engine knock using time--frequency distributions," *36th Midwest Symp. Circuits and Systems*, vol. 1, pp. 360--363, 1993.

[16] B. Samimy and G. Rizzoni, "Time-frequency analysis for improved detection of internal combustion-engine knock," *Proc. IEEE-SP Intl. Symp. Time-Freq./Time-Scale Analy.*, Philadelphia, PA, pp. 178-181, 1994.

[17] Y. Shin, J-J. Jeon and S. Spooner, "Pseudo-Wigner-Ville distribution and its application to machinery condition monitoring," *Proc. 47th Meeting of the Mechanical Failures Prevention Group*, pp. 235-250, 1993.

[18] J. Taylor, "Phase contributes to accurate diagnostics of rotating machinery," *Proc. 48th Meeting of the Mechanical Failures Prevention Group*, pp. 21-32, 1994.

[19] W. Wang and P. McFadden, "Early detection of gear failure by vibration analysis — I. Calculation of the time-frequency distribution," *Mech. Syst. Sig. Proc.*, vol. 7, no. 3, pp. 193-203, 1993.

[20] W. Williams in "Methods and Applications for Time-Frequency Signal Anaysis," by L. Cohen, W. Williams and M. Amin. Course Notes for University of Michigan Summer Conference, Aug 15-17, 1994.

[21] Y. Zhao, L. Atlas and R. Marks, "The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals," *IEEE Trans. ASSP*, vol. 38, no. 7, pp. 1084-1091, 1990.

Bearing Monitoring Using Reduced Interference Distributions

William J. Williams*
Electrical Engineering and Computer Science Department
Christopher K. H. Koh and Jun Ni
Mechanical Engineering and Applied Mechanics Department
University of Michigan
Ann Arbor MI 48109

January 16, 1995

Abstract

The vast majority of the powerful and effective algorithms in signal processing start with the assumption of stationarity. In addition, the deterministic portion of the signal is often assumed to be composed of complex exponentials which are the solutions to linear time-invariant (LTI) differential equations. These assumptions are often valid enough to yield good results when the signals and systems involved result from idealized engineering design which often assures compliance with these conceptualizations. Many signals do not comply with these assumptions, however, resulting in disappointment when conventional techniques are used. We now have at hand time-frequency (t-f) analyses which can provide new insights into the nature of nonstationary signals. This paper describes some results using Reduced Interference Distributions (RIDs) in the analysis of signals recorded from bearings. It is shown that RID analysis of these signals results in new insights, confirm some earlier hypotheses and have the potential of providing early diagnosis of developing failures.

Key Words: Bearings; Machine monitoring; Time-frequency analysis; Spectral analysis; Time-series analysis

1 Introduction:

Signals of practical interest often do not conform to the requirements of realistic application of Fourier principles which are so useful in many engineering applications. It is often assumed that signals are stationary so that the power spectrum correctly reflects the character of the signal in the frequency domain. Many signals often do not conform to these requirements of "stationarity". Their frequency content changes rapidly with time. Such signals are said to be nonstationary even if the time course of the frequency change is perfectly well known and can be expressed analytically.

A sinusoidal signal which exhibits an increasing (or decreasing) frequency with time is often called a chirp. Such signals can result from systems which are linear, but exhibit time-varying systems parameters. The differential equations describing these systems are not linear time-invariant, but exhibit time-varying coefficients and can properly be called linear time-varying systems.

It has been quite difficult to analyze signals such as chirps satisfactorily using conceptualizations based on stationarity. The spectrogram represents an attempt to apply the Fourier transform (FT) for a short-time analysis window, within which it is hoped that the signal behaves reasonably according to the requirements of stationarity. By moving the analysis window along the signal, one hopes to capture and track the variations of the signal spectrum as a function of time. The spectrogram has many useful properties including a well developed general theory. The spectrogram often presents serious difficulties when it is used to analyze rapidly varying signals, however. If the

*Supported in part by ONR Contract no. N00014-90-J-1654 and NSF Grant BCS 9110571

analysis window is made short enough to capture rapid changes in the signal it becomes impossible to resolve signal components which are close in frequency within the analysis window duration. Watkins [24] has provided an enlightening view of the limitations of the spectrogram when it is used to analyze marine mammal sounds.

The Wigner distribution (WD) has been employed as an alternative to overcome some of the shortcomings of the spectrogram. The WD was first introduced in the context of quantum mechanics [26] and revived for signal analysis by Ville [23]. The WD has many important and interesting properties [7]. It provides a high resolution representation in time and in frequency. In addition, the WD has the important property of satisfying the time and frequency marginals in terms of the instantaneous power in time and energy spectrum in frequency. While the WD certainly provides unambiguous high resolution t-f representations of nonstationary monocomponent signals, such as chirps, its representations of multicomponent signals often are not useful. These representations contain interference (cross-terms) in different regions of the t-f plane resulting from interactions between signal components. Cross-terms are a hindrance to interpretation since they carry redundant information and may obscure primary features of the signal. An excellent discussion on the geometry of interferences has been provided by Hlawatsch [13] and Flandrin [11].

Both the spectrogram and the WD are members of Cohen's Class of Distributions [6]. Cohen has provided a consistent set of definitions for t-f distributions which has been of great value in guiding and clarifying efforts in this area of research. Desirable properties of a distribution and the associated kernel requirements are easily determined using these concepts. A comprehensive review by Cohen [7] provides an excellent overview of t-f distributions and results using them. This latter paper addresses a specific subset of t-f distributions belonging to Cohen's class. These are the time shift and frequency shift invariant t-f distributions. For these distributions a time shift in the signal is reflected as an equivalent time shift in the t-f distribution and a shift in the frequency of the signal is reflected as an equivalent frequency shift in the t-f distribution. The spectrogram, the WD, and the RID all have this property.

Different members of Cohen's class can be obtained by substituting their respective kernel functions into the general formulation. Recently Choi and Williams introduced a new time-frequency distribution, the Exponential Distribution (ED), with kernel $\phi_{ED}(\theta,\tau) = e^{-\theta^2\tau^2/\sigma}$, where σ is a kernel parameter ($\sigma > 0$) [4]. The ED overcomes several drawbacks of the spectrogram and WD, providing representations with high resolution and suppressed interferences. The RID is an improved and generalized version of the ED [27, 28, 30].

The spectrogram has been a useful tool in speech and many other areas. A local spectrum can be found by time windowing a portion of a signal and then Fourier transforming it. This yields the Short Time Fourier Transform (STFT). Squaring the magnitude of this result yields the power spectrum. By moving the center of the window along the signal one obtains a time-varying power spectrum. This is the spectrogram. It can be shown that the spectrogram has a kernel that suppresses cross-terms, but also smears the autoterms. It has a number of other undesirable qualities as well. Its primary asset is that it is non-negative. Daubechies [9] has pointed out that the STFT bears a relationship to the wavelet transform. In the wavelet transform, the window varies with frequency. It is thought by some that the wavelet transform escapes from the well-known uncertainty principle which demands a tradeoff between window duration and frequency resolution. This is not true. The wavelet transform scales the window with frequency such that it is more narrow at high frequncies and more broad at low frequencies, preserving the uncertainty trade-off (often in a desirable manner, but sometimes to the detriment of the analysis). In fact, there is a wavelet analog to the spectrogram, called the scalogram [18]. There are several members of Cohen's class (the RID among them) that escape the rigidity of this trade-off and provide high resolution for a variety of signal components.

2 Time-Frequency Distributions:

Some theoretical underpinnings for understanding t-f distributions will be presented in this section. We will show the important issues can be readily understood in a manner similar to the relationships between a signal, its FT, and a filter applied to that signal.

2.1 Ambiguity Function Relationships:

The key to understanding t-f relationships and manipulations is a thorough understanding of the ambiguity domain. Let $Z(w)$ be the FT[1] of the signal $z(t)$:

$$Z(\omega) = F[z(t)] = \int f(t)e^{-j\omega t}dt \tag{1}$$

and

$$z(t) = F^{-1}[Z(w)] = \frac{1}{2\pi}\int Z(\omega)e^{j\omega t}d\omega \tag{2}$$

Let $R_z(t,\tau)$ be the *instantaneous autocorrelation* of a complex signal $z(t)$, defined as:

$$R_z(t,\tau) = z(t+\tau/2)z^*(t-\tau/2) \tag{3}$$

The Wigner distribution of $z(t)$ is defined as the FT of $R_z(t,\tau)$ with respect to the lag variable τ.

$$W_z(t,\omega) = F_\tau[R_z(t,\tau)] \tag{4}$$

Similarly, but with a different physical meaning, the symmetrical ambiguity function (AF) is defined as the Inverse Fourier transform (IFT) of $R_z(t,\tau)$ with respect to the first variable.

$$A_z(\theta,\tau) = F_t^{-1}[R_z(t,\tau)] \tag{5}$$

Thus, $W_z(t,\omega)$ and $A_z(\omega,\tau)$ are related by the two-dimensional (2-D) FT.

$$W_z(t,\omega) = \iint A_z(\theta,\tau)e^{-j(t\theta+\omega\tau)}d\theta\, d\tau \tag{6}$$

These relationships may be combined with Eq. (1) to show that $C_z(t,\omega;\phi)$ may be found by

$$C_z(t,\omega;\phi) = \iint \phi(\theta,\tau)A_z(\theta,\tau)e^{-j(t\theta+\omega\tau)}d\theta\, d\tau \tag{7}$$

Thus, while $W_z(t,\omega)$ may be found from the symmetric ambiguity function by means of a double FT, any member of Cohen's Class of Distributions may be found by first multiplying its kernel, $\phi(\theta,\tau)$, by the symmetric ambiguity function and then carrying out the double FT. The generalized ambiguity function, or $\phi(\theta,\tau)A_z(\theta,\tau)$ [8] is a key concept in t-f which aids one in clearly seeing the effect of the kernel in determining $C_z(t,\omega;\phi)$.

2.2 The Exponential Distribution:

The ED is an attempt [4] to improve on the WD. Its kernel is designed to encompass most of the auto-terms in the ambiguity domain, while placing cross-terms in its attenuation region. The ED kernel is designed to have a unity value along the τ and θ axes in the ambiguity domain and to fall off rapidly away from these axes. The unity value along the axes ensures many desirable properties of the WD, but the attenuation away from the axes reduces the size of the cross-terms. The σ parameter in the kernel of the ED may be varied over a range of values to obtain different trade-offs between cross-term suppression and high auto-term t-f resolution. In fact, as σ becomes very large the ED kernel approaches the WD kernel. This provides the best resolution but the cross-terms become large and approach WD cross-terms in size.

[1] The range of integrals is from $-\infty$ to ∞ throughout this paper.

2.3 Design Procedures for Effective RID kernels:

Incorporating the idea of interference reduction employed in the ED) with other desirable properties , we have introduced a new class of t-f distributions, called RIDs. Requirements for the RID and the RIDs properties are quite similar to the WD. The RID shares many of the desirable properties of the WD, but also has the important reduced interference property as well [14, 30]. The ED is a precursor to the RID class of distributions.

The RID kernel should follow

$$|\phi(\theta,\tau)| \ll 1 \quad \text{for} \quad |\theta\tau| \gg 0.$$

We have proposed the following procedure to design a RID kernel [14].

Step 1. Design a primitive real-valued function $h(t)$ that satisfies the following:

H1: $h(t)$ has unit area, i.e., $\int h(t)dt = 1$.

H2: $h(t)$ is a symmetrical function of time, i.e., $h(-t) = h(t)$.

H3: $h(t)$ is time-limited on [-1/2, 1/2], i.e., $h(t) = 0$ for $|t| > 1/2$.

H4: $h(t)$ tapers smoothly towards both ends so that its frequency response eventually declines and has little high frequency content. That is, $|H(\theta)| \ll 1$ for $|\theta| \gg 0$, where $H(\theta)$ is the FT of $h(t)$.

Step 2. Take the FT of $h(t)$, i.e.,

$$H(\theta) = \int h(t)e^{-j\theta t}dt.$$

Step 3. Replace θ by $\theta\tau$ in $H(\theta)$:

$$\phi(\theta,\tau) = H(\theta\tau).$$

This design procedure is sufficient to guarantee many important properties. There is a great deal of flexibility in choosing the primitive $h(t)$ so that desirable features of the signal are enhanced and cross-terms and noise can be diminished. Conceptually, the idea is rather like choosing a filter impulse response, $h(t)$, so that a desirable filter, $H(\omega)$, will result. However, now h(t) expands into a two-dimensional filter in τ and θ. One may use an $h(t)$ that yields a good filter in the conventional sense, e.g., maximally flat or equal ripple. The RID kernel will also inherit these desirable properties.

It is beyond the scope of this paper to describe discrete computational techniques for RIDs. We note, the Binomial distribution [30] is a particularly attractive discrete form which can be computed very efficiently. It is a member of a discrete class of t-f distributions which can be designed to reduce cross-terms and avoid aliasing [15]. With modern computer hardware it is likely that such distributions can be computed in real-time for many real-world applications.

2.4 Applications of Cohen's Class:

There are many applications of Cohen's class of distributions. It is beyond the scope of this article to provide a comprehensive treatment of their use. The spectrogram has been used in many applications. We note that the distribution of Zhao, Atlas and Marks[32] is a member of Cohen's class which had early application to machine monitoring. The RID was first applied in biological signal applications [31, 22, 25] where it has proved particularly valuable in studying signals with mixed tonal, impulsive and time-frequency concentrated components. The RID has also been applied to a number of mechanical signals in the automotive and manufacturing areas. This paper presents the opportunity to report some of those results and indicate the promise therein.

3 Vibration Monitoring of Machinery:

Vibration monitoring is widely used to monitor the condition of a process or equipment. An issue to be considered is the location and proximity of the transducer to the origin of the vibration. Although it is ideal to be able to measure the vibration directly at the source, it may not be possible

to realize this in practice due to the design and construction of the machinery. In many cases, there is a multitude of transmission paths between the source and the transducer, which transfer functions may severely modulate the final signal reaching the transducer. In addition, the vibration generated by other moving parts of the machinery may be superimposed on the signal of interest as noise. The signal-to-noise (SNR) ratio, if low enough, may render the vibration readings useless.

A variety of techniques based on time and frequency domain principles [2, 10] have been widely used to extract information from the vibration signal. In essence, the information yielded by these methodologies is encoded in a set of performance parameters believed to be most representative of the condition of the machinery. Although the parameters are usually very accurate in representing the state of the machinery, it does not often provide an intuitive grasp of the prevailing situation, nor does it offer any insight into the time-varying nature of these signals. The "waterfall" plot, which is nothing more than a three-dimensional representation of the spectrogram, has long been considered as the de facto method of studying the time-varying characteristic of many process signals in industry. However in recent years, the development of time frequency theory in the signal processing field has heralded a new wave of machinery monitoring methodologies based on the Wigner distribution [33] and Wavelet transforms [16]. In the following section, vibration monitoring of bearing using the RID will be used to mark yet another milestone in the successful application of time-frequency analysis of nonstationary process signals.

3.1 Bearing Vibration Monitoring:

Bearings are found in almost every type of machinery. Although they are relatively inexpensive and maintenance free under normal operating condition, several factors from lubrication [5] to changing load conditions [19] may severely shorten its service life. For this reason, the study and early detection of bearing failure have attracted much attention from both industry and academia [1, 3, 5, 17, 20, 21]. There are basically two main approaches to this problem, (i) the study of lubricating oils and greases, and (ii) vibration monitoring. Although the value of tribological analysis is acknowledged, it does not fall into the context of this paper and only vibration monitoring will be discussed.

The range of frequencies and amplitude of vibrations emitted by bearings at different stages of its life cycle is well-documented [1], [20, 21]. Bearings can also fail gradually or abruptly; and two identical bearings may have different failure modes and life cycles depending on their operating conditions.

There were two motivations for this study. The first was the need to improve on existing failure detection techniques. Since the success of detection algorithms depends largely on how well the mechanism of failure is understood, the RID presented an opportunity to "look" beyond the limited capabilities of the spectrogram. The other motivation was the advancement made in material science, where in recent years, ceramics have rapidly gained popularity as an alternative to steel. The ball-bearing industry is no exception. The results in the next section is probably the first known analysis of vibration of ceramic bearings. The time-frequency representation of the vibration signals is currently being used to assist in the study of failure modes in ceramics.

4 Vibration Signals from Defective Bearings:

When a ball rolls over a localized defect, a shock pulse is generated. This shock pulse is a wide band impulse and of very short duration in comparison with its repetition rate. This repetition rate is also known as the characteristic defect frequency or impact rate. The amplitude of the pulse is a measure of the shock intensity and hence dependent on the rotational speed of the rotor, the load conditions and the size of the defect. The wide band energy of the shock pulse often sets off a number of resonances in the bearing system and neighboring structure. More than one resonance may be excited and it is likely to find several natural frequencies falling within the same range of the spectrum.

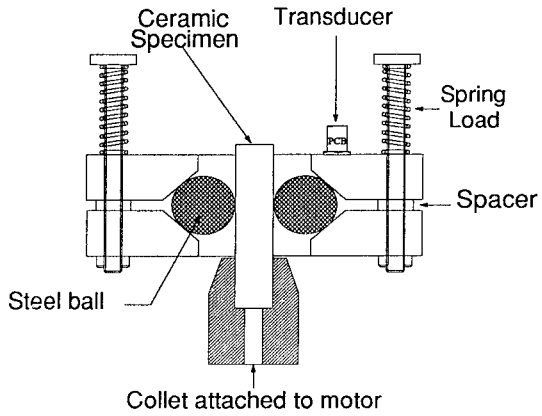


Figure 1: Schematic of fatigue tester loading assembly.

The detection of bearing failure by frequency spectrum analysis is confined to magnitude and impact frequency information. [21] showed that there is a difference between the magnitude and frequency of surface-induced vibration (due to the manufacturing process) and defect-induced vibration. However very little can be deduced about the dynamic behavior of the bearings under both conditions from the frequency spectrum.

5 Experimental Protocol:

The experimental rig used in this experiment is a bench tester designed specifically for testing the rolling contact fatigue life of bearing materials. This is a more economical and efficient method of studying the properties of bearing material using a test specimen with a simple cylindrical geometry with the flexibility of varying the bearing load. A full description of this test rig can be found in [12]. The specimen rod measuring $\frac{3}{8}$ in. diameter and 3 in. in length, is held vertically by a precision collet connected to a direct drive electric motor operating at 3600 rpm. Three steel balls of radius $\frac{1}{2}$ in., equally spaced by a bronze retainer, provide the radial loading on the cylindrical rod through the thrust loading of the spring assembly schematically shown in Figure 1. The spring load can be preset to any desired load by adjusting the amount of compression.

Two accelerometers are mounted at right angles to each other to measure the horizontal and vertical vibrations from the steel balls. When a preset vibration level is detected, indicating a surface crack or fatigue spall, the motor is stopped automatically and the test is concluded. An hour meter records the duration of the test.

Lubrication is applied throughout the test using a drip feed. The steel balls used in the test have been abrasive blasted with a surface roughness of 3.5 μin. The number of stress cycles per revolution between the balls and specimen rod is 2.389 and is a function of motor speed, cup angle, ball diameter, rod diameter and contact diameters. The hertz stress is 786 000 psi based on an individual spring load of 57 lb.

The accelerometer used in this test is a PCB 308B2. The transducer signals are amplified by the Kistler 5004 charge amplifiers before terminating at the input channels of a TEAC RD135T digital audio tape (DAT) recorder. The recording bandwidth for 2-channel recording is 48 kHz. The data was further downsampled to 12 kHz for analysis.

6 Discussion of the Results:

Accelerometer signals were examined at the beginning of the test and just before failure in order to examine the character of these signals in "Normal" and "Failing" conditions. It is often difficult to see the precursors of failure in the time series, or in the power spectrum of these signals, though some hints are available. The RMS value of the signal rises rapidly at failure, but it is difficult to predict failure far ahead of time using this measure. Impulsive components begin to appear in the signal near failure, but these components are difficult to sort out from other signal components. Figure 2 shows time series, power spectra and time-frequency distributions for a bearing in the Normal condition and in the Failing condition. One can see that the time series becomes more impulsive near failure and there are differences in the power spectra as well. The Binomial distribution provides a dramatic picture of the differences in signal structure in the two conditions, however. The Binomial distribution obtained during the Normal state of the bearing shows some structure, but the energy patterns as a function of time and frequency are somewhat variable. The dominant pattern is a horizontal orientation of energy consistent with a generally tonal structure. As the bearing approaches Failure, there is an increase in the vertical structuring of the Binomial distribution as well, indicating increasing impulsive components. Energy becomes more concentrated into discrete components in general during failure. The Binomial results are obviously much more information rich with regards to the dynamics of the signal components than is the case for either the time series or the power spectra.

The Wigner distribution result is shown for comparison in Figure 2. One can see that there is much more interference in this case, making visual interpretation more difficult. Figure 3 shows the result for two different spectrograms, one with a short time window and the other with a long time window. Here, one can easily appreciate the liabilities of the spectrogram. The time-frequency representation is badly blurred in both cases. The short window result emphasizes the impulsive components and the long window result emphasizes the tonal components, but both are badly blurred. The full detail available in the time-frequency distributions of Figure 2 are difficult to appreciate in the gray scale images. Two additional time series are derived from the results shown in Figure 2 and are shown in Figure 3 as well. A slice along time at a particular frequency (4500Hz as shown by the dashed line in Figure 2) for both the Binomial and Wigner distributions shows some of the detail available. One can see that the impulsive behavior is captured very well by the Binomial distribution, but is almost completely obscured by interference in the Wigner result.

These impulses are much clearer in the Binomial slice than in the original time series. The time between these impulses was measured and was found to be very consistent with the idea that the impulses are produced by the balls hitting small defects. There are approximately 2.4 of these events per revolution, which is quite consistent with theoretical value of 2.389 stress cycles per revolution [12]. There are many other fascinating details in the time-frequency results which need more testing to appreciate. The results shown in this paper are for only one bearing. The vertical accelerometer results are similar to the horizontal results, but exhibit some important differences as well. It is not possible to discuss more subtle effects which have been observed within the scope of this paper. In general, nearing Failure, the time-frequency structure becomes more ordered, becoming arranged in a more horizontal-vertical structure. This indicates a stronger contribution from impulsive and tonal components as opposed to the more disordered "noise-like" structure of the Normal state. Objective measures which take advantage of these observations are being developed at this writing. One possible measure is based on Rényi Generalized Information [29]. This is a method for determining the uncertainty inherent in time-frequency distributions.

7 Conclusions:

It is not the purpose of this paper to introduce a new technique for failure detection in bearings. Rather, it is to present an innovation to the study of the dynamic behavior of bearings from the time-varying characteristics of the vibration generated. The limitations of the spectrogram and "waterfall" plot to adequately represent time-varying signals is no longer a hindrance to the study of the dynamic changes in manufacturing processes and machinery. Although time and frequency domain techniques will continue to play a dominant role in process monitoring, the development of

time-frequency theory has paved the way for new methods of process monitoring. Detection algorithms based on the time-frequency properties of signals are beginning make their mark in many applications and there appears to be a great deal of promise in this application as well.

The authors thank the NTN Technical Center (USA) Inc. at Ann Arbor for their assistance and the use of the data presented in this paper.

References

[1] Berggren, J. C., "Diagnosing Faults in Roller Element Bearings," *Vibrations*, vol. 4, no. 1, pp. 5-14, March 1988.

[2] Braun, S., "Mechanical Signature Analysis : Theory and Applications," *Academic Press*, 1986.

[3] Ching, L., "On-line Bearing Condition Monitoring," *PhD Thesis*, University of Wisconsin-Madison, 1986.

[4] Choi, H. I., and Williams, W. J., "Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernels," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-37, no. 6, pp. 862-871, 1989.

[5] Choy, F. K., Braun, M. J., and Hu, Y., "Nonlinear Transient and Frequency Response Analysis of a Hydrodynamic Journal Bearing," *Trans. of the ASME, J. Tribology*, vol. 114, pp. 448-54, 1992.

[6] Cohen, L., "Generalized Phase-Space Distribution Functions," *J. of Math. Phys.*, 7, pp. 781-786, 1966.

[7] Cohen, L., "Time-Frequency Distributions - A Review," *Proc. IEEE,* vol. 77, no. 7, pp. 941-981, July 1989.

[8] Cohen, L., and Posch, T.E., "Generalized Ambiguity Function", *IEEE Int. Conf. Acous., Speech, Sig. Proc.*, pp. 1033-1036, 1985.

[9] Daubechies, I., "The Wavelet Transform, Time-Frequency Localization and Signal Analysis", *IEEE Trans. Information Theory,* vol. IT-36, no. 5., pp. 961-1005, 1990.

[10] Dimarogonas, A. D., and S.Haddad, "Vibration for Engineers," *Prentice Hall*, 1992.

[11] Flandrin, P., "Some Features of Time-Frequency Representations of Multicomponent Signals," *IEEE Int. Conf. Acous., Speech, Sig. Proc.*, vol. 41B, pp. 4.1-4.4, 1984.

[12] Glover, D., "A Ball-Rod Rolling Contact Fatigue Tester," *American Society for Testing and Materials*, Standard Technical Publication 771, pp. 107-124, 1982.

[13] Hlawatsch, F., "Interference Terms in the Wigner Distribution," in *Digital Signal Processing - 84*, V. Cappellini and A. Constantinides, Eds., pp. 363-367, North-Holland, 1984.

[14] Jeong, J. and Williams, W. J., "Kernel Design for Reduced Interference Distributions," *IEEE Trans. Sig. Proc.*, vol. 40, no. 2, pp. 402-412, 1992.

[15] Jeong, J., and Williams, W. J., "Alias-free Generalized Discrete-time Time-frequency Distributions," *IEEE Trans. Sig. Proc.*, vol. 40, no. 11, pp. 2757-2765, 1992.

[16] Li, C. J., and Ma, J, "Bearing Localized Defect Detection Through Wavelet Decomposition of Vibrations," *Sensors and Signal Processing for Manufacturing*, Winter Meeting of the ASME, pp. 187-196, 1992.

[17] Nishio, K., et al, "An Investigation of the Early Detection of Defects in Ball Bearings by Vibration Monitoring," *ASME Design Engin. Tech. Conf.*, May 1979.

[18] Rioul, O., and Flandrin, P., "Time-Scale Energy Distributions: A General Class Extending Wavelet Transforms," *IEEE Trans. Sig. Proc.,* vol. 40, no. 7, pp. 1746-1757, 1992.

[19] Scheithe, W., "Better Bearing Vibration Analysis," *Hydrocarbon Processing (International edition),* July 1992. '92

[20] Su, Y-T., Sheen, Y-T., and Lin, M-H., "Signature Analysis of Roller Bearing Vibrations: Lubrication Effects," *Proc. Inst. Mech. Eng. Part C, J. of Mech. Eng. Sc.,* vol. 206, no C1, pp. 193-202, 1992.

[21] Su, Y-T., and Sheen, Y-T., "On the Detectability of Roller Bearing Damage by Frequency Analysis," *Proc. Inst. Mech. Eng. Part C, J. of Mech. Eng. Sc.,* vol. 207, no C1, pp. 23-32, 1993.

[22] Tyack, P. L., Williams, W. J. and Cunningham, G. S., "Time-frequency Fine Structure of Dolphin Whistles," *2nd Intl. Conf. on Time-frequency and Wavelet Techniques,* pp. 18-20, Victoria, B. C., Oct. 1992.

[23] Ville, J., "Theorie et Applications de la Notion de Signal Analytique," *Cables et Transmissions,* vol. 20A, pp. 61-74, 1948.

[24] Watkins, W. A. "The Harmonic Interval Fact or Artefact in Spectral Analysis of Pulse Trains," *Marine Bio-acoustics,* vol. 2, pp. 15-43, 1966.

[25] Widmalm, S. E., Williams, W. J. and Zheng, C. S, "Time frequency distributions of TMJ sounds," *J. of Oral Rehabilitation,* Vol. 18, pp. 403-412, 1991.

[26] Wigner, E., "On the Quantum Correction for Thermodynamic Equilibrium," *Phys. Rev.,* vol. 40, pp. 749-759, 1932.

[27] Williams, W.J. and Jeong, J., "New Time-frequency Distributions : Theory and Applications", *IEEE Int. Symp. Circuits and Systems,* vol. 2, pp. 1243-1247, May 1989.

[28] Williams, W.J. and Jeong, J., "New Time-frequency Distributions for the Analysis of Multi-component Signals," *Advanced Algorithms and Architectures for Signal Processing IV : SPIE Proceedings*, vol. 1152, pp. 483-495, 1989.

[29] Williams, W. J., Brown, M. L., and Hero, A. O., III, "Uncertainty, Information and Time-Frequency Distributions," *Advanced Signal Processing Algorithms, Architectures and Implementations II,* SPIE vol. 1566, pp. 144-156, 1991.

[30] Williams, W. J., and Jeong, J., "Reduced Interference Time-Frequency Distributions", Chapter 3 in: *Time-Frequency Signal Analysis* B. Boashash, Ed., Longman and Cheshire - J. W. Wiley, 1992.

[31] Zaveri, H. P., Williams, W. J., Iasemidis, L. D., and Sackellares, J. C., "Time-Frequency Representations of Electrocorticograms in Temporal Lobe Epilepsy", *IEEE Trans. on Biomed. Engineering,* vol. 39, no. 5, pp. 502-509, 1992.

[32] Zhao, Y., Atlas, L., and Marks, R., "The Use of Cone-Shaped Kernels for Generalized Time-Frequency Representations of Nonstationary Signals," *IEEE Trans. Acoust. Speech, Sig. Proc.,* vol. ASSP-38, no. 7, pp. 1084-1091, 1990.

[33] Zheng, K., and Whitehouse, D. J., "The Application of the Wigner Distribution Function to Machine Tool Monitoring," *Proc. Inst. Mech. Eng. Part C, J. of Mech. Eng. Sc.,* vol. 206, no C4, pp. 249-264, 1992.

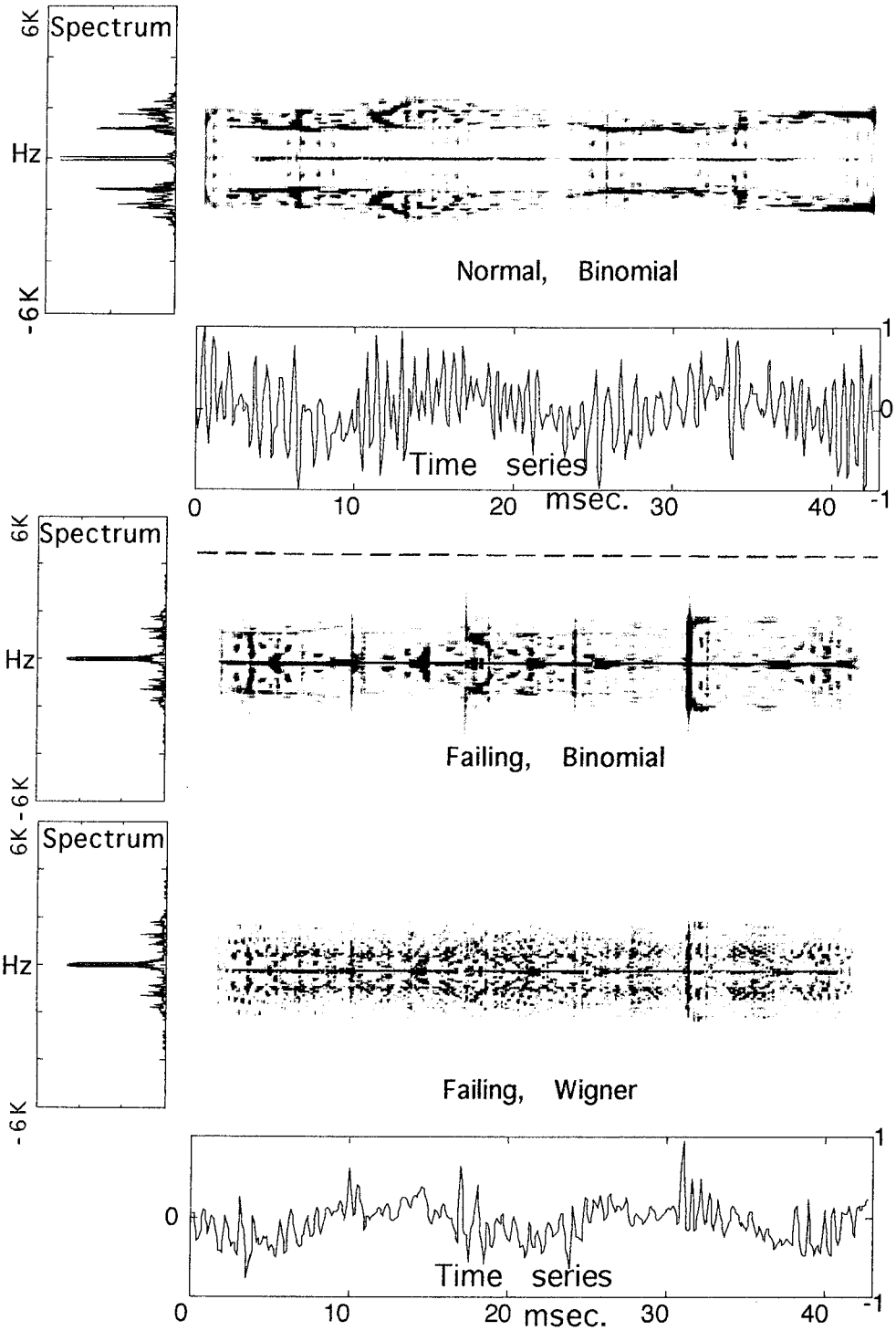


Figure 2. A comparison of a normal bearing and failing bearing using the Binomial and Wigner distributions.

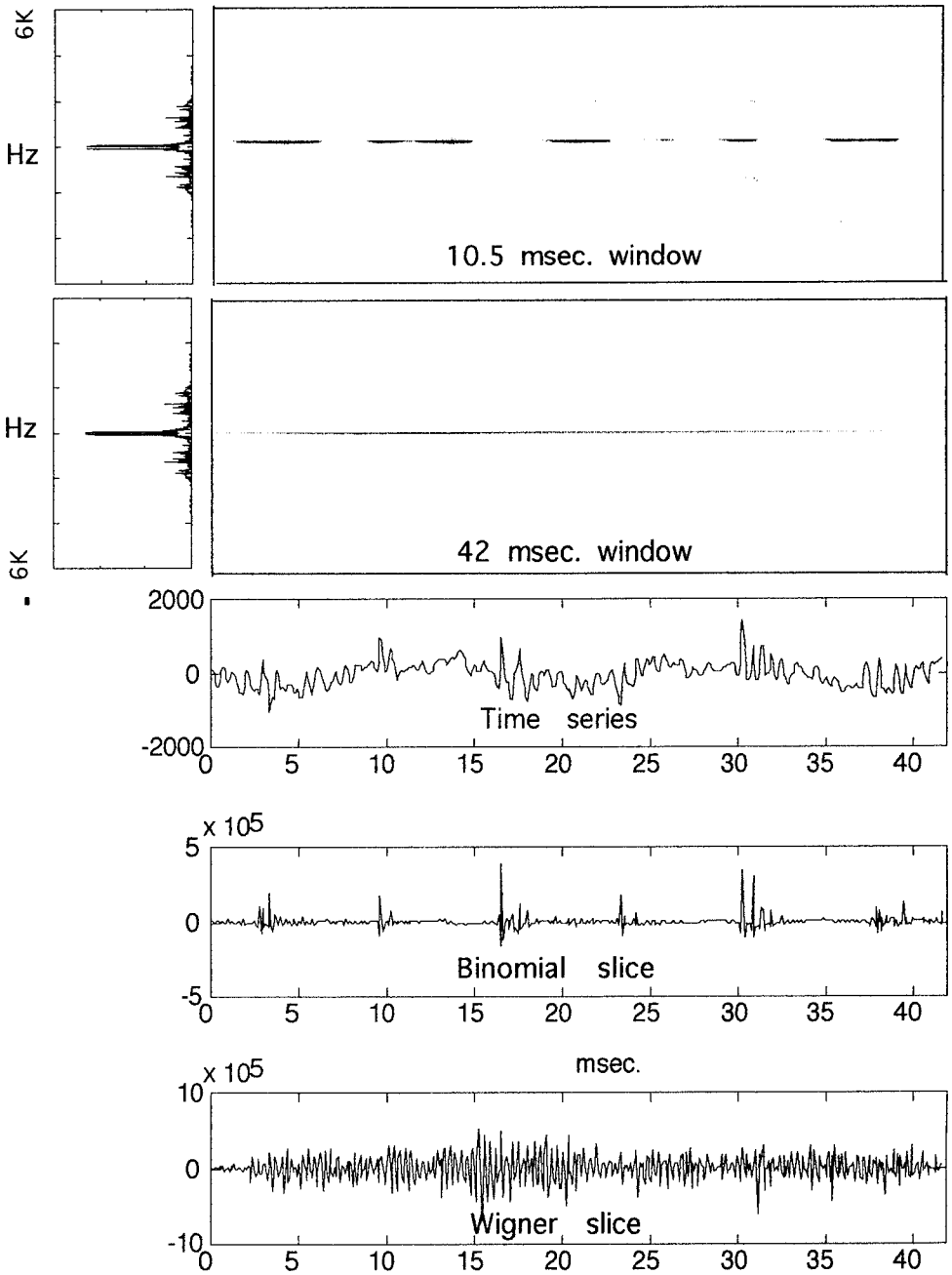


Figure 3. Failing bearing. A comparison of a short and long window spectrogram for the failing bearing. Time series and power spectra are shown. Three time series shown: the signal; a frequency slice for theBinomial distribution; a frequency slice for the Wigner distribution at 4.5KHz.

TIME-FREQUENCY ANALYSIS OF CRANKSHAFT SPEED DATA IN A LAMBORGHINI V-12 ENGINE

Bahman Samimy
Giorgio Rizzoni

Department of Mechanical Engineering and Center for Automotive Research
The Ohio State University
206 West 18th Avenue
Columbus, OH 43210
(614) 292 - 3331
email: samimy.4@osu.edu and rizzoni.1@osu.edu

Abstract

This paper presents the application of some modern signal processing methods to the analysis of angular velocity signals in a rotating machine for diagnostic purposes. The signal processing techniques considered in this paper include: Short-time Fourier transform (STFT) and joint time-frequency representation (TFR).
Various algorithms are employed to process shaft angular velocity data measured from an internal combustion engine, with the intent of analyzing and detecting engine misfire. The results of these analyses show that these algorithms have potential for on-board diagnostic application in passenger and commercial vehicles, and more generally for failure detection of other classes of rotating machines.

Key Words: Engine Misfire; Misfire Detection; Time-Frequency Analysis

1 Introduction

In recent years considerable interest has been placed on the detection of engine misfire. As part of the California Air Resources Board on-board diagnostics regulations for 1994 model year vehicles [1], misfire should be monitored continuously by the engine diagnostic system. It is expected that the next generation of on-board diagnostics regulations will demand monitoring of partial misfire as well. This paper describes the application of a misfire detection method based on a measurement of crankshaft angular velocity to a high-performance automotive engine.
The development of the misfire detection algorithm considered here starts from the Fourier analysis based method originally proposed and by Ribbens and Rizzoni [2, 3, 7], and

also studied and improved upon by others ([4], for example). This paper illustrates an extension of their method using time-frequency analysis techniques to a 12-cylinder high performance engine (5.7-liter Lamborghini engine).
It should be noted that in a 12-cylinder engine it may be very difficult to identify as many as two misfiring cylinders even for an experienced test driver. Thus, the importance of this particular application for the engine manufacturer goes beyond meeting the CARB and EPA requirements.
This paper summarizes the signal processing algorithms and the experimental work performed to carry out an initial validation of the method on a high-performance twelve-cylinder engine connected to an eddy-current brake. In this first phase of the research the experimental tests were carried out considering one or few misfires in a cycle. Work is currently in progress to perform chassis dynamometer and road tests. The work reported in this paper is still in the research stages. Further work must be done before it can be implemented in a production unit.

2 Method

It is well known that instantaneous fluctuations of the indicated torque produced by each cylinder as a result of the combustion pressure force in the cylinders cause fluctuations in the angular velocity of the crankshaft. The references (e.g., [2]) describe a method based on the dynamic relationship between torque and angular velocity fluctuations in a multi-cylinder spark ignition (SI) engine. Various researchers have shown that the torque waveform can be reconstructed from the crankshaft angular velocity signal by means of this relationship, and consequently that the reconstructed torque waveform can be used to obtain an indicator of engine performance, and in particular of engine misfire. This fact has been proven most convincingly by Ribbens and Park in their recent works [5, 6], which describe extensive road testing of the method for engines equipped with V-6 and 6-in-line engines. The method was recently tested on a Lamborghini V-12 engine with favorable, though not conclusive results [9]. The works mentioned above are based on a conventional Fourier analysis of the crankshaft angular velocity data. In the present paper, we explore the possibility of using time-frequency analysis to improve on the detection of engine misfire, and propose a misfire detection method based on time-frequency domain analysis of the crankshaft angular velocity signal. The idea underlying the method is that the occurrence of a misfire will cause a brief transient to occur in the signal. This transient will manifest itself in the form of a shift in the energy of the signal towards lower frequencies (sub-harmonics of the engine firing frequency), due to the subcyclic nature of the misfire event. The duration of these event may be limited to a fraction of one engine cycle if a single cylinder misfires only once; therefore the ability to perform time-frequency localization may be essential in the development of a highly sensitive misfire detector. A note is in order at this point regarding the domain of the signal and of its transform: the engine angular velocity signal is quasi periodic, with period equal to one engine cycle (two crankshaft revolutions). Therefore, it is natural to sample the signal in the crank angle domain, and the corresponding frequency variable is *orders of engine cycle.* As a consequence, time-frequency analysis applied to the above mentioned data is in fact better denoted crank angle - order of engine cycle analysis, or, from the symbols used to denote these variables, *theta - lambda* analysis.

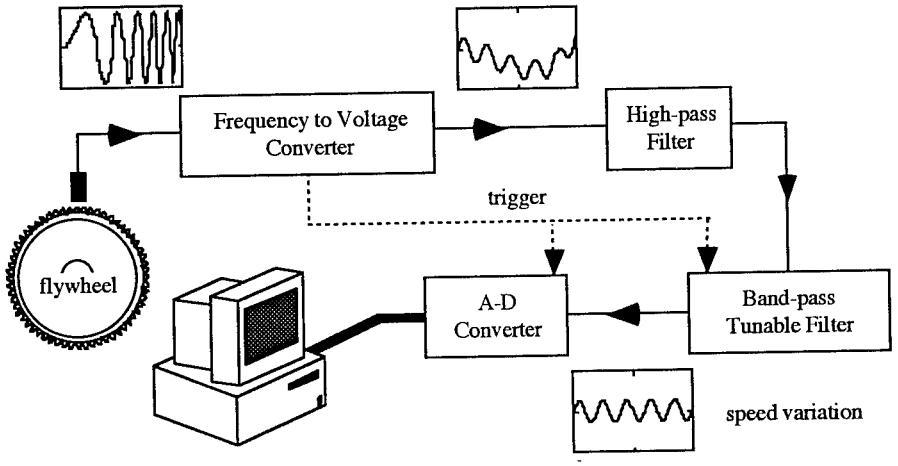


Figure 1: Data acquisition scheme

2.1 Data acquisition system

To perform the analysis described in the preceding section, flywheel angular velocity data were acquired from a Lamborghini V-12 engine; the data were supplied by Lamborghini Automobili, and processed at ENEA with the collaboration of the Istituto di Macchine of the Universita' di Bologna; the data acquisition procedures and the experiments are described in greater detail in [9]. Figure 1 depicts the instrumentation used in the experiments. Figure 2 depicts a section of an acceleration-decelaration test from which the data analyzed in this paper were extracted. The data was originally acquired with a crank angle resolution of 210 samples per engine cycle, or approximately one sample for each 3.4 degrees of crankshaft rotation. This resolution is greater than what is needed for misfire detection, and the original data was decimated by a factor of 5 to reduce the data processing requirements. The resulting resolution is approximately one sample every 17.1 crank angle degrees.

3 Time-Frequency Representations

3.1 Time and Frequency Domains

The Fourier transform provides a one-to-one relationship between the time domain and frequency domain representations of a signal:

$$X(f) = \int_{-\infty}^{+\infty} x(t)exp(-j2\pi ft)dt$$
$$x(t) = \int_{-\infty}^{+\infty} X(f)exp(-j2\pi ft)df \quad (1)$$

It is well known that time and frequency domain analysis provide two complementary perspectives on the energy content of a signal. However, the Fourier transform does not permit simultaneous analysis of the two domains: although $X(f)$ provides us with spectral components of a signal $x(t)$, it does not provide us with the time localization of spectral components.

The Time-Frequency representation (TFR) of a signal characterizes the signal in the time-frequency plane by simultaneously representing the signal energy as a joint function of time and frequency; using such techniques a more revealing picture of the temporal localization of the signals' spectral content may be obtained. Time-Frequency representations can be used for non-stationary signals in order to see how a signal's spectrum change in time; they can also be used for signal detection or serve as basis for coding and synthesis. In this paper we are interested in exploiting joint time-frequency information for the purpose of detecting occasional misfire events in an internal combustion engine.

In general, there are two kind of TFRs, namely linear and quadratic, or bilinear. The more common linear representations are the short-time Fourier transform (STFT) and the wavelet transform (WT). Often used quadratic representations are: the Wigner distribution (WD); the ambiguity function (AF); and various classes of quadratic time-frequency representations. The idea of a time-frequency representation is to map a one dimensional signal which is function of time, x(t), to a two-dimensional function of time and frequency, i.e., $T_x(t, f)$. This is conceptually similar to a musical score with time running along one axis and frequency along the other axis [11]. Time-frequency representations have been applied to identification, analysis, estimation of non-stationary or time-varying signals. Three-dimensional plots of TFR surfaces can be used to analyze how the spectral content of a signal varies with time, and time-frequency domain based detection method can assist in identifying non-stationary events, such as failures.

3.2 Linear Time-Frequency Representations

It is well known that Fourier transform does not show the time localization of frequency components. However, the time localization can be obtained by pre-windowing the signal x(t) into short segments. The short-time Fourier transform (short-time spectrum, STFT) of a signal is defined as follows:

$$STFT_{x_W}(t, f) = \int x(t_1)w^*(t_1 - t)exp(-j2\pi ft_1)dt_1 \quad (2)$$

This means that STFT at time t is the Fourier transform of the signal $x(t_1)$ multiplied by a shifted analysis window $w^*(t_1 - t)$ centered around t. Since multiplication by a relatively short window $w^*(t_1 - t)$ suppresses the signal $x(t_1)$ outside the analysis window, the STFT is a local spectrum of $x(t_1)$ around the analysis time t. It is evident that STFT is linear and is complex valued in general. Also, it is evident that the STFT of a given signal is highly influenced by the choice of the window. The single biggest limitation in the use of STFT is the inherent trade off between time and frequency resolution [11].

3.3 Bilinear Time-Frequency Representations

In recent years, a considerable amount of work has been done in the area of time-frequency representations (TFRs). the tutorial papers by Cohen [10] and Hlawatsch [11] provide an excellent review of the subject. A special class of TFRs which has received a considerable amount of attention in the area of signal processing is the class of representations which satisfy the basic property of *time-frequency shift invariance* [11]: if the signal is delayed in time and/or modulated in frequency, then its TFR will be shifted by the same time delay and/or modulation frequency.

$$\tilde{x}(t) = x(t - t_0)e^{j2\pi f_0 t} \Rightarrow T_{\tilde{x}}(t, f) = T_x(t - t_0, f - f_0) \tag{3}$$

Many of these representations have been derived in different fields using different approaches with the aim of understanding what a time-varying spectrum is. Cohen [10] pointed out that there is an underlying structure among all these representations, and formulated a unified approach to study their properties. Each member of the class shares the general form of

$$T_x(t, f; \Phi) = \frac{1}{4\pi^2} \iiint \Phi(\mu, \tau)x(\mu + \frac{\tau}{2})x^*(\mu - \frac{\tau}{2})e^{-j\xi t - j2\pi f t + j\xi\mu} d\mu d\tau d\xi \tag{4}$$

where $x(\mu)$ is the time-domain signal, $x^*(\mu)$ is its complex conjugate and $\Phi(\xi, \tau)$ is a so-called kernel function which characterizes the particular representation. More recently, there has been an increasing interest in the use of TFRs for the analysis of mechanical systems [13, 8, 14].

3.4 Signal Adaptive Time-Frequency Representations

A significant performance improvement may often be achieved by using a signal-adaptive kernel. Every bilinear time-frequency representation corresponds to a fixed kernel function that controls its cross-terms suppression properties. The selection of a fixed kernel may limit the performance of certain classes of signals in particular those that have multicomponent signals. Baraniuk and Jones [15] has proposed using a signal dependent kernel which is based on optimization criteria. Of course, the resulting signal-adaptive time-frequency representation is no longer quadratic.

4 Experimental Results and Analysis

Tests were conducted both at steady state and during transient condition. For the steady state tests, four speed were chosen (1,800 rev/min, 3600 rev/min, 5400 rev/min, and 6600 rev/min). The transient test consisted of an acceleration - deceleration run starting from a low-load condition at 1500 rev/min and varying the load an the throttle opening to reach a maximum engine speed and then reach to a lower speed.

Misfire was induced by cutting off the fuel to one cylinder. This resulted in a short sequence of misfires before the engine resumed its normal operation. A section of the acceleration test, ranging from approximately 5,000 to 6,000 rev/min is shown in Figure 2. Note the presence of some slow oscillations superimposed on much higher frequency oscillations; the former are due to the transient response of the dynamometer control system as a consequence of the misfire, while the latter are the velocity fluctuations that

form the basis of the misfire detection method proposed here. The data segment shown in Figure 2 includes three misfire sequences, occurring approximately as the engine passes through 5,200 rev/min, 5,550 rev/min, and 5,750 rev/min, respectively. The tail end of the transient due to an earlier misfire event is also seen at the beginning of the ramp. A simple high-pass filter (fourth order Butterworth) was sufficient to remove the undesired effects of the dynamometer transient response; the cutoff frequency of the filter was significantly lower than the engine cycle frequency; thus, the filtering did not affect the spectrum of interest. Figures 3 through 6 depict the time frequency representation of the crankshaft angular velocity signal obtained using the four different methods mentioned earlier. Although the black and white contour plots do not permit clear identification of the amplitude, it is evident that each of the four techniques succeeds in providing a clear representation of the effects of a misfire . The horizontal (crank angle axis) corresponds to the rotation of the crankshaft, where 42 samples correspond to one engine cycle. Misfire is induced by shutting off an injector in the sixth cycle, and takes full effect in the seventh cycle, continuing through the tenth cycle. The STFT clearly indicates a significant increase in signal energy after the onset of misfire in the spectrum at and just above the order of engine cycle. Since misfire is a "once per cycle" event, this is a reasonable result. The WD reveals a similar result, but the effect of cross-terms is highly visible. The CWD effectively removes cross terms when compared to the WD, especially in the non-misfiring cycles. Similar results are also obtained with the adaptive kernel method. Although the limited scope of this study did not permit investigating the design of a detector based on these concepts, it is clear that any of the above mentioned techniques would lead to successful detection of the misfire event. Work is currently in progress towards the design of time-frequency detectors, and additional experimental data is being acquired in vehicle to further validate these and other misfire detection methods.

5 Conclusions

This paper has demonstrated that various time-frequency inspired signal processing techniques permit the detection of engine misfire in a high-performance multi-cylinder engine through analysis of the crankshaft angular velocity waveform. Although the results presented in this paper are only preliminary, we believe that these methods offer great potential for the development of advanced diagnostic techniques for rotating machinery.

Acknowledgment

The authors gratefully acknowledge the collaboration of M. Ceccarani and S. Mazzetti of Lamborghini Automobili, S. Agata Bolognes, Italy, P. Azzoni and G. Cantoni, ENEA, Dipartimento Innovazione, Bologna, Italy, G. Minelli and D. Moro, Instituto di Macchine, Università di Bologna, Bologna, Italy and the support of the National Science Foundation under grant No. DDM-9157211.

References

[1] California Air Resources Board, "Technical Status Update and Proposed Revisions to Malfunction and Diagnostic System Requirements Applicable to 1994 and Subsequent

California Passenger Cars, Light-Duty Trucks and Medium-Duty Vehicles - (OBD II)" CARB Staff Report, 1991.

[2] G. Rizzoni, "A Dynamic Model for the Internal Combustion Engine", Ph.D. Dissertation, University of Michigan, Ann Arbor, MI, 1986.

[3] G. Rizzoni and W. Ribbens, " Crankshaft Position Measurement for Engine Testing, Control and Diagnostic", 1989, SAE Technical Paper 890885.

[4] Y. Shiao and J. Moskwa, "Misfire Detection and Cylinder Pressure Reconstruction for SI Engines", 1994, SAE Technical Paper 940144.

[5] W. B. Ribbens and J. Park, "Road Test of a Misfire Detection System", 1994, SAE technical paper 940975.

[6] W. B. Ribbens and J. Park, "Application of Neural Networks to Detection Misfires in Automotive Engines", ICASSP, 1994.

[7] F. Connolly, G. Rizzoni, "Real Time Estimation of Engine Torque for the Detection of Engine Misfires", Transaction of the ASME, Journal of Dynamic Systems Measurement and Control. Vol. 116, No. 2, December 1994.

[8] Härle and N. and Böhme, "Detection of Knocking for Spark Ignition Engines Based Structural Vibrations", Proc. ICASSP, 1987.

[9] P. Azzoni, G. Cantoni, G. Minelli, D. Moro, G. Rizzoni, M. Ceccarani, S. Mazzetti, "Measurement of Engine Misfire in a Lamborghini V-12 Engine Using Crankshaft Speed Fluctuations", SAE technical paper 950837.

[10] L. Cohen, "Time-Frequency Distribution- A review" In Proceedings of the IEEE, Vol. 77, July 1989.

[11] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and Quadratic Time-Frequency Signal Representations", SP Magazine, April 1992.

[12] H. I. Choi and W. J. Williams, "Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernels", IEEE-SP, vol. 37, June 1989.

[13] M.J. Dowling, "Application of Non-stationary Analysis to Machinery Monitoring", Proc. ICASSP, 1993.

[14] B. Samimy and G. Rizzoni, " Time-Frequency Analysis for Improved Detection of Internal Combustion Engine Knock", Proc. IEEE-SP Symposium on TFTSA, 1994.

[15] R.G. Barniuk and D.L. Jones, " A Radially-Gaussian, Signal-Dependent Time-Frequency Representation", Proc. ICASSP, 1991.

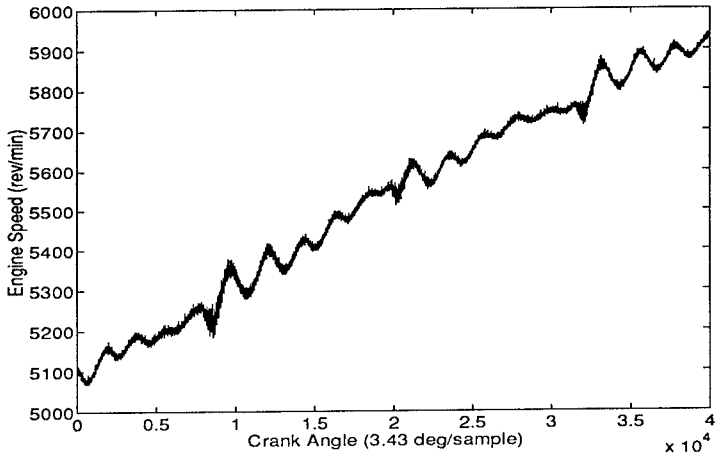


Figure 2: Crankshaft angular velocity during acceleration (190 combustion cycles)

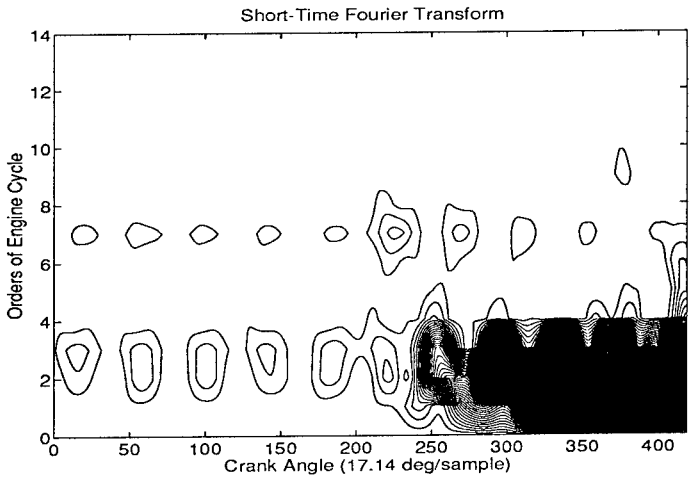


Figure 3: STFT of ten combustion cycles with misfire starting in cycle six

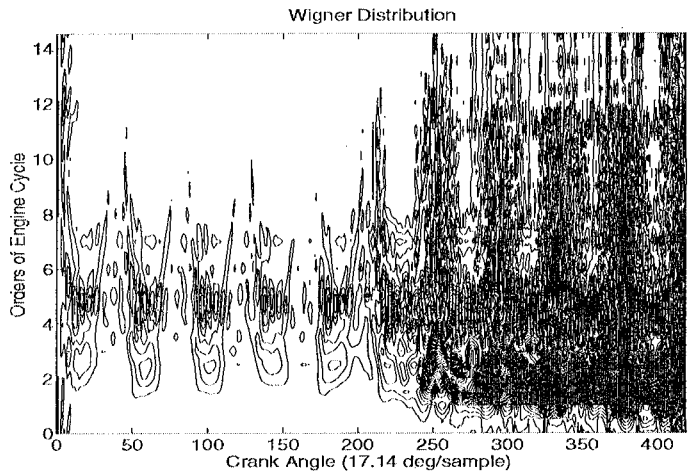


Figure 4: WD of ten combustion cycles with misfire starting in cycle six

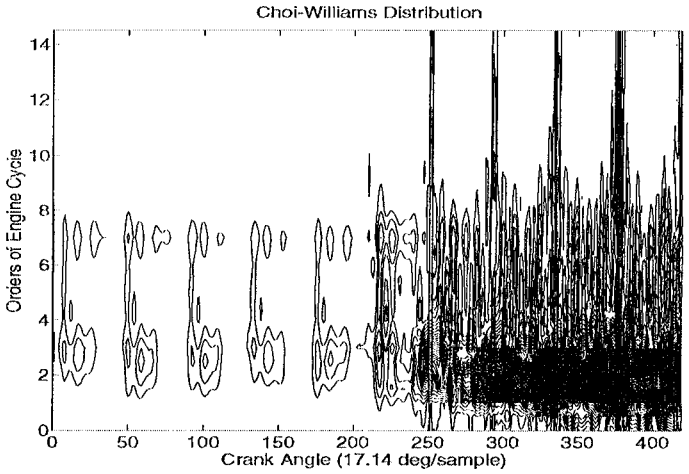


Figure 5: Choi-Williams distribution of ten combustion cycles with misfire starting in cycle six

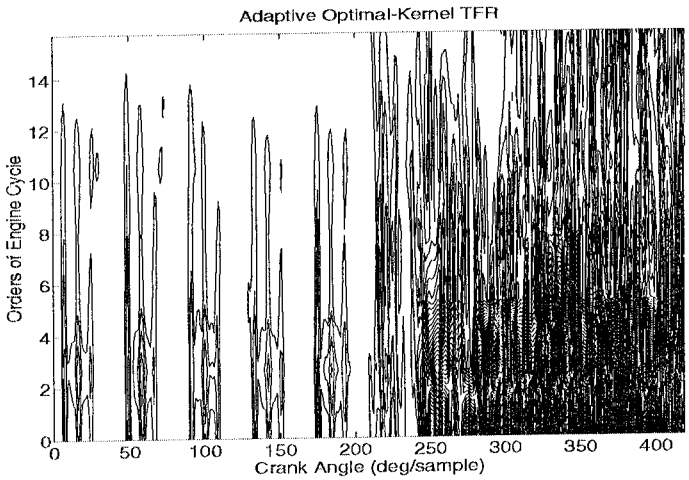


Figure 6: Adaptive optimal-kernel time-frequency representation of ten combustion cycles with misfire starting in cycle six

TIME-FREQUENCY DOMAIN REPRESENTATION FOR APPLICATION TO MACHINE CONDITION MONITORING

Young S. Shin and John E. Harding

Department of Mechanical Engineering
Naval Postgraduate School
Monterey, California 93943

Abstract: Time-frequency domain representation is needed to characterize the nonstationary noise and vibration signatures. Both Pseudo Wigner-Ville distribution and wavelet transform were applied to and analyzed the vibration data taken from the reciprocating air compressor. Artificial faults were introduced in the signatures and analyzed for the detection of the faults in the signatures.

Key Words: machine condition monitoring, nonstationary signatures, pseudo Wigner-Ville distribution, wavelet transform

Introduction: When the machine noise and vibration signature is stationary process such as the data taken from rotating machinery, frequency domain analysis is usually sufficient to characterize the data. However, when the signature is nonstationary process such as the data taken from transient or reciprocating machinery, the vibration characteristics change with time and time-frequency domain representation is needed to fully analyze the data. Both Pseudo Wigner-Ville distribution and wavelet transform were investigated, applied to and analyzed the vibration data taken from the reciprocating air compressor. Artificial faults were introduced in the reciprocating air compressor data and analyzed for the detection of the faults in the nonstationary signatures.

Pseudo Wigner-Ville Distribution: Pseudo Wigner-Ville Distribution (PWVD) is a three dimensional (time, frequency and amplitude) representation of a signature and is ideally suited for describing transient or other nonstationary phenomena. The Wigner Distribution function (WDF) [Wigner, 1932] was first introduced by Wigner in 1932 and later Ville [1948] proposed the use of the analytic signal before computing WDF to avoid aliasing problem, resulting Wigner-Ville Distribution (WVD) function. The WDF is expressed as,

$$w(t,\omega) = \int_{-\infty}^{\infty} s^{*}(t-\frac{\tau}{2})\, s(t+\frac{\tau}{2})\, e^{-j\omega\tau}\, d\tau \qquad (1)$$

where w(t,ω) is WDF, s(u) is the time signal and s*(t) is its complex conjugate. Two pronounced characteristics of the WDF should be noted. First, the WDF of multifrequency signals generates the cross terms. Second, the WDF may have negative amplitudes caused by interference due to the presence of these cross-terms. To suppress the interference components of the WDF, a sliding Gaussian window function was applied in the time-frequency domain and the WDF obtained with a window function is called, PWVD, the Pseudo Wigner-Ville Distribution function. Shin and Jeon [1993] investigated PWVD, and applied to analyze both stationary and nonstationary signals. They also discussed PWVD as a useful tool for machinery condition monitoring.

Wavelet Analysis: Research has been recently focused on time-frequency domain with the modeling of a new decomposition of a signal into a family of functions which are the dilations and translations of a unique function called a wavelet. Instead of portraying a signal into harmonic functions ($e^{i\omega t}$ in Fourier transform), the signal is presented into a series of orthogonal basis functions of finite length. Each wavelet is located at a different position on the time axis. At the finest scale, wavelets may be very short and, at a coarse scale, they may be very long. Alternatively very small disturbances in a record of acoustic and machinery vibration signatures can be easily characterized from a wavelet map in which the mean-square value of the time record is shown over wavelet scale and position. One important property of a wavelet transform is its ability to characterize easily the local regularity of a function. The continuous wavelet transform is expressed by,

$$y(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \overline{\psi(\frac{t-b}{a})}\, s(t)\, dt \qquad (2)$$

where y(a,b) is wavelet transform, ψ is an analyzing wavelet, a represents a time dilation, b a time translation, and bar for complex conjugate. The normalization factor $1/\sqrt{a}$ is perhaps most effectively visualized as endowing $|Y(a,b)|^2$ with units of power/Hertz [Shensa, 1992].

A modulated Gaussian wavelet ψ originally proposed by Morlet et al. [1989] is expressed by,

$$\psi(t) = e^{i\omega_0 t}\, e^{-t^2/2} \qquad (3)$$

This modulated Gaussian wavelet was used in this paper. In the continuous wavelet, the family is considered,

$$\psi^{a,b}(t) = |a|^{-1/2}\ \psi\left(\frac{t - b}{a}\right) \tag{4}$$

where $b \in \mathbf{R}$, $a \in \mathbf{R}+$ with $a \neq 0$, and ψ is admissible. **R** denotes the real line and **R+** denotes the positive real line. The discrete form of Eq.(4) is expressed as shown in the following equation [Mallat, 1989]

$$\begin{aligned}\psi_{m,n}(t) &= a_0^{-m/2}\ \psi\left(\frac{t - nb_0a_0^m}{a_0^m}\right) \\ &= a_0^{-m/2}\ \psi(a_0^{-m}t - nb_0)\end{aligned} \tag{5}$$

For the computation efficiency, we assume that $a = 2^m$, that is, $a_o = 2$, where m is termed the octave of the transform. This means that the frequency resolution of wavelet has an octave band. The integral Eq.(2) yields a wavelet series as a following equation by using Eq.(5) and taking b to be a multiple of a.

$$y(2^m, n2^m) = \frac{1}{\sqrt{2^m}} \int_{\infty}^{-\infty} \overline{\psi\left(\frac{t}{2^m} - n\right)}\, s(t)\, dt \tag{6}$$

Discretizing the integral in Eq.(6),

$$y(2^m, n2^m) = \frac{1}{\sqrt{2^m}} \sum_k \overline{\psi\left(\frac{k}{2^m} - n\right)}\, s(k) \tag{7}$$

Octave m is only output every 2^m samples. In this form the resulting algorithm will not be translation invariant [Mallat, 1989]. The discrete wavelet transform is highly not invariant under translations. In practice one does not use an infinite number of scales, but cuts off very low and very high frequencies. Shin and Jeon [1994] discussed the performance of the wavelet transform using simple examples from the standpoint of detection of the abnormality and fault contained in the signatures.

Vibration Test Data: The vibration test data of shipboard high-pressure air compressor was obtained and analyzed using both PWVD and wavelet transform. The vibration data was collected by the transducer mounted on the free end of the compressor. The measured vibration data used in the analysis is shown in Figure 1. The sampling frequency was 7,500 Hz.

Artificial Fault Simulation: To create deformities in the measured vibration data an artificial fault signal was introduced. Because the vibration test data was captured using rotating machinery analysis techniques, the major timing events and crank-angle position were never established. These shortcomings prevented an artificial fault signal that would represent poor operating conditions of a particular compressor mechanical components, such as a loose cylinder wrist pin, to be incorporated into the measured data. The fault signals that were introduced into the measured vibration test data were made up of a series of instantaneous impulses (spikes) of constant amplitude at equidistant spacing along the time axis. These simple fault signals were created to correspond to the signal time span of the high sampling frequency data for the high-pressure air compressor. Three different levels of instantaneous impulses (low, medium and high) were added to the vibration test data and are shown in the Figure 2.

Discussions: First 3-dimensional representation (time, frequency and amplitude) of the measured vibration data was plotted in Figure 3 for both PWVD and wavelet transform. In the PWVD plot, Figure 3(a), the dominant peaks at the low frequency are well defined and its harmonics are shown clearly. Figure 3(b) shows the 3-D wavelet transform. As can be seen in Figure 3, PWVD has a higher resolution than the wavelet transform. However, the computational time for PWVD is extremely higher than the wavelet transform. For PWVD, the amplitude, $w(t, \omega)$ is computed at every Δt, time step and $\Delta\omega$, frequency step. For wavelet transform, the amplitude, $y(t, \omega)$ is computed at every wavelet level (m), and 2^m time points in each wavelet level. PWVD appears to be very well defined in low frequency range and is rather insensitive in high frequency region, which may be caused by the application of sliding window function on time-frequency domain. For wavelet transform, it appears to be opposite.

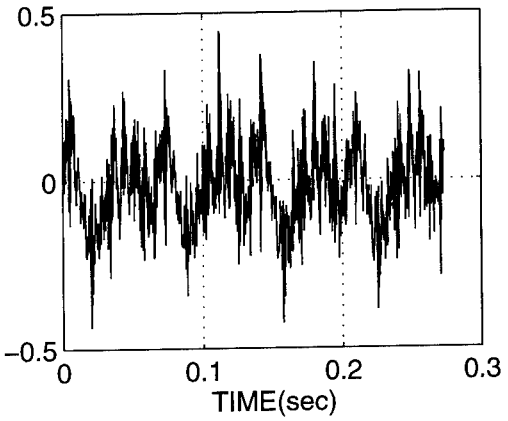
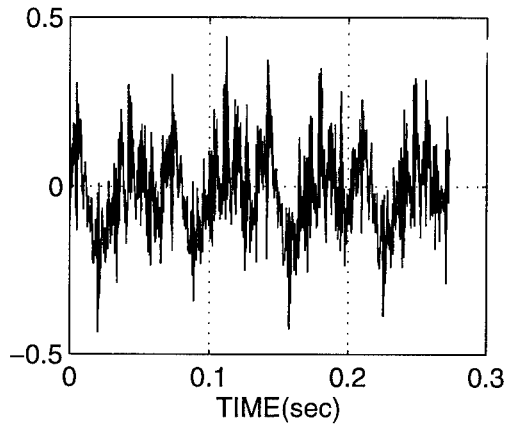
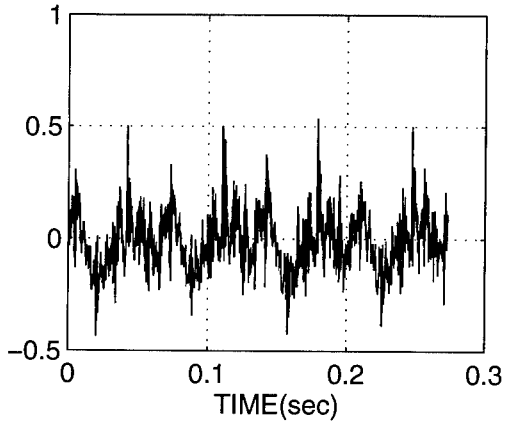


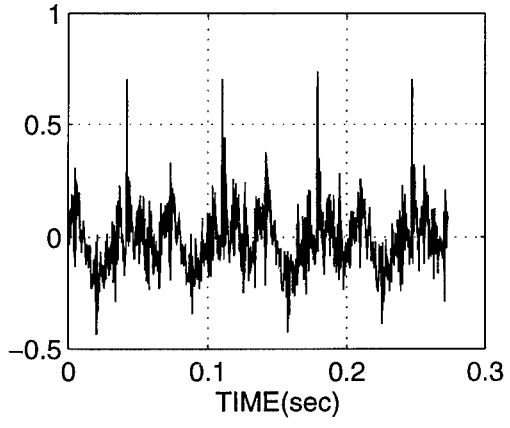
Figure 1. The Measured Vibration Test Data



(a) Low Level Fault

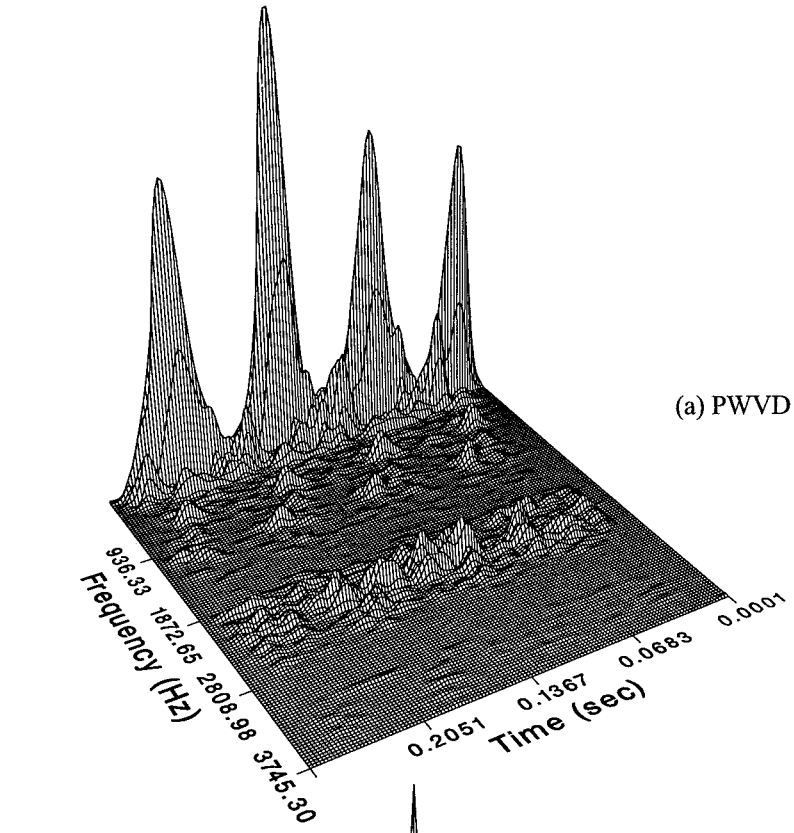


(b) Medium Level Fault

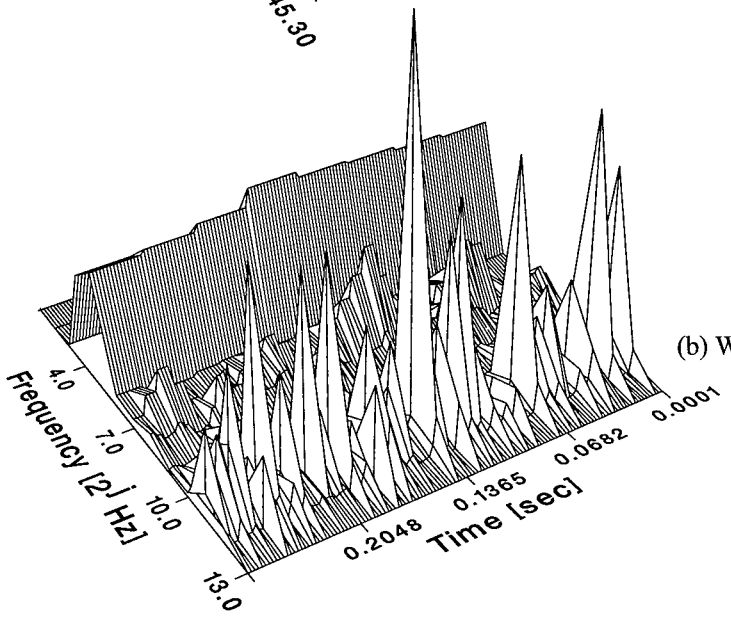


(c) High Level Fault

Figure 2. The Measured Vibration Data with Artificial Fault Simulation



(a) PWVD



(b) Wavelet Transform

Figure 3. Three-Dimensional Representation of PWVD and Wavelet Transform for the Measured Vibration Data

2-D PWVD Map: 2-D PWVD maps were plotted in Figure 4 for the vibration test data and the same with three different levels of the artificial faults. PWVD map for low level fault [Fig. 4(b)] show slight change in contour. However, for medium and high levels of faults [Figs. 4(c) and (d)], it distinctly shows the change in frequencies at each time point where the fault is introduced. It appears at lower one-third frequency range in the map.

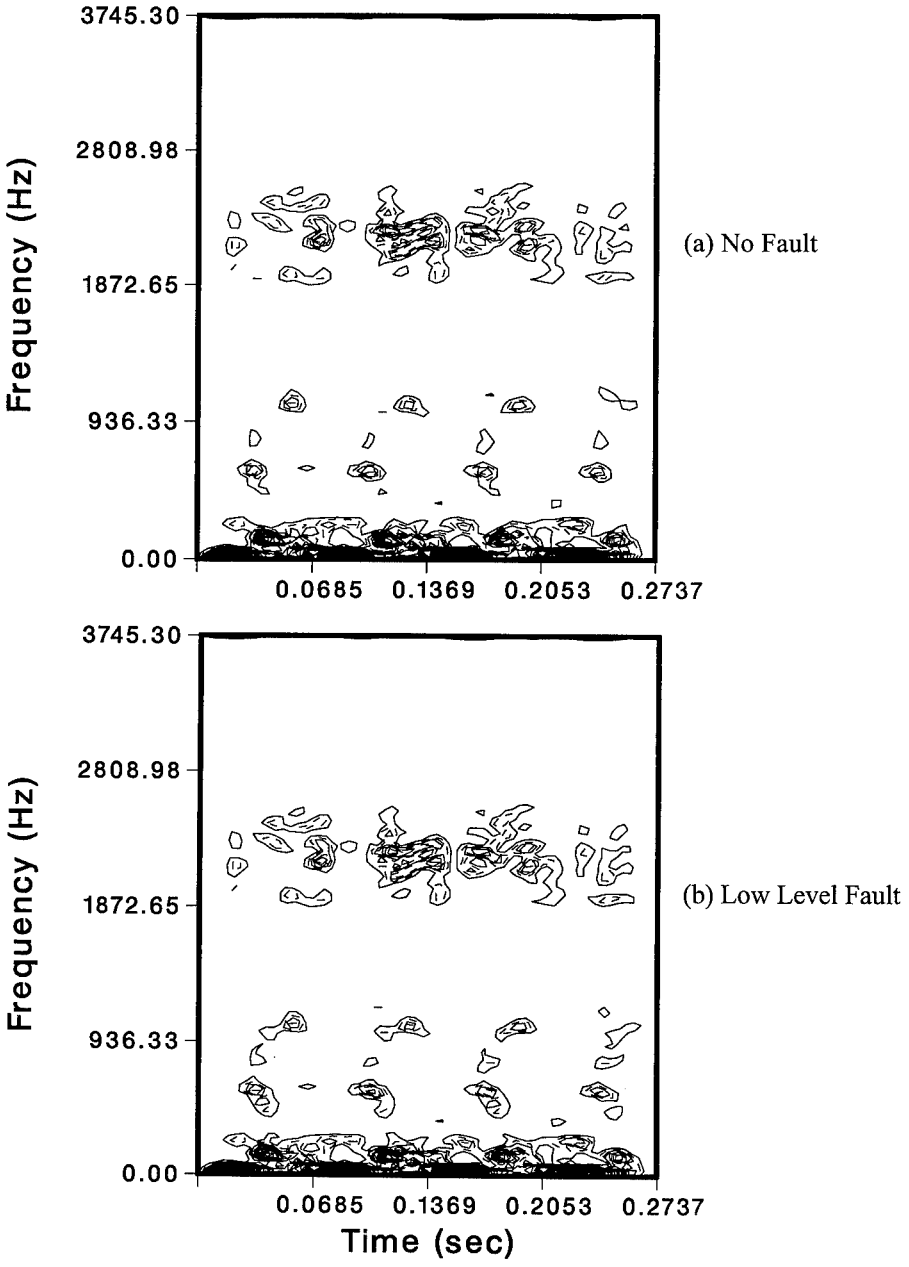
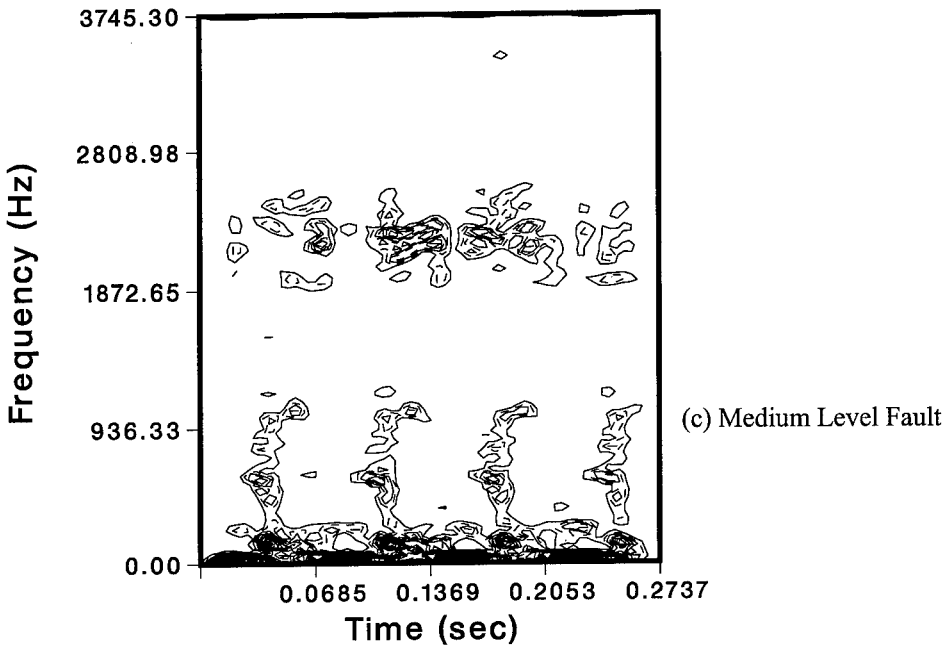
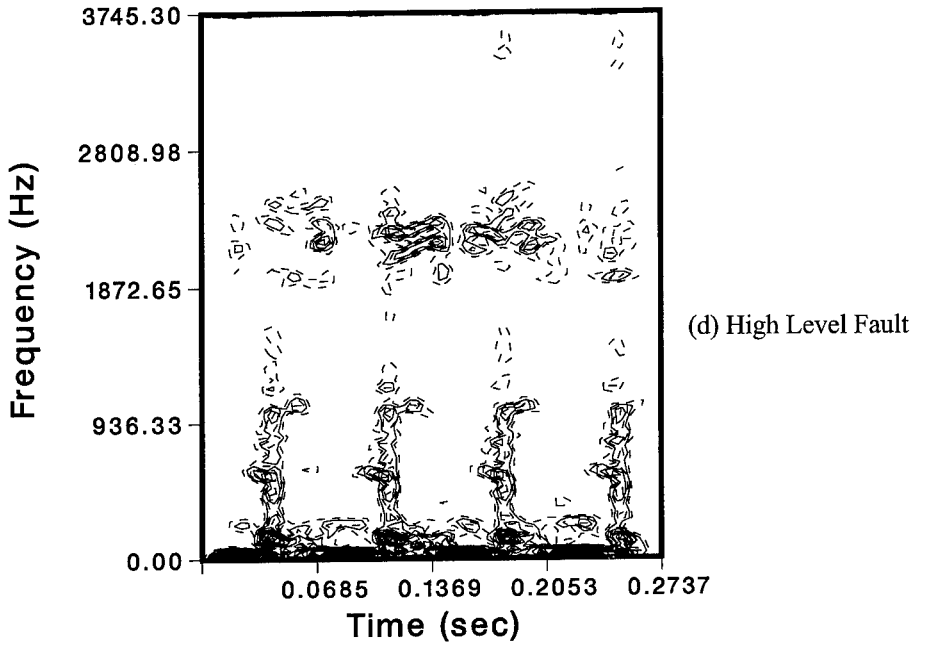


Figure 4. 2-D PWVD Maps for the Vibration Data with and without Artificial Faults



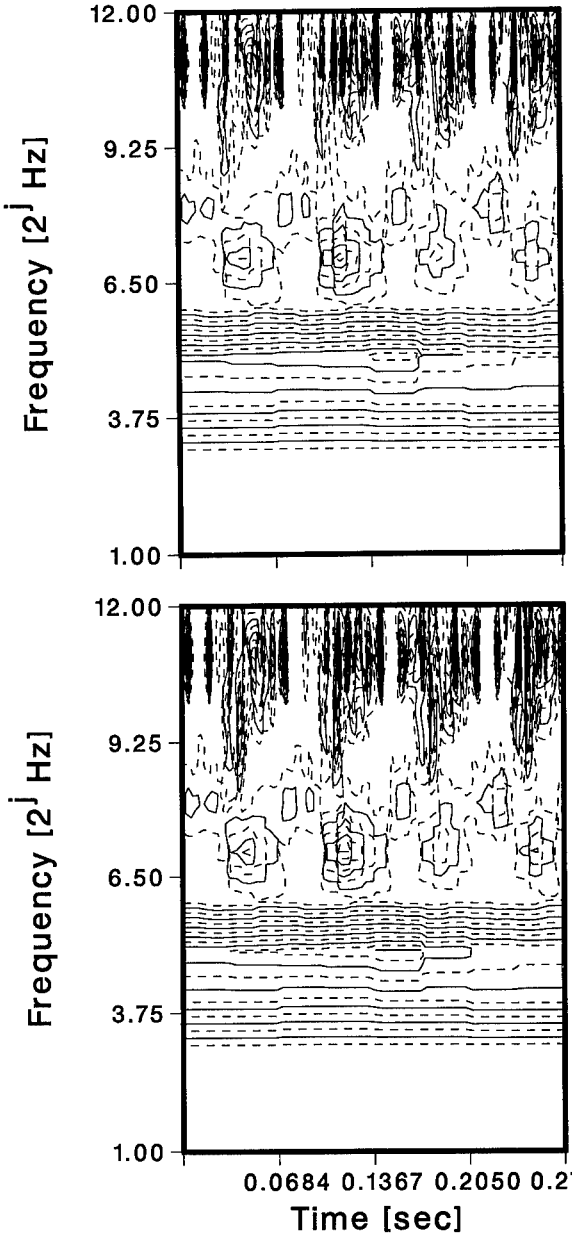
(c) Medium Level Fault



(d) High Level Fault

Figure 4. 2-D PWVD Maps for the Vibration Data with and without Artificial Faults

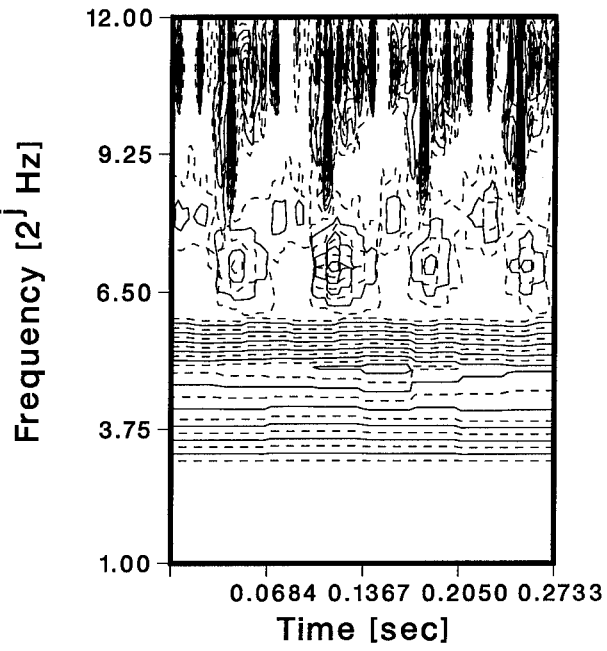
2-D Wavelet Transform Map: 2-D wavelet transform maps were plotted in Figure 5 for the vibration data and the same with three different levels of the artificial faults. Wavelet transform map for low level fault [Fig. 5(b)] shows somewhat significant change in contour. However, for medium and high levels of faults [Figs. 5(c) and (d)], it distinctly shows the change in frequencies at each time point where the fault is introduced. It appears at higher frequency range in the map.



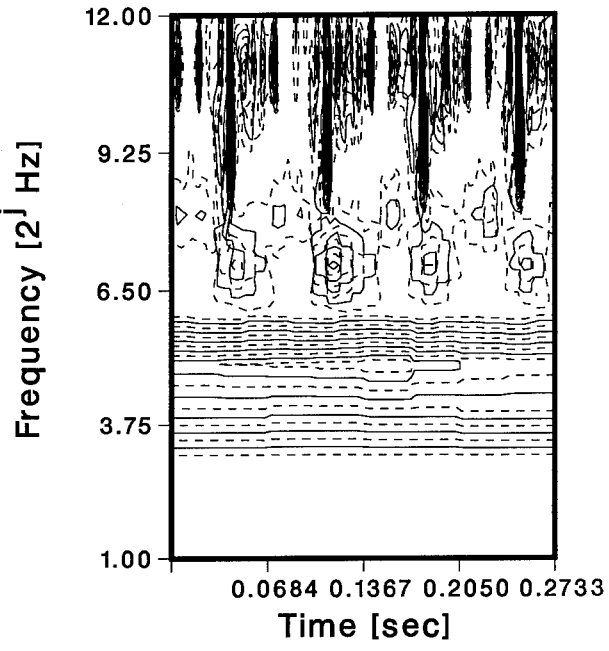
(a) No Fault

(b) Low Level Fault

Figure 5. 2-D Wavelet Transform Maps for Vibration Data with/without Artificial Faults



(c) Medium Level Fault



(d) High Level Fault

Figure 5. 2-D Wavelet Transform Maps for Vibration Data with/without Artificial Faults

Conclusions: The following conclusions may be drawn : (a) Both PWVD and wavelet transform are well suited for portraying the transient or nonstationary noise and vibration data in time-frequency domain, (b) Wavelet transform has a great advantage of detecting the small local disturbance of signature, especially in high frequency region, and (c) PWVD characterizes the signatures well in lower frequency range and is suitable detecting abnormality in lower frequency range.

Acknowledgement: We would like to express our sincere appreciation to Mr. Lynn Hampton of Naval Sea Systems Command Detachment, PERA(CV) to sponsor the research program and for providing machine vibration data. Additionally, we'd like to thank Mr. J. K. Kim for additional computation of PWVD and wavelet transform.

References:

Mallat, S. G., 1989, "Multifrequency Channel Decompositions of Images and Wavelet Models," *IEEE Trans. Acoust. Speech and Signal Processing*, Vol. 37, No. 12, pp. 2091-2110.

Molet, J., et al., 1989, "Wave Propagation and Sampling Theory - Part I, II," *Geophysics*, Vol. 47, No. 2, pp. 203-236.

Shensa, M. J., 1992, "The Discrete Wavelet Transform Wedding the A Trous and Mallat Algorithms," *IEEE Trans. on Signal Process.*, Vol. 40, N0. 10, pp. 2464-2482.

Shin, Y. S., and Jeon, J. J.1993, "Pseudo Wigner-Ville Time-Frequency Distribution and Its Application to Machinery Condition Monitoring," *Journal of Shock and Vibration*, Vol. 1, No. 1, pp. 65-76.

Shin, Y. S., and Jeon, J. J., 1994, "Wavelet Analysis of Nonstationary Acoustic and Vibration Signatures," *Proceedings of inter-noise 94*, Vol. 3, pp. 1903-1906.

Ville, J., 1948, "Theorie et Applications de la Notion de Signal Analytique," *Cables et Transmission*, Vol. 2a, pp. 61-74.

Wigner, E., 1932, "On the Quantum Correction for Thermodynamic Equilibrium," *Physics Review*, Vol. 40, pp. 749-759.

Time-Frequency Analysis of a Cam Operated Pump

Robert A. Rohrbaugh

Naval Surface Warfare Center
Detachment Puget Sound
Bremerton, WA 98314

and

Leon Cohen[1]

Hunter College and Graduate Center of CUNY
695 Park Ave.
New York, NY 10021

ABSTRACT

We apply time-frequency methods for the analysis of a cam operated pump. We show that these methods reveal features that are not seen by traditional methods, such as the power spectrum. These features may be used for early detection of faults and as supplemental information to assess the condition of the pump.

1. INTRODUCTION

The two historic approaches to machine monitoring have involved examining the time waveform and power spectrum of the machine signal[19]. These two methods are capable of identifying many problems associated with machines. The basis of these approaches is that when a machine is significantly changed from its normal state the vibrational waveform and spectrum will be different. However what is crucial in machine assessment is the earliest possible indication that the machine condition is changing. Recent advances in signal processing methods indicate that one can enhance our ability to detect changes earlier than by traditional spectral methods[8,11–18] Furthermore, there is evidence to show that in addition to early detection, the features being identified are good indicators and classifiers of particular faults. These new methods are based on time-frequency analysis, a methodology which describes how frequencies change in time. In brief, the basic reason why time-frequency analysis reveals features that are not seen by the power spectrum is this: when a machine starts to go bad it produces transients in time of such short duration that traditional analysis averages them so that they do not appear clearly in the spectral lines. As will be shown and discussed in detail the new time-

[1]Work supported in part by the NSA HBCU/MI program and the PSC-CUNY Research Award Program.

frequency methods reveal them effectively.

In Section 2 we summarize the basic time-frequency approaches and explain by way of examples why the new time-frequency methods reveal considerably more than standard spectral analysis. In Section 3 we apply them to a cam operated pump.

2. SPECTRAL ANALYSIS AND TIME-VARYING SPECTRAL ANALYSIS

One of the standard means for machine condition assessment is by observing the presence of assumed stationary frequency components, noting the amplitude of these components with respect to a reference vibration level, and comparing the levels of these frequency components with those of previously acquired spectra. The machine analyst may then gain insight into various physical aspects of the machine. For example, the presence of certain frequency components not usually seen in the spectrum of a machine may indicate degradation of a bearing, or a relatively high rotational component may signify that the shaft is out of balance. Analysis of the power spectrum has proven to be effective at identifying many problems associated with rotating machinery or machinery with constant frequency operating parameters (i.e. transformers). However, spectral analysis may not provide sufficient insight to uniquely identify problems associated with impact-like forces, ringing or chirp transients, sudden increases and decreases in tones or broadband noise, and other short-duration forcing functions. Defects that cause such forces may be associated with pulley assemblies, impellers, cavitation, resonance, steam tones, and certain shaft and bearing problems. Time domain information may be useful for these problems, but is limited as a stand-alone analysis tool. We now briefly review the basic ideas of time-frequency analysis to show why the power spectrum may not be sufficient in analyzing phenomenon for machine assessment.

For a signal, $s(t)$, the intensity or energy per unit time is $|s(t)|^2$. The spectrum of the signal is given by

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int s(t)\, e^{-j\omega t}\, dt \tag{2.1}$$

and the intensity per unit frequency, $P(\omega)$, which is called the energy density spectrum is[2]

$$P(\omega) = |S(\omega)|^2 \tag{2.2}$$

The energy density spectrum is an indication of the frequency existing for the *duration* of the signal. It does not indicate how the frequencies are changing in time. Over the past fifty years methods which show how the spectrum is changing in time have been developed and in the past ten years considerable advances have been made. The methods developed are called time-frequency analysis[1,4–7,9,10,20]. The fundamental idea is to devise a joint density of time and frequency where

[2]Technically a distinction should be made between the energy density spectrum and the power spectrum. When we consider deterministic signals, they are the same. The phrase "power spectrum" is usually reserved for the stochastic case where ensemble averaging has been done. Since we are dealing with deterministic signals only we shall use them interchangeably.

at each time one can determine what frequencies are occurring. The joint time-frequency density, $P(t,\omega)$, is determined by the signal and, ideally, should satisfy the individual densities of time and frequency, that is the marginal densities

$$\int P(t,\omega)\,d\omega \;=\; |\,s(t)\,|^2 \tag{2.3}$$

$$\int P(t,\omega)\,dt \;=\; |\,S(\omega)\,|^2 \tag{2.4}$$

One should also be able to use $P(t,\omega)$ to calculate joint time-frequency averages in the same manner as any other joint density.

The earliest method that was developed is the spectrogram. The idea of the spectrogram is simple and effective. If one is interested in the frequencies that exist at a certain time, one takes a small piece of the signal around that time and neglects or significantly suppresses the signal for times far away from the time of interest. This is accomplished mathematically by windowing the signal with a function, $h(t)$, centered at t, to produce a modified signal,

$$s_t(\tau) \;=\; s(\tau)\,h(\tau - t) \tag{2.5}$$

where τ is the running time and t is the momentarily fixed time of interest. By making the window narrow, $s_t(\tau)$ emphasizes the signal around the time t and suppresses it at other times. Now, the spectrum of this modified signal is

$$S_t(\omega) \;=\; \frac{1}{\sqrt{2\pi}}\int e^{-j\omega\tau}s_t(\tau)\,d\tau \tag{2.6}$$

$$\;=\; \frac{1}{\sqrt{2\pi}}\int e^{-j\omega\tau}s(\tau)\,h(\tau - t)\,d\tau \tag{2.7}$$

and is called the short-time Fourier transform. The energy density spectrum at time t is hence

$$P_{SP}(t,\omega) \;=\; |\,S_t(\omega)\,|^2 \;=\; \left|\frac{1}{\sqrt{2\pi}}\int e^{-j\omega\tau}s(\tau)\,h(\tau - t)\,d\tau\right|^2 \tag{2.8}$$

and is called the spectrogram. Since the modified signal emphasizes the signal at the time of interest, the expectation is that the spectrum of the modified signal will reveal the frequencies around that time. The spectrogram is the most widely tool used for time-frequency analysis and is very often effective. However it has a fundamental limitation. If one wants to get finer time localization a very narrow window must be used. However, using a very narrow window produces a smeared out spectrum that bears almost no relation to the properties of the signal. Because of this, other time-frequency densities have been proposed. In 1966 it was shown that any time-frequency representation can be obtained from the general form[2, 4, 5]

$$C(t,\omega) \;=\; \frac{1}{4\pi^2}\iiint s^*(u - \tfrac{1}{2}\tau)\,s(u + \tfrac{1}{2}\tau)\,\phi(\theta,\tau)\,e^{-j\theta t - j\tau\omega + j\theta u}\,du\,d\tau\,d\theta \tag{2.9}$$

where $\phi(\theta,\tau)$ determines specific properties of the representation and is known the kernel function. The advantage of the kernel method is that desirable properties of the distributions are reflected as simple properties of the kernel. Fixing the

kernel fixes the distribution. In the past few years significant advances have been made in finding kernels which produce distributions that reveal considerably more then the spectrogram. Of particular note are the methods developed by Choi and Williams,[1] and Zhao, Atlas, and Marks,[20] Loughlin, Pitton and L. Atlas,[9] Jeong and Williams,[7] and Cunningham and Williams.[6]

If the kernel is a pure function, that is, if it is not a functional of the signal, then the densities are said to be bilinear because the signal appears twice in Eq. (2.9). It is known that bilinear distributions cannot both satisfy the marginals and be manifestly positive. However one can construct distributions that satisfy the marginals and positively.

Positive Time-Frequency Distributions

A method for constructing manifestly positive distributions was given by Cohen and Posch[3]. A important practical implementation has recently been developed by Loughlin, Pitton, and Atlas.[10] It is based on a minimum cross entropy scheme (MCE). One starts with an estimate of the density, $P_0(t,\omega)$ and typically a spectrogram is used. The cross entropy is defined

$$\Delta = -\iint P(t,\omega)\log\frac{P(t,\omega)}{P_0(t,\omega)}\,dt\,d\omega \tag{2.10}$$

One maximizes Δ with the constraints of the marginals, and other possible desirable constraints. We have found that these manifestly positive densities are very effective for analyzing machine data and the results we present are based on it.

Spectral analysis and time-varying spectral analysis

When one uses the energy density spectrum to study a signal, that is generally called spectral analysis. As mentioned above, the spectrum tells us what frequencies the signal contains. But it does not tell us when those frequencies existed. A joint time-frequency density gives the frequencies for each time. If one sums over all time then one should get the frequencies that exist irrespective of time. That is the significance of the marginal conditions. Therefore, many different variations of occurrence of frequencies can give, essentially, the same energy density spectrum. Hence the energy density spectrum may not be a good discriminator. To illustrate this fundamental idea we give some examples of signals, their energy density spectrum, and the time-frequency plot of each signal.

In Fig. 1 we show three signals whose power spectrum is approximately the same. Nonetheless the time frequency structure is significantly different and can be ascertained simply. In Fig. 2 we give two signals, again whose spectrum is roughly the same but the time-frequency structure is significantly different. These examples illustrate that while the energy density spectrum of the signals may be the same, indicating that for the duration of the signal both had the same frequencies, the temporal occurrences of the frequencies may be dramatically different.

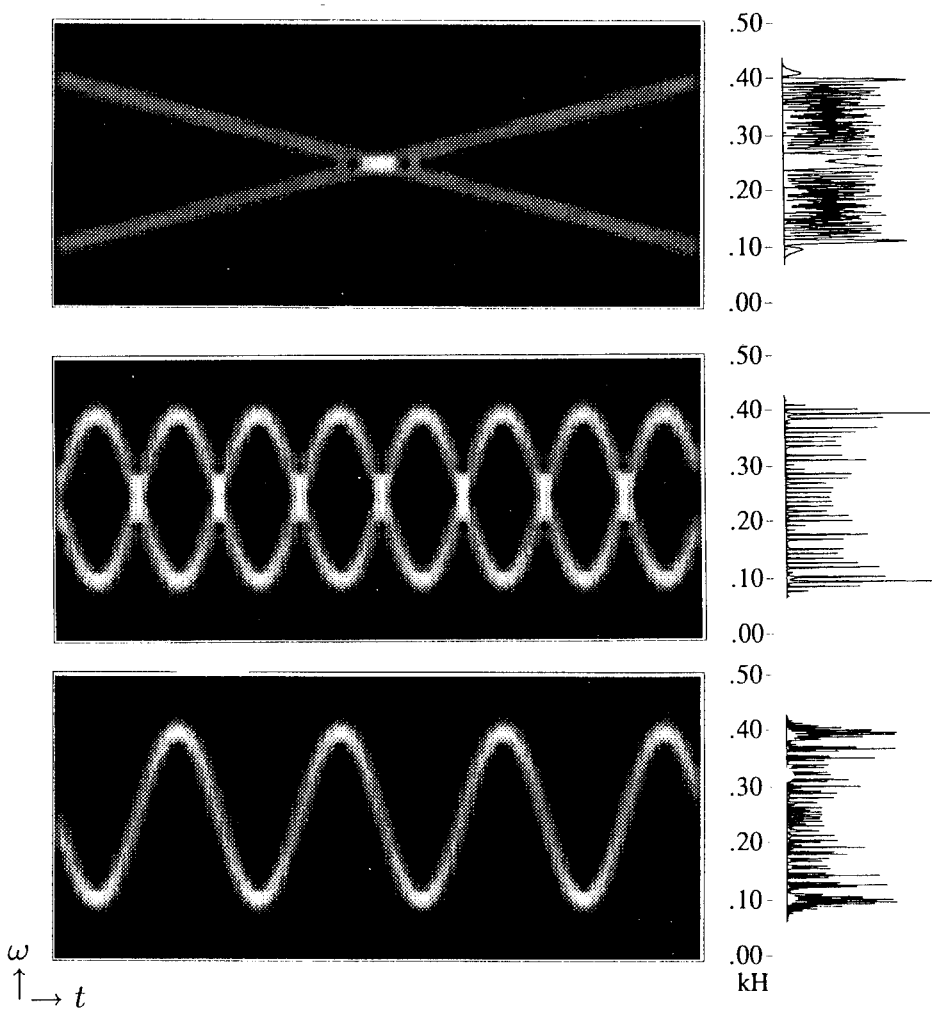


Fig. 1. An example of three signals whose spectra are approximately the same but the time-frequency behavior is very different. The main plot is the time frequency distribution. The plot on the right is the spectrum.

3. CAM OPERATED PUMP

A special challenge to traditional analysis methods is presented by reciprocating machines such as internal combustion engines, or in our example, a cam-operated pump. The normal operating condition of a reciprocating pump involves forces generated by periodic acceleration of masses (plungers) with corresponding increasing and decreasing fluid pressures over short periods of time. Variation of these short-time duration forces from a baseline condition may be indicative of impending fault.

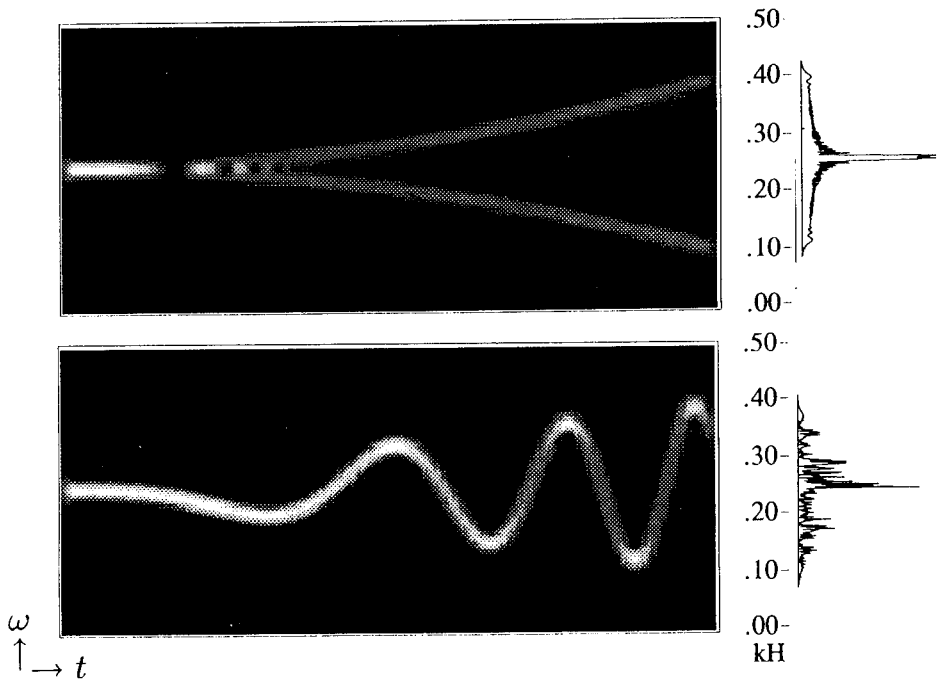


Fig. 2. Another example where the spectra are approximately the same but the time-frequency behavior is very different.

The technique used for comprehensive condition assessment of a reciprocating machine should allow an inspection of energy over the duration of the shortest operating cycle. This obviously precludes stationary analysis methods such as a the power spectrum, although such methods are still useful for analysis of forces that are relatively stable over short time periods (shaft and pulley rotation, certain bearing frequencies, etc.). Time domain methods are often used to analyze the signal over short time periods but the lack of frequency information may inhibit identification of individual machine components. If both methods are used , the power spectrum may be used to identify frequency components of individual machine parts, but the information may still not be easily related to the forces causing the amplitude variation of the time domain signal. It follows that an alternative to time domain or frequency processing would be a joint distribution of time and frequency. A time-frequency distribution offers the capability of observing how the frequency structure of a signal changes with time. In the case of fault analysis, a machine that is degrading may exhibit changes in amplitude of frequency components, modulation, or the appearance of new frequency components. The resulting problem is to find changes in frequency that are associated with a particular fault and the optimum time observation period required to identify the degradation.

Experimental setup

The pump test stand is illustrated in Fig. 3. The pump was seated on resilient mounts and drew water via a suction isolation valve that was elevated approx-

imately 6 feet above the floor. The pump was belt-driven by an electric motor mounted to the foundation. The power end of the pump contained four elliptic cams that pushed four plungers into valve chambers. The motion of the plungers in the chambers caused fluid to be drawn into and then discharged from the chambers. The discharge end of the pump was connected to the tank by a throttle valve that was used to adjust discharge pressure during the test.

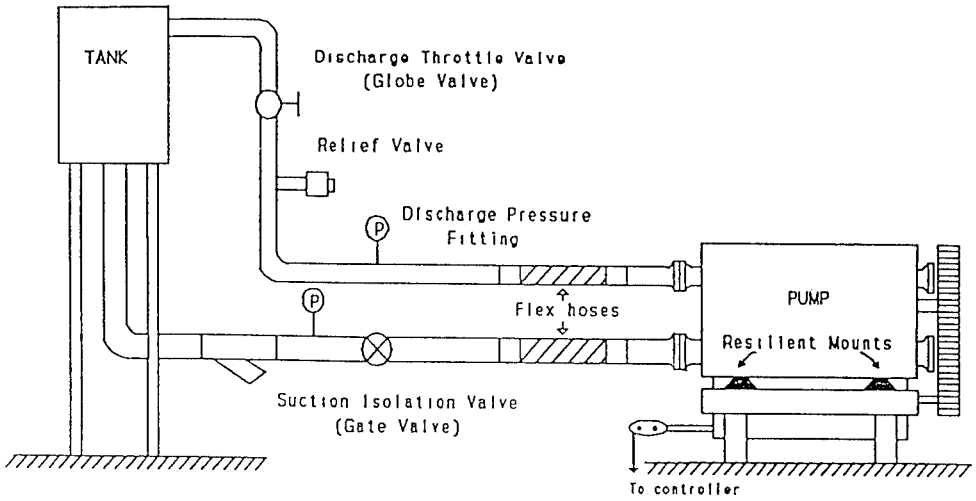


Fig. 3. Pump test stand

The sensors used to collect information from the test stand included accelerometers mounted on the pump and a current probe that was clamped to one phase lead of the motor to sense changes in motor current characteristics. An alternative current analysis system was used to measure the amplitude and phase modulation of the 60 Hz line frequency. The information from the sensors was controlled by a Zonic Workstation 7000 and Digital Equipment Corporation VAX 4000. The data was recorded with a 28 channel Honeywell 101 tape recorder operating at a speed of 30 inches per second. After the baseline data had been collected, tests were initiated to simulate two common pump faults.

Baseline condition

The pump was originally run in a baseline (nondefective) condition with varying water temperatures and discharge pressures. A globe valve was used to throttle discharge pressure from no load (discharge valve open) to the most stable operating pressure of 300 psig, which was used for data acquisition.

The baseline condition of the pump for the amplitude modulated current signal is presented in Fig. 4 (a-d), where in (a),(b),(c) we have used a narrow band, medium band and wide band spectrogram. Fig. 4d gives the positive MCE distribution. Clearly, considerably greater detail is revealed by the positive MCE distribution.

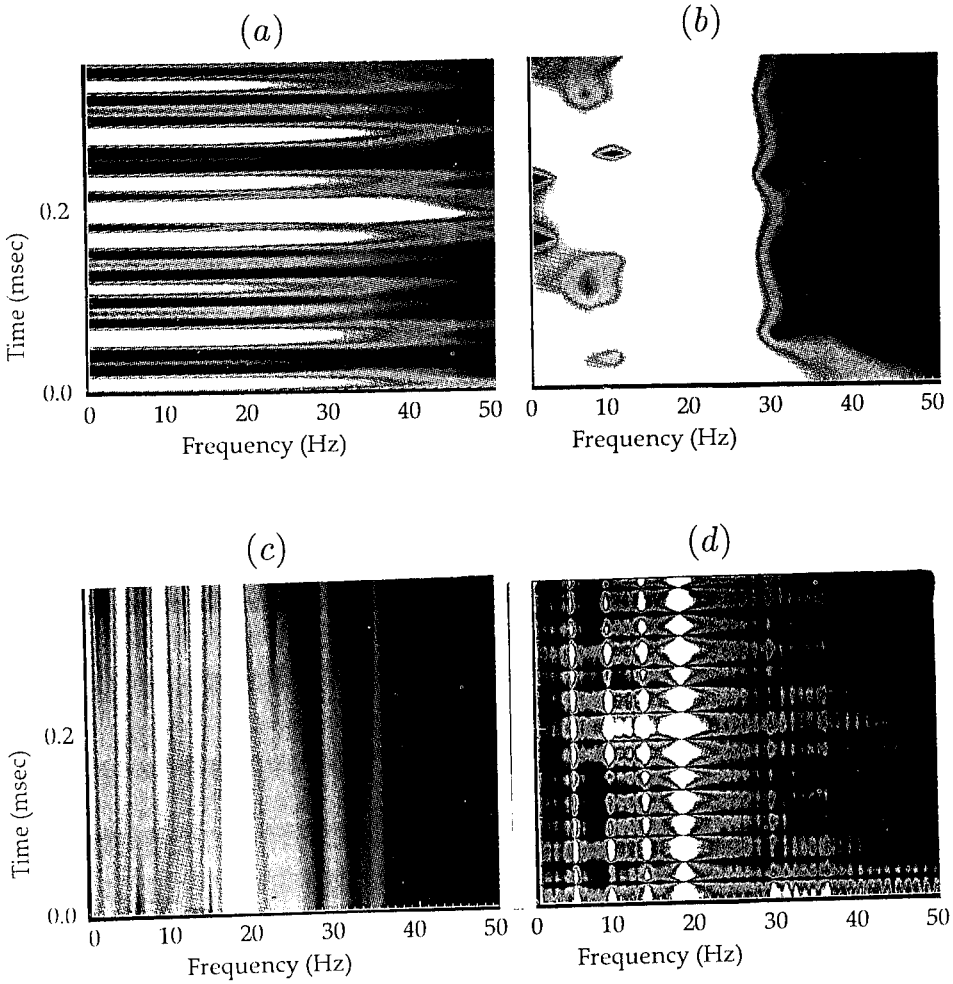


Fig. 4. The baseline condition of the amplitude modulated motor current signal. In (a), (b), (c), we have used the narrow, medium and wide band spectrogram respectively. In (d) we show the MCE distribution. The spectrogram can not simultaneously give fine time and frequency localization. The MCE can.

The predominant forcing functions are revealed as pump rotation (5.5 Hz), the second and third harmonics of pump rotation, and the plunger frequency (four times pump rotation) at 18 Hz. Notice at the plunger frequency of 18 Hz that the energy associated with the motion of the individual plunger is not uniform. Measurements taken during the test suggest that these vibrations are a result of the plungers producing nonuniform discharge pressures. A "perfect" pump would exhibit identical time-frequency energy characteristics for each plunger.

Leaking Discharge Valve Defect

The first fault simulation involved inserting a defective discharge valve into the pump. The seating surface of the valve had been filed to simulate uneven wear.

After data acquisition, the leaking discharge valve was replaced with the original nondefective valve.

The leaking discharge valve defect (Fig. 5) resulted in relatively pronounced differences in data acquired with the motor current and vibration sensors. Compared to the baseline MCE plot for the accelerometer data taken from the cylinder head (Fig. 5a), note the increase in levels of the 5.5 Hz pump rotation fundamental and second and third harmonics (Fig. 5b). In addition, an increase in broad band energy is observed between pump rotation at 5.5 Hz and the second harmonic at 9 Hz. This broadband energy is excited at a rate (215 Hz) equivalent to the rotational speed of the pulley that connects the pump to the motor.

Compared to the baseline phase-modulated current data for this defect (Fig. 6a), the level of the 18 Hz plunger frequency was significantly reduced (Fig. 6b). One effect of this reduction in energy is that the primary source of modulation is now the 215 Hz pulley frequency instead of the 18 Hz plunger frequency. As a result of the new dominant source of modulation, the phase-modulated current data assumes an appearance similar to that of the vibration data.

Over-tightened packing defect

For the second fault simulation packing for one of the four plungers was overtightened until an audible rub was detected. The packing was then loosened until the rubbing noise was just barely audible. Data was then collected for the second induced fault. Neither of the faults were aurally detectable. The over-tightened packing defect (Fig. 7) resulted in differences that were apparent in data collected from the cylinder-head accelerometers and motor current sensor. At first glance, the vibration data of the overtightened packing defect (Fig. 7b) appears almost identical to that of the defective discharge valve. However closer inspection reveals that the increase in broadband energy that is modulated by the 215 Hz pulley frequency now occurs between the second (9 Hz) and third (13.5 Hz) pump rotation harmonics. This contrasts with the same type of increase in broadband energy that occurs between the fundamental and second harmonic in the case of leaking discharge valve.

Similar to the vibration data, the amplitude modulated current data of the overtightened packing defect (Fig. 8b) exhibits an increase in second (9 Hz) and third (13.5 Hz) pump rotation harmonics, while fundamental pump rotation (5.5 Hz) remains unaffected. However, in contrast to the vibration data, the plunger frequency remains the dominant source of modulation.

4. CONCLUSION

We have shown that for the specific instance of a cam-operated pump, time-frequency analysis provided detail not only to observe the simulated faults, but also that the pump was not operating perfectly in its baseline condition. Differences between the baseline and defective conditions were easily observed for the vibration and motor current signals. In addition, amplitude modulation of narrow

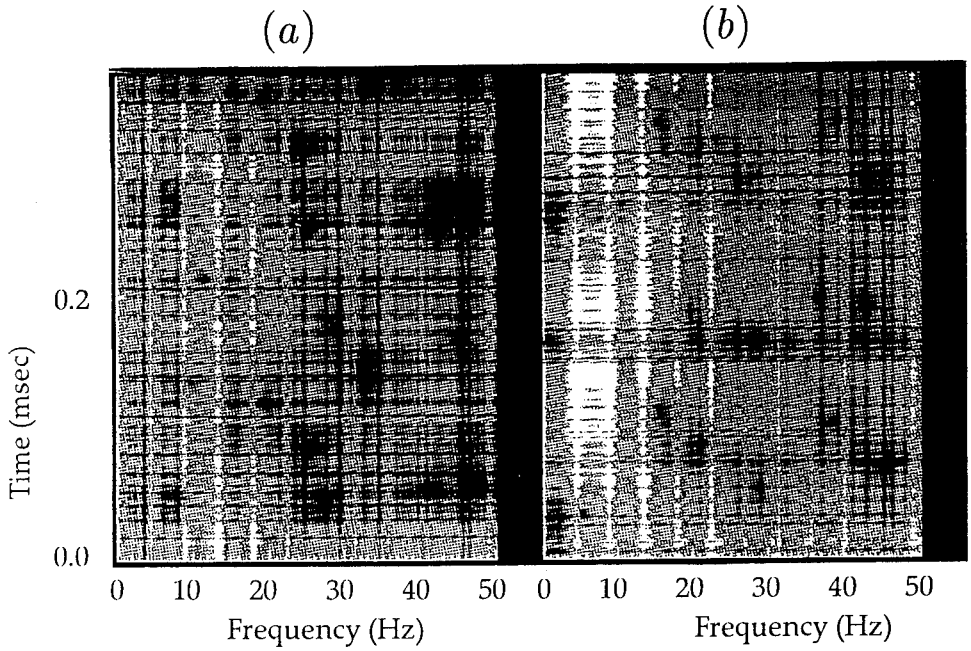


Fig. 5. Leaking discharge valve defect using vibration signal. (a) and (b) are the baseline and defect distributions respectively.

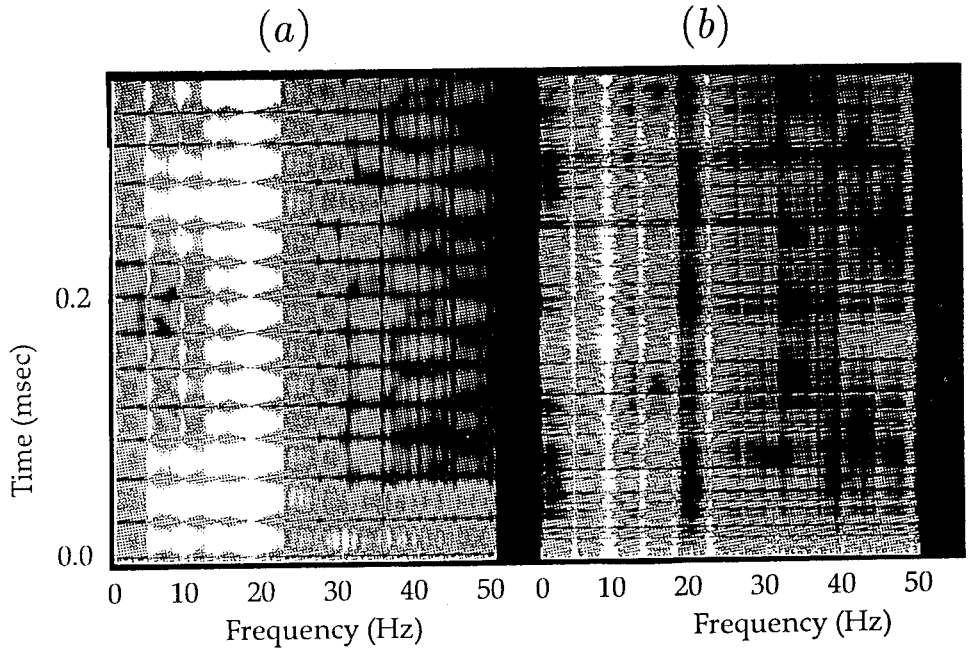


Fig. 6. Leaking discharge valve defect using phase modulated current data. (a) and (b) are the baseline and defect distributions respectively.

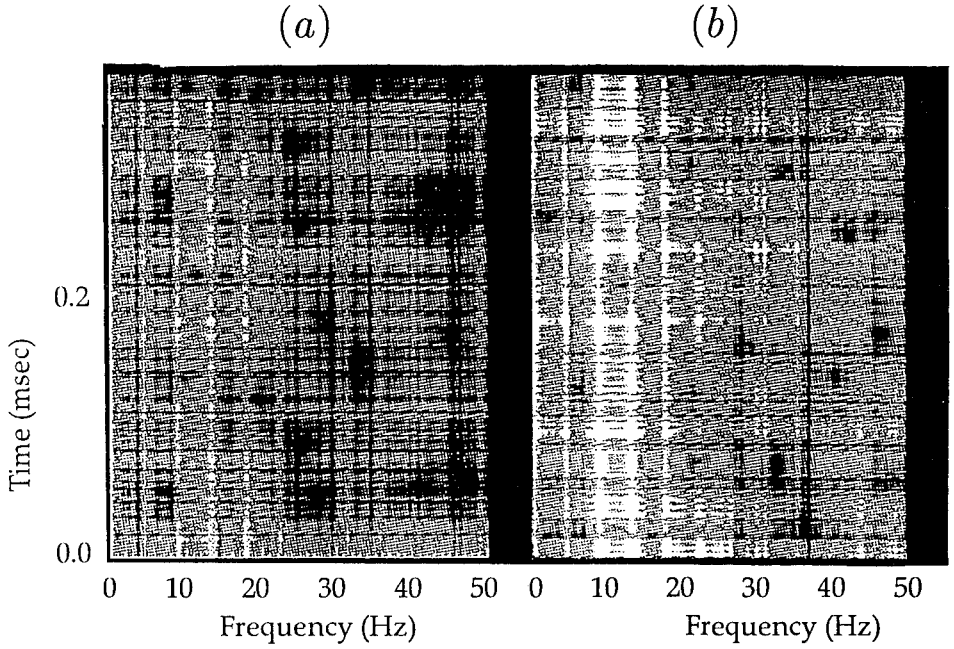


Fig. 7. Overtightened packing defect using vibration signal. (a) and (b) are the baseline and defect distributions respectively.

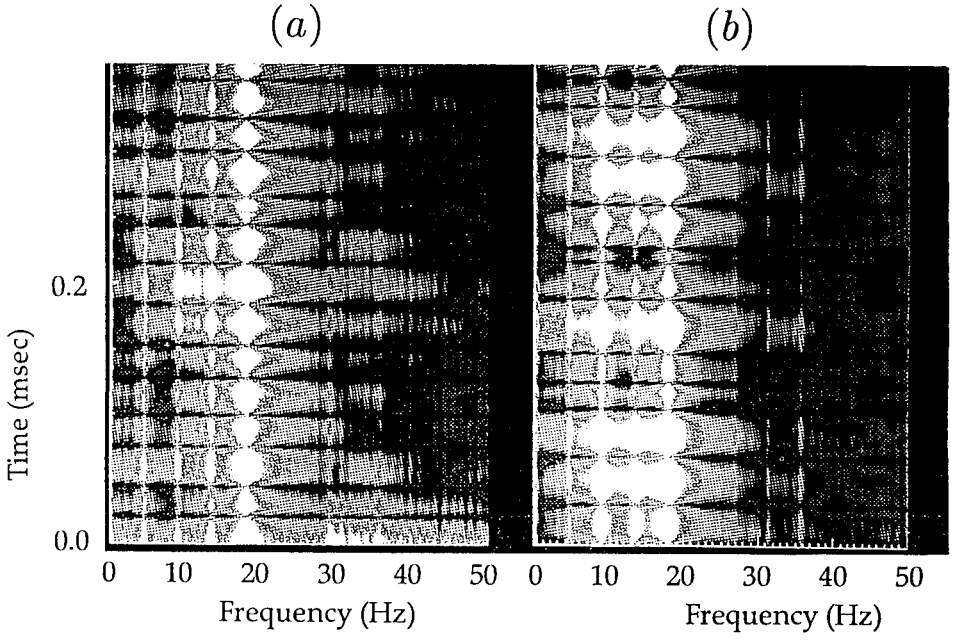


Fig. 8. Overtightened packing defect using phase modulated current data. (a) and (b) are the baseline and defect distributions respectively.

band frequency components was identified, which would be a very difficult task using conventional analysis methods.

An important point is that the time-frequency distribution we have used is manifestly positive. This opens up the possibility of comparing regions in the time-frequency plane, that is, comparing time-frequency energy levels of a machine to previously recorded levels, similar machines, and calibrated reference levels for trending. The ability to compare regions in time-frequency opens up many possibilities because it gives us a greater base for classification, since we are no longer just comparing changes in the spectrum which is a one dimensional function. It is very likely that faults can be characterized better not only by their change in frequency but also by the time of occurrence of the change. Therefore, two faults that both produce change in certain frequency ranges may possibly be different if they produce them in a somewhat different time interval of the cycles involved. Time-frequency analysis may therefore become an important tool in a comprehensive machinery condition assessment program.

Acknowledgment. The authors would like to express their appreciation to Myron Aldridge and Richard Young of Naval Sea Systems Command and Douglas Henry of NSWC for supplying the data on the cam operated pump.

REFERENCES

[1] H. I. Choi and W. J. Williams, "Improved time–frequency representation of multi-component signals using exponential kernels," *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 37, pp. 862–871, 1989.

[2] L. Cohen, "Generalized phase–space distribution functions," *Jour. Math. Phys.*, vol. 7, pp. 781–786, 1966.

[3] L. Cohen and T. Posch, "Positive time–frequency distribution functions," *IEEE Trans. ASSP*, vol. 33, pp. 31–38, 1985.

[4] L. Cohen, "Time–Frequency Distributions – A Review," *Proc. of the IEEE*, vol. 77, pp. 941–981, 1989.

[5] L. Cohen, *Time-Frequency Analysis*, Practice–Hall, 1995.

[6] G. S. Cunningham and W. J. Williams. "High–Resolution Signal Synthesis for Time–Frequency Distributions," *Proc. IEEE ICASSP– 93*, vol. 4, pp. 400–403, 1993.

[7] J. Jeong and W. J. Williams, "A new formulation of generalized discrete–time – frequency distributions," *Proc. IEEE ICASSP– 91*, pp. 3189–3192, 1991.

[8] M. D. Ladd, and G. R. Wilson, "Proportional Bandwidth Properties of Fault Tones in a Ball Bearing System", in: *Proc. of 1994 Asilomar Conference on Signals, Systems and Computers*, to appear, 1994.

[9] P. Loughlin, J. Pitton and L. E. Atlas, "Bilinear time–frequency representations: new insights and properties," *IEEE Trans. Sig. Proc.*, vol. 41, pp. 750–767, 1993.

[10] P. Loughlin, J. Pitton and L. E. Atlas, "Construction of positive time–frequency distributions," *IEEE Trans. Sig. Proc.*, vol. 42, pp. 2697-2705, 1994.

[11] D. Rock, D. Malkoff, and R. Stewart, "AI and Aircraft Health Monitoring", *AI Expert*, vol. 8, pp. 29-35, 1993.

[12] R. Rohrbaugh, "Application of time–frequency analysis to machinery condition assessment, *Proc. 27th Asilomar Conf. on Sigs., Syst. and Comps.*, vol. 2, pp. 1455–1458, 1993.

[13] R. Rohrbaugh, "Advanced Time-Frequency Analysis: Theory and Application to Machinery Condition Assessment," *Naval Surface Warfare Center*, Technical report SADP-U93/00039-715, 1993.

[14] G. Rizzoni and X. C. Chen, "Detection of Internal Combustion Engine Knock Using Time–Frequency Distributions," *36th. Midwest Symposium on Circuits and Systems*, vol. 1 pp. 360–363, 1993.

[15] B. Samimy and G. Rizzoni, "Time–Frequency Analysis for Improved Detection of Internal Combustion Engine Knock," *Proc. IEEE–SP Int. Sympos. Time–Frequency and Time–Scale Analysis*, to appear, 1994.

[16] B. Samimy and G. Rizzoni, "Experimental Studies of Internal Combustion Engine Knock Detection," *SAE 1995 International Congress and Exposition*, to appear, 1995.

[17] Y. S. Shin, J. J. Jeon, and S. G. Spooner, "Pseudo Wigner–Ville Distribution and its Application to Machinery Condition Monitoring", *Proceedings of the 47th Meeting of the Mechanical Failures Prevention Group*, pp. 235–250, 1993.

[18] W. J. Wang, and P. D. McFadden, "Early detection of Gear Failure by Vibration Analysis – I. Calculation of the Time Frequency Distribution", *Mechanical Systems and Signal Processing* vol. 7, pp. 193–203, 1993.

[19] V. Wowk, *Machine Vibration*, Mcgraw–Hill, 1991.

[20] Y. Zhao, L. E. Atlas, and R. J. Marks, "The use of cone–shaped kernels for generalized time–frequency representations of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1084–1091, 1990.

DETECTION, MONITORING AND RESPONSE

Cochairmen: S. Nils Straatveit
Purvis Systems, Inc.

James E. Birdsall
SynEx, Inc.

FAILURE ANALYSIS OF A COMPRESS MELT UNIT FROM THE CONTROL SYSTEM PERSPECTIVE

John Dentler
AmDyne Corporation
1121 B Benfield Blvd.
Millersville, MD 21108-2541

INTRODUCTION

The Navy Compress Melt Unit (CMU) prototypes designed and built by Westinghouse MTD encountered a self destructive catastrophic failure which was not predicted by the Failure Modes Effects and Criticality Analysis (FMECA). The primary cause of the failure was found to be improper material used for the actuator plate. The actuator plate was redesigned and made of the proper material to prevent repetition of this failure on future units.

NAVSEA and NSWC commissioned a second "Poor Man's FMECA" from the perspective of the host service system, control system and sensor suite. This analysis focuses on potential modes of self destruction. As a result of this analysis, the control system was modified to improve the protection of the machine.

A set of tests were proposed and executed to answer the questions posed in the second analysis to ensure that the modified control system protects the machine as much as possible. The control system was further revised as a result of these tests.

During the redesign stage, consideration was given to using a smaller motor for the drive system, and simplifying the machine by removing the mechanical brake which is interlocked with the drive motor. These revisions were tested and not implemented.

This paper:

1. Describes the CMU and its operating modes and catastrophic failure mode.
2. Describes the steps in the Host System Service, Control System, and Sensor Suite Analysis approach.
3. Summarizes the Test Procedures, Results and the Motor Current Data which was used to check and revise setpoints in the control system. Of specific interest are the data taken when:
   a. The ram stalled against the door with, and with out the motor current trip. This data shows how the motor current linearly reflects the system response in terms of plate position and spring compression. It is included to validate motor current as an accurate measure of load in the system.
   b. The Lower Lead Screw Grade 8 bolt was damaged in the eject mode. This data shows how the revised motor current setpoint will prevent this damage
   c. The lead screw pushed through the bottom plate of the Lower Support Housing. The breakage occurred with such ease that the lower support housing design was modified to eliminate the bottom plate.
   d. The Ram Tee Plate and bolts were slightly damaged in the retract mode with the motor stalled (motor current sensor bypassed) and Ram in mid position.

1. Background Data:

A. How the CMU Operates

The Compress Melt Unit (CMU) compresses and heats shredded waste plastic. The process reduces large bags of food contaminated waste plastic to a stable disk, 22 inches in diameter and 2 to 4 inches thick. The ship stores the disks in odor barrier bags until safely off loaded.

The CMU has a cylindrical chamber with a ball screw driven ram for compression, and ejection. A patented limit switch and actuator plate system regulates compression pressure. The door, ram and chamber walls are heated during compression. The disk is cooled for safe handling prior to ejecting.

Figure 1 shows both conceptual and functional views the Navy Compress Melt Unit.

Figure s 2a through f show the steps for using the unit to reduce the volume and stabilize the plastic waste.

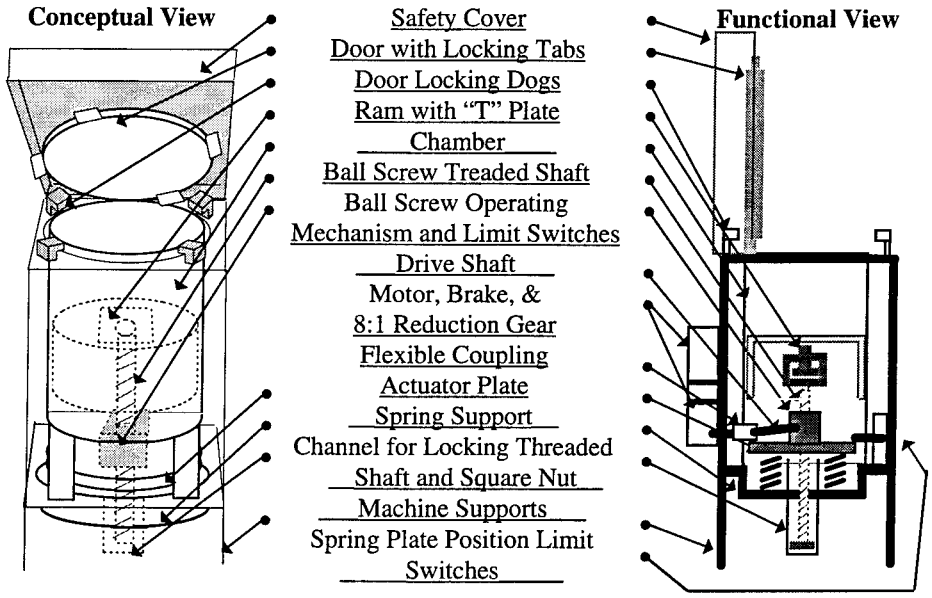


Figure 1: Conceptual and Functional views of the Navy Compress Melt Unit. An open foundation supports the unit. A motor driven ball screw compresses plastic between a ram and a door. The door is locked in place under locking dogs. Electric heaters and cooling coils are imbedded into the castings for the ram and door. A removable jacket with heaters and cooling coils surrounds the chamber.

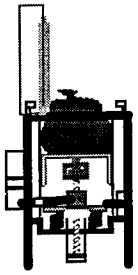


Figure 2a. The operator loads bags of shredded plastic into the chamber with the ram retracted. The springs are relaxed.

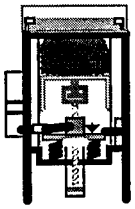


Figure 2b. The operator closes the door and the safety cover. A pneumatic cylinder rotates the door to engage the latching mechanism. The operator presses the "process button" to begin compressing and heating the waste plastic. The ball screw drives the ram forward. The door, ram and chamber heaters energize and heat to their designated values.

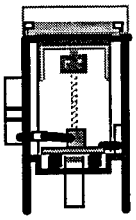


Figure 2c. The unit simultaneously heats and compresses the waste plastic. As the ball screw compresses the plastic, the spring plate depresses and activates the limit switch de-energizing the motor. If the plastic expands while heating, the actuator plate depresses against the springs. Sufficient expansion will activate a second limit switch, which drives the Ball Screw in the reverse direction to relieve pressure. As the plastic softens, the spring plate rises, and releases the pressurizing limit switch. The motor drives forward to compress the plastic. Resistance Temperature Detectors (RTD's) sense the temperature and feed a Programmable Logic Controller (PLC). Thermal switches backup the RTD control system. The unit heats the plastic for a designated period of time, then cools the plastic for safe handling.

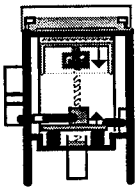


Figure 2d. After processing, the motor reverses to relieve the pressure on the plate, Ram and Door. A pneumatic cylinder rotates the door to the unlatched position.

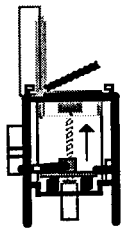


Figure 2e. The operator opens the safety cover and the door, and presses the eject button. The drive system moves forward so that the top of the ram is above the chamber wall. The operator removes the processed plastic disk.

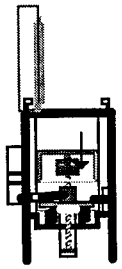


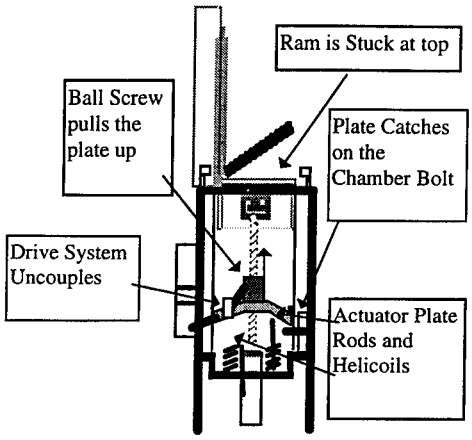
Figure 2f. The operator retracts the ram by pressing the retract button. The operator cleans and prepares the chamber with mold release for processing the next batch of plastic.

Figures 2 a through f: Operation of the Navy Compress Melt Unit.

B. Catastrophic Failure Mode

Food waste and plastic accumulates between the ram the chamber. If allowed to cool, the accumulated waste binds the ram to the chamber. The procedure for operating the machine requires the operator to remove the plastic disk, clean the top and exposed sides of the ram, retract the ram, and clean the chamber prior to shutting down the machine.

If these procedures are not followed, and the ram is left in the eject position after processing, the ram will be bound at the top of the chamber by the cooled food and plastic glue. The next step in the process is to retract the ram to the load position. The drive system distorted the spring plate when attempting to retract the bound ram. The actuator plate rods pulled out of their helicoils. The plate caught on bolts which hold the chamber in place and bent. The plate distorted far enough to separate the drive system at the flexible coupling. By design, the failure point should have been at the T plate which holds the ram onto the ball screw.



If procedures are followed, the binding occurs in the fully retracted position. From this position, the ram drives up, compressing the springs and stressing the machine where it is designed for strength.

The strength of the combined plastic and food had not been anticipated by the designer. The control system and sensor suite had not been designed to detect the binding which develops between major cleanings, and the predicted failure points had not been tested.

The control system was revised to:

1. Preheat the chamber prior to any movement
2. Test for a jam condition prior to retracting
3. Alert the operator of a "Ram Jam" Condition
4. Incorporate a motor current sensor which backs up the jam test and limit switches and alerts the operator to perform a major cleaning which require ram removal.

In parallel with these revisions, an analysis was performed to predict failure modes and the adequacy of the control and sensor suite. Based on the analysis, and testing:

1. The setpoint for motor current in eject was revised.
2. A failure prone structural part was eliminated from the system
3. The circuit breaker feeding the system was changed to include a shunt trip device to prevent fire should a heater controller fail to open, and the control system was changed to support this breaker.
4. A level switch was added into the closed loop cooling system logic to protect the pump should a leak occur.

One predicted problem was deemed acceptable for operator response:

1. A loss of Control System Air disables the cooling system and prohibits rotation of the door to either the locked or unlocked position. The operator is warned if the cooling system is experiencing a problem and warned if the door fails to operate as expected.

2. Approach for Determining Failure Modes from the Host Service System, Control System and Sensor Suite Perspective.

This approach to protecting a system from self inflicted failures begins by examining the processes and conditions that energize and de-energize the main power devices in the system. The first step establishes a list of main power controllers, and lists the response of each controller for each operational mode. The second step assumes the relays and sensors associated with these main controllers fail to protect the system as designed. Assuming these failure occur, then predict the mechanical or electrical components that are stressed for each failure. (Motor Overloads trip, a mechanical part breaks, or heaters cause a fire.) The third step checks the proposed layers of protection provided by the sensors and control system against these failures. This analysis should reveal the single point failure conditions where catastrophes occur. Incorporation of additional sensors and controllers, or mechanical fuses should prevent the most serious failures.

Summarizing: For each main power controller:

1. List each mode of system operation.
2. Estimate what happens if any controller fails to turn off or turn on in each mode. Determine which events are not inherently fail-safe, and which components are stressed.
3. List the sensors that turn on and turn off each controller in each mode. Determine if these sensors provide adequate layers of protection to prevent damage.
4. Where the sensor suite and control system fail to provide more than a single layer of protection, determine the appropriate solution such as:
   a. Design a mechanical fuse that is relatively easy to repair.
   b. Ignore the problem because protection against this casualty is not cost effective.
   c. Determine that adequate operator training will allow safe operation.
   d. A simple routine maintenance check will minimize the likely hood of this particular casualty.
   e. Modify the sensor suite or control system to protect the system from this casualty or combination of casualties.

A similar approach is valid for determining the system response to a loss, and subsequent recovery, of services provided to the equipment from the host platform (ship, building, facility, etc.). Host services include electricity, low and high pressure air, fresh water, salt water, drainage, supply air, exhaust piping, air conditioning and heating.

Most systems are protected from a loss and subsequent recovery of electrical power by Low Voltage Protection (LVP) controllers. (The device stays off until intentionally restarted.) The system response to loss and recovery of other services such as Control Air and cooling water should be just as safe if required. The specific services used by the system are known early in the design phase. Analysis starts with listing each service and combination of services that could fail during each mode of operation. Determine if the system responds in a safe condition if the service fails. Determine the sensor suite required to protect the operator or the machine from a service failure? Assume the service is restored after the failure with no operator action. Does the system respond in safely? Follow through the casualty with no sensors, and the planned sensor suite. The analysis will show where increasing or reducing sensors is appropriate to ensure economical, fail safe, operation. The analysis should continue as the design evolves to incorporate more or less services from the host.

Summarizing: For each service provided by the host platform:

1. List the Modes of operation.
2. Estimate what happens if any service fails during any mode of operation. Determine which events are not inherently fail-safe.
3. Consider the effects of an unexpected restoration of each service after failure. Determine which events are not inherently fail-safe.
3. If a Sensor is designed to measure the condition of a service, determine the failure mode of the sensor and if the failure is safe or not.
4. Where the sensor suite and control system fail to provide more than a single layer of protection, determine the appropriate solution such as:
    a. Design a mechanical fuse that is relatively easy to repair.
    b. Ignore the problem because protection against this casualty is not cost effective.
    c. Determine that adequate operator training will allow safe operation
    d. A simple routine maintenance check will minimize the likely hood of this particular casualty.

e. Modify the sensor suite or control system to protect the system from this casualty or combination of casualties.

There are some subtle points which must be considered when performing this analysis:

1. There may be many sensors and many setpoints that turn off or turn on a particular controller. However, a single point failure could block all of the control signals if they all use the same relay, control line, coil, springs, and contacts in the controller. Whether or not a single point failure in any of these components will result in a catastrophic, failure, or an acceptable failure, is a question for the designer to resolve.
2. Motor Overloads may not adequately protect a Mechanical system. Properly sized overloads will typically carry Locked Rotor Current for 20 seconds or more. The large margin allows starting transients and motor operation over a wide range of temperatures. The system must protect itself from unacceptable mechanical failures should the controller fail in the energized condition and stress the system with a locked rotor torque.

3. The ultimate failure mode in any system with heaters is a catastrophic fire. The designer should install a second device to interrupt power in series with the normal controller to avoid a catastrophic single point failure of the heater controller.

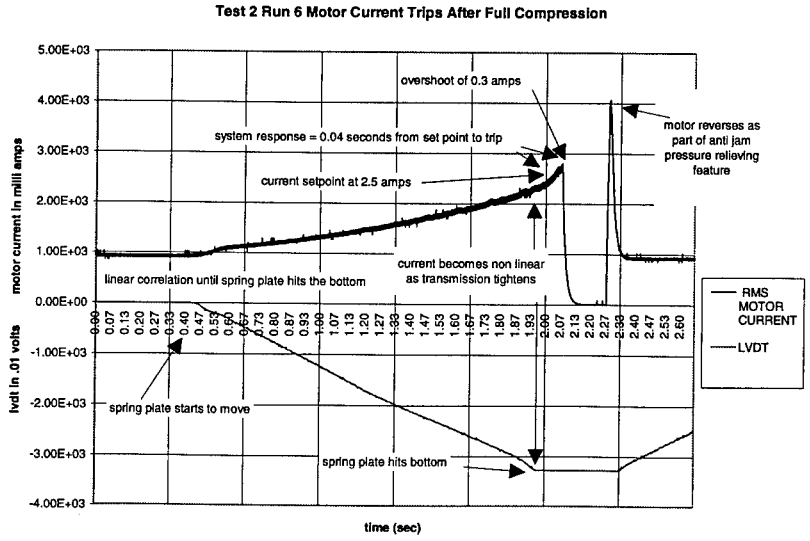
4. Test Summary, Procedures, and Results

Test 1. Several Food Contaminated Disks were processed to test Food Stresses in Retraction Direction. After an initial build up, the jam test was tested against "Cold" food to Ensure that all added control changes act to alert the operator and protect the Machine prior to Machinery Damage:

Results of Test 1. Heating the Chamber appeared to increased the number of bricks processed between cleaning with cold shut downs (6 to 13)
The Jam which drives the Ram forward successfully detected the binding and alerted the operator to clean the CMU prior to any failures.
Motor Current successfully detected jams when the limit switch failed to respond.

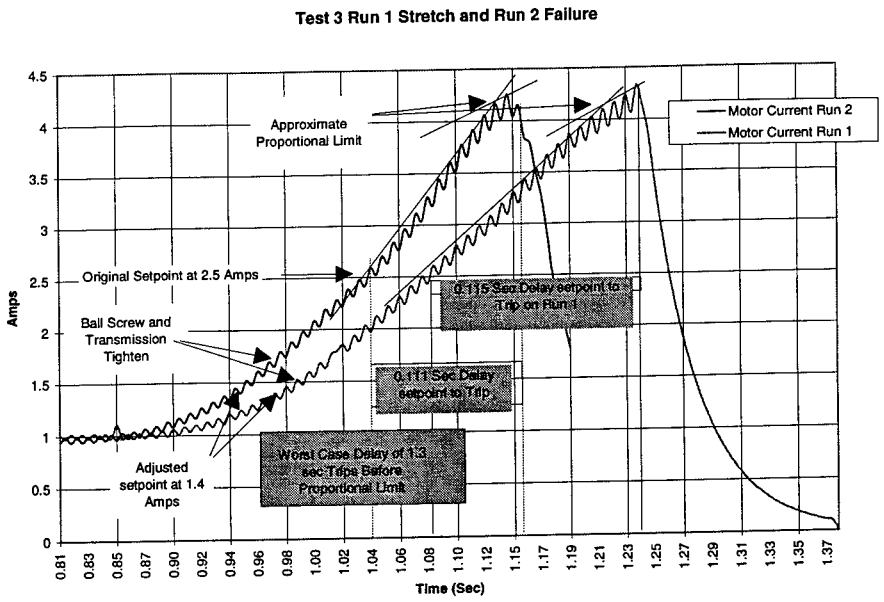
Test 2. Sensor Failures in the Forward Direction against the Door: The LS4 and 5 interlock was tested to indicate Failed Limit Switch.
The over current protection set point was tested to protect the machine in the event of simultaneous LS4 and LS5 Failures.

Results of Test 2. The Eject Limit Switch was checked and found to not trip triggered when driving to a stall against the closed door.
The system was tested and found strong enough to stall the motor and trip overloads when compressing against the door.
The motor current linearly tracts as the actuator plate is compressed, then rapidly rises when the actuator plate is fully depressed as shown below:



Test 3. Sensor Failures in the Forward Direction while Ejecting

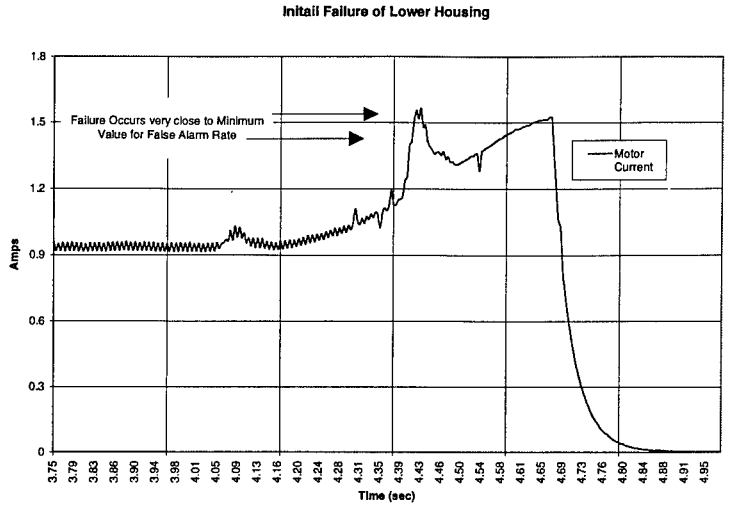
Results of Test 3. The motor current setpoint in the Forward Direction was tested to determine if damage occurred should fail while ejecting through an Open Door. Damage did occur, and the motor current setpoint was reset to low enough to prevent exceeding the proportional limit of the Grade 8 bolt as shown below:



Test 4. Retraction with a failed Retract Limit Switch

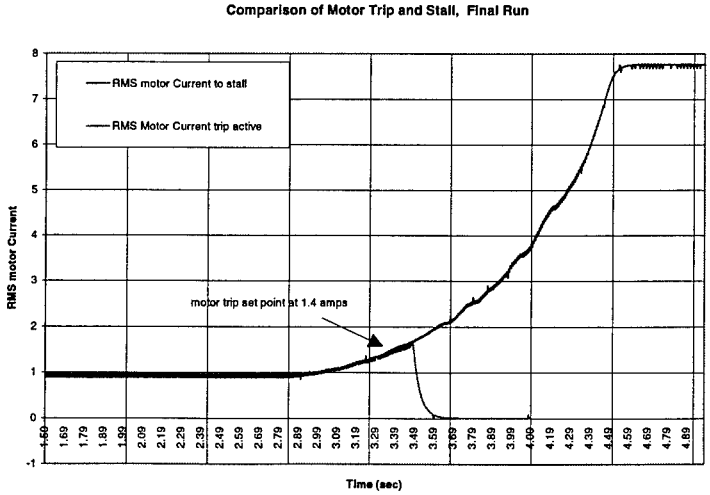
The Ram was retracted with the retract limit switch bypassed.

Results of Test 4. The bottom of the ball screw punched through the lower casting without tripping the motor current setpoint. (The current exceeded the setpoint for a short burst, however damage occurred before the sensor could respond.) The design was changed to eliminate the end plate from the lower casting, thus improving cleaning accessibility of the machine. Once the casting was broken, the motor was safely stalled by driving the bottom of the Tee plate assembly against the top of the ball screw housing.



Test 5. Sensor Failure while Retracting: Motor Current Failure with a Blockage in the Reverse Direction. The Ram was stalled in the Reverse direction, by placing Aluminum blocks on the inside of the chamber supports which stopped the ram prior to reaching the full retract position.

Results of Test 5. The Motor Current sensor stopped the motor prior to Damage. When the Motor Current Sensor was bypassed, the stainless steel bolts at the top Tee plate were slightly damaged (the Allen head sockets were distorted, but the bolts held and are still being used).



Conclusions:

The type of Failure Modes Effects and Criticality Analysis typically performed on a developmental system may skip some very important issues dealing with the main controllers, sensor suites, and service systems. Typically, systems are broken into components rather than viewed as systems with a total system response.

The Guide Lines listed in this report give potential customers of such failure analyses some serious questions which should be answered during the design phase of a typical system.

The testing shows that a typical RMS. current sensor into a programmable will have between 0.1 and 0.2 seconds delay after reaching the trip setpoint until power is disconnected from the system.

After performing the analysis and subsequent testing on the Compress Melt Unit, the system developer has

1. Assured himself that at least two problems must occur simultaneously to create a catastrophic failure due to self destruction.

2. That the Compress Melt Unit is as robust as economics will allow.

COMPENSATING MACHINING ERRORS WITH ACTIVE STRUCTURES

Michael D. Bryant, Richard Friedrich, T. S. West, and Benito Fernandez
Department of Mechanical Engineering
University of Texas
Austin, Texas 78712-1063

Abstract: In lathes, a carriage that supports the tool post slides across two very straight and flat parallel tracks or ways. After several years wear, the ways develop a concave and uneven profile which induces normal, roll, and pitch errors of the tool post with respect to the workpiece. A mil of this and other "slop" can seriously degrade cutting performance and misalign the machine. Components can be replaced or refurbished, but this is very expensive and of no use for worn machines still operating.

Presented is the development of active structures to correct these and other errors such as vibrations. Sensors appropriately mounted monitor errors and over time a neural network "learns" the error profiles. Structure types include a table with active legs and a truss with active links: active elements deform to compensate errors. The table structure is mounted on a lathe carriage; roll, pitch, and vertical displacement errors can be compensated by adjusting the leg lengths. A two member active truss, mounted atop the table, can move a point within the tabletop plane. Statically, all structures are very stiff. During operation, a neural controller constantly ``relearns" and compensates the structures as errors accumulate from aging and wear.

Key Words: Active structures; Machining; Magnetostriction; Neural Networks; Terfenol; Vibration control

Introduction: The ways of a machine tool need only wear a few thousandths of an inch before medium precision lathes become troublesome and high precision lathes useless. Replacement or refurbishment of machine tools or components is expensive. This article offers a more economical solution wherein an active structure adaptively compensates for errors due to wear, manufacturing, calibration, and vibrations.

Principle components of an engine lathe are the bed, headstock, tailstock, spindle gearbox, feed gearbox, apron and carriage. Parts are often mounted on a cast iron bed. The ways are inverted v-shaped track structures on the upper surface of the bed. The carriage rides over the ways to traverse the lathe longitudinally (along the z direction of figure 1). Dovetail ways on the top of the carriage allow the mating cross-slide to travel perpendicular to the spindle axis (along the r direction of figure 1). The compound rest consists of a swivel base and tool post slide and is used for cutting tapers and chamfers. The tool post, mounted atop the tool post slide, supports and carries the cutting tool.

Sliding contact between the carriage and ways unevenly wears the ways. Measurements in figure 2 on a Boye & Emmes 16 inch engine lathe with badly worn ways shows the effects of wear on traversal along z of the tool carriage. As the carriage approaches a worn location, it rolls, pitches and/or descends as the carriage rides over the contour of the ways. An error map of the wear pattern as a function of location along the z-axis is plotted in figure 2. A capacitance gauge with 0 to 50 micron range and 1 micron resolution measured vertical tool movement Z versus z with respect to a reference bar of tool steel, ground flat and mounted on the bed parallel to the ways. Angulations of roll R and pitch P of the tool

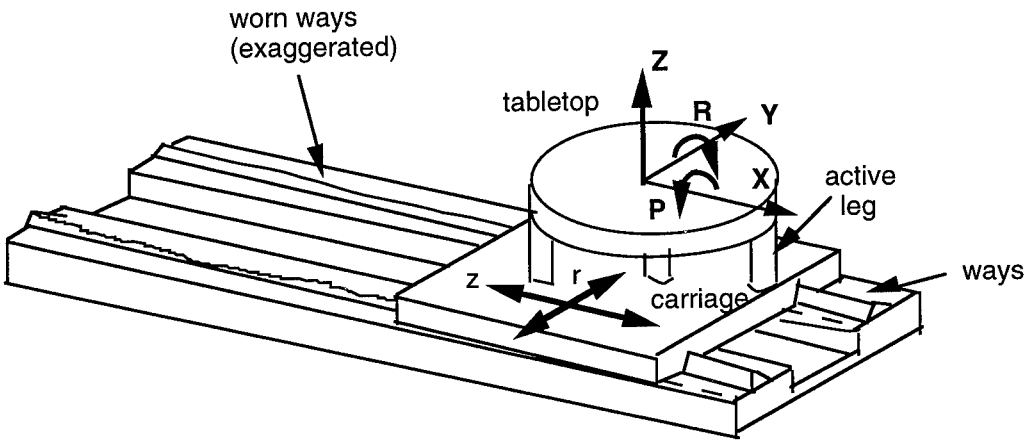


Figure 1: A tool carriage slides along two parallel ways (tracks). Uneven track wear displaces the carriage (normal Z, roll R and pitch P) and misaligns the upper surface with respect to the chuck. A table structure (with the toolpost mounted on top) with three actively strainable legs compensates these errors and corrects tool post alignments. Additional structures (truss) can compensate other errors (X and Y).

holder with respect to the lathe were measured by two 0.001 degree resolution inclinometers mounted on the tool carriage with their axes directed along the X and Y axes (see figure 1). These inclinometers established a reference plane with a normal vector parallel to the gravitational field vector.

In figure 2, the concave worn profile W = Z (z) is deepest (almost 60 μm) near the middle of the 0.62 m carriage range traversal. The reference point z = 0 is at the headstock. Measurements over several to and fro excursions of the carriage along the bed were conducted to check repeatability and to detect possible hysteresis patterns in the motion. The to and fro curves in figure 2 nearly superimposed (the thick lines represent many to and fro measurements), indicating that the traversal direction has small effect. The inclinometer data indicates that pitch and roll are roughly related.

Permissible bed wear on general purpose lathes is about 0.02 mm per 1000 mm of length. Permissible height differences in the ways is about 0.05 mm per 1000 mm of length. Wear accumulates through use and is exacerbated by chips and other small hard particles abundant in a machine shop. Although cross slide wear can be compensated by adjusting tapered gibs, there are no tapered gibs on the carriage to compensate worn ways. In addition, total wear of the ways depends on location: a single adjustment could not correct along the entire length of the ways. Expensive repair procedures first builds up the worn ways (using welded metal) and then scrapes them to within tolerance.

Other errors are present. Rotational error components in machine tool slides give rise to translational errors directed along X and Y (figure 1) different for every point on the body. Rotational errors, due to the machine tool not being collinear with the machine scale, are proportional to the perpendicular distance between the tool point and the machine scale. Vibrations degrade machining accuracy and/or mars surface finish. Vibrations can originate externally or be part of the machining process (chatter). To minimize vibrations, lathes are usually massive and stiff. However, a worn spindle bearing or an inadequately supported cutter may permit tool chatter vibrations.

Instead of replacing or refurbishing the machine tool, this article outlines features of an active structure that supports, guides, and corrects alignments of a cutting tool to

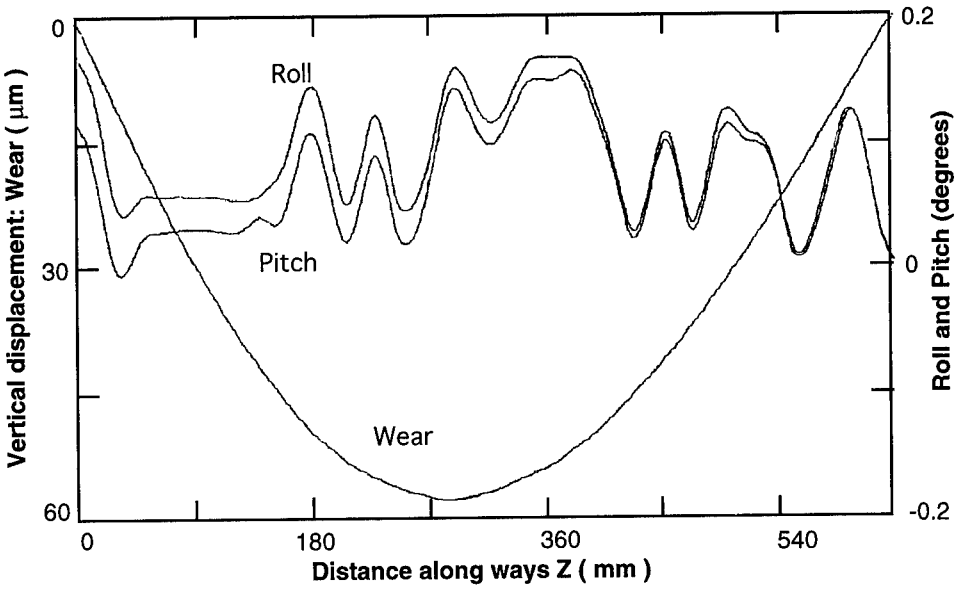


Figure 2: Wear (μm) versus distance z (mm) along the ways of an aged lathe. The concave curve is the worn profile; the two other traces represent tool carriage roll (degrees) and pitch (degrees). The point z = 0 is at the headstock.

compensate errors on worn or new lathes. The structures, assembled from magnetostrictive actuators and controlled by Artificial Neural Networks, retrofit to existing machine tools. The structures can actuate and correct misalignments through degrees of freedom orthogonal to and outside the reach of the two degrees of freedom usually present on a lathe. The structures can adapt to changes in the machine tool and upgrade precision and extend machine tool life, since even severe misalignments caused by component wear---which would normally junk the machine---can be corrected.

Active Structures Background: A very stiff, linear motion actuator (forces always along the same axis) utilizing giant magnetostrictive strain terfenol-d (a mixture of terbium, dysprosium, and iron) was developed into a basic structural element [1],[2]. Passively this terfenol actuator/structural element could support large loads (kN's). By applying 0 to 20 A currents to the roughly 200 turn coil that surrounds the magnetostrictive rod within each actuator, precise displacements of one to hundreds of microns with forces of hundreds of Newtons can be synthesized over DC through 20 kilohertz bandwidths.

Structures can be assembled from the actuator/structural elements. One basic structure built from these elements is a three-legged table [3] depicted in figures 1 and 3; actuator legs support a very stiff tabletop. Another [2] pins two linear actuators together to form a planar truss (figure 4). The pin joints, which press fit teflon coated pins into teflon coated holes, transmit axial forces--not moments--between elements while exhibiting almost free rotation about the pins, with only sub-micron pin/hole clearances (backlash). The three legged table compensates normal, roll and pitch errors Z, R and P; the two member truss compensates translations X and Y within the plane of the truss. If the truss was mounted atop the table, together these structures could compensate five degrees of freedom (3 translations + 2 rotations) at the tool post, both statically and dynamically (DC to 20 kHz). If needed, a

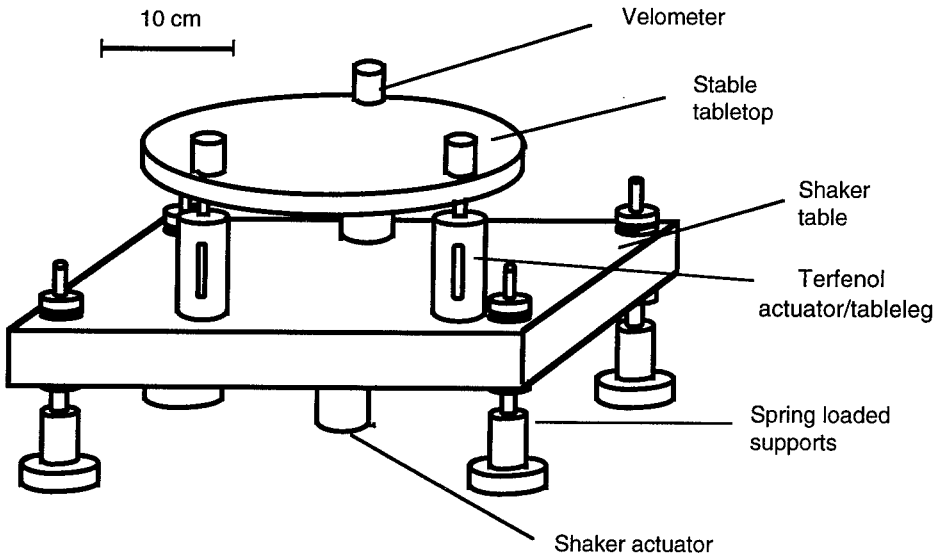


Figure 3: A three legged table with legs that are also magnetostrictive actuators. Velometers sense tabletop disturbance motions supplied by the underlying shaker system and actuator legs provide compensating motions.

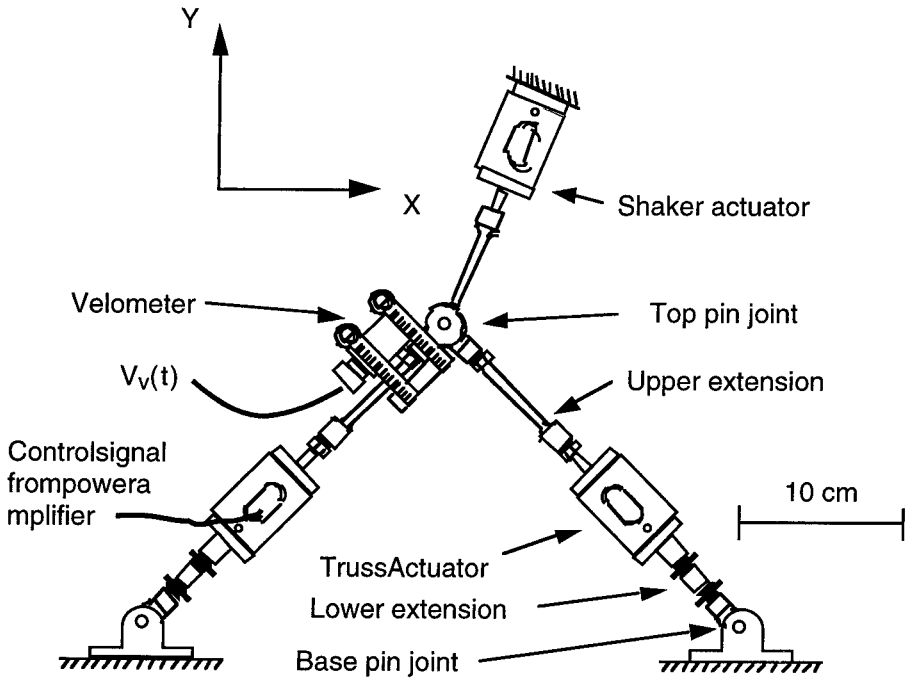


Figure 4: Two member active truss. Truss links are magnetostrictive actuators pinned together. Velometers sense truss link disturbance motions supplied by the shaker actuator and terfenol actuator truss links synthesize in plane compensating motions X and Y of the top pin joint.

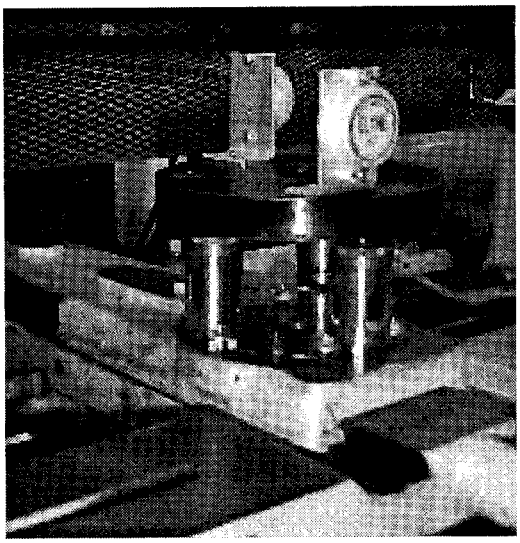


Figure 5: Photograph of the table structure (with toolpost removed) mounted atop the tool carriage. The three active legs are below the tabletop and the two inclinometers are above.

sixth degree of freedom---rotations about the vertical Z---could be compensated by a pair of actuators rotating the tool post about the vertical.

Apparatus and Design Details: A photograph of the table mounted atop the tool carriage is shown in figure 5. Design details are contained in reference [4]. The table was designed to deflect only a few microns under a 4.4 kN load, or about 2% of the maximum wear in figure 2. To insure that the cutting edge of the tool remained on the original centerline of the lathe (collinear), the actuators were approximately the same height as the compound rest.

Static Compensation: As the carriage moved along z, control signals $C_i = C_i(z)$ were dialed in to each of the three tableleg actuators to synthesize corrective leg displacements. To within the resolutions of the capacitance gauge and inclinometers, the structure easily compensated all static errors Z, R and P (figure 2) at the toolpost for the tool carriage positioned at the headstock (z = 0), center, and tailstock (z = 0.62 m).

Dynamic Compensation: To gauge the overall controls problem inherent in the active structure and carriage system, frequency response measurements of vertical displacement Z (see figure 1)----activated by applying a sinusoidal current to the activation coil of one of the tableleg actuators----were conducted at different locations z along the ways. The Bode plots of figure 6 show dynamic response near the headstock z = 0 (figure 6a) and tailstock z = 620 mm (figure 6b).

In both of figures 6, the magnitued peak near 200 Hz indicates a moderately damped natural frequency. Numerous smaller peaks (one at 45 Hz not shown) at various frequencies suggest other resonant modes. In figure 6b compared to figure 6a, there are more resonant peaks as well as a drop in the coherence of the calculation. One possible explanation is the that carriage and way surfaces at the headstock (figure 6a) have worn together to form profiles that better conform than at the tailstock. A test eliminated this degree of freedom: two large C-clamps restrained rocking of the carriage relative to the

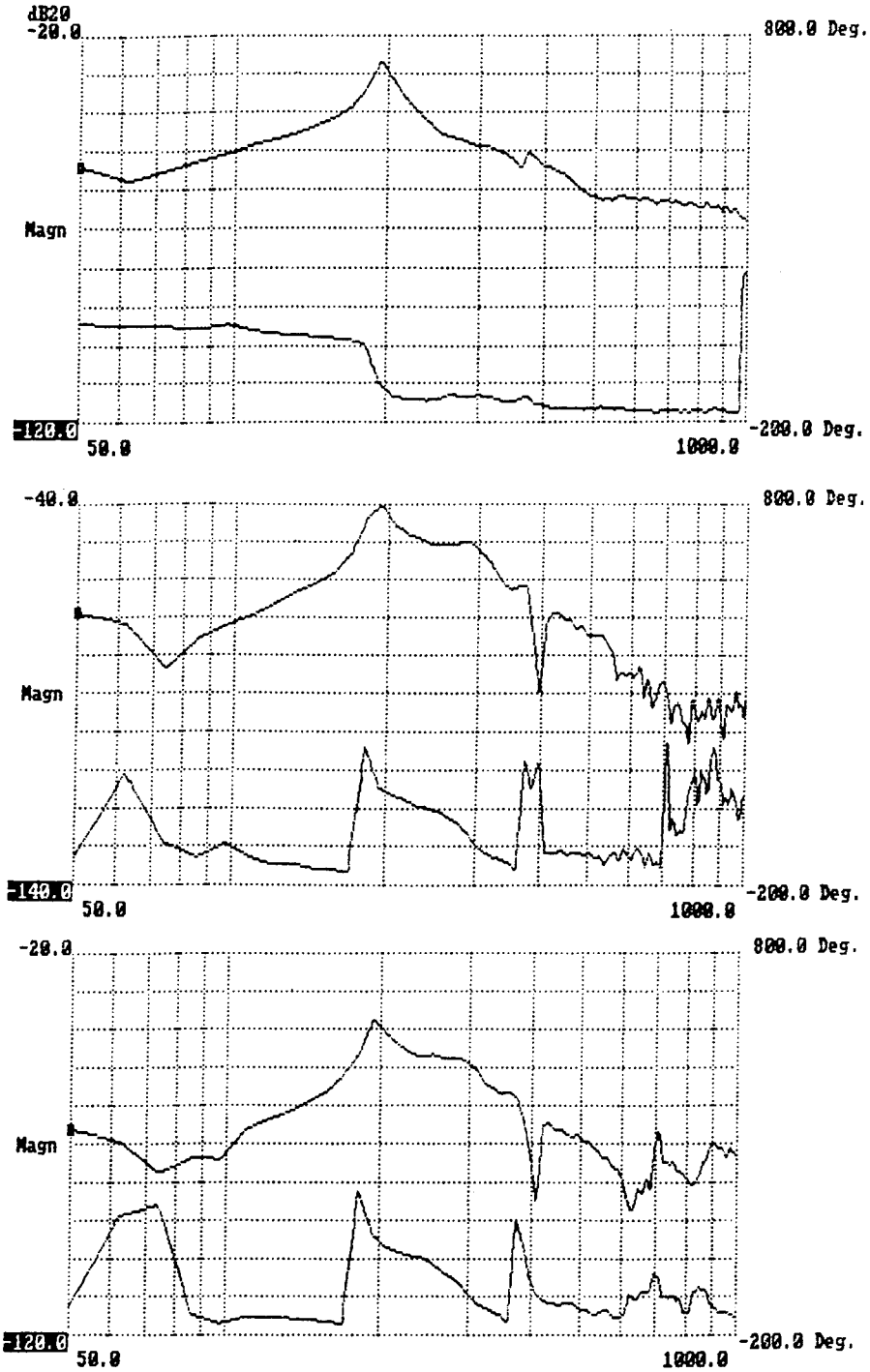


Figure 6: Bode magnitude and phase plots of dynamic response a.) near the headstock z = 0, b.) near the tailstock z = 620 mm, and c.) near the tailstock and clamped. The input was current (A) to the energizing coil and the output was displacements (μm).

lathe bed with the carriage positioned at the tailstock. Results (see Figure 6c) show fewer resonant peaks, reduced phase activity, and increased coherence at higher frequencies.

Controllers: Figures 2 and 6 collectively suggest the difficulty of the controls problem for the active structure. Figure 2 shows a concave depression in the ways with a maximum depth of about 58 microns, almost beyond the stroke capability of the magnetostrictive actuators. Figures 6 show that the carriage and structure dynamics depend on position z: higher order resonant frequencies vary with position along the ways. Reinspection of figure 2 reveals similar depression depths at headstock (z = 0 mm) and tailstock (z = 620 mm), but different inclinations. Different inclinations at different positions z suggests rocking of the carriage, which kinematically permits other resonant frequencies and modes.

An effective controller must adjust to the variable dynamics suggested by figures 2 and 6. Furthermore, although slow, wear increases with time: this will change the way contours and the carriage and structure dynamics with time. The controller must adapt to dynamics that change with position and time. With neural network based controllers, many system variables need not be quantified or known. A neural network learns the system dynamics, and adaptively tunes gains and/or restructures the controller. In this case, during operation the controller would learn the output signal(s) $C_i = C_i(z)$ required at a given carriage location to drive the error vector, E(Z)----the difference between actual and desired tool post/carriage positions-----to zero. Controller inputs (see figure 1) are: table pitch P about the X-axis, table roll R about the Y-axis, table elevation Z, bed angle about Z-axis, and table locations X and Y. Finally, other measurements such as lathe temperature could permit the controller to adapt to thermal strains. Controller outputs (three voltages) would energize the actuators.

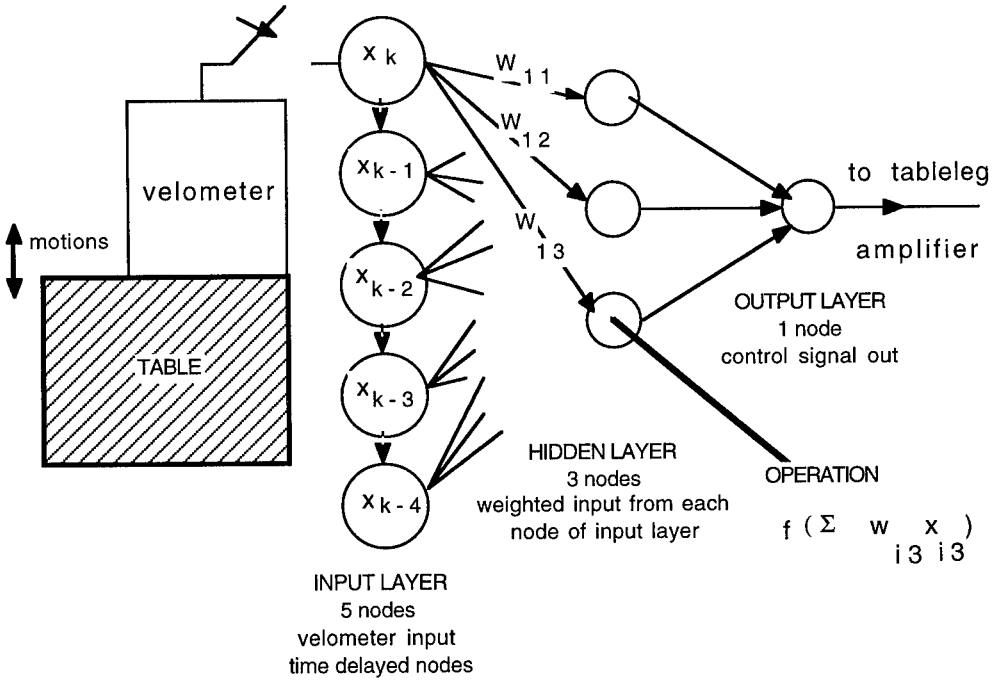


Figure 7: Separate neural controllers, each with five input nodes, three hidden nodes, and a single output node, learn the coupled table and tableleg dynamics and over time constructs control signals that attenuate disturbance vibrations imposed on the table of figure 3.

Dynamic Compensation and Active Vibration Attenuation: In a related vibration attenuation experiment performed with the table [3], stray vibrations of hundreds of microns from a shaker foundation beneath the table (see figure 3) were sensed by three velometers (time integrated accelerometers which sensed vertical velocity) mounted on the tabletop directly above the tablelegs. Three digital signal processors received the velometer signals, processed a control algorithm, and outputted three control signals $C_i = C_i(z)$ which were power amplified and then applied to the tableleg excitation coils. The resulting anti-phase vibrations produced within the table leg actuators attenuated vibrations and corrected three degrees of freedom (vertical displacement Z, roll R, and pitch P) at the tabletop. Each tableleg controller operated independently of the others. Control algorithms based on PID yielded 70 percent (10 dB) overall attenuation (30 μm reduction in amplitude) over a 100 to 1000 Hz range. Self tuning adaptive control by neural networks [3] applied individually to each leg reduced disturbances at the tabletop by 18 dB without cross-talk, despite severe dynamic coupling introduced by the rigid tabletop. Figure 7 diagrams the control layout. Although each neural controller only had 5 time-delayed nodes in its input layer, 3 hidden layer nodes, and a single output node, good decoupling and attenuation occurred over a 500 Hz bandwidth. It was felt that more nodes (limited by processor speed) in the neural controller would improve attenuations. The two actuators in the truss successfully positioned and stabilized the top pin joint at a point within an X-Y plane [2], despite large force disturbance vibrations (100 N's) and dynamic coupling. Velometers sensed disturbances. Attenuations were similar to the table.

Discussion, Summary and Conclusions: Presented were active structure concepts (table and truss) to correct various errors in turning processes. When passive (turned off), the structures are very stiff and machine performance is acceptable. Under control (turned on), structures actively deform through several degrees of freedom (DOF) not reachable by the machine tool mechanism to correct misalignments, compensate machine "slop," upgrade accuracy, and attenuate chatter and other vibrations. The structures would upgrade precision and extend machine tool life, since even severe misalignments could be corrected.

The present table structure can correct three degrees of freedom (elevation, roll and pitch) over a DC to 20 kHz bandwidth. Static tests demonstrated the table capable of compensating position dependent static errors such as wear of the ways. Vibration attenuation experiments performed on a similar structure suggest the structures capable of quenching machining vibrations such as chatter. Present and future developments include:

a. apply more sophisticated neural controllers to the table during turning,
b. combine the structures of figures 3 and 4 to implement 5 degrees of freedom.

Acknowledgements: Support from the National Science Foundation Tribology program (grant MSS-9302210) is gratefully acknowledged.

References:

[1] Hiller, M.W., Bryant, M.D., and Umegaki, J., "Attenuation and Transformation of Vibration Through Active Control of Magnetostrictive Terfenol," J. Sound & Vibration, Vol. 134, pp. 507-519, 1989.

[2] West, T.S., "Active Control of Vibration in a Truss Structure Using Magnetostriction," M.S. thesis, Mechanical Engineering, University of Texas at Austin, 1991.

[3] Bryant, M. D., B. Fernández, N. Wang, V.V. Murty, V. Vadlamani, and T. S. West, "Active Vibration Control in Structures Using Magnetostrictive Terfenol with Feedback and/or Neural Network Controllers," J. Intelligent Material Systems and Structures, Vol. 4, pp. 484-489, 1993.

[4] Friedrich, R. "Compensating for Machining Errors with Active Structures," M.S. thesis, Mechanical Engineering, University of Texas at Austin, 1994.

FIRM: A RULE-BASED DIAGNOSTIC NEURAL NETWORK

Benjamin Hellstrom
Analysis & Technology, Inc
Century Building, Suite 1250
2341 Jefferson Davis Highway
Arlington, VA 22202

James E. Birdsall
SynEx, Inc
5342 Echols Avenue
Alexandria, VA
22311

Richard Colsher
EPRI M&D Center
3 Industrial Highway
Eddystone, PA 19022

Abstract

Rule-based and neural network systems are currently the most promising vehicles for machine diagnosis. These two approaches share a common abstract framework. This paper compares the approaches in the context of this framework and discuss the strengths and weaknesses of each. In particular, the inability of a rule-based system to self-learn from its mistakes and the neural network's lack of innate knowledge of well-characterized machine malfunctions. A new consolidation is presented based on techniques for embedding rule-based knowledge in neural networks. The resulting rule-based network overcomes both problems by providing neural networks with a-priori expert knowledge while maintaining the ability to self-learn and thereby improve the breadth and accuracy of that knowledge. The system is named **FIRM**[1] and, under the supervision of EPRI[2], is being implemented to diagnose the turbine and motor-driven boiler-feedpumps of four large electric-power utilities.

Key Words: Hybrid systems; machine diagnosis; neural-networks; rule-based systems

Introduction

Computerized machine diagnosis has, in the past decade, enjoyed a surge in interest. Without doubt, this movement has been motivated by the

[1] FIRM is "Failure Introspection in Rotating Machinery."
[2] Electric Power Research Institute

availability of inexpensive computers and mircosensors, together with economic pressures to extend equipment life and avoid unnecessary maintenance. Although refinements in diagnostic hardware have progressed rapidly, there is still no unified systems approach to machine diagnosis. Two promising paradigms, knowledge-based inference and artificial neural networks, carry complementary strengths and weaknesses. These paradigms share a common abstract framework of knowledge representations, knowledge acquisition mechanisms, and decision engines. Their concrete realizations and operational characteristics are, however, widely disparate.

Knowledge Representation

Rule-based systems represent "permanent" knowledge by a hierarchy of labeled facts about known objects. For efficiency, inheritance mechanisms are commonly utilized. Logical rules and their numerical parameters comprise another form of permanent knowledge. "Transient" knowledge is represented by a possibly large number of "state variables" that may include current inputs as well as asserted conditions. Novel conditions may be automatically asserted by non-deterministic arbitration of sets of applicable rules. These novel conditions may be saved as "new" knowledge. In all of these cases, the knowledge is labeled, and accessible to the knowledge engineer. Even "new" knowledge is expressed as constructions of explicitly stored facts.

Neural networks store information as large numbers of correlation statistics or "connection weights." For certain network models, each statistic can be viewed as supporting or disputing evidence relating two hypotheses. For example, a neural network input "neuron" may represent the amplitude of a vibration's primary harmonic. An output "neuron" may represent an asserted state of machine imbalance. If a large primary amplitude is statistically consistent with the hypothesis that the machine is imbalanced, then a large positive correlation statistic will, over time, develop. If a large amplitude is inconsistent with the hypothesis, then a large negative correlation statistic will develop. If the amplitude is unrelated to the hypotheses, then a small correlation statistic, i.e., near zero, develops. For some neural network models, additional auxiliary neurons[3] may be provided to promote the discovery of very complex statistical relationships. Depending on the neural network model, learned knowledge may be extremely difficult to probe or express in a form that is understandable.

Knowledge Acquisition

Constructing a rule-based system requires the intensive, combined effort of domain experts and knowledge engineers. The relationships between tens or

[3] So-called "hidden units."

hundreds of sensed quantities, or features, and each output failure-class must be well-understood by the domain expert, communicated to the knowledge engineer, and coded as rules. Although experts may easily examine failure symptoms and diagnose machines correctly, introspection and the formalization of "intuition" are skills that are likely to be less-developed. The overall effort is costly, may span several years for a single machine, and is not readily extendible to new, different machines.

Rule-based systems are limited by the experience of human domain experts who, like all humans, are constrained to a single thread of cognitive processing. Humans approach high-dimensional pattern recognition problems hierarchically and sequentially. A few observations are used to form a hypothesis which is then validated or invalidated by additional observations. The human expert may examine several sensor features and conclude that each feature, in sequence, is within its norm. Although each individual feature may be within its norm, it is entirely possible for the combination of features to be outside of the joint norm. Humans are poor at discriminating broad distributions of slight deviations and good at discriminating small numbers of pronounced deviations. Rule-based systems that mimic human cognition inherit this limitation.

Neural networks observe and construct internal "models" of their inputs and their interrelationships during a learning process. They mimic the reflexive, subconscious sensory processing of biological nervous systems and are consequently adept at processing extremely high-dimensional data. Neural networks process *combinations* of features in parallel. They can respond to slight deviations from the learned model and can detect anomalous patterns well below the threshold of human perception. They are not well suited to cognitive tasks requiring long chains of inference. Although certain neural network models can be trained for inference, the procedure is arduous and inefficient. With exhaustive training data, including statistics from failing machinery, *supervised* neural networks can readily be taught to perform diagnostic tasks. *Unsupervised* learning, without human guidance and without failure data, is sufficient for early failure detection. For failure diagnosis, a domain expert is needed to label, or "tag" the machine state associated with each new anomalous behavior. Failure labeling can be done "off-line" at the convenience of the expert.

For both rule-based and neural network systems, sensitivity is a critical issue. Machine failures rarely come into existence instantaneously and without warning. During failure evolution, the configuration of sensor features progresses, in time, from a "healthy" region of the feature space, through an "ambiguous" region, and finally into a "failure" region. The precise placement of decision boundaries or thresholds between these regions determines whether the computerized diagnostician can spot the failure at its

earliest formative stage or only after significant progression. Since human decision making relies on few strong indicators rather than many weak indicators, and since indicator strength correlates with failure progression, it is likely that neural networks will outperform human experts (and rule-based systems) at detecting the earliest formative stages of machine failure.

Decision Engines

The differences between rule-based and neural network decision engines are more significant from the perspective of computational efficiency than from that of computability. Both have equivalent computability sets. It is possible to write a rule-based system to perform low-level pattern recognition, though such a system would have tens of thousands of rules and would operate on knowledge for which there is no clear acquisition approach. Similarly, there are classes of recurrent neural networks that can be organized to perform sequential inference, though the majority of neurons are idle during computation.

Hybrid rule-based neural network systems have been proposed with neural networks performing low-level audio or image-recognition processes and rule-based systems performing semantic or image-understanding tasks. Although this approach has merit, in this paper we present an alternative - a rule-embedded neural network.

A Rule-Embedded Neural Network

We have developed a diagnostic system of self-organizing neural networks for estimating probability density in the machine-feature space (Figure 1). A neural network is automatically constructed for each disparate machine state, including "NORMAL" operation. The density curves for the NORMAL class are readily computed after observing a normally operating machine. Unfortunately, without direct observation of machine failure, the neural network can not characterize regions of the feature space that correspond to new classes of machine failure. Two approaches for embedding rule-based knowledge are considered.

The first approach is simple, but has limitations. Whenever the maximum density estimate returned by all neural networks falls below a threshold, a new failure-class is asserted and a new network is created to recognize the class. The vector is then presented to a stand-alone rule-based system for diagnosis. The rule-based system returns a diagnostic label that is used to tag the newly-created network. Subsequently observed vectors that are within the receptive field of the new network elicit the return of the stored tag as a diagnosis. Without rule-embedding, existing neural networks self-learn to

detect anomalous machine behavior. The new neural network is taught, by the rule-based system, to correctly *diagnose* the behavior.

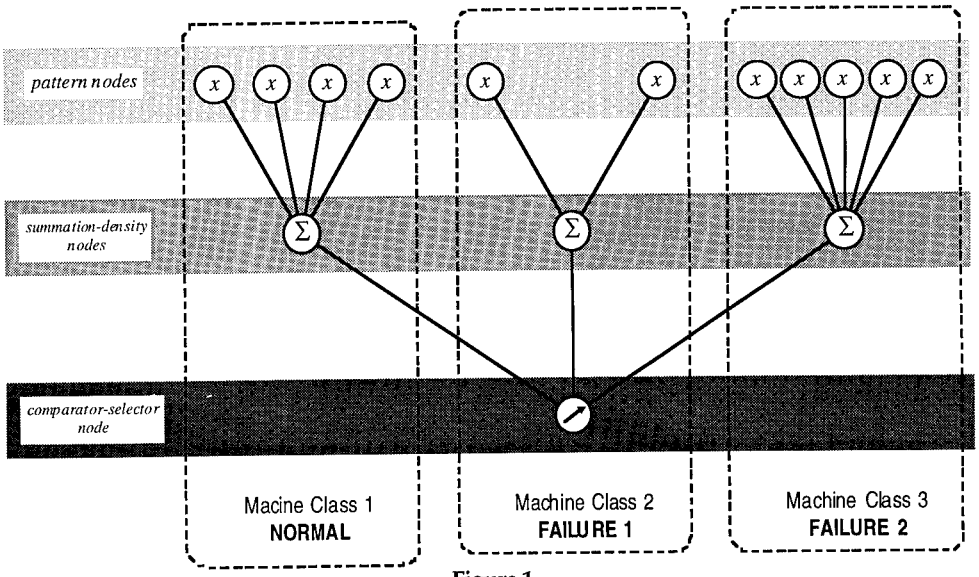


Figure 1
FIRM Network Architecture

This is clearly an expensive solution since both neural network and rule-based diagnostic systems must be independently developed. More importantly, the rule-based system is still limited by human perception. It is likely that many anomalies will be detected by the neural network for which the rule-based system does not concur that any fault exists. In these situations, the neural network can be given the benefit of the doubt and can assert that a "indeterminate" faults exist. If we assume that a fault truly exists, then the rule-based system must be able to diagnose the most-likely fault.

The second approach, which we have chosen to pursue, is to "pre-train" neural networks on NORMAL vectors that are synthetically distorted to elicit fault diagnoses from the rule-based system. Given insufficient populations of true fault data, the pre-trained networks can be used to return "rule-based" density estimates. The network that returns the largest density estimate represents the diagnostic class. The tantalizing aspect of this approach is that the mechanisms for network adaptation, i.e., learning, can be readily applied to the "pre-trained" classes. This allows the network to correct or refine its embedded rules to accommodate the peculiarities of its specific machine as well as numerical error in the rules. Initially, the network uses its embedded rules for diagnosis. As failures occur, the rules evolve to reflect the historical experience of the particular network.

FIRM System Description

As previously stated, the **FIRM** diagnostic system is a dynamic set of sub-networks that is subject to both connection and architectural-level adaptation. Each sub-network is responsible for approximating the density estimate of a particular class of machine failure and is labeled or "tagged" with the name of the class, e.g., "NORMAL", or "MISALIGNMENT." A sub-network consists of two layers of nodes - pattern nodes and summation/density nodes. The density estimates are weighted by the a-priori probability of each failure class and then compared. Pattern nodes are unsupervised, radial-basis neurons with hyper-elliptical receptive fields whose dimensions are derived from the statistical variation of observed features. A variety of competitive and stochastic training algorithms have been tested. Of these, "top-N-takes-all" has been found to return the smallest error with the value of "N" a function of the feature-space dimensionality. The output of neuron (c,i), i.e., pattern node i in sub-network c, in response to input vector $\mathbf{x}$, is given by -

$$g_{ci}(\mathbf{x}) = n_i^c \exp\left(-\frac{1}{2} \sum_{j=1}^{p} \left(\frac{x_j - \mu_{ij}^c}{\sigma_{ij}^c} \right)^2 \right)$$

Where n_i^c, μ_{ij}^c, and σ_{ij}^c are the numbers of feature vectors that have been "won" by the node (in competition), and the mean and standard deviation of the won vectors' j^{th} dimension of the feature space respectively. The j^{th} component of the feature vector and dimensionality of the feature space are given by x_j and p respectively. Clearly, a multi-dimensional Gaussian distribution is modeled by a single pattern node. Distributions of arbitrary complexity, including multimodals, are modeled by collections of pattern nodes.

We have previously referred to the "receptive field" of a neuron. This is an imprecise terminology. The receptive field of neuron (c,i) is the subspace $\mathbf{x} \in \mathcal{R}^p$ for which $g_{ci}(\mathbf{x})$ is "greatest." While "greatest" is now a term of imprecision, it is safe to say that if $(\forall j)\left[\left| x_j - \mu_{ij}^c \right| < \sigma_{ij}^c \right]$ then $\mathbf{x}$ is in the receptive field of neuron (c,i).

Summation/density nodes integrate pattern node outputs and normalize the sum to a density estimate. The summation node for class c has the response -

$$f_c(\mathbf{x}) = \left(\left(\sum_{i=1}^{m_c} n_i^c \right) \sqrt{2\pi} \right)^{-p} \sum_{i=1}^{m_c} \left(\prod_{j=1}^{p} \sigma_{ij}^c \right)^{-1} g_{ci}(\mathbf{x}).$$

where m_c is the number of pattern nodes that have been allotted to represent the distribution of feature vectors for class c. The only quantity that is not subject to adaptation is p, the system dimensionality.

The **FIRM** system is implemented under Windows™† on a 486-based PC.

System Adaptation

The system is trained in two distinct phases, initialization, and steady-state, or "monitoring." Prior to initialization, the machine is observed for a fixed period of time and features, derived from sensor data[4],are assembled and cached to computer disk. During initialization, the cached data is iteratively processed as the receptive field centroids and diameters of a fixed number of NORMAL pattern nodes are optimized. The diagnostic system then commences to monitor the machine. During the monitoring phase, features are presented to all sub-networks[5] and responses are compared. If the maximum density estimate from all summation nodes falls below a fixed threshold, a novel machine-state is asserted and a new class, "UNKNOWN," is created together with a new pattern node. If the largest density estimate is not emitted from the NORMAL sub-network, an alarm is sounded and the tag of the sub-network with strongest response is output as the diagnosis. An independent decision hierarchy determines when it is necessary to create new pattern nodes for existing classes. In any case, if neither pattern node nor new class is created then the strongest-responding pattern nodes are tuned to respond more strongly to the feature vector[6].

Rule Embedding

Rule embedding is the expression of innate knowledge within the structure of the network architecture. It is done after initialization and prior to monitoring.

---

† Windows™ is a registered trademark of Microsoft Corporation.

[4] For our experiments, turbine-driven boiler feedpumps were studied. 115 features were extracted (i.e., p=115) including 19 vibration amplitudes from each of 4 bearing probes and 2 accelerometers together with machine speed.

[5] Initially, there is just one sub-network for "NORMAL" machine operation.

[6] This form of Hebbian learning, together with a mechanism for "forgetting" allows the FIRM system to track transient feature vector distributions associated with normal machine aging.

Even "small" neural diagnostic systems operate in feature spaces with, by human standards, "large" dimension. The human vibration expert typically uses simple, general rules-of-thumb that are referenced from NORMAL machine characteristics. For example, a rotating machine might be diagnosed as MISALIGNED when the 1×, 2×, 3×, and 4×, bearing-proximity harmonic amplitudes have characteristic ratios of 20: 70: 5: 5. The rule-of-thumb only specifies the relationships between a few of the p dimensions of $\mathbf{x}$. For example, the rule does not specify the 7× harmonic amplitude taken from a housing-mounted accelerometer. The neural diagnostic system is designed to operate in *all* p dimensions. The omission of details concerning the majority of feature components implies that those components are not significant with respect to the misalignment hypothesis. This insignificance can be treated by two methods - the response of the summation node, $f_c(\mathbf{x})$, can be altered to ignore temporarily insignificant dimensions or "pseudo-feature" components, selected to lack statistical significance, can be substituted for unimportant features. For simplicity, and due to the ready availability of machine statistics for our application, we chose the latter approach.

In the case of the rule for misalignment, the **FIRM** system iterates through the cache of NORMAL features. Those features that are not referenced in the rule for misalignment are unchanged. Those features that represent amplitudes of the 1×, 2×, 3×, and 4×, bearing-proximity harmonics are summed, and reapportioned to the proper ratio, 20: 70: 5: 5, thereby retaining the same total power. The original statistical variation *between* these key components can also be retained by adding appropriate noise. Once a distorted vector has been synthesized, it is presented to all sub-networks. If one or more sub-networks respond with a density estimate above the alarm threshold, the ambiguity is brought to the attention of the designer and human expert. This procedure is repeated with different cached observation vectors until the centroid and diameter of the receptive field of the new pattern node(s) for MISALIGNED stabilize.

Ambiguity does not disrupt system operation since the diagnosis is derived from the tag of the sub-network with strongest output - and there will only be one maximum. If there is significant overlap of the receptive fields of the pattern nodes associated with the new rule and some existing class, the receptive field of the existing class will be trimmed at the site of conflict.

Neural network memories of embedded rules are treated just as self-learned memories. These memories are tagged to indicate their source, e.g., "MISALIGNED, RULE-BASED" and to distinguish them from self-learned memories. There are two key features that distinguish networks embedded with rules from rule-based systems. First, precision is not critical. If the ratio of the aforementioned harmonics is 10: 75: 10: 5, and the vector continues to

fall below the density threshold for the NORMAL machine class, the most-likely abnormal class will be asserted. If the 20: 70: 5: 5 rule sufficiently discriminates the misalignment condition from other abnormal conditions a MISALIGNED condition will still be hypothesized.

Second, and more importantly, the memory of the embedded rule is subject to adaptation in the case that the machine exhibits the symptoms of the rule. The memory of the malfunction is refined with experience while the system makes a correct diagnosis. Consider the following analogy - We desire to train a child to discriminate apples from green vegetables. We presume that the child, having been raised in a small tribal village north of the Arctic circle, has eaten fish all his life and has never seen a fruit or vegetable. Furthermore, we have green vegetables but have no apples. Instead, we have a small red rubber ball. During training, the child is shown the vegetables which are properly identified as vegetables. Subsequently, the child is shown the rubber ball and told that it is an apple. After training, the child is presented with a real apple and asked to identify it based on what he has been taught. Since the apple is most similar to the red ball, the child properly classifies it as an apple. It does not matter that the ball is not really an apple. Our objective, that of teaching the child to discriminate between apples and vegetables, has been achieved. This analogy illustrates the first point, above; the network can generalize imprecise embedded memories.

Let us continue. While the child is being tested and continues to properly discriminate apples and vegetables, a crate of real apples becomes available to us. We commence to teach the child using the real apples instead of the rubber ball. Subsequently, over a period of time[7], the child learns that bumps, stems, skin blemishes, and variations in color are all consistent with apples. Furthermore, after seeing many apples, the memory of the rubber ball fades and eventually disappears. After this additional training, the child's mental "model" for apples is substantially refined. This analogy illustrates our second point - while correctly identifying machine malfunction, the neural network refines its memories to become consistent with the true feature/sensor characteristics of malfunctions as they become available.

We have successfully embedded and tested, in the **FIRM** system, rules for rubbing, unbalance, and misalignment. Doubtless, there are many complex rules that may not appear to lend themselves to this simple approach. These rules may require new features, e.g. eccentricity of shaft orbit. Rules that are expressed as elaborate boolean expressions are encoded as measures of "fuzzy" membership[8]. The system is easily extended to accommodate these features.

---

[7] In the neural network context, the rates of adaptation and memory loss are controllable.

[8] For example, the expression $(x<a) \wedge (y \geq b)$ can be encoded as the feature $\max(a-x,0)\cdot\max(y+\varepsilon-b)$ for suitable selection of ε. This measure smoothly approaches 0 as its condition ceases to be met.

Conclusion

We have discussed the strengths and weaknesses of rule-based and self-learning neural networks for machine diagnosis and have presented a mechanism for combining properties of both. By embedding innate, rule-based knowledge as neural network memories, the network is endowed with the capability of classifying malfunctions that are not in its prior learned experience. Additionally, those embedded rules are made available for subsequent adaptation and improvement. Work remains to accommodate the large numbers of complex rules found in a typical rule-based system. In the future, we hope to devote additional effort in this direction.

Acknowledgment

The authors would like to acknowledge their sponsors, Joseph Carbonara of Consolidated Edison, Mark Richardson of NYSEG, Dennis Lincolns of TVA, and Tom Fritsch of the EPRI Maintenance & Diagnostic (M&D) Center. Without their support and professional guidance the **FIRM** project would not be possible.

IMPROVED FAULT IDENTIFICATION USING MULTISENSOR WAVELET BASED DIFFERENTIAL FEATURES

Jose E. Lopez
James C. Deckert
Robert R. Tenney
ALPHATECH, Inc.
50 Mall Road
Burlington, MA 01803-4562

Abstract: Vibration analysis remains the cornerstone of the vast majority of predictive maintenance methodologies. Predictive maintenance schemes provide for substantial economic returns across a number of domains. The primary beneficiaries are industries with high capital equipment investments and the military who are currently working with a rapidly aging installed equipment base. Improvements, even small improvements in predictive maintenance systems can translate into substantial savings for a given application. This paper presents an investigation for improving the performance of a wavelet-based fault detection and identification system (FDI). Improvement is achieved via the use of multisensor wavelet-based differential features augmentation. This method is applied to testbed data sets from an intermediate helicopter transmission. Explanation of the methods along with single channel results are given for the base FDI system. The differential feature augmentation method is explained and the results from the enhanced FDI system are presented.

Key Words: equipment diagnosis; fault detection; helicopter transmissions; predictive maintenance; state-of-health monitoring; vibrating systems; wavelets

Introduction: Thinning profit margins, shorter time to market, increased global competition are putting tremendous pressures on management to increase the effective operating life of existing equipment while simultaneously driving up product quality and improving overall efficiency. In the military domain decreased budgets, cancellation of important new systems have produced pressures to efficiently sustain a high state of readiness in the face of rapidly aging platforms [1]. Predictive maintenance or effectively condition-based maintenance provides one avenue for decreasing these pressures.

The goals of any predictive maintenance system is to anticipate or provide reliable indicators concerning a machine's condition. Vibration monitoring remains the primary component of many predictive maintenance systems. The basic atoms in such vibration monitoring schemes consist of reliable fault detection and identification (FDI) methods [2]. New technology, such as wavelet-based methods, can drive the development of substantially improved FDI methods, which are robust, adaptable, and readily applied across a broad range of platforms.

This paper presents an investigation for improving the performance of a wavelet-based FDI system. A brief overview of this wavelet-based FDI system is provided in the next section. More extensive details of this system are available in [3]-[6]. Improving this base FDI system is achieved via the use of multisensor wavelet-based differential feature augmentation. To illustrate the improvement achieved through the use of this method, testbed data sets from an intermediate helicopter transmission are analyzed using the base system and the enhanced system. Single channel results are given for the base FDI system. The differential feature augmentation method is explained and the results from the enhanced FDI system are then presented.

Base System Overview: Figure 1 provides a simplified block diagram of a wavelet-based FDI system where projections of the system dynamics are extracted in the form of wavelet features using a data-driven approach and classification is performed in an adaptive fashion. This general methodology proceeds in the following way. A wavelet feature processor extracts from the transduced signals associated with the system a set of robust projections of the system dynamics. Features extracted using this processor, possibly coupled with additional feature post-processing, will form a time series that captures important system dynamics in the time-scale plane. Prepared data can be used to extract features and train the classifier for the various fault conditions. When training is complete, features extracted from operational data will be used for detection and classification purposes.

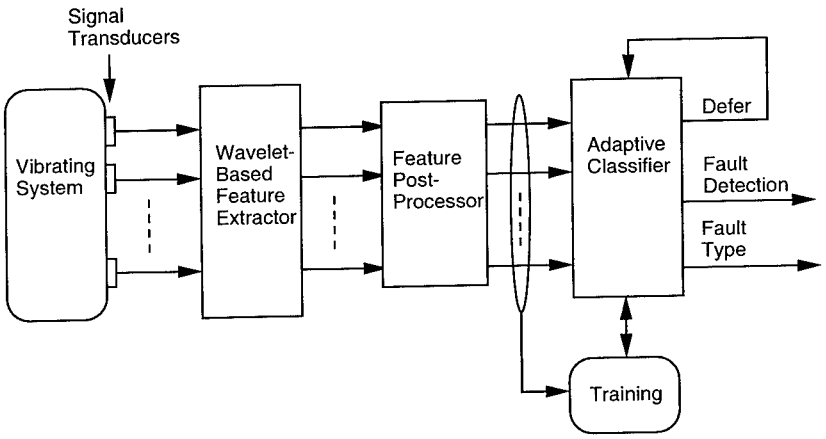


Figure 1. Basic Wavelet-based FDI System

The ALPHATECH's approach to designing the feature extractor relies on development of a special class of continuous wavelet transforms (CWT) which have semi-infinite support in the time domain and can be implemented using real-rational, causal wavelet transfer functions [7]-[9]. This provides for the capability to reduce to practice a FDI methodology which can be applied in real-time monitoring applications. Figure 2 illustrates the CWT for two different cases (normal and one fault case) of intermediate gearbox data. Clearly visible in the wavelet mesh associated with the tooth cut fault is the lost of wavelet spectra regularity found in the normal wavelet mesh, particularly in the range of 4 to 16 kHz. This illustrates the sort of discrimination afforded by wavelet techniques and provides the motivation for developing a set of methods which effectively exploit these wavelet discrimination properties for purposes of FDI. The ALPHATECH feature extractor consists of algorithms that are both adaptive and non-adaptive wavelet transform methods. Applying the adaptive versus non-adaptive methods is dependent on the vibrational system under test (i.e., narrow-band vibrating systems versus broad-band systems). All of the technical details of the wavelet-based feature extraction and subsequent identification results are available in [3]-[6].

Formulating Wavelet-Based Differential features: To improve upon fault identification the use of multiple sensor wavelet extraction is being investigated. The essential idea is that complex vibrational systems and their associated operating environments will benefit from the use of incorporating information from more than one sensor in the FDI system [10]-[11]. The first application of these methods stem from the use of multiple sensor differential wavelet feature formulation. The technique involves identifying a primary sensor site from which wavelet features will be extracted. Using a secondary sensor site (or possibly multiple sensor sites), wavelet features at various scale

settings can be differenced against the primary sensor values. These differential features are then conjoined to the original primary feature vector increasing its dimension. This enhanced set of feature vectors is then used as the basis for classification.

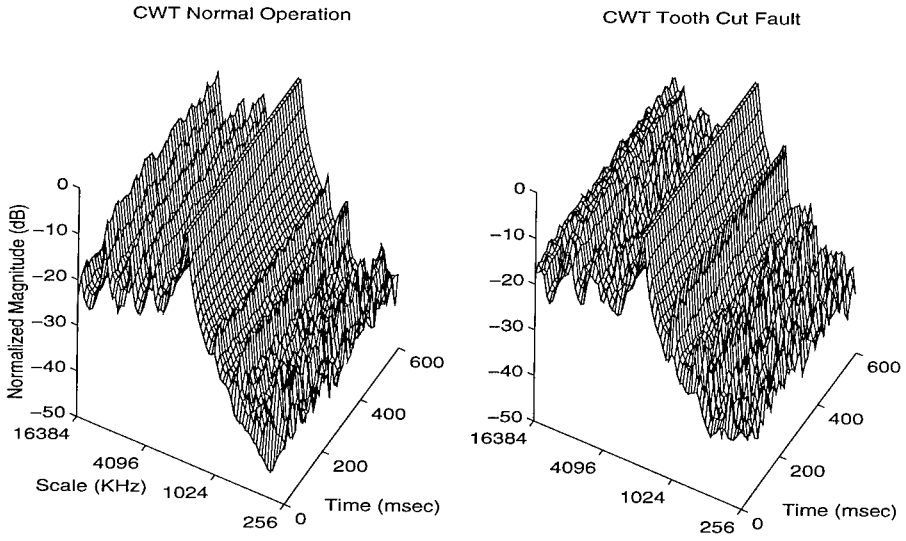


Figure 2. Continuous Wavelet Transforms of Normal and Fault Data

Hollins Data Set: Application of this methods was applied to two channels of a intermediate gearbox (NSTC, Hollins TH-1 data set) and the results were compared against results obtained using single channel data and no differential wavelet-feature augmentation. The NSTC Hollins data set is a set of seeded fault data measurements from a TH-1 42 degree intermediate gear box. Vibration readings were recorded from two accelerometers (channels 5 & 6 oriented with and orthogonal to the bearing load zones) mounted on the gearbox output end for six separate fault conditions. The data set consisted of the following fault classes: 1) normal (nd) operating condition 2) inner race (ir): nick in ball bearing 3) outer race (or): nick in ball bearing 4) rolling element (re): flattening of a ball in a ball bearing 5) gear spall (sp): roughening/pitting of a gear 6) tooth cut (tc): top half of gear tooth removed.

Adaptive wavelet feature extraction was employed [3], [5] which allowed tracking of the fundamental mesh frequency and avoided the need for using an auxiliary tachometer signal to accommodate small variations in input shaft RPM. A non-linear feature transformation was used and resulted in 14 curvature/power features being extracted for the signal channel case and 29 curvature/power (15 of which were differential features) for the multiple channel case. Vibration data from sensor 5 was used for feature extraction in the single channel case and wavelet features extracted from vibration data of both sensor 5 and 6 (sensor 5 primary signal source and sensor 6 secondary signal source) was used for the formulation of multi-sensor differential feature vectors.

Single Channel Performance: Previous work [3], [6] has indicated that excellent fault detection and very good identification is possible with the Hollins data set by using wavelet-based features extracted from a single channel. Figure 3 and Figure 4 give the performance results using a trained minimum-error-rate classifier. The classification rule

used to classify the feature vectors is based on a Bayes decision rule to minimize the risk associated with employing a symmetrical loss function [12]. In other words, minimizing the risk that the assigned fault class for a given feature vector is incorrect is analogous to the Bayesian problem of assigning the fault class based on the maximum posterior probability that the assigned fault class is indeed the correct class given the feature vector. A underlying assumption made (for purposes of constructing a discriminant function) is that the feature vectors are multivariate and normally distributed. This assumption is of course not perfectly valid for the case of gear-box data. However for purposes of this investigation we were interested in a comparison between wavelet-based FDI with and without wavelet differential feature augmentation. Hence a baseline trainable classifier such as the aforementioned Bayesian classifier does the job. In other studies where the goal revolved around directly achieving the best classification more sophisticated methods involving neural networks have been employed (see [3], [5], [6]).

The minimum-error-rate classification can be achieved using the following discriminant function:

$$g_i(x) = \frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) + \frac{d}{2}\log 2\pi + \frac{1}{2}\log|\Sigma_i| - \log P(\omega_i) \qquad (1)$$

where x is the feature vector, μ_i is the sample mean, Σ_i is the covariance matrix derived from the training data, d is the dimension of the feature vector, and $p(\omega_i)$ is the a-priori probability of the occurrence of fault class ω_i. Ignoring for the moment deferral mechanisms based on outlier and nearest neighbor strategies, classification is achieved via:

$$\min[g_i(x)] \quad i = 1,2,\cdots n \qquad (2)$$

for n classes when presenting feature vector x for classification.

Performance results (Figure 3 and 4) are given in the form of a confusion matrix with the estimated classes given along the top (columns) of the matrix and the correct classes given along the left hand side (rows) of the matrix. To read these tables one examines a row (for example the row corresponding to fault "or"). All numbers in a given row correspond to the actual fault class listed on the left hand side of the table. The distribution of the numbers across a given row indicate the number of feature vectors identified (classified) as the particular estimated class shown in the associated column heading. Ideally, for perfect identification one wants all the numbers occurring in the main diagonal positions.

The last column of this matrix is a deferral column. When deferral is turned on in the statistical classifier numbers in this column refer to the number of deferrals that occurred for feature vectors from a given fault class. A deferral occurring indicates that the classification system (due to various deferral parameters which are user assignable) is detecting some ambiguities associated with the feature vector currently under classification. Rather than assign a fault class to the feature vector in this ambiguous situation the system defers classification and the next feature vector is examined for classification. Since features where extracted every 10ms with the Hollins data, a deferral essentially implies a delay of 10ms to the next feature vector. The two deferral mechanisms used consisted of an outlier deferral number (based on dimension of the feature vectors—essentially the larger the dimension of the feature vector, the larger the outlier deferral number) and a nearest-neighbor deferral number (based on a percentage of how close the nearest likelihood result is to the minimum result).

As can be seen from Figs. 3 and 4 for this single channel example, the fault detection is flawless (i.e. detecting deviation from normal operating condition). What remains a

problem area is errors associated with fault identification. In a condition based maintenance system fault identification or localization can precipitate substantial economic savings by directing the maintenance schedule and identifying with high confidence the subsection of the platform requiring overhaul. Any improvement via enhanced identification algorithms can significantly drive up the overall performance of any condition based maintenance strategy. The next two subsections indicate the improved results obtainable for fault identification through the use of differential feature vectors.

Estimated Classes

Correct Classes	nd	tc	ir	re	or	gs	def
nd	1200	0	0	0	0	0	0
tc	0	1193	0	0	1	6	0
ir	0	0	1200	0	0	0	0
re	0	0	0	857	340	3	0
or	0	0	0	53	1138	9	0
gs	0	0	0	2	8	1190	0

The normal case is "nd".

Prob. of false alarm = 0; prob. of missed detection = 0;

prob. of misclassification = 0.05861; prob of deferral = 0

Figure 3. Classification Results Single Channel Features, No Deferral

Estimated Classes

Correct Classes	nd	tc	ir	re	or	gs	def
nd	1083	0	0	0	0	0	117
tc	0	1152	0	0	0	1	47
ir	0	0	1151	0	0	0	49
re	0	0	0	785	275	2	138
or	0	0	0	31	1079	6	84
gs	0	0	0	2	4	1164	30

The normal case is "nd".

Prob. of false alarm = 0; prob. of missed detection = 0;

prob. of misclassification = 0.04458; prob of deferral = 0.06458

Figure 4. Classification Results Single Channel Features, with Deferral

Differential Feature metrics: Before examining the improved performance obtained using differential features an analysis is conducted on the single channel wavelet features and multichannel differential wavelet features to provide a basis for quantifying the improved results based solely on the construction of the feature vectors and the fault

classes hoping to be separated. The method used is to apply the Fisher Linear Discriminate which provides a one dimensional metric (linear functional) to the multidimensional feature vectors so that a comparison can be performed concerning the pairwise separation achievable between two fault classes [13]. Two versions of this metric are performed. The first version calculates the ratio of between class scatter to overall class scatter on a pairwise basis. This results in a real number indicating the "best" overall separation between fault classes projected to one dimension. The best way to think of this metric is to visualize finding a line through multidimensional feature space for which features projected on to this line result in the maximum separation between pairs of fault classes. This calculation is carried out as follows:

$$J(w) = \frac{w^T S_B w}{w^T S_w w} \tag{3}$$

where m_i is the mean of the feature vectors of the respective fault class, and

$$S_w = S_1 + S_2 \qquad S_i = \sum_{x \in H_i} (x - m_i)(x - m_i)^T \tag{4}$$

$$w = S_w^{-1}(m_1 - m_2) \tag{5}$$

where H_i refers to a set of feature vectors associated with the particular fault class indexed by i, and matrix S_B is given by

$$S_B = (m_1 - m_2)(m_1 - m_2)^T \tag{6}$$

The metric as given here provides the best overall separation between pairs of classes projected to one dimension.

The second version of a Fisher Linear Discriminate calculates on a feature element by feature element basis the maximum separation between a pair of fault classes. This version is useful in identifying a specific feature element responsible for large separations between fault classes. However, since this second version is carried out only over the feature vector elements the resulting separation numbers will be smaller than the numbers produced in the first version and may not necessarily provide insight into improved performance obtained via the use of larger dimension feature vectors consisting of similar components in the smaller dimensional feature vector.

This index is computed by selecting a different w than given in Eq. 6, so one can calculate the largest separation achievable between pairs of classes based on feature-vector elements. This formula is achieved by the rewriting the index given by Eq. 3 in the following manner:

$$\max_{w \in \{n_i\}} \sqrt{J(w)} \qquad where \quad i = 1,2,\cdots N \tag{7}$$

and $\{n_i\}$ is a set of N standard basis vectors for feature vectors of dimension N. That is, n_i is a column vector of N elements all of which are zero except for the i^{th} element, which takes on the value of one. In effect the line chosen to examine separation between pairs of fault classes is essentially the coordinate axis of each element of the feature vector. The feature-vector element producing the greatest separation between the two classes is chosen to represent the Fisher index via Eq. 7. The more familiar representation of Eq. 7 for two fault classes labeled with subscripts 1 and 2 is

$$\max_{i\in[1,N]} \frac{|m_1(i) - m_2(i)|}{\sqrt{\sigma_1^2(i) + \sigma_2^2(i)}} \tag{8}$$

where i takes on values from 1 to N and refers to the individual feature elements in a N dimensional feature vector. In order to be consistent with Eq. 8, when using Eq. 3 in the display tables the actual value shown is the *square root* of Eq. 3.

Figure 5 gives the Fisher separations for the single channel gearbox data. These tables were compiled using features vectors extracted every 10 ms from 14 seconds of vibration data for each fault class. The main diagonal of these tables contain zero indicating of course no separation between the same fault class. Usually, any number above 3 indicates a reasonably good linear separation. A problem area as identified by these metrics for the single-channel case is going to be separating the fault classes "re" and "or" which has a overall maximum fisher separation of 2.311.

Figure 6 gives the Fisher separations for the multichannel wavelet based differential feature vectors. Once again these tables where compiled using features extracted every 10 ms from 14 seconds of vibration data. The primary wavelet features were obtained from the original features extracted for sensor 5. Differential feature elements conjoined with the primary wavelet feature vector elements were constructed by differencing wavelet features extracted from sensor 5 against features extracted from sensor 6.

The maximum separations of Fig. 6 are significantly improved via the use of differential feature elements over the maximum separations observed in Figure 5 for the single channel case. A couple notable improvements are "gs" verses "or". Based on these numbers one expects to see reductions in the number of misclassifications between "gs" and "or". And the problem area "re" verses "or" has also an improved maximum separation index by about 15.2%.

Performance Results: Figure 7 gives the classification results for the multichannel, wavelet based differential features without deferral. Figure 8 gives these results with deferral (outlier deferral parameter = 5.762 and nearest neighbor parameter = .01). The overall improvement (via reduced probability of misclassification—see Figs. 3 and 4) in identification using multichannel, differential feature vectors is 17.3% without deferral and 42.3% with deferral. As indicated by the separation metrics of Figures 5 and 6 improvement was obtained in identifying the faults "gs" verses "or" and "re" verses "or". These improved results are observed by comparing the confusion matrices of Figs. 7 and 8 against those of the single channel confusion matrices, Figs. 3 and 4.

Maximum separations:

	nd	tc	ir	re	or	gs
nd	0	9.876	6.928	10.09	7.514	9.313
tc	9.876	0	7.819	5.117	5.072	4.765
ir	6.928	7.819	0	7.808	6.16	7.018
re	10.09	5.117	7.808	0	2.311	3.746
or	7.514	5.072	6.16	2.311	0	3.381
gs	9.313	4.765	7.018	3.746	3.381	0

Maximum separation for a feature element:

	nd	tc	ir	re	or	gs
nd	0	6.388	5.261	6.941	4.661	5.908
tc	6.388	0	4.076	3.554	2.67	2.786
ir	5.261	4.076	0	3.953	3.628	3.645
re	6.941	3.554	3.953	0	1.992	2.002
or	4.661	2.67	3.628	1.992	0	1.753
gs	5.908	2.786	3.645	2.002	1.753	0

Figure 5. Fisher Separations for Single Channel Features

Maximum separations:

	nd	tc	ir	re	or	gs
nd	0	10.99	8.383	10.78	8.315	11.43
tc	10.99	0	8.774	6.515	6.144	7.117
ir	8.383	8.774	0	8.874	7.466	9.583
re	10.78	6.515	8.874	0	2.726	5.329
or	8.315	6.144	7.466	2.726	0	5.542
gs	11.43	7.117	9.583	5.329	5.542	0

Maximum separation for a feature element:

	nd	tc	ir	re	or	gs
nd	0	6.388	5.261	6.941	4.661	5.908
tc	6.388	0	4.632	3.554	2.67	2.786
ir	5.261	4.632	0	3.953	3.628	3.645
re	6.941	3.554	3.953	0	2.155	2.002
or	4.661	2.67	3.628	2.155	0	2.523
gs	5.908	2.786	3.645	2.002	2.523	0

Figure 6. Fisher Separations Multiple Channel, Differential Features

Estimated Classes

Correct Classes	nd	tc	ir	re	or	gs	def
nd	1200	0	0	0	0	0	0
tc	0	1200	0	0	0	0	0
ir	0	0	1200	0	0	0	0
re	0	0	0	883	304	13	0
or	0	0	0	32	1168	0	0
gs	0	0	0	0	0	1200	0

The normal case is "nd".

Prob. of false alarm = 0; prob. of missed detection = 0;

prob. of misclassification = 0.04847; prob of deferral = 0

Figure 7. Classification Results Multiple Channel, Differential Features, No Deferral

Estimated Classes

Correct Classes	nd	tc	ir	re	or	gs	def
nd	1006	0	0	0	0	0	194
tc	0	1108	0	0	0	0	92
ir	0	0	1071	0	0	0	129
re	0	0	0	740	171	4	285
or	0	0	0	10	1050	0	140
gs	0	0	0	0	0	1121	79

The normal case is "nd".

Prob. of false alarm = 0; prob. of missed detection = 0;

prob. of misclassification = 0.02569; prob of deferral = 0.1276

Figure 8. Classification Results Multiple Channel, Differential Features, With Deferral

Conclusions: The conclusions to be drawn are that the directional nature of the fault vibrations impacting the orthogonally aligned sensors contained enough variation to provide appreciable improvement in fault identification using the differential wavelet-based feature formulation method. This provides evidence that improved identification is possible through the use of multisensor, wavelet-based differential feature vectors. Indicators of potential performance can be achieved by analyzing the differential feature vectors via the overall maximum Fisher Linear Discriminant. Baseline classification using wavelet feature vectors augmented by differential wavelet-based feature elements continue to achieve enhanced performance through the use of deferral mechanisms in the classification process.

References:

1. Chamberlain, M. (Captain, USN, (Ret.)),"U.S. Navy Pursues Air Vehicle Diagnostics Research," *Vertiflite*, Vol. 40, No. 1, March/April 1994, 16-21
2. Willsky, A.S., "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, November 1976, 601-611.
3. Lopez, J.E., Tenney, R.R., and Deckert, J.C., "Wavelet Feature Extraction For Real-Time Neural Network Condition Based Maintenance," *Proc. US Navy SBIR Conference on Neural Network Applications*, Arlington, VA, pp. 109-121, 16-17 June 1994.
4. Lopez, J.E., Deckert, J.C., Tenney, R.R, "Condition-Based Machinery Maintenance: Interim Report," ALPHATECH TR-663, August 1994.
5. Lopez, J.E., Deckert, J.C., Tenney, R.R, "Wavelets, Helicopters, and Maintenance" ALPHATECH TR-670, September 1994.
6. Lopez, J.E., Tenney, R.R., and Deckert, J.C., "FDI Using Real-Time Wavelet Feature Extraction," IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, Philadelphia, PA, pp. 217-220, 25-28 October 1994.
7. Daubechies, I., "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE Trans. Inform. Theory*, 36, 1990, 961-1005
8. Mallat, S.G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. PAMI*, 11, July 1989, 674-693.
9. 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.
10. Cempel, C., "Vibroacoustical Diagnostics of Machinery: An Outline," *Mechanical Systems and Signal Processing*, Vol. 2, No. 2, 1988, pp. 135-151.
11. Smith, J.D., *Gears and Their Vibration*, MacMillan Press Ltd.,1983.
12. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley & Sons, 1968.
13. Duda, R., and Hart, P., *Pattern Classification and Scene Analysis*, John Wiley & Sons, 1973.

Acknowledgments: This work was supported by the Office of Naval Research under contract N00014-93-C-0077 and by the Advanced Research Projects Agency under contract DAAH01-93-C-R297.

STRUCTURAL HEALTH MONITORING OF LARGE STRUCTURES

Hyoung M. Kim* and **Theodore J. Bartkowicz†**
McDonnell Douglas Aerospace, Houston, Texas 77059

Suzanne Weaver Smith‡
University of Kentucky, Lexington, Kentucky 40506

David C. Zimmerman§
University of Houston, Houston, Texas 77204

ABSTRACT: This paper describes a damage detection and health monitoring method that was developed for large space structures using on-orbit modal identification. After evaluating several existing model refinement and model reduction/expansion techniques, a new approach was developed to identify the location and extent of structural damage with a limited number of measurements. A general area of structural damage is first identified and, subsequently, a specific damaged structural component is located. This approach takes advantage of two different model refinement methods (optimal-update and design sensitivity) and two different model size matching methods (model reduction and eigenvector expansion). Performance of the proposed damage detection approach was demonstrated with test data from two different laboratory truss structures. This space technology can also be applied to structural inspection of aircraft, offshore platforms, oil tankers, bridges, and buildings. In addition, its applications to model refinement will improve the design of structural systems such as automobiles and electronic packaging.

Key Words: Damage detection; health monitoring; modal testing; model correlation; nondestructive evaluation; space structure; and system identification.

*Senior Principal Engineer, Space Station Division
†Engineer Specialist, Space Station Division
‡Assistant Professor, Department of Engineering Mechanics
§Associate Professor, Department of Mechanical Engineering

Presented at the Dual-Use Space Technology Conference, NASA-JSC, Publication No. 3263, Houston, TX, Feb. 1994. Copyright ©1995 by McDonnell Douglas Corporation. Published by Society for MFPT with permission.

INTRODUCTION: Space structures intended for long-term operation such as the International Space Station Alpha (ISSA) may be damaged by micrometeoroid and orbital debris impacts, long exposure to the space environment, fatigue, or accidents with spacecraft docking and berthing operations. Comprehensive on-orbit inspection of the ISS structure using traditional non-destructive evaluation (NDE) methods requires extensive extravehicular activity (EVA) and does not always reveal structural anomalies. Robotic inspections are also possible, but there are limitations on areas of the spacecraft which can be inspected visually by robots or EVA astronauts.

Offshore structures for oil and gas production also must survive for thirty years or more in remote and hostile environments. They are subject to damage from tropical storms, hurricanes, corrosion, fatigue, or docking accidents. Offshore structures are inspected below the waterline by divers who, much like EVA astronauts, have limitations on the time their life support systems can sustain them for structural inspection. Robotic or human visual inspections are hampered by murky waters, barnacles, and other subsea plant life attached to the structure.

Vibration measurements have been used with system identification algorithms to produce quantitative damage detection results [1]. A math model of the undamaged structure, usually correlated with test data of the undamaged structure, is used with vibration information measured from the damaged structure. This damage detection approach is equivalent to verification of structural properties in specific locations (called model refinement). This non-traditional use of vibration measurements shows promise for application to on-orbit, remote NDE for both manned and unmanned space structures.

The proposed on-orbit damage detection method monitors the location and extent of structural damage through periodic system identification during the service life of the spacecraft. Modal parameters such as natural frequencies and mode shapes from a damaged structure can be related to the system mass and stiffness matrices of the undamaged structure. The feasibility of identifying mode shape, frequency, and damping of large space structures through on-orbit modal tests has been addressed in previous studies [2]. In addition, the cross-correlation random data processing strategy was developed to identify modal parameters using ambient responses from spacecraft [3].

However, the outcome of the research will have much wider applications. Offshore structural inspections can benefit from the same technology which utilizes vibration responses from the structure to locate the specific area and extent of suspected damage. Oil tankers, bridges, and buildings can also utilize the same technology for structural inspection. Serious accidents can be avoided through the use of a real-time, early warning system implemented by a small number of sensors on those existing structures. In addition, applications in model refinement will improve the design of structural systems such as automobiles and electronic packaging.

This paper describes a damage detection and health monitoring method that was developed for large and complex structures with a limited number of measurements. Theoretical background is provided for the optimal-update method, design sensitivity analysis technique, and a model reduction/expansion technique. In the new two-step damage detection approach, a general area of structural damage is first identified and, subsequently, a specific damaged structural component is located. Performance of the proposed damage detection approach was demonstrated with testing and analysis of the NASA-LaRC 8-bay and McDonnell Douglas 10-bay truss structures. Further detailed theoretical and technical information can be found in References [4, 5, 6].

INITIAL DAMAGE DETECTION: System identification techniques can be directly applicable to structural damage detection and categorized into the following four classifications: classical approach [7], design sensitivity approach [8], eigenstructure assignment approach [9], and optimal-update approach [10]. The optimal-update approach adjusts system mass and/or stiffness matrices to match the test frequencies and mode shapes by solving a constrained optimization problem. In addition to minimum changes of the initial system matrices, other physical properties can be imposed such as symmetry, positive definiteness, and structural connectivity. By comparing the adjusted model to the initial model, structural damage can be identified and located. This approach was selected for initial damage detection because it is applicable to large and complex structures.

Theoretical Background: Figure 1 shows an analysis flow diagram for the initial damage detection approach. An analytical (undamped, linear) model of

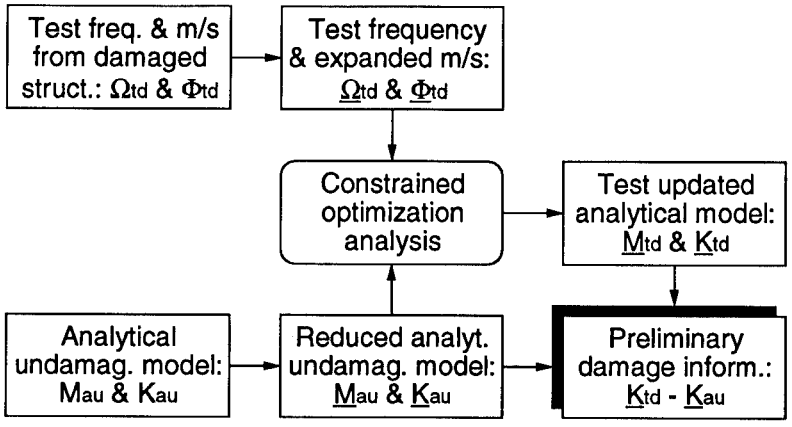


Figure 1: An Analysis Flow Diagram for the Initial Damage Detection Approach.

the undamaged structure can be generated using the finite element method and written as

$$M_{au}^{n\times n}\ddot{x}_{au}^{n\times 1} + K_{au}^{n\times n}x_{au}^{n\times 1} = 0 \tag{1}$$

where subscripts a and u represent analytical and undamaged, respectively, and superscript n represents n-degree of freedom (dof). Frequencies and mode shapes of the damaged structure obtained from modal testing will satisfy the following modal equations

$$-M_{td}^{m\times m}\Phi_{td}^{m\times p}(\Omega_{td}^{p\times p})^2 + K_{td}^{m\times m}\Phi_{td}^{m\times p} = 0 \tag{2}$$

where subscripts t and d represent test and damaged, respectively; and superscripts m and p represent the number of measurements and the number of identified modes, respectively ($m \ll n$).

To compare the undamaged analytical model and damaged test model for damage detection, they must have the same number of dofs. Therefore, the order of the analytical model (system mass and stiffness matrices) must be reduced or the order of test results (mode shapes) must be expanded. A new reduction/expansion approach was developed by selecting an intermediate dof set, i, which is small enough for computational efficiency but large enough to describe the physical model in sufficient detail to locate the damage. The orders of equations 1 and 2 will become the same, allowing comparison to locate the structural damage:

$$\hat{M}_{au}^{i\times i}\ddot{\hat{x}}_{au}^{i\times 1} + \hat{K}_{au}^{i\times i}\hat{x}_{au}^{i\times 1} = 0 \tag{3}$$

$$-\hat{M}_{td}^{i\times i}\hat{\Phi}_{td}^{i\times p}(\Omega_{td}^{p\times p})^2 + \hat{K}_{td}^{i\times i}\hat{\Phi}_{td}^{i\times p} = 0 \tag{4}$$

where ($\hat{}$) indicates new variables through the coordinate transformations [4].

The basic optimal-update approach, developed by Baruch and Bar Itzhack [10], adjusts the analytical stiffness matrix using the modal parameters obtained from modal testing. The undamaged analytical stiffness matrix is adjusted by using the measured frequencies and mode shapes to estimate the damaged stiffness matrix, i.e. (superscripts are dropped)

$$\begin{aligned} &\text{minimize} \quad \theta = \tfrac{1}{2}\|\hat{M}_{au}^{\frac{1}{2}}(\hat{K}_{td} - \hat{K}_{au})\hat{M}_{au}^{-\frac{1}{2}}\| \\ &\text{subject to} \quad \hat{K}_{td}^{T}\bar{\Phi}_{td} = \hat{M}_{au}\bar{\Phi}_{td}\Omega_{td}^2,\ \hat{K}_{td} = \hat{K}_{td}^{T} \end{aligned} \tag{5}$$

where $\bar{\Phi}_{td}$ is the orthogonalized modal matrix. Solutions to Equation 5 result in

$$\begin{aligned} \hat{K}_{td} = \hat{K}_{au} - \hat{K}_{au}\bar{\Phi}_{td}\bar{\Phi}_{td}^{T}\hat{M}_{au} - \hat{M}_{au}\bar{\Phi}_{td}\bar{\Phi}_{td}^{T}\hat{K}_{au} \\ +\hat{M}_{au}\bar{\Phi}_{td}\bar{\Phi}_{td}^{T}\hat{K}_{au}\bar{\Phi}_{td}\bar{\Phi}_{td}^{T}\hat{M}_{au} + \hat{M}_{au}\bar{\Phi}_{td}\Omega_{td}^2\bar{\Phi}_{td}^{T}\hat{M}_{au} \end{aligned} \tag{6}$$

By comparing differences in the stiffness matrices of the test-adjusted *damaged* model and the initial analytical *undamaged* model, i.e. $\hat{K}_{td} - \hat{K}_{au}$, the location and extent of structural damage can be identified.

Case Studies: Performance of the proposed initial damage detection approach was demonstrated using test data from the NASA-LaRC 8-bay Dynamic Scale Model Technology (DSMT) model. Figure 2 shows a schematic of the cantilevered truss structure which was fully instrumented at 96 translational dofs and tested for 15 different damage conditions. Full instrumentation allowed the study of

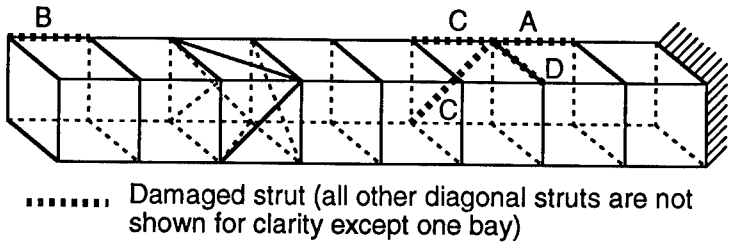


Figure 2: A Schematic of the NASA-LaRC 8-Bay DSMT Model.

sensor placement, model reduction, and eigenvector expansion techniques. Types of damage include full, multiple, and partially damaged struts from (easy-to-locate) high-strain longerons to (difficult-to-locate) low-strain battens. During the study, several techniques were employed to enhance the damage detection capability: sensor placement, identification of bad sensors, an orthogonal procrustes (OP) technique [11] for mode shape expansion, and a parallel technique based on minimum rank perturbation theory (MRPT) [12].

The performance improvement from each enhancement for all 15 damage cases was examined and quantified [6]. Table 1 shows part of the damage analysis results, with two different enhancements, for four damage cases in which the stiffness loss of a removed member is 100%. For this study, five test modes at all 96 instrumented dofs were used without model reduction or eigenvector expansion. The baseline approach located two of five damaged members. The enhanced baseline process using iterative sparsity preservation improved the performance considerably except for the case with damage in a low-strain batten. The use of MRPT successfully located all the damaged members and provided an improved assessment of the stiffness loss in each case.

DETAILED DAMAGE DETECTION: The optimal-update approach, even with enhancements, identifies only a general area of structural damage when the instrumentation is limited. The design sensitivity approach provides damage detection information directly related to physical properties of the individual structural members by minimizing analytical/modified and test modal parameters. Without *a priori* information, however, this method requires extensive computational resources for large and complex structures. This approach was selected to identify a specific damaged structural component based on the initial damage

Case	Damage Type	Baseline	Enhanced Baseline	Enhanced MRPT
A	high-strain longeron	*	46%	91%
B	low-strain longeron	4%	37%	77%
C	multiple members - longeron/diagonal	*/21%	64/48%	95/93%
D	low-strain batten	*	*	62%**

* not located ** located with difficulty

Table 1: Damage Detection Results with Enhancements for the DSMT Model.

detection results.

Theoretical Background: Figure 3 shows an analysis flow diagram for the detailed damage detection approach. The design vector, which includes physical

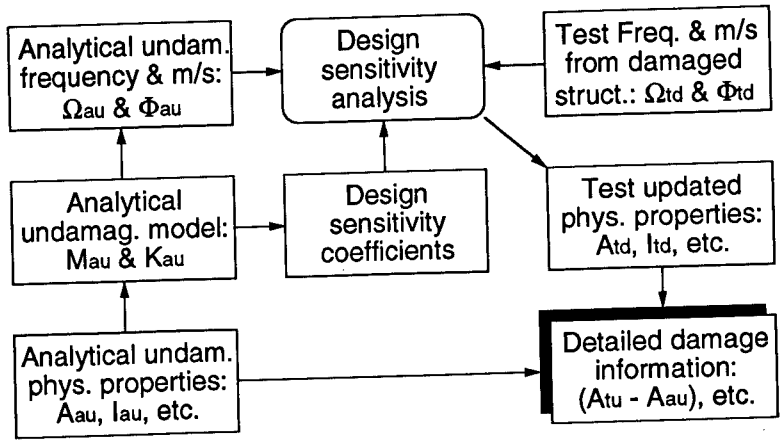


Figure 3: An Analysis Flow Diagram for the Detailed Damage Detection Approach.

properties such as cross-sectional area and moment of inertia, is defined as

$$\boldsymbol{u} = [u_1\, u_2\, \cdots\, u_d]^T \in \Re^{d\times 1} \tag{7}$$

The state vector, which includes frequencies and mode shapes, is defined as

$$\boldsymbol{v} = [v_1\, v_2\, \cdots\, v_s]^T \in \Re^{s\times 1} \tag{8}$$

Using linear perturbation theory, the change of state vector, $\Delta \boldsymbol{v}$, due to the normalized change of design vector, $\Delta \boldsymbol{u}$, can be approximated without solving the eigenvalue problem, i.e. [5]

$$\Delta \boldsymbol{v} \doteq Q \, \Delta \boldsymbol{u} \tag{9}$$

where Q is called a sensitivity matrix and defined as

$$[Q_{ij}] = \left[\frac{\partial v_i}{\partial u_j}\right] \in \Re^{s \times d} \tag{10}$$

The modified state vector, $\boldsymbol{v}^m \in \Re^{s \times 1}$, with respect to the change in design vector can be calculated using Equation 9

$$\boldsymbol{v}^m = \boldsymbol{v}^o + \Delta \boldsymbol{v} \doteq \boldsymbol{v}^o + Q \, \Delta \boldsymbol{u} \tag{11}$$

where $\boldsymbol{v}^o \in \Re^{s \times 1}$ is the original state vector calculated from the undamaged analytical model.

The state vector error, $\boldsymbol{e}^s$, and the magnitude of design vector change, $\boldsymbol{e}^d$, are defined by

$$\boldsymbol{e}^s = \left|\frac{\boldsymbol{v}^t - \boldsymbol{v}^m}{\boldsymbol{v}^t}\right| \in \Re^{s \times 1} \text{ and } \boldsymbol{e}^d = \left|\frac{\Delta \boldsymbol{u}}{\boldsymbol{u}^o}\right| \in \Re^{d \times 1} \tag{12}$$

where $\boldsymbol{u}^o \in \Re^{d \times 1}$ is the original design vector used in the undamaged analytical model and $\boldsymbol{v}^t \in \Re^{s \times 1}$ includes frequencies and mode shapes obtained from modal testing of the damaged structure. The optimal design vector can be found by minimizing the difference between the test results and the modified analytical results with least changes to the analytical model, i.e.

$$\text{minimize} \quad \theta = \boldsymbol{g}^T \boldsymbol{e}^s + \boldsymbol{h}^T \boldsymbol{e}^d \tag{13}$$

where $\boldsymbol{g} \in \Re^{s \times 1}$ and $\boldsymbol{h} \in \Re^{d \times 1}$ are weighting vectors for the state vector error and design vector change, respectively. Now the damaged structural component can be identified by examining the resulting design vector change. During the study, several techniques were employed to enhance the damage detection capability such as using a multiple set of design sensitivity coefficients.

Case Studies: Performance of the proposed detailed damage detection approach was demonstrated using test data from the 5.1-meter long hexagonal truss (HexTruss) which has limited instrumentation. Figure 4 shows the location and sensing direction of 48 accelerometers mounted on the 10-bay (5 full- and 5 half-hexagonal bays) HexTruss which was tested for 20 different damage conditions. A finite element model of 432 dofs was created for the undamaged configuration for use in the damage detection analysis. Analysis results for one damage case,

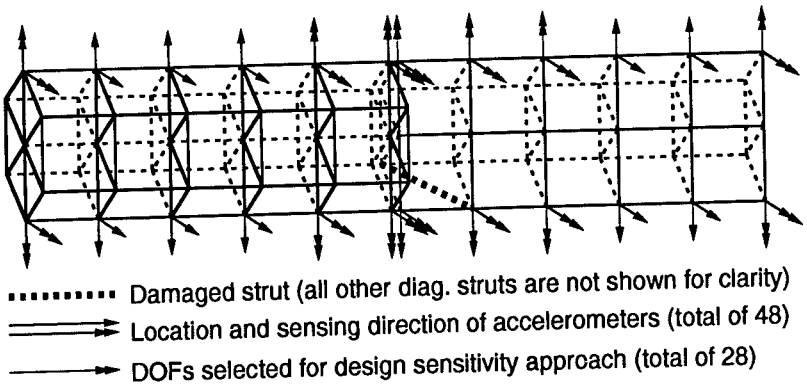


Figure 4: A schematic of the 10-Bay HexTruss.

which had a broken diagonal truss element in the half-hexagonal section, are discussed in the following.

For initial damage detection, test mode shapes of the intentionally damaged HexTruss were expanded to 144 dofs from 48 measured dofs using the dynamic expansion method [13]. The undamaged analytical system matrices were reduced to the same 144 dofs from 432 dofs using the Guyan reduction method [14]. Nine frequencies and mode shapes of the damaged structure along with the undamaged analytical model were used to detect and locate the structural damage using the simple Baruch method. The initial damage detection approach located the damaged truss bay, but did not locate the specific damaged truss member due to the limited number of measurements.

For detailed damage detection, based on the initial damage information, all 22 truss members contained in the suspect truss bay were selected (out of 224 members in the entire truss) as potential damaged components. This resulted in 88 design variables for consideration as potential damaged properties, out of at least 896 physical properties without preselection. Mode shapes were included in the state vector, because the specific damaged member could not be identified with frequencies alone. To further reduce the computational effort, two modes and 28 measurements were selected for the detailed damage detection among nine modes and 48 measurements used for the initial damage detection, respectively. As shown in Figure 5, the damaged member of the HexTruss was identified within three iterations.

CONCLUSIONS: A damage detection and health monitoring method for large and complex structures with limited instrumentation was presented. The method first identifies a general area of structural damage using optimal model update and a hybrid model reduction/eigenvector expansion technique. Then, it locates

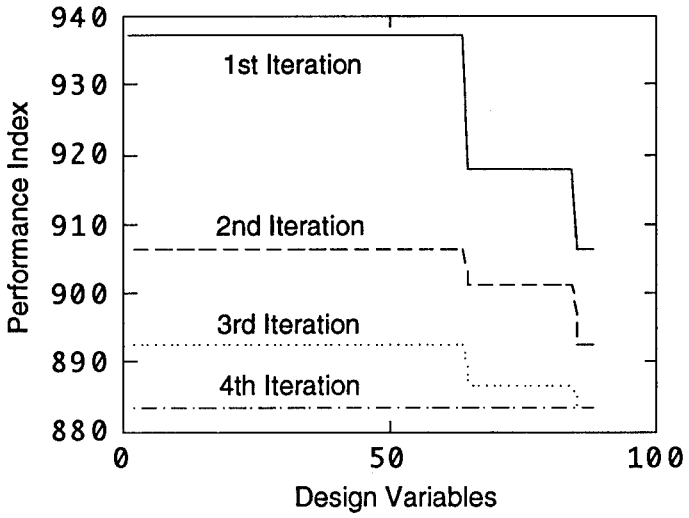


Figure 5: Convergence of Performance Index for the HexTruss.

a specific damaged structural component using a design sensitivity technique based on the initial damage information. Through the experimental validation, it was found that the proposed two-step damage detection approach can locate the specific damaged component of large structures with limited instrumentation. The same technology developed for space structures has a potential for structural inspection of aircraft, oil platforms, oil tankers, bridges, and buildings.

ACKNOWLEDGEMENT: The authors wish to express their sincere thanks to H.H. Doiron of MDA for many helpful discussions, T.A.L. Kashanggaki of NASA-LaRC for providing the DSMT test results, and D.A. VanHorn of MDA for extracting modal parameters from the HexTruss test data.

REFERENCES:

[1] Zimmerman, D.C. and Smith, S.W., "Model Refinement and Damage Location for Intelligent Structures," *Intelligent Structural Systems*, H.S. Tzou and G.L. Anderson, Editors, pp. 403–452, Kluwer Academic Publishers, The Netherlands, 1992.

[2] Kim, H.M. and Doiron, H.H., "On-Orbit Modal Identification of Large Space Structures," *Sound & Vibration*, Vol. 26, No. 6, pp. 24–30, June 1992.

[3] Kim, H.M., VanHorn, D.A., and Doiron, H.H., "Free-Decay Time-Domain Modal Identification for Large Space Structures," *J. GUIDANCE*, Vol. 17, No. 3, pp. 513–519, May 1994.

[4] Kim, H.M. and Bartkowicz, T.J., "Damage Detection and Health Monitoring of Large Space Structures," *Sound & Vibration*, Vol. 27, No. 6, pp. 12–17, June 1993.

[5] Kim, H.M. and Bartkowicz, T.J., "A Two-Step Structural Damage Detection Approach with Limited Instrumentation," to be presented at the *35th AIAA SDM Conference*, Hilton Head, SC, April 1994.

[6] Zimmerman, D.C., Smith, S.W., Kim, H.M., and Bartkowicz, T.J., "An Experimental Study of Structural Damage Detection Using Incomplete Measurements," to be presented at the *35th AIAA SDM Conference*, Hilton Head, SC, April 1994.

[7] West, W., "*Structural Fault Detection of a Light Aircraft Structure Using Modal Technology,*" JSC Loads and Structural Dynamics Branch Report, Apr. 1988.

[8] Flanigan, C.C., "Test/Analysis Correlation Using Design Sensitivity and Optimization," *Aerospace Technology Conference and Exposition*, SAE Paper No. 871743, Long Beach, CA, Oct. 1987.

[9] Minas, C. and Inman, D.J., "Correcting Finite Element Models with Measured Modal Results Using Eigenstructure Assignment Methods," *Proc. 6th IMAC*, pp. 583–587, Feb. 1988.

[10] Baruch, M. and Bar Itzhack, I.Y., "Optimal Weighted Orthogonalization of Measured Modes," *AIAA Journal*, Vol. 16, No. 4, pp. 346–351, Apr. 1978.

[11] Smith, S.W., "Simultaneous Expansion and Orthogonalization of Measured Modes for Structure Identification," *Proc. 31st AIAA SDM Conference - Dyn. Spec.*, pp. 261–270, Apr. 1990.

[12] Kaouk, M. and Zimmerman, D.C., "Structural Damage Assessment Using a Generalized Minimum Rank Perturbation Theory," *Proc. 34th AIAA SDM Conference*, La Jolla, CA, pp. 1529–1538, Apr. 1993.

[13] Berman, A. and Nagy, E.G., "Improvement of a Large Analytical Model Using Test Data," *AIAA Journal*, Vol. 21, No. 8, pp. 1168–1173, Aug. 1983.

[14] Guyan, R.J., "Reduction of Stiffness and Mass Matrices," *AIAA Journal*, Vol. 3, No. 2, pp. 380, 1965.

EXPERIMENTAL STUDY OF EMBEDDED ABSOLUTE EFPI FIBER OPTIC STRAIN SENSORS FOR CIVIL STRUCTURE APPLICATIONS

M.J. de Vries, V. Bhatia, T. Tran*, K.A. Murphy, R.O. Claus

Fiber & Electro-Optics Research Center, Bradley Department of Electrical Engineering
Virginia Polytechnic Institute & State University, Blacksburg, VA 24061-0111
(703) 231-7203, (703) 231-4561 (fax)

* Fiber & Sensor Technologies, Inc.
250 Arbor Drive NE, Christiansburg, VA 24073
(703) 382-7556, (703) 382-8072 (fax)

Absolute extrinsic Fabry-Perot interferometric (AEFPI) optical fiber strain sensor elements and collocated reference foil strain gages were individually attached to composite tendons, used for prestressing concrete, prior to pretensioning the tendon. The fiber sensors and foil strain gages were attached to the tendon prior to applying any loads, since it is important to accurately monitor strain throughout the entire loading process. Attaching fiber sensors to a strand that has been loaded can be extremely dangerous. Surface treatments prior to sensor attachment as well as imperfections in the strand itself, can cause the strand to snap and result in serious personal injury. This paper discusses a fiber attachment method that allows strains on the order of 12,000 μstrain to be measured accurately, without degrading the quality of the glass-fiber. Quantitative strain measurements obtained with both the fiber sensors and the foil strain gages are presented. Results show that fiber optic sensors, when properly attached, can provide accurate information regarding structural integrity.

Key Words; Bridges; Civil structures; Fiber optic sensors; Nondestructive evaluation; Strain gauges

Introduction

Optical fiber sensors, unlike conventional electrically-based sensors, are fabricated with high-strength silica which will not corrode or be affected by electromagnetic

interference. Their small size and geometric flexibility allows them to be unobtrusively embedded or surface-attached to the host structure. Fiber optic sensors have been used for the quantitative, non-destructive evaluation of advanced materials and structures for over ten years.[1,2] Fiber sensors have been demonstrated to be feasible for the measurement of internal material changes during fabrication, the in-service lifetime measurement of strain, temperature, vibration, and other physical perturbations, and the eventual detection of damage or property degradation.[3,4] Recent advances in fiber optics communication technology and the widespread use of fiber communications have made available a large selection of reliable and inexpensive opto-electronic devices and components for use in sensors and sensor instrumentation support systems.

In this paper we describe the application of a state-of-the-art fiber optic sensing system to the measurement of strain on rebar and prestressing strands used to reinforce concrete.

Prestressed Concrete

Concrete, though strong in compression, is quite weak in tension, and compressive stress on it can be used to counterbalance any tensile force due to loading, which might lead to either cracking or deflection. Although prestressed concrete is more expensive to manufacture than reinforced concrete, the long term savings due to reduced maintenance and extended lifetime are making it a popular choice for use in long span structures, such as, segmental bridges.[5] The Philadelphia Walnut Lane Bridge was one of the first major structures in the United States that utilized prestressed concrete.

In pre-tensioning prestressed concrete, a tendon is tensioned before the concrete is poured and the prestress is transferred to the concrete after it has cured by releasing the tension on the tendon. If this prestressing force is applied along the axis of the structure, the procedure is termed linear prestressing. In linear prestressing it is often required to determine the axial strain on the tendon during the initial procedure of pre-tensioning, so the required longitudinal force to achieve maximum concrete strength can be accurately determined.

This measurement process should be economical and reliable, and should possess the appropriate resolution. Moreover, the sensing process should not in any way alter the properties of either the tendon or the concrete. Conventional techniques for this purpose involve the use of foil strain gages, which are not only expensive to use, but are also known for their failure rate in high strain environments. We discuss the absolute extrinsic Fabry-Perot interferometer fiber optic sensor for monitoring strain in pre-tensioned tendons used to prestress concrete.

The AEFPI System

The extrinsic Fabry-Perot interferometer (EFPI)[6,7] consists of two optical fibers enclosing an air cavity inside a hollow core fiber, as shown in Figure 1. Light is coupled into the lead-in fiber using a 3-dB optical coupler. The two reflections - sensing and reference - interfere at the detector face and produce a sinusoidal output intensity pattern on the digital oscilloscope screen (Figure 2). This intensity, I, is dependent on the length, d, of the air cavity between the two fibers inserted from either end of the hollow core fiber,

$$I = I_0 \cos\left(\frac{4\pi}{\lambda} d\right), \tag{1}$$

where, I_0 is the maximum value of the output intensity and λ is the laser diode center wavelength. Any external perturbation, like strain and temperature changes, modulates the value of the gap length, d, which in turns leads to variations in the output intensity, I.

This change in output intensity of a conventional EFPI system is non-linear corresponding to the magnitude of the parameter being measured. Moreover, if a change in direction of the applied strain or any other perturbation occurs at the maxima or minima of the sinusoidal transfer function curve, the system is unable to detect it. Additionally, complex and expensive fringe counting techniques have to be employed for post-processing of the acquired data.

The absolute EFPI (AEFPI) system[8] was designed to specialize the EFPI system for long-term field use. It utilizes the concept of white light interferometry in that a broadband source, like a super luminescent diode (SLED), is employed, instead of a laser diode. The signal processing is done using a spectrometer followed by a personal computer as depicted in Figure 3. Two wavelengths, λ_1 and λ_2 (Figure 4), which are 2π out of phase are determined from the spectrometer and a dedicated computer code determines the gap separation, d, unambiguously using,

$$d = \frac{(\lambda_1 \lambda_2)}{2(\lambda_2 - \lambda_1)}. \tag{2}$$

Other methods proposed to extract absolute information from an EFPI cavity are difficult to implement because they involve the use of two laser diodes or additional phase modulation equipment. The AEFPI system on the other hand, possesses the capability for absolute, on-line, real-time acquisition and signal processing of data without the necessity of expensive laser diode and fringe counting equipment. This system is self calibrating and possesses a large dynamic range for perturbation measurement. Moreover, unlike in the conventional EFPI system, the shift in the source spectrum no longer leads to degradation in the overall performance. The AEFPI system can detect temperature changes and strain as small as 2 °C and 100 με

(micro-strain), respectively. Since the system is not differential in nature, it does not require a reference measurement to be made prior to the actual sensing procedure, and an absolute measurement can be obtained every time the power to the system is turned on.

Experimental Results

To demonstrate feasibility, an AEFPI sensor was attached to a 15 inch section of steel rebar which was subsequently embedded in a small concrete block (9 x 4 x 1.5 inches) to form a small reinforced concrete specimen. The cured concrete specimen was subjected to a three point bend test. This was done to obtain a qualitative assessment of the strain field induced in the concrete specimen due to three point loading.

Figure 5 shows the increase in gap separation of the Fabry-Perot cavity as the load on the concrete was increased with the specimen in the three point bend configuration. The loading function is given in arbitrary units and the gap separation is in micrometers. Given the gage length of the AEFPI sensor, strain information can be derived from the change in gap separation. As seen from

Figure 5, the gap separation and hence the strain induced in the concrete specimen increases as a linear function of the increasing load in a three point bend test.

To obtain quantitative strain measurements, a load frame experiment was conducted. The concrete specimen with embedded AEFPI sensors was loaded in an Instron Model 1331 load frame. The specimen was subjected to a series of incremental static compressive loads. The load was applied in steps of 100 lbs up to a maximum of 1000 lbs. The wavelength domain information was recorded using an ANDO 6310 optical spectrum analyzer and the change in gap separation was calculated mathematically using the wavelength data. Three loading tests were conducted for incrementally increasing the load. The data obtained from the two tests is shown in Figure 6.

The next experiment was performed at the Turner Fairbanks Federal Highway Administration research facility in McLean, Virginia. The reinforcing of concrete was implemented using two pre-tensioned strands of different materials. A high finesse EFPI was surface attached to the prestressing strand. In order to enable survivability of the sensors in high strain environments, the input fiber was epoxied to the hollow core fiber while the output/reflective fiber was allowed to move freely inside the hollow core fiber by only epoxing it to the prestressing strand. The gage-length, L, was determined by measuring the distance between the point where the input fiber was epoxied to the hollow core fiber and the point where the output fiber was attached to the prestressing rod. Since the strain was not directly transferred to the fiber glass, the authenticity of the measurement process was maintained even when the load on the prestressing strand exceeded 10,000 lbs. The fiber sensors were

attached so that the axis of the strand and that of the hollow core fiber were parallel to each other. Five residual strain gages were also attached at different locations on each of the prestressing tendon.

One end of the prestressing tendon was kept anchored while the other end was loaded so that the strand was axially strained with increasing load. The load was increased in steps of approximately 2,000 lbs. The measured strain is plotted versus the applied load in Figure 7. The final value of strain measured by the AEFPI system was 12,255 $\mu\varepsilon$, which was within 1 % of the value measured by a collocated strain gage. The strain (ratio of the change in the air-gap, Δd, and the gage length, L) from the AEFPI system was obtained real-time and on-line, while the output waveforms were stored for post-processing. It is interesting to note that while two of the strain gages experienced catastrophic failure at high values of loads, the fiber sensors performed as well as they did at lower values of loads. For the second fiber strand also, both the conventional EFPIs survived loads greater than 10,000 lbs while at least one strain gage suffered permanent damage. This might suggest the application of EFPI sensors in monitoring strain in harsh environments where conventional techniques might not prove to be practical.

Conclusions

We have demonstrated the survivability of EFPI fiber optic sensors in conditions where the axial strain exceeds 10^5 $\mu\varepsilon$. The AEFPI system was shown to detect strain within 1 % of the value measured by a collocated conventional foil strain gage in strands used for prestressing concrete. Such fiber optic sensors could be used in applications requiring real-time, absolute measurement of parameters such as axial strain and temperature change in the rapidly developing field of the so-called "smart" materials and structures. Fiber optic sensors might, in the long run, reduce the cost of monitoring and maintaining civil structures, such as highways and bridges, by providing accurate information about structural degradation prior to any major catastrophe.

Future work in the development of the AEFPI system will include the improvement of the strain rate of the system and investigation of techniques to multiplex a large number of sensors using a single system.

Acknowledgements

We would like to acknowledge support from the Virginia Center for Innovative Technology, Fiber & Sensor Technologies Inc., and Dr. Charles McGogney, Dr. William Wright, Dr. Eric Munley, and Carol Thompson at the Federal Highway Administration in McLean, Virginia.

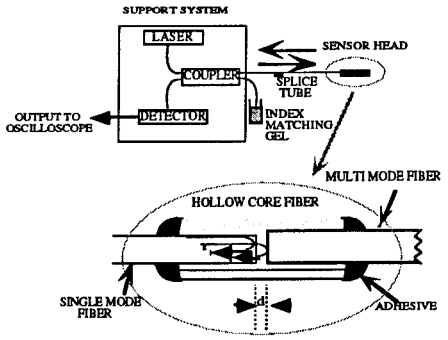


Figure 1. The extrinsic Fabry-Perot interferometer (EFPI) system.

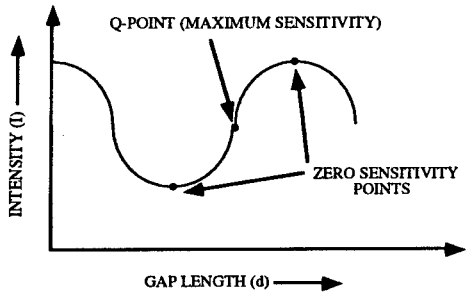


Figure 2. The EFPI transfer function curve.

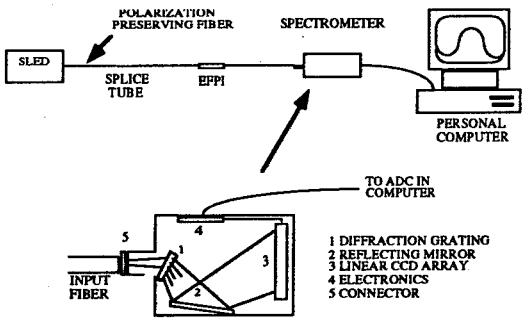


Figure 3. Setup of the Absolute EFPI (AEFPI) system.

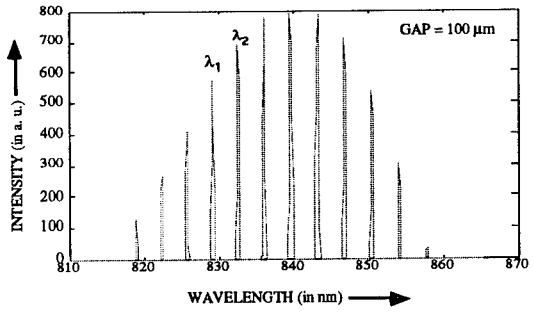


Figure 4. Typical output of the AEFPI system.

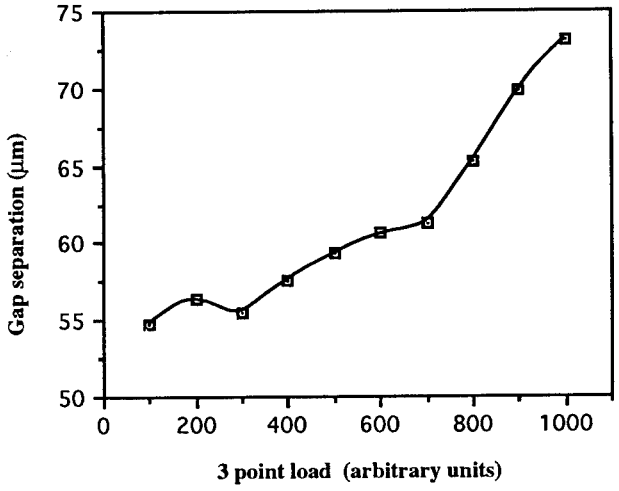


Figure 5. Gap separation in AEFPI sensor as a function of increased three point bending.

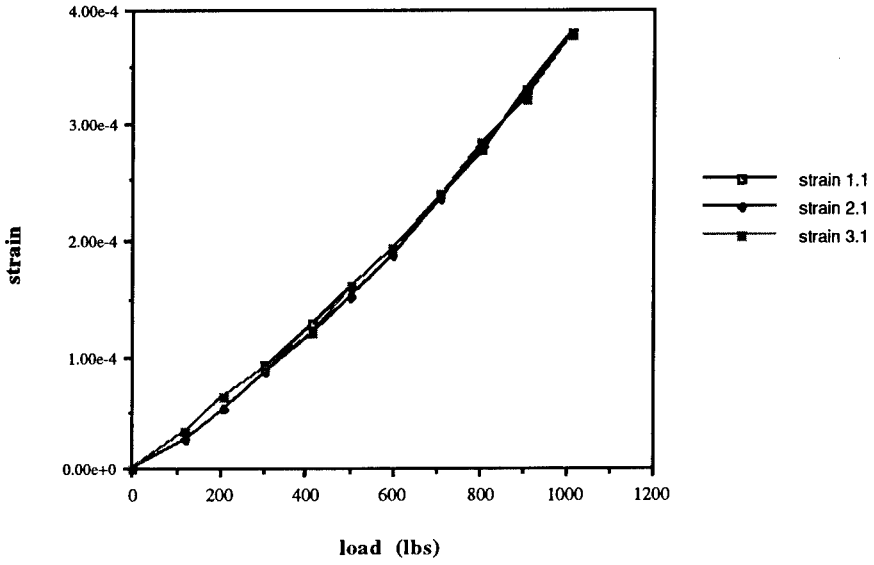


Figure 6. Strain measured by the AEFPI sensor as a function of increasing load on the reinforced concrete specimen.

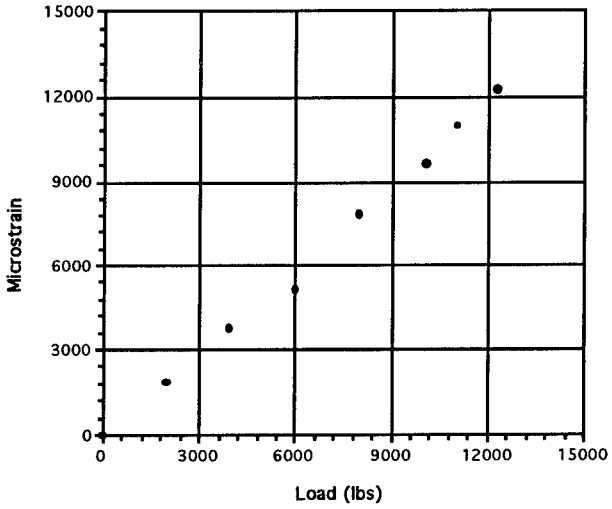


Figure 7. Measured strain versus applied load.

References

[1] J. Dakin, B. Culshaw, Optical Fiber Sensors: Principles and Components, Artech House, Boston, MA 1988.

[2] T.G. Giallorenzi, J.A. Bucaro, A. Dandridge, G.H. Sigel, J.H. Cole, S.C. Rashleigh, R.G. Priest, "Optical fiber sensor technology," IEEE J. of Quantum Electronics, vol. 18, pp. 626 - 665, 1982.

[3] E. Udd, Overview of fiber optic applications to smart structures, 1988, New York: Plenum.

[4] R.O. Claus, Fiber Optic Sensor-Based Smart Materials and Structures, IOP Publishing, Bristol and Philadelphia, April 1992.

[5] E. G. Nawy, Prestressed Concrete, Prentice Hall, New Jersey, 1989.

[6] R. O. Claus, M. F. Gunther, A. Wang, and K. A. Murphy , "Extrinsic Fabry-Perot sensor for strain and crack opening displacement measurements from -200 to 900 °C," *Smart Materials and Structures*, vol. 1, pp. 237-242, 1992.

[7] K.A. Murphy, M.F. Gunther, A. Vengsarkar, C. Koob, R.O. Claus, "Extrinsic Fabry-Perot fiber optic sensor," U.S. Patent No. 5,301,001.

[8] T. A. Tran, J. A. Greene, M. S. Miller, V. Bhatia, K. A. Murphy, A. Wang, and R. O. Claus, "Absolute strain measurements using the extrinsic Fabry-Perot interferometer," *FEORC Fiber Optics Review Conference*, Blacksburg, VA, April 1993.

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" April 25-27, 1972 — AD 750 411

Meeting No 18 "Detection, Diagnosis and Prognosis" November 8-10, 1972 — AD 772 082

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) April 23-25, 1975 — PB 248 254

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) November 3-5, 1976 — PB 271 743

Meeting No 26 "Detection, Diagnosis and Prognosis" (NBS SP 494) May 17-19, 1977 — PB 272 848

Meeting No 27 "Product Durability and Life" (NBS SP 514) November 1-3, 1977 — PB 281 878

Meeting No 28 "Detection, Diagnosis and Prognosis" (NBS SP 547) November 28-30, 1978 — PB 297 399

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

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS None
2a SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) MFPT 49	5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A
6a NAME OF PERFORMING ORGANIZATION Vibration Institute 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) 6262 S. Kingery Hwy Willowbrook, IL 60514	7b. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.
11 TITLE (Include Security Classification) Life Extension of Aging Machinery and Structures	
12 PERSONAL AUTHOR(S) Compiled by Henry C. Pusey and Sallie C. Pusey	
13a TYPE OF REPORT Proceedings 13b. TIME COVERED FROM 95/1/1 TO 95/4/20	14. DATE OF REPORT (Year, Month, Day) 95/4/20 15. PAGE COUNT 436
16 SUPPLEMENTARY NOTATION	

17 COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
			Machinery, structures, condition monitoring, failure analysis, diagnostics, sensors, signature analysis, life extension, oil analysis and nondestructive evaluation

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This document is the proceedings of the 49th meeting of the Society for Machinery Failure Prevention Technology (MFPT) which was held in Virginia Beach, Virginia on April 18-20, 1995. There are session papers on Failure Analysis, Diagnostics, Life Extension, Sensors Technology, Time Frequency Analysis and Detection, Monitoring and Response.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL Peter Schmidt	22b. TELEPHONE (Include Area Code) (703) 696-4362 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

☆ U.S. Government Printing Office: 1985—607-044