AD-A285 785

ADVANCED MATERIALS AND PROCESS TECHNOLOGY FOR MECHANICAL FAILURE PREVENTION

This Document Contains Missing Page/s That Are Unavailable In The Original Document

Reproduced From Best Available Copy

(D)

ADVANCED MATERIALS AND PROCESS TECHNOLOGY FOR MECHANICAL FAILURE PREVENTION

Proceedings of the 48th Meeting
of the
Mechanical Failures Prevention Group

Wakefield, Massachusetts
April 19-21, 1994

DTIC
ELECTE
OCT 28 1994
S B D

Compiled by
Henry C. Pusey
and
Sallie C. Pusey

DTIC QUALITY INSPECTED 2

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

A Publication of the

Vibration Institute
A NOT-FOR-PROFIT CORPORATION

94-33416

45987

9410 21 116

Copyright © 1994 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

APPLICATIONS OF DIAGNOSTICS

NONDESTRUCTIVE TEST AND EVALUATION

PREFACE

The 48th Meeting of the Mechanical Failures Prevention Group(MFPG) was sponsored by the Office of Naval Research(ONR), Arlington, VA; Naval Surface Warfare Center(NSWC), Annapolis, MD; the Naval Civil Engineering Laboratory(NCEL), Port Hueneme, CA; the U.S. Army Research Laboratory(ARL), Watertown, MA and the Vibration Institute. The conference was hosted by ARL and held April 19-21, 1994 at the Colonial Hilton Hotel in Wakefield, Massachusetts. Meeting management, program coordination, and proceedings compilation were by the Vibration Institute. The Nondestructive Testing Information Analysis Center (NTIAC) assisted the sponsors in developing the program and promoting the conference. MFPG Council Member and Host Representative **Victor K. Champagne** chaired the Opening Session; Council Chairman **G. William Nickerson** was CoChairman. The CoChairmen of the Technical Sessions are identified on the title pages for each section in these proceedings. The MFPG Council and the MFPG Program Committee Members are listed separately.

We were fortunate to have **Professor Thomas W. Eager** from the Massachusetts Institute of Technology as our Keynote Speaker. Doctor Eager's paper, along with two Opening Session papers (Speakers **Craig Rogers** and **James Taylor**) and a Plenary Paper (Speaker **T. James Du Bois**), are included in the **FEATURED PAPERS** section of these proceedings. Four distinguished invited speakers presented excellent papers that are not included in the proceedings. **RADM Andrew A. Granuzzo**, Commander of the Naval Safety Center, spoke on *Safety and Mechanical Diagnostics* and **Dr. John P. Gudas** from NIST discussed *Materials Research and Development in the Advanced Technology Program*. On the first afternoon, **Dr. Richard D. Sisson,Jr.** from Worcester Polytechnic Institute gave a Plenary Address on *Hydrogen Embrittlement of Electroplated Fasteners-Physical Metallurgy and Mitigation Methods*. The final Plenary Lecture on *The Technology Reinvestment Project and the Implications for Materials Related R & D* was presented by **Dr. C. Robert Crowe** of the Advanced Research Projects Agency.

The MFPG Technical Program also included three Mini Courses, four formally scheduled Technical Committee meetings and a final afternoon Panel/Workshop on *Machinery Diagnostic Technology: What is Most Useful, What is Most Useless and What is Missing?* The Workshop organizers were **G. W. Nickerson** and **Paul. L. Howard.** The Mini Courses presented were as follows:

1. *Failure Analysis:* **Mr. Cedric D. Beachem**, U.S. Naval Research Laboratory, Washington, DC

2. *Failure Mechanisms in Polymers*: **Dr. Rudolph J. Scavuzzo**, University of Akron, Akron, OH

3. *Understanding the Technology Transfer Programs and Your Opportunities in Them*: **Dr. Gary W. Carriveau**, NTIAC/TRI, Austin, TX

I wish to call attention to two papers that were not listed in the Technical Program but which are included in these Proceedings. These were to have been listed as **Supplementary Papers** which were to be presented should space become available and were to be published in any case. I apologize to the authors for my oversight. The papers are: (1) *The Automated Transformer Oil Monitor (ATOM)* by **G. J. Gorton**, published in the **Machinery Monitoring and Diagnostics Session**, and (2) *On the Quantum Electrodynamic Origins of "Scatter" in the Mechanical Properties of Metals* by **M. Ensanian**, published in the **Failure Analysis Session**.

In selecting the 1994 THEME, the MFPG Council recognized that the White House Technology Reinvestment Project (TRP), properly executed, can stimulate a healthy, competitive economy while maintaining a strong national defense using the most advanced, affordable military systems. It was also clear to the Council that the many technical disciplines essential to achieving mechanical failure prevention or reduction are also essential to the success of the TRP. It was hoped that MFPG 48 would not only present new technical information on advanced materials and process technology that contributes to mechanical failure prevention, but also provide information on how conference participants can benefit from government/industry cooperative programs. Those who attend the conference are the final judges of how well these goals were achieved.

The Mechanical Failures Prevention Group was organized in 1967 under the sponsorship of the Office of Naval Research. The MFPG was formed for the express 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 MFPG will continue to act as a focal point for any technological developments that contribute to mechanical failure reduction or prevention. For this reason MFPG 48 was the occasion for the first scheduled open meetings of the four Technical Committees. All interested conference registrants were invited to attend. Perhaps as a result of these meetings specific goals and objectives for each committee can be established and subcommittee chairmen can be identified and appointed to provide leadership for the MFPG effort. The Chairman of each Technical Committee will be asked to summarize the results of MFPG 48 deliberations and discuss their ongoing program at MFPG 49.

Those interested in working on any of the Technical Committees should contact the appropriate committee chairman. The committee chairmen are listed in these Proceedings as members of the MFPG Council.

On behalf of Dr. Eshleman and the Vibration Institute, I want to thank our sponsors, NTIAC and the MFPG Council for their cooperation in organizing and conducting the 48th MFPG 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 engaged in failure avoidance technology.

Henry C. Pusey
Executive Secretary

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist Avail and/or Special
A-1

MFPG COUNCIL

Chairman
G. William Nickerson
Pennsylvania State University
Applied Research Laboratory
PO Box 30
State College, PA 16804

Vice Chairman
MarjorieAnn E. Natishan
Naval Surface Warfare Center
Code 2814
Annapolis, MD 21402-5067

Past Chairman
Henry R. Hegner
ECO, Inc.
1036 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
Naval Civil Engineering Lab.
MC L72
Port Hueneme, CA 93043

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

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

Sensors
Henry R. Hegner
ECO, Inc.
1036 Cape St. Claire Center
Annapolis, MD 21401

TECHNICAL PROGRAM COMMITTEE

Chairman
MarjorieAnn E. Natishan
Naval Surface Warfare Center
Code 2814
Annapolis, MD 21402-5067

Members
Robert A. Bayles
Naval Research Laboratory
Code 6327
Washington, DC 20375-5000

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

Janice Garfinkle
Naval Air Warfare Center-AD
Code SR 13
Lakehurst, NJ 08733-5059

Henry R. Hegner
ECO, Inc.
1036 Cape St. Claire Center
Annapolis, MD 21401

Howard A. Gaberson
Naval Civil Engineering Lab.
MC L72
Port Hueneme, CA 93043

Ruben A. Lebron, Jr.
Naval Air Warfare Center-AD
Code SR 13
Lakehurst, NJ 08733-5059

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

Ex-Officio
Paul L. Howard
Paul Howard Enterprises
1212 Clearbrook Road
W. Chester, PA 19380

G. William Nickerson

Henry C. Pusey

FEATURED PAPERS

Opening Session and Plenary

AVOIDING FAILURES IN ADVANCED STRUCTURAL MATERIALS

T.W. Eagar, Co-Director Leaders for Manufacturing Program
POSCO Professor of Materials Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract: The dramatic growth of materials science and engineering over the past two decades has made it possible to engineer new materials with specialized properties in a very short period of time. We have changed our philosophy from "designing with materials" to "designing of materials." Indeed, we conceive many new products and systems before the necessary materials are available; we assume that the materials properties which are required can be developed. It is correct that often materials can be designed to meet these applications, but frequently we find that the cost of these specialized materials is excessive. In addition, as materials become ever more specialized, the market volume decreases which further increases the costs. As new designs require the operation of the materials closer to their limits, the failure rates increase. While the growth of materials science and engineering has brought, and will continue to bring, major advances in our standard of living, it is essential that we use sound judgment in selecting the new technologies which are most promising. Reliable use of advanced structural materials will require improvements in every facet of design, production and use of the product.

Key Words: Advanced structural materials; materials design; product design; failure rates; product failures; properties comparison; advanced materials.

Introduction: Currently, there is intense worldwide interest in advanced materials and processing. The Office of Technology Assessment has selected advanced materials, technology, bio-technology, and information technology as the keys to future industries with a growth in the materials market of $100-billion projected over the next decade. The development of these materials has been linked to our national competitiveness and our future standard of living (ref. 1).

Many research findings extol the benefits of newly developed polymers, inter-metallics, ceramics, and composites, with predictions that these new materials will change our lives significantly in the years to come. Many companies have reduced their research efforts in traditional materials and have diversified into these advanced materials. Others, constrained by an increasingly competitive international trade climate, have greatly reduced or eliminated their research and development activities altogether. Some have predicted a continual erosion of the market for traditional materials in favor of advanced materials. The rationality of this prediction is examined in this article.

Advanced materials have exciting properties but their commercialization often is stalled due to several pitfalls including biased presentation of materials properties, failure to consider the implications of materials substitution, and ignoring production-scale processing considerations. All of the factors that influence the substitution of a new material for a traditional one must be examined carefully to derive an accurate prediction of the potential of the new materials. For example, in the commercial-aircraft industry, where the consequences of a system failure can be very serious, there is considerable reluctance to incorporate advanced materials without extensive testing and prototyping. In other industries, such as those governed by industrial codes that have

been built up through years of experience, adoption of a new material can move at a glacial pace. Also, it is important to realize that in most applications, a traditional material can perform as well as an advanced material -- often at a lower overall cost. An examination of the methods used to assess a new material's performance during product development illustrates why some new materials are commercially successful while others are not.

The purported potential of an advanced material can be misleading for many reasons. Some benefits are related to the exuberance of the materials scientist in extolling his/her accomplishments, while others are related to failure to consider the many dimensions involved in substituting one material for another, including fair property comparisons and whether pilot-scale materials can be manufactured and fabricated cost effectively on a production basis. The real potential of an advanced material is often not apparent due to the incomplete assessment of its capabilities. The pitfalls involve:

- Discussion of property improvements only on a relative (rather than absolute) scale
- Use of inappropriate or carefully chosen adjectives that bias the merits of the material
- Discussion of the material one-dimensionally
- Comparison of the properties of a new material with the current (as opposed to potential) properties of a traditional material
- Ignoring market volume
- Ignoring processing considerations

Discussion of property improvements on a relative rather than absolute scale occurs when the materials scientist attempts to place the new development in its most favorable light. The reported improvements in the toughness of quartz glass illustrates this point. A recent research report claimed an increase of 100% in the fracture toughness of quartz glass. At first glance, the improvement appears to be a remarkable advance; however, closer evaluation of the report reveals that the improved glass now has a fracture toughness equivalent to one-half of that of gray cast iron. Thus, the new material hardly is a candidate for use in critical structural applications. Consideration of the absolute value of the toughness was disregarded in favor of the dramatic relative gain in performance.

Both strength and toughness are important properties in structural applications because high strength permits more economical use of the material, while toughness provides resistance to fracture. Metals are extensively used as structural materials because they have exceptional combinations of strength and toughness. By comparison, although glass and other ceramics have exceptional strength, they are also very brittle. This is why ceramics are used only in applications where loads are primarily compressive. Dramatic increases have been made in the relative toughness of ceramics, but on an absolute scale, there still is a very long way to go before they are considered for use in structural applications.

The use of inappropriate or carefully chosen adjectives to describe advanced materials can lend an aura of significant practical achievement to a material's properties when the development may only have reached the stage of scientific curiosity. Examples of such terms include: "tough or toughened" structural ceramics; "conductive" polymers, having a specific electrical conductivity higher than that of copper; and "high-strength" composites and fibers, having specific strengths greater than that of steel.

Each of these claims could be accurate. However, the important question is: What difference does the statement really make? Advanced ceramics may be tougher than their predecessors, but they hardly can be classified as tough materials, at least not in the sense that a designer or materials engineer understands the word. In the case of conductive polymers, while they are not insulators, the implication that their electrical properties exceed those of copper is misleading. It is true only if specific properties (property/material density) are compared. Although the best conductive polymers have an electrical conductivity that is within an order of magnitude of the conductivity of copper, so does mild steel. Yet, steel wire has never been considered as an alternative material to copper based on steel's electrical properties. Specific conductivity is not the critical measure of performance. If it was, aluminum which has a specific conductivity two times greater than that of copper, should be a viable candidate as a substitute for both copper and conductive polymers (which it is not).

A one-dimensional discussion of materials possibly is the most widespread method of praising new materials. A good illustration involves new high-temperature superconductors (HTSCs). The critical temperature of these ceramic materials was raised more than six-fold over a period of months in early 1987, and there were many claims that our lives would be improved within a few years. Unfortunately, this has not occurred.

A designer of a high-field-strength superconducting magnet must consider at least four materials properties: critical temperature, critical magnetic field, critical current density, and tensile strength. With respect to critical temperature, the new superconductors exceed the most optimistic hopes of only a few years ago. HTSCs also perform very well with respect to critical magnetic field. However, HTSCs missed the required level of performance for critical current density by six orders of magnitude initially. More recently, oriented crystals have demonstrated properties close to what is necessary at zero magnetic field. Unfortunately, there is an interaction between the critical current and the applied magnetic field that comes into play as well. Thus, at a useful magnetic field, even these high-current-density superconductors do not have superconducting properties.

Tensile strength has been totally ignored to date. The windings of a high-field-strength magnet experience extremely high tensile hoop stresses; up to 690 MPa (100×10^3 psi) in some instances. The brittleness of ceramic HTSCs poses an obstacle to their use, at least for the near future.

Another danger in assessing advanced materials is to compare the properties of a new material with the current properties of a traditional material. The danger in using this approach is assuming that the properties of traditional materials will not be improved further, especially when industries are faced with new competition.

The area of soft-magnetic materials illustrates this point. Nearly 20 years ago, a University of Pennsylvania researcher discovered that amorphous Fe-Si-B alloys had only one-tenth the magnetic core losses of crystalline Fe-Si alloys used in electrical transformers. Although the amorphous alloys cost two to three times as much as the traditional alloys, a ten-fold improvement in amorphous-alloy properties looked very promising. Based on the potential benefits of amorphous alloys, a 23-million kg/yr (50 million lb/yr) production facility was commissioned. However, manufacturers of crystalline Fe-Si alloys decided not to give up a $1 billion market easily. Increased R&D of the traditional alloys over the past 20 years resulted in a three-fold improvement in performance properties. While competition is stiff today, the future is

uncertain. Traditional-alloy producers conceivably could recapture the entire market if they could improve the properties of their materials to a level that surpasses that of amorphous alloys. After all, nothing more can be done to improve the structure of amorphous materials.

Ignoring market volume can be disastrous in assessing the commercial potential of advanced materials. For example, estimated world markets (in sales) and annual growth in the year 2000 for advanced ceramics, composites and semiconductors are $5 billion, $15 billion, and $100 billion and 20, 10, and 5%, respectively. While ceramics and composites markets have high growth rates, part of the rapid growth is due to a small initial base. Even at a meager 2% annual growth over the next decade, the increased volume in the already huge steel market will exceed the combined increase in the advanced-ceramics, composites, and semiconductor markets.

Ignoring materials processing may be the most perilous pitfall in predicting the success of advanced materials. For example, while materials science has progressed to the stage where carbon can be converted into diamonds and lead into gold in the laboratory, the diamonds and gold produced are more expensive than what is found in the natural state. None of the new materials that are developed will be useful unless they can be manufactured economically. Unfortunately, development of economical processing is not considered a scholarly activity in much of the scientific community. Many of the brightest scientists refuse to work in this area because of its "low stature." Even worse is that some of these individuals argue that such efforts are unworthy of substantial government or industry support. It is not that we cannot afford research and development on processing and fabrication. Instead, financial resources are directed in other areas. There are indications that this situation is changing, but the change must occur more rapidly if we are to be competitive in a global market.

Challenges in the Use of Advanced Materials: The major question in use of advanced materials is whether these materials can be used cost effectively. Tremendous efforts have been made in improving the processing, structure, properties and performance of new materials, or as the Japanese often call them, "high function materials". Figure 1 shows how these features have changed over time. More elaborate processing produces more complex structures which are tailored to specific applications in which the material is pushed to the limit of performance. We have changed from design <u>with</u> available materials to design <u>of</u> materials for specific applications.

As an example, consider the production of aircraft structures. Fifty years ago, designers selected from available materials consisting of wood, canvas and aluminum. Today, designers dream of the Orient Express, which must endure surface temperatures of 1500°C, in addition to being lightweight, having high strength and resistance to hydrogen degradation. Advanced intermetallics and composite materials must be developed to meet the design rather than the design being tailored to the properties of available materials. Unfortunately, assessment of the reliability of these materials is very difficult and limits the use of these materials in new designs.

A major factor in the reliability of all materials is the quality of the joints in the manufactured product. Joints are ubiquitous in nearly all products. As an example, I often ask my students to think of the largest stand-alone manufactured product that does not contain a joint. For many years, the best response was a cast iron frying pan, but recently one of my faculty colleagues took up the challenge; he thought of an anvil. The point of this exercise is to emphasize that every manufactured product contains

joints and that the quality of the product is directly related to the quality of the joint. Even further, when teaching fracture, another colleague is known to have said, "Something will not fail unless it has been welded!". While this statement appears to be a terrible indictment of welding, there is some truth to it Welds are often the weakest part of the structure and are generally located at the most highly stressed locations.

The cost of many of these new materials is so high and their properties are so specialized that they will only be used where they are essential. As a result, products will contain more joints and a greater fraction of these will be between dissimilar materials. This will only compound problems of quality and reliability in the final product. The common design rule of eliminating all possible joints is being violated at an increasing rate. Due to a desire to use the minimum amount of these costly, high function materials, the joints are being placed in more aggressive environments. The properties of the joint are pushed to the limit.

In many cases, designers expect the joint to match the properties on either side of the dissimilar joint. If this were an easy task, one would not need to produce a dissimilar materials joint. One could make the entire part from the joint material, if such a materials with maximal values of all properties were available! Some designers assume too much of joining technology. Rather, the solution to the use of many new materials lies in improved design which limit stresses placed on the joints. One challenge for joining engineers is development of new design rules which reduce the risk of failure at the joints. It is no longer possible to select the joint configuration or joining process as an after-thought to the design. Joining technology must become an integral part of the product design.

With increased sophistication in our design techniques, we are able to use materials closer to their limits. Stated more simply, we are reducing the size of our traditional factors of safety. This is fine as long as we fully understand all of the failure modes of the product. Unfortunately, the increasing sophistication of the structure of advanced materials means that they have even more failure modes than traditional materials. For example, the delamination and microbuckling failure modes of composites are much more complex than homogeneous materials. This requires more qualification testing than for traditional materials. The problem is compounded by the fact that the smaller market for high function materials must absorb these higher development and qualification costs. The result is that more and more designers are walking away from many of the specialized advanced materials in favor of the simpler traditional materials. It is likely that this trend back to lower cost, higher reliability traditional materials will continue. The prediction that advanced structural materials will replace most of our traditional materials is false. In the majority of commercial applications, the advanced materials simply do not meet the reliability to performance ratio at an acceptable cost.

We must change the way we manufacture materials. Figure 2 represents the old paradigm of manufacturing. A process contained inputs and outputs. The inputs were inspected prior to use in order to ensure conformance to a standard. After the process was complete, inspectors checked the function of the product and scrapped or repaired any defective parts.

In the new paradigm which is evolving, the input materials are received from a pre-qualified supplier; hence, no incoming inspection is necessary. The process is no longer a simple black box. It is the heart of the quality engineer's job. Rather than inspect the finished product, the quality engineer must sense the process, feed sensed data into a process model, and develop a control methodology which can modify the

process to produce acceptable parts see Figure 3. If this sequence of sensing, modeling and controlling is working properly, there is no need for outgoing inspection while scrap and repair are minimized. A process running at this level can be modified to meet not only the explicit requirements of the customer, but can be adapted to meet the latent requirements for which the customer does not yet know there is a desire. In addition, the reduction of waste makes the process environmentally sound. The modern quality engineer measures the process, not the product.

Although quality engineers have always been in the business of measurement, the type of measurements which are needed are changing and it is necessary to make a science of this new work. As Lord Kelvin said:

> *I often say that when you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.*

The quality engineer must put numbers on both the process and the product. Unfortunately, most people tend to measure what they can measure most easily - not what needs to be measured. Traditionally, materials were sold based upon size, thickness, weight, density or the like. These require measurements of length, width, weight and volume. Today, materials are sold based upon structure and properties as well as form and size. These new features require measurements of internal geometry, processing conditions or specific properties. Sensors are not necessarily available for these purposes. It is up to the quality engineer to develop new sensing technologies.

In addition, much of our materials manufacturing in the past was batch processed. Today's higher productivity requires that continuous flow processing dominates many industries. This requires "measurements on the move" as materials will no longer sit still in order to be inspected. With increasing value added per part, the rate of sensing must increase both spatially and temporally, while maintaining ever tighter environmental restrictions. A potential solution to these challenges lies in microsensor technology, which may make it possible to have many low cost sensors distributed over a wide area.

All of this will require quality engineers to work closely with people in other disciplines such as materials engineers, designers, industrial engineers, physicists, chemists and the like. It will be necessary to have access to the processing equipment in full scale production. This will require new partnerships and new methods of cooperation.

Learning from our Experience: The previous sections dealt with challenges in the designing and manufacturing of advanced materials, but improvement in the reliability of advanced materials requires feedback from the users of advanced materials. Traditionally, much of this feedback has been collected as case histories of failure analysis or has evolved in various codes and standards. There are problems with both of these approaches. The case histories are too empirical and have never been collected in an organized database. Development of the codes and standards moves at a glacial pace and will not provide rapid enough feedback in a world which is changing at an accelerating rate. We must develop new methods of improving the dissemination of the knowledge gained by use of these new materials. One impediment which should be addressed, perhaps through legislation, is the reluctance to publicly report the causes of failures due to fear of increased exposure to litigation. There is much knowledge available that is hidden due to fear of having the information used against the producer of the material.

Conclusion: Earlier, I noted that one of my colleagues tells his students that something will not fail unless it has been welded. As a welding engineer when I first heard this statement, I was offended, but I quickly changed my attitude to a more positive one -- as long as welds keep failing, I have a job. This can be extrapolated to everyone performing failure analysis of advanced structural materials. With all of the challenges in design, production and use of these materials, we have every reason to expect a long and prosperous career in failure analysis and prevention.

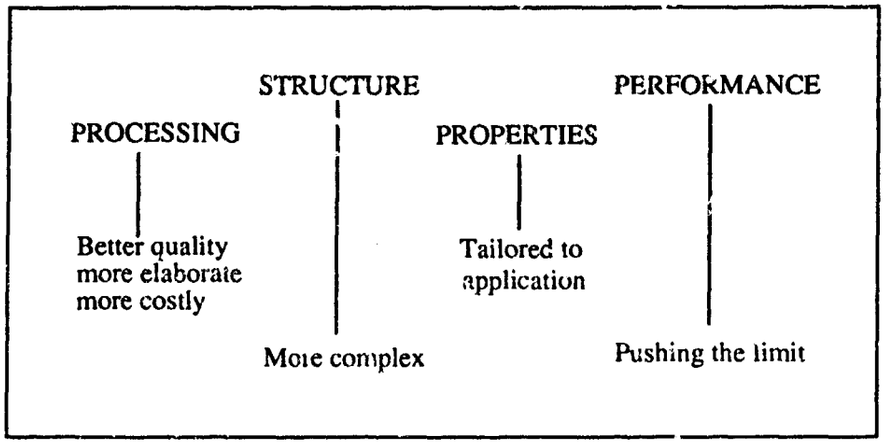


Figure 1 Advanced materials create more complexity and specificity in every phase of materials science and engineering.

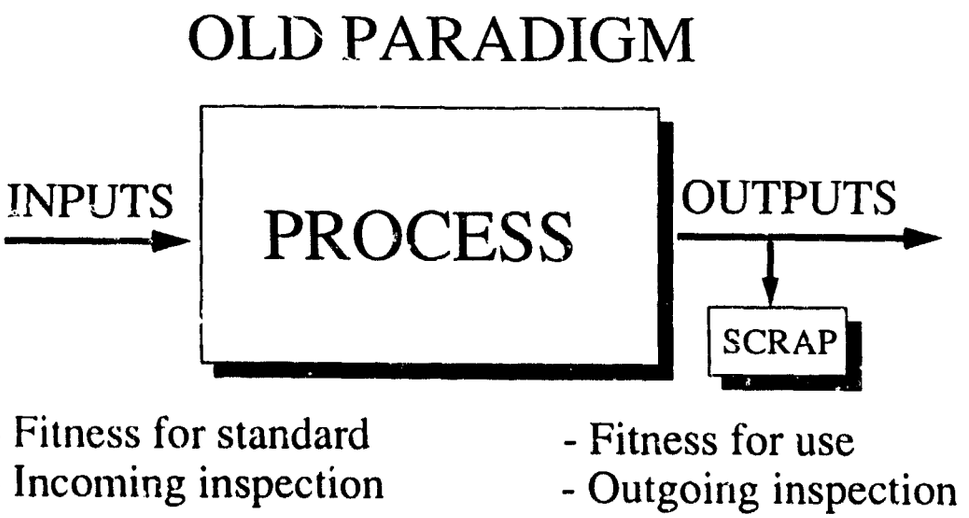


Figure 2 Traditional paradigm for quality control of a manufacturing process

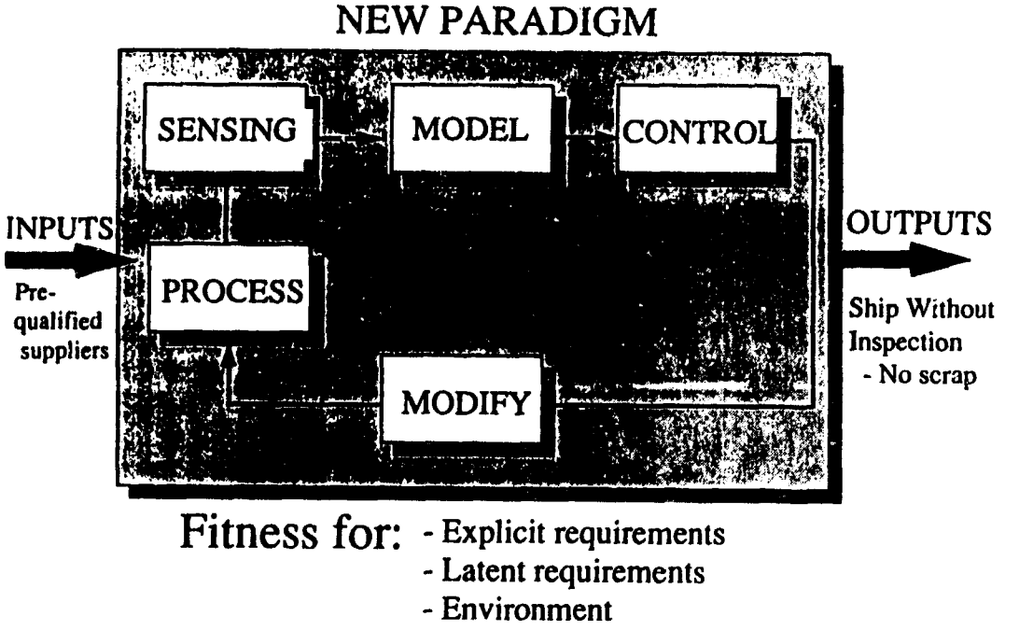
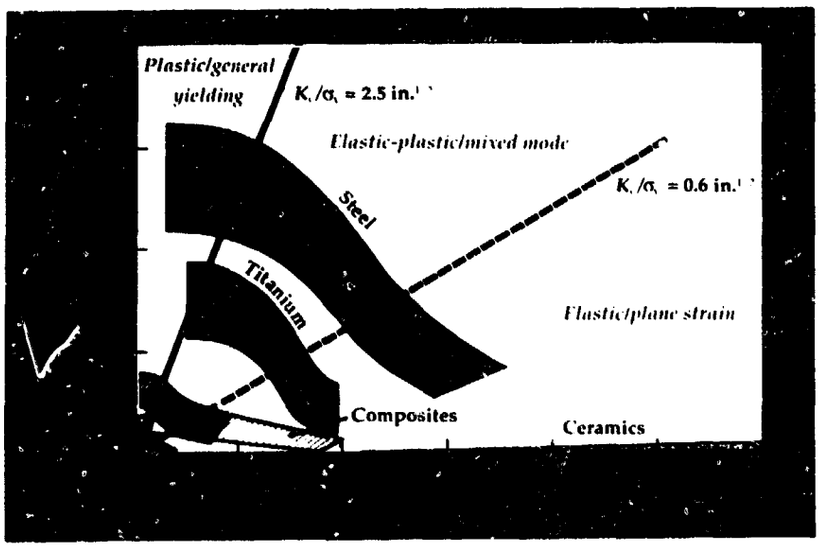


Figure 3 Evolving paradigm for control of product quality through improved process sensing and control

Reference: 1. Chaudhari, P., and Flemings, M.C. *Materials Science and Engineering for the 1990s: Maintaining Competitiveness in the Age of Materials.* National Academy Press, National Academy of Sciences, Washington, D.C. 1989.

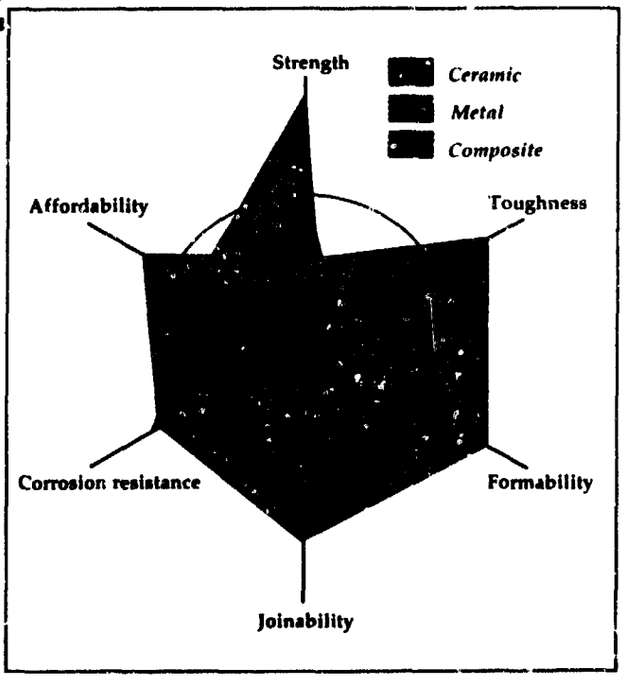
ADDENDUM

GUIDANCE ON MATERIALS SELECTION FOR DESIGN APPLICATIONS



Metals are widely used in structural applications because they provide a good combination of strength and toughness. By comparison, ceramics have exceptionally high strength, but are very brittle. This is why ceramics are used only in applications where the loads primarily are compressive. The three regions defining elastic and plastic behavior are for 25-mm (1 in.) thick materials. The transition areas shift with both thinner and thicker materials.

In a given application, materials selection generally is based on properties/characteristics that could be of critical or of secondary importance to the part design. The dimensionality of a materials selector's choices is illustrated by plotting six primary materials-selection factors (strength, toughness, formability, joinability, corrosion resistance, and affordability) along the spokes of a wheel. The values are normalized so the circle represents a specified property level. In the hypothetical situation shown, the metal offers good toughness, formability, joinability, and affordability, but lower strength than ceramic. An intermediate set of properties is obtained by forming a metal-ceramic composite. However, the gain in strength of the composite results in a degradation of the other original metals properties. While composites can provide properties that are unattainable in a homogeneous material, severe penalties are assessed, in terms of the other properties, for the enhanced performance in one dimension.



SMART STRUCTURES: ON-LINE HEALTH MONITORING CONCEPTS AND CHALLENGES

Zaffir Chaudhry and Craig A. Rogers
Center for Intelligent Material Systems and Structures
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061-0261

Abstract: Utilization of the actuator, sensor materials and the associated smart structures concepts offer true hope for the realization of on-line health monitoring concepts; in particular, the non-destructive evaluation of structures *via* "vibration signature" analysis. What the smart structures solid-state actuators and sensors have to offer different from the shakers and accelerometers typically used for this work is their compactness, light-weight, low-power consumption, ease of integration into critical structural areas, ease of activation through electrical signals, higher operating frequency, use of the same material for both activation and sensing i.e., collocated actuator-sensor, and low cost. Beyond obtaining the vibration signature or metrics associated with near and far field excitation, and the necessary signal processing, to identify damage location and size, there are two common approaches: i) model-based, and ii) non-model-based. The model-based methods rely on the system identification concepts, whereas non-model-based methods rely on "open-loop" systems or adaptive and learning approaches like artificial neural networks and artificial intelligence. Both approaches have their merits and limitations, and are in the development phases. The application of these techniques to a real time structure is yet to be demonstrated. Some of the challenges associated with these techniques are 1) the ability to find smaller damage like delaminations or cracks, 2) reducing the size of the signal processing equipment, 3) the ability to distinguish the effect of normal operating conditions like changed mass or boundary condition on the vibration signature from that of damage, 4) the ability to filter global vibration effects from the local excitation.

Key Words: Damage detection; Health monitoring; Non-destructive evaluation (NDE); Smart structures.

Introduction: Great interest presently exists in the engineering community to develop methods with which complex structures can easily be analyzed for defects without being taken out of service. The traditional damage identification techniques which originated with visual inspections of structural components are time consuming and are difficult, if not impossible, for components that are not accessible. Most importantly, visual inspection provides no comprehensive assessment of the integrity of the structure. Other methods that have been employed include radiographic, X-ray, acoustic emission, and ultrasonic measurements. However, these methods provide only local information and cannot give any indication of the structural strength on a system level.

The non-destructive evaluation (NDE) techniques based on vibration signature analyses that exist and those that are being developed for this purpose utilize many common features. The basic principle in an evaluation first involves obtaining a vibration signature of the structure. This signature is created by shaking or vibrating the structure with some system of actuators. The vibration patterns of the structure are then recorded by a network of sensors attached to the structure. This vibration signature shows the response of the structure over a wide range of frequencies and thus provides information about the magnitude of the oscillations, the resonant frequencies, the phase differences between force and response, and other valuable data. Comparisons of the vibration signatures between an "undamaged" structure and a test structure can be used to make an evaluation of the state of the test specimen. Vibration signatures of a structure taken at two different times can also be compared to look for changes in behavior that indicate the possibility of damage. Once the behavior of the structure is known, the vibration data can be processed using a variety of NDE techniques, each of which performs different operations and provides different insight into potential problems in the structure. The goal for NDE in the future includes the ability to determine if damage has occurred in the structure, where the damage is located, and what type of damage is present. Not all of these capabilities are possible with the techniques that currently exist. For that reason, research into the development of new analysis approaches and improvement of existing ones is ongoing.

It is important to mention that vibration signature analysis has long been used as a diagnostic tool for fixed rotating machinery such as pumps, compressors, fans, motors and shafts. In the simplest case, a change in the vibration amplitude is often indicative of a damaged rotor blade or a bad bearing. Acceptable levels of vibration are often included in the specification sheets supplied by the manufacturer. An increased vibration level is only an indication of a problem; relating it, or the vibration signature, to a specific fault is much more complex. And in this regard, the problems associated with the techniques that are being investigated as part of this research effort are similar.

Actuators and sensor materials: Smart structures concepts utilizing solid state actuators and sensors greatly facilitate the process of obtaining the vibration signature. What distinguishes these solid state actuators from conventional hydraulic, electrical, and electromechanical actuators, and makes them especially attractive for smart structures is their ability to change their dimensions and properties without utilizing any moving parts. These actuator materials contract and expand just like the muscles in the human body. When integrated into a structure (either through embedding or through surface bonding) they apply localized strains and directly influence the extensional and bending response of the structural elements. Because of the absence of mechanical parts, they can be easily integrated into the base structure. Integration within the structure ensures an overall force equilibrium between the forcing actuator and the deforming structure, and thus precludes any rigid body forces and torques. In addition, typically these actuators are compact, light-weight, and consume low-power and can be easily located in those critical areas of the structure which need continuous monitoring.

One of the actuator/sensor materials which is particularly suitable for this operation is the piezoelectric (PZT). Piezoelectricity describes the phenomenon of the generation of

an electric charge in a material when subjected to a mechanical stress (direct effect), and conversely, a mechanical strain in response to an applied electric field. The direct piezoelectric effect has often been used in sensors like an accelerometer; however, until recently, the converse effect has been restricted to ultrasonic transducers. It is only recently that researchers in the area of structural control have taken notice of the very desirable features of piezoelectric actuators and have started using them for many structural control applications. Within the linear range, they produce strains which are proportional to the applied electric field/voltage. These features make them very attractive for many structural control applications. The use of PZT actuators can be divided into two broad categories: linear actuators and actuators used for structural control applications. In the first category, the PZT actuators, arranged in the form of stacks, are used pretty much like shakers or conventional hydraulic or electrical actuators. In structural control applications, the actuators are typically embedded within the structure and they apply localized strains which can be used to directly control structural deformations. In such a configuration, the actuation does not require a back reaction and does not produce any rigid body modes. For inducing flexural vibrations, for instance, two actuators can be bonded to the upper and lower surface of the structure. When actuated out of phase, the upper actuator expands and the bottom contracts, causing bending.

Because of the direct and converse effect that exists in piezoelectric materials, it can just as easily be used for both actuation and sensing, also referred to as the collocated actuator-sensor. An explanation of this effect follows. When bonded or attached to a structure, and driven by a fixed alternating electric field, it excites and induces vibrations in the structure. Meanwhile, the resultant vibrational response, which is characteristic of the particular structure, modulates the current flowing through the PZT. This modulation is representative of the degree of the mechanical interaction between the collocated actuator-sensor and the structure at different frequencies. If the variation in the current can be monitored, then the electric admittance, defined as the ratio of the resulting current to its energizing voltage, can provide the same information as, and serves the same purpose as the conventionally known transfer function between force and velocity. Utilizing the same material for both actuation and sensing not only reduces the number of sensors and actuators, but also reduces the electrical wiring and associated hardware.

Model-based methods: The concept of model-based vibration signature analysis originated in the great body of work performed mostly in the 1970's and 1980's concerning system or structural identification. Structural identification as a methodology is concerned with the ability to identify a structure by comparing experimental data to that produced by an analytical model. Most often, the analytical behavior is insufficient to accurately predict structural behavior, and model refinement is necessary. This inadequacy often occurs because of material variations, structural nonlinearities, damping, etc., not modeled. The classical approach to structural identification involves direct comparison of the measured eigenvalues, eigenvectors, and mode shapes (termed eigenstructure) to analytical vibrational behavior. Lack of agreement indicates the need for model improvement. Use of time-domain behavior was also investigated as an identification tool. However, these methods required storage of vast amounts of data and

relied on often noisy measurements and thus were of questionable reliability. Structural identification as a precursor to damage detection is most concerned with three different activities: system parameter representation, model improvement, and structural reconstruction.

System parameter representation is a process in which there is an attempt to relate the measured eigenstructure to a pre-selected set of structure parameters often including material modulus and density, structural geometry, etc. These often non-linear relationships between structural behavior and the parameter set can be determined using local linearization techniques incorporated into an iterative optimization algorithm as was done by Adelman and Haftka (1986). Work was also done by Creamer and Hendricks (1987) to develop techniques for relating system parameters to experimental behavior, specifically natural frequencies. Once the structural behavior is quantified, a sensitivity analysis can be performed to gain an improved understanding of how each of the elements in the parameter set affect the structural behavior. This sensitivity analysis determines how a small variation in each element of the parameter set affects the eigenstructure, thus allowing for an appropriate adjustment of the parameter set elements in order to reproduce the experimental behavior. This is the first form of model improvement. The use of a sensitivity analysis is beneficial in comparison to techniques that require a complete dynamic analysis at every iteration in that a sensitivity analysis only requires differentiability of the behavior relationships with respect to structure parameters and does not require lengthy or numerous calculations required for a complete analysis.

Other work done involving model improvement has centered on developing methods in which the eigenstructure could be used to modify an analytical finite element model in such a way that the behavior of the adjusted model would imitate that of the actual structure. Presently, there exist two methods for such model improvement: eigenstructure assignment, and optimal-update. Eigenstructure assignment approaches the issue of eigenstructure alteration from a controls perspective. Controls literature has long demonstrated the ability to modify the eigenstructure of a system with the addition of a feedback controller whose gains could be modified so that the desired, closed-loop behavior was obtained. Eigenstructure assignment uses a "pseudo-controller" that uses feedback of the structural behavior to instill the experimental eigenvalues into the analytical model. However, no actual controls hardware is necessary for this operation. The optimal-update method uses a constrained optimization problem to find an adjusted mass, stiffness, or damping matrix that reflects the experimental eigenstructure and also is the most similar to that of the analytical model.

The last topic in the research area of structural identification is known as the "inverse process". This process is the reconstruction of the structural parameters from only the experimental eigenstructure. Gladwell, England, and Wang (1987) successfully showed the reconstruction of a Euler-Bernoulli beam using spectral data from three different imposed boundary conditions. Specifically, the cross-sectional properties and moment of inertia were found from the natural frequencies of the beam. Later work in structure reconstruction, performed by Shen and Taylor (1991), extended the concept with the inclusion of identification of modeled damage from only spectral data.

The developments and continued efforts in structural identification provide valuable tools for creation of damage identification techniques. Structural identification has allowed for the development of improved models of structures which can be used as a comparison tool with experimental eigenstructure recorded from structures with "known" damage. Thus, the changes in behavior caused by specific types of damage can be more easily seen. More importantly, the cause-and-effect relationships between damage and the corresponding changes in behavior can be better investigated.

The use of time responses has also been attempted as a tool for damage assessment, but because of the large amount of data storage required and the questionable accuracy of noisy signals, it has not gained as much attention as other methods such as the use of frequency domain information. Kircher (1976) investigated the use of vibration signatures of buildings taken at different times to detect structural changes, indicating the possibility of earthquake damage. Other early work using vibration signatures attempted to simulate many types of damage and record the vibration signatures for each of those instances. This required the storage of large amounts of data, did not allow quantification of different degrees of damage or combinations of damage types, and required knowledge of the type of damage present. Cawley and Adams (1979) successfully related changes in structural stiffness due to damage to the changing of natural frequencies. Cawley et al. showed that the changes in natural frequencies of a beam could be related to the degree of damage and distance to the damage from the vibration signature measurement point. An error function was used to sort the locus of possible damage points and determine where the damage existed in the structure. Problems with the method developed by Cawley involved the dependency on highly accurate frequency information and often one mode of vibration would provide slightly erroneous information, causing the algorithm to fail. Also, the method was only able to detect damage at one site in the structure. Another significant damage identification technique developed by Shen and Taylor (1991) used the "inverse approach" of structural reconstruction to find crack locations and sizes in beams. The cracks were modeled as part of the beam structure and an optimization procedure with an initial guess of crack magnitude and location was able to determine the actual crack parameters by minimizing the difference between the experimental and analytical eigenstructure. The method was shown to be limited by the precision of the crack model.

Non-model-based methods: The non-model-based methods are fast gaining popularity. The primary reasons for this are: i) the lack of dependence on complex analytical models of the structural system and ii) the ability to learn on-line. Artificial neural networks and inductive learning methods are receiving the most attention amongst non-model-based methods. A brief discussion of each follows.

Artificial neural networks: The artificial neural networks (ANN) functions as a computational model loosely based on the neural construction and operation of the human brain; it is set up as a collection of highly interconnected simple processing units (neurons) that can be trained to recognize known input. The connections between the neurons have numerical values which represent the strengths of those connections called "weights." The knowledge or learning ability of the ANN lies in the formulation of those connection weights or neurodes. Training is performed by a process known as

"supervised learning" where the network is presented with sets of data, each of which corresponds to a different state of structural health. The input to the network is important in that it must contain the information that accurately represents the situation that is desired. The training results in a network that not only interprets the training input correctly but also exhibits the ability to interpolate and generalize to unknown input. It is this training/learning ability that enables us to use the ANN in order to avoid developing complicated analytical models for the structure and each different damage situation that may develop. Neural networks possess the ability to identify and imitate the often non-linear mapping between input and output seen in the behavior of complex structures. The neurode or connection strengths inside the network can be seen as the relative cause-and-effect relationships which relate damage to changes in structural behavior. Typical back-propagation network construction consists of neurode layers including an input layer, hidden layers, and an output layer. The neurodes act as transfer functions and neurode output is gained from applying this transfer function to the summation of all the inputs to that neurode. These inputs are the outputs of previous neurodes multiplied by the "trained" connection strengths. Thus, each input is able to spread its influence, more or less depending on the connection strengths, on every higher level neurode.

If the vibration signature from the actual structure is used for the training of the ANN, then the noise and other spurious events become a part of the ANN model. This adds to the robustness of the identification system; whereas, noise can quickly render a model-based system useless.

Inductive learning methods: The artificial intelligence (AI) technique used for investigation of damage identification is known as "rule induction" or inductive learning. Inductive learning is defined as: "A computer learning from examples, or learning inductively, consists of hypothesizing general rules from specific examples of information provided to it." Thus, machine learning (AI) can be utilized to make some form of judgment about the status of newly-presented data by using rules formulated from previously-seen data. The benefits of a machine learning approach lie in that no model is required to make an assessment of the structure being analyzed. Instead of trying to define all the parameters of the structure, as a model improvement scheme does, AI only attempts to recognize relationships between inputs and outputs, and uses those to "learn" or define similar behavior patterns.

Discussion: The compact actuator and sensor materials can facilitate the acquisition of structural vibration signature. Perhaps the biggest challenge in the realization of a vibration signature based damage detection system lies in the ability to separate the changes in the vibration signature due to normal usage from that due to damage. The experiments in the laboratory are often done under very controlled repeatable conditions to ensure that the only changes are due to the self-created damage. Truss structures are a favorite because different members can be conveniently removed and then replaced. A missing member in a truss structure can easily be found by either a model-based or a non-model-based method, but such extensive damage is rarely what we want to find. Except space structures, where the structural properties are relatively constant over time, all other structures experience changed mass, damping, and boundary conditions due to

normal usage. Simulation of all possible normal usage changes and their effect on the vibration signature is impossible to store so as to distinguish them from damage. Thus, techniques which allow for the localization of the actuation and sensing area need to be developed. In truss type structures, the localization has been shown to be possible at higher frequencies.

The other challenge lies in reducing the signal processing equipment. Frequency domain vibration signature relies heavily on complex digital signal processing hardware and software which adds considerably to the cost and the weight of the overall system. Analog signal processing techniques need to be investigated. Analog signal processing cannot provide the detailed frequency domain information, but that may not be necessary. A possible option is to use the root-mean-square value of the sensor output over a selected frequency range as an indication of the condition of the structure. Analog signal processing can be used for the initial qualitative assessment of damage. Once damage has been detected then for detailed analysis and damage location, sophisticated hardware may be used with the same sensors. For cost effective real-time implementation, alternative signal processing methods such as the one discussed above need to be explored and developed.

Another challenge is the ability to find minor damage such as cracks and delaminations. The model-based methods are severely limited in this area. Analytical methods of predicting the changes in modal parameters or response characteristics become dubious with more complex structures, and at best, only large amounts of damage can be predicted with such methods. Obviously, the frequency range required to find minor damage is much higher than that used for the typical system identification work, and the solid state actuators/sensors have much to offer in this regard.

The list of challenges listed above is by no means exhaustive. There are many other issues which require investigation. These practical issues will require concurrent investigation in order to produce simple, comprehensive, and valid techniques for complex structures requiring evaluation.

References:

1. Cawley, Adams, 1979, "The Location of Defects in Structures From Measurements of Natural Frequencies," Journal of Strain Analysis, Vol. 14, No. 2, 1979.

2. Shen, Taylor, 1991, "An Identification Problem for Vibrating Cracked Beams," Journal of Sound and Vibration, Vol. 150, 1991.

3. Gladwell, England, and Wang, 1987, "Examples of Reconstruction of an Euler-Bernoulli Beam from Spectral Data", Journal of Sound and Vibration, Vol. 119, 1987.

4. Shen, Taylor, 1991, "An Identification Problem for Vibrating Cracked Beams," Journal of Sound and Vibration, Vol. 150, 1991.

5. Adelman, Haftka, 1986, "Sensitivity Analysis of Discrete Structural Systems," AIAA Journal, Vol. 24, No. 5, May 1986.

PHASE CONTRIBUTES TO ACCURATE DIAGNOSTICS OF ROTATING MACHINERY

James I. Taylor
Vibration Consultants, Inc.
5733 South Dale Mabry Highway
Tampa, FL 33611

Abstract: If you want accurate diagnostics of rotating machinery problems, you must pay the price. The price is hardware, software, technology, and skilled people. The people involved must understand, obey, and use basic physics. It is fact that basic physics of electronics, mechanics, and chemistry are very closely related. This is why we can use a transducer to convert mechanical motions, pressure fluctions, and airborne sound into an electronic signal and use that signal to identify problems. Every educated person should know this. Yet, once the electronic signal is produced, we seem to forget everything. Behavior becomes neanderthal; we measure some pseudo amplitude, and plot trends; we use seemingly good alternatives, such as spike energy, shock pulse, envelope detection, demodulators and cepstrum to replace knowledge and physics. This paper is a result of 20 years of research and development using on-line operating machines, basic physics, and empirical measurements. The importance of the time domain signal and the phase relationship between the fundamental and harmonics is stressed. The time signal is required because it contains information not available in the frequency spectra. The phase relationship between the fundamental and harmonics is required to identify which problems are loose and those that are eccentric.

Key Words: Analyzers; diagnostics; FFT; harmonics; modulation; phase; physics; sidebands; spectra; technology; time signal; truncation.

Introduction: Phase relationships have been used for years to obtain intelligence in sophisticated applications. In radar, phase is used to distinguish between fixed and moving targets. In space, phase is used for Space Object Identification (SOI). Phase is also used to identify inoperative space vehicles, sometimes called space junk. Phase is widely used in radio, television, and telecommunications. In fact, these industries could not function without phase. Phase is also required to accurately diagnose machinery problems. Accurate diagnosis may require a frequency spectrum on several frequency ranges and a zoom on various frequency windows. The time signal for various time periods is equally important. Since all events are not present in every memory period, several memory periods may be observed before the desired event is captured. The time signal and frequency spectra are required to obtain necessary information. The frequency spectra measures frequencies and the amplitude is averaged. The time signal contains phase, peak, and peak-to-peak amplitude of each cycle, truncation, pulses, amplitude and frequency modulation, etc. These bits of information are used during the Fast Fourier Transform (FFT). However, once the transform is completed, they are discarded. This explains why the same frequency spectra is obtained from several

Copyright © James I. Taylor, 1994. All rights reserved. No part of this publication may be reproduced in any form or by any means without permission from the author.

different looking time signals. These bits of information are also required to accurately diagnose problems.

How frequencies are generated: Some machine problems generate discrete frequencies. For example, pure imbalance generates a frequency at rotor speed. Each cycle should be the same amplitude, and sinusoidal. Gearmesh frequency is the number of teeth on the gear times speed. Gearmesh frequency should also be sinusoidal. Any distortion of the time signal is caused by the introduction of some harmonic content. Some machinery problems generate harmonics. Rotor looseness is a good example. If the rotor is loose and not restrained by belts or some other device, it will generate harmonics of rotor speed. The number of harmonics is one indication of the amount of looseness. For example, a fluid film bearing on a turbine may have 0.004" or 0.005" clearance depending on shaft size. The spectra from the turbine may contain the fundamental and 4 or 5 harmonics. When the clearance increases the amplitude and number of the harmonics could increase. Other machinery problems generate a distorted waveform and when an FFT is performed the harmonic content that would cause the distortion is produced. Misalignment, bent shaft, some forms of looseness, and some eccentric gears are good examples. Each cycle of gearmesh frequency is a profile of two teeth meshing with due respect to transfer function. Since the same two teeth will not mesh again until one cycle of the hunting tooth frequency is completed, each memory period of gearmesh frequency could be different.

The same frequencies can be generated by more than one equipment problem. For example, imbalance, bent shaft, and machine looseness can generate the fundamental. Looseness and bent shaft can also generate the second harmonic of rotor speed. The second harmonic of gearmesh frequency is generated by not enough backlash and too much backlash or oscillating gears. Rotor looseness can generate the fundamental and several harmonics. Loose gears can generate several harmonics of gearmesh frequency. Loose gears also cause modulation. Pulses can also cause harmonics of the pulse repetition rate.

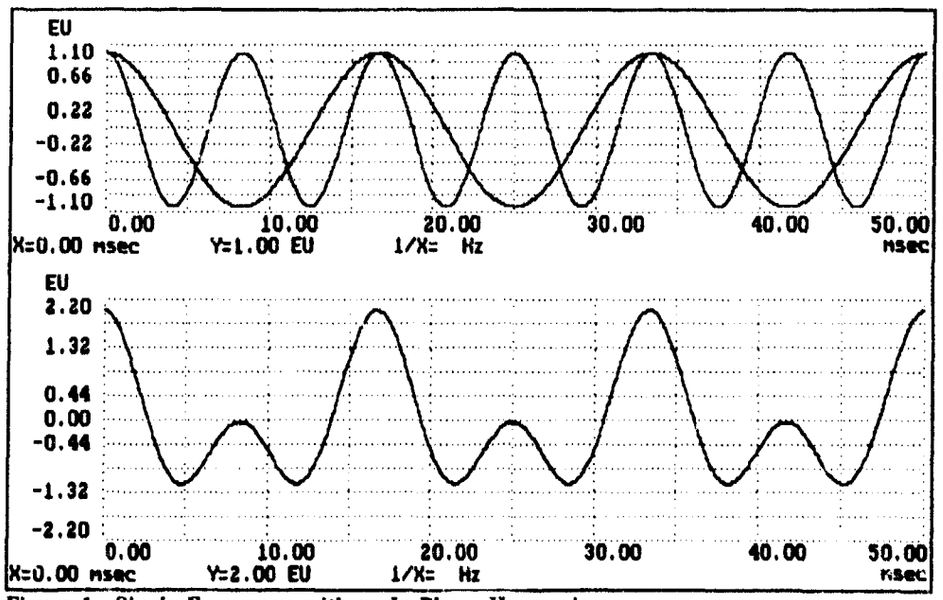


Figure 1. Single Frequency with an In-Phase Harmonic.

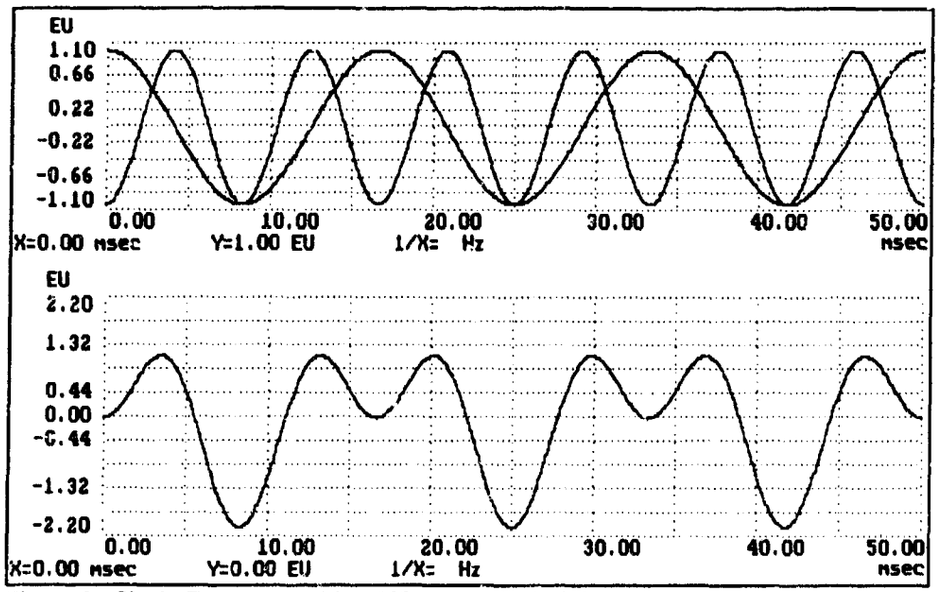


Figure 2. Single Frequency with a 180 Degree Out-of-Phase Harmonic.

As indicated above, eccentric and loose gears cause modulation. As a result, gearmesh frequency is often modulated by the speed of one or more gears. Gearmesh frequency can also be modulated by 2, 3, 4, etc. times gear speed. Sometimes the modulations generate sidebands on the high side, low side, and both sides of gearmesh frequency. The amplitude of the sidebands can be higher on either side of gearmesh frequency. The location and amplitude of the sidebands is determined by the modulation process, type of modulation, phase relation between the carrier and modulator, type of defect, and defect seriousness.

Generally, eccentric problems are an in-phase condition. For example, a bent shaft can cause imbalance and the signal may be distorted. The bend in the shaft is in-phase with the imbalance. Therefore, the time signal will be truncated on the negative-going half or the in-phase condition. The FFT of this signal will contain a fundamental and second harmonic. Eccentric gear problems are also an in-phase or constant phase condition. Therefore, these problems cause the amplitude of the sidebands on the high side of gearmesh frequency to be higher in amplitude. Looseness and looseness-type problems such as excessive backlash are an out-of-phase or changing-phase condition. For example, if the rotor is loose and is not restrained by drive belts, couplings, etc., the rotor may hit and bounce as it rotates. Since only the looseness controls the hitting and bouncing, the harmonics will be out-of-phase with the fundamental, or at least have a different phase each revolution or a changing phase relationship. The same rational applies for the second harmonic of a loose machine.

How Frequencies Mix. The top graph in Figure 1 contains the time signals of a fundamental and a second harmonic. The two signals are equal in amplitude and in-phase. The bottom graph contains the composite mixed signal. When the two signals are in-phase, the amplitudes add. When they are out-of-phase, the amplitudes subtract. The composite signal is truncated on the negative side and the positive side is higher in amplitude. The second harmonic in Figure 2 is 180 degrees out-of-phase with the fundamental. This causes truncation on the positive portion of the signal. All machines, frequencies, signals, etc., must obey the laws of physics. From this we can conclude that truncation at the top of the signals

is an out-of-phase condition, and truncation at the bottom of the signals, is an in-phase condition. If the two signals are 90 or 270 degrees out-of-phase, truncation will not occur. As the phase relation moves from 90 to 270, some truncation will occur on the top half of the signal. As the phase relation moves from 270 to 90 degrees, some truncation will occur on the bottom half of the signal. Of course, the signals are in-phase when maximum truncation occurs at 0 degrees, and out-of-phase when maximum truncation occurs at 180 degrees, based on the cosine function. Both of these time signals will produce the same spectrum of the fundamental and a second harmonic similar to the spectra in Figure 3.

In a perfect set of meshing gears, a low amplitude of gearmesh frequency could be generated. Each cycle would be sinusoidal and the same amplitude. When imperfection in the gears is introduced, the amplitude of gearmesh frequency increases, and modulations occur at the speed of the offending gear(s). These modulations can occur at 1, 2, 3, etc., times gear speed. In this case, the gearmesh frequency is the carrier and the gear speed is the modulator. If the modulator is sinusoidal, simple double sideband (DSB) amplitude modulation occurs. For example, if a 27-tooth gear is in mesh with a 61-tooth gear and the speed of the 27-tooth gear is 29.6 Hz, the gearmesh frequency would be 27 x 29.6 = 799.2. Figure 4 contains the time signal and frequency spectra that could be produced. The two sidebands at 29.6 Hz are in-phase with the carrier. It may be difficult for a real set of gears to generate this signal, however, it does prove some points. First, the amplitude of the sidebands is determined by the percentage of modulation. If only one sideband was present, the percentage of modulation would be 0.1 ÷ 0.3 = 33%. Since both sidebands are in-phase with the fundamental, the amplitudes add. Therefore, the percent of modulation is (0.1 + 0.1) ÷ 0.3 = 66 percent. The percentage of modulation can also be calculated from the time signal.

$$\frac{A \text{ max} - A \text{ min}}{A \text{ max} + A \text{ min}} = \frac{0.50 - 0.01}{0.50 + 0.01} = \frac{0.4}{0.6} = 66\%$$

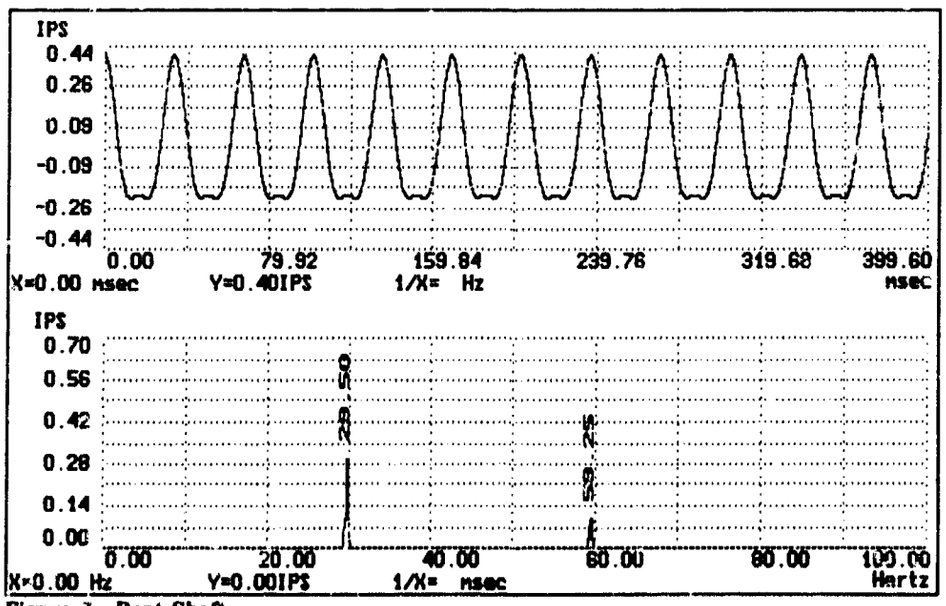


Figure 3. Bent Shaft.

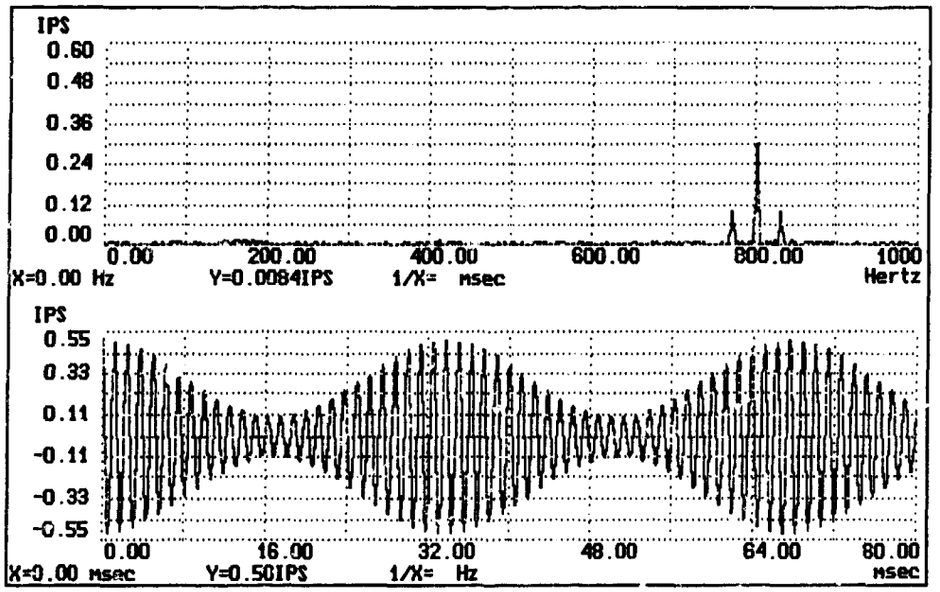


Figure 4. Double Sidebanded Amplitude Modulation.

Second, frequency identifies the problem; amplitude and harmonic content indicates problem severity. Gearmesh frequency modulated by gear speed identifies the problem gear. The phase relationship identifies looseness and/or eccentricity. If this signal was demodulated, the carrier would be lost and the modulator would be retained. However, one times gear speed, and gearmesh frequency modulated by gear speed are completely different problems. Phase is also lost with demodulation. In other words, demodulation and envelope detection destroys the intelligence required for accurate diagnosis.

Over modulation occurs when the amplitude of the modulator is greater than the amplitude of carrier. In such cases, the sideband may be higher than the carrier. Over modulation causes phase reversal, and results in distortion of the modulator or envelope. Simple envelope detectors and/or demodulators may not recover the complete envelope or modulator when phase reversal occurs. This point may be moot because a demodulated signal is meaningless when diagnosing rotating machinery problems. However, distortion of the modulator may cause additional sidebands.

Single sideband (SSB) modulation approaches the modulation caused by some gear problems. SSB produces an upper single sideband (USSB) or a lower single sideband (LSSB). If the phase is negative at the summing point, the LSSB is produced. If the phase is positive at the summing point, the USSB is produced. Pure SSB modulation may not be possible in rotating machinery because of phase changes, distortion, over modulation, and noise. The type of modulation often occurring in machines is vestige sideband modulation (VSB). This means the sidebands will be higher on the upper or lower side of the carrier. For example, an eccentric gear is normally in-phase with the gearmesh frequency. The envelope or modulator is gear speed. This results in VSB with the upper sidebands amplitude higher than the lower sidebands. The lower sidebands are caused by some out-of-phase or changing phase condition. A loose gear is normally out-of-phase with gearmesh frequency. However, gear speed could be in-phase with gearmesh frequency some of the time. When gear speed is in-phase with gearmesh frequency, VSB modulation can occur. In this case, the amplitude of

the USB will be less than the LSB. The distortion in the envelope/modulator caused by looseness, eccentricity, and noise can cause several sidebands to occur on both the upper and lower side.

The mathematics that prove the above theory have been omitted because of length. However, the analytical theory is well documented in several Communication System textbooks. The third edition of Communication Systems, by A. Bruce Carlson published by McGraw-Hill is very good. The reader should know the signals generated by rotating machines. Otherwise it may be difficult to recognize which communication principals apply to vibration data.

The first case history concerns a gear with high places. The gear speed is about 2 Hz and the gear has 84 teeth. Gearmesh frequency is 2 x 84 = 168 Hz. The gear is constructed with five (5) spokes, and has a high place at each spoke. This problem is fairly common in gears with spokes. When these gears are mounted on a shaft with an interference fit, the expansion is greater at the spoke. This expansion often causes the gear to be out of tolerance at those locations.

Figure 5 contains the frequency spectra from the gear. Frequency is plotted on the x-axis and amplitude is plotted on the y-axis. The spectral line at 10 Hz is five (5) times gear speed. The spectral line at 168.6 Hz is gearmesh frequency within measurement accuracy. The gearmesh frequency has upper sidebands at 178 and 188 Hz. Lower sidebands are present at 158 and 148 Hz. However, the amplitude of the LSB is much lower than the USB. The 10 Hz modulations are five times gear speed. This simple problem does not require data on other frequency ranges or a zoom. However, a time domain signal is required.

The time signal in Figure 6 produced the above spectra. Figure 7 contains one revolution of the gear. The individual cycles of gearmesh frequency are clear in this signal. Please note the time signal is truncated on the negative half. This indicates an in-phase condition. As the teeth near the spokes go into mesh, the amplitude increases and then decreases as the teeth

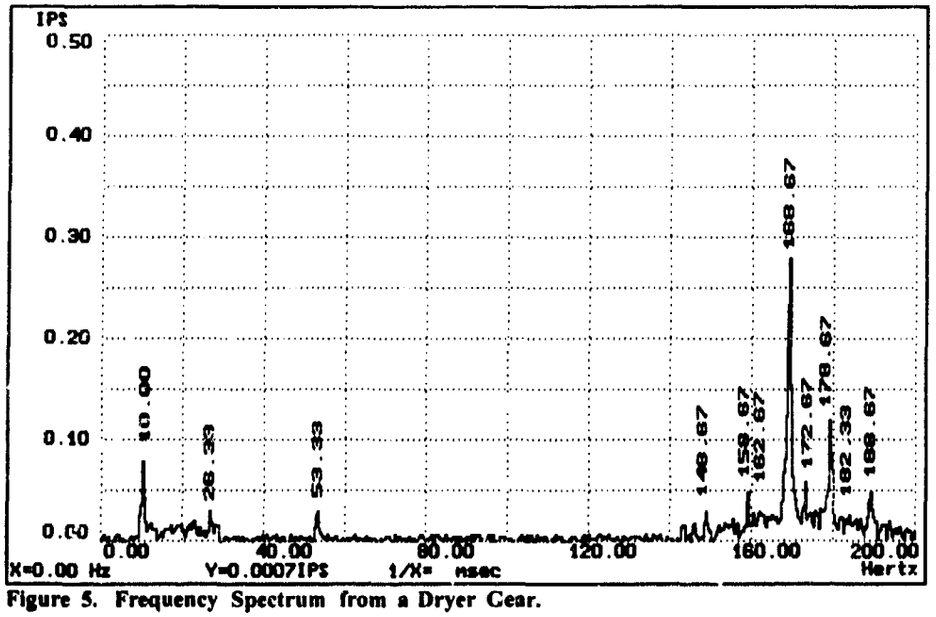


Figure 5. Frequency Spectrum from a Dryer Gear.

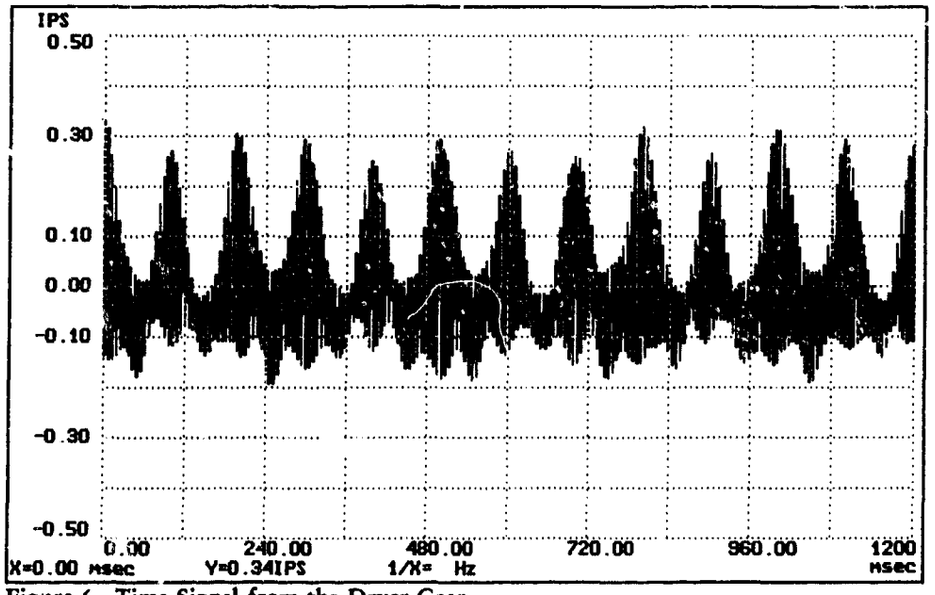


Figure 6. Time Signal from the Dryer Gear.

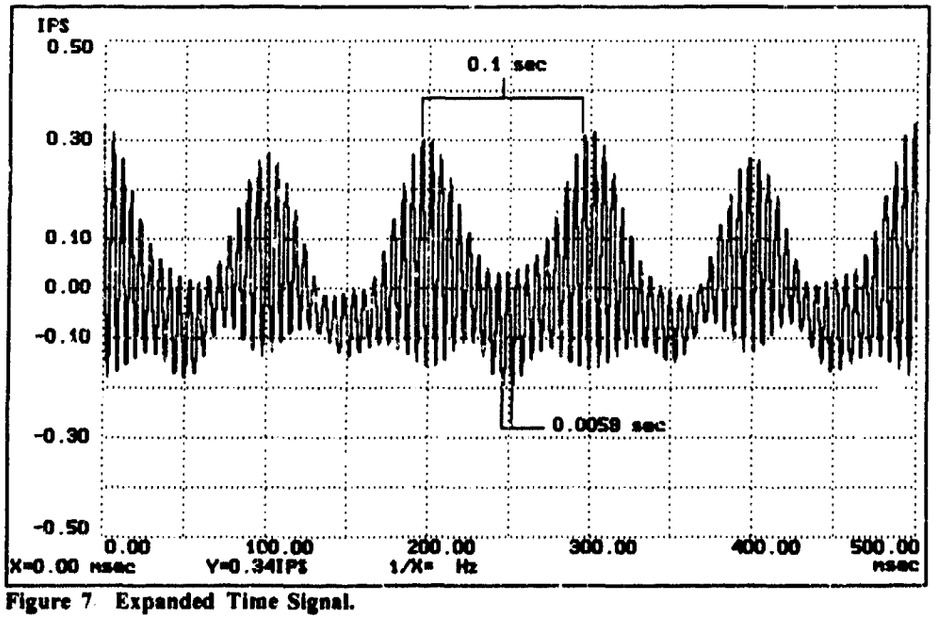


Figure 7 Expanded Time Signal.

at the spoke move out of mesh.

This writer believes that gears with spokes should be avoided. If these gears must be used, pitch line runout must be checked again after the gear is installed on the shaft.

The second case history concerns loose gears with eccentric teeth. The gear has 126 teeth and the speed is 2.82 Hz. The meshing gear (2) has 81 teeth and a speed of 4.38 Hz. Figure 8 contains the output of the Gears software program for evaluation of gears. This product indicates the gears have an improper ratio because the two gears have a common factor of 9. If the improper ratio was the problem, we would expect to see frequencies at 1/9 the gearmesh frequency or 39.42 Hz and harmonics. However, this is not the case.The spectrum in Figure 9 was taken from the gears. The spectral line at 177.5 Hz is half gearmesh frequency. Please note the amplitude is 0.06 IPS. This frequency can be caused by eccentric gears with a common factor of 2, 4, or 6, etc. This is not the case here. Half gearmesh frequency can also be generated when some teeth are eccentric. For example, every other tooth could be wider or thinner, and/or every other tooth could have pitch line runout. One group or several groups of teeth could have these or similar eccentricities. The spectral line at 355 Hz is gearmesh frequency within measurement accuracy. The wide band of spectral lines around 522 to 537 Hz is 1.5 times gearmesh frequency. The spectral line at 210 Hz is 2 times gearmesh frequency. The widebanded spectral lines and the white noise indicate looseness.

A zoom is required for accurate diagnosis. Figure 10 contains a 40 Hz window around gearmesh frequency. The amplitude of the lower sidebands at gear 1 speed (2.8 Hz) is higher than the upper sidebands. This indicates gear 1 is loose. Gear 1 could also be eccentric because of the higher harmonic content of the USB. The upper sideband at the speed of gear 2 (4.38 Hz) is higher than the LSB. This indicates gear 2 is eccentric.

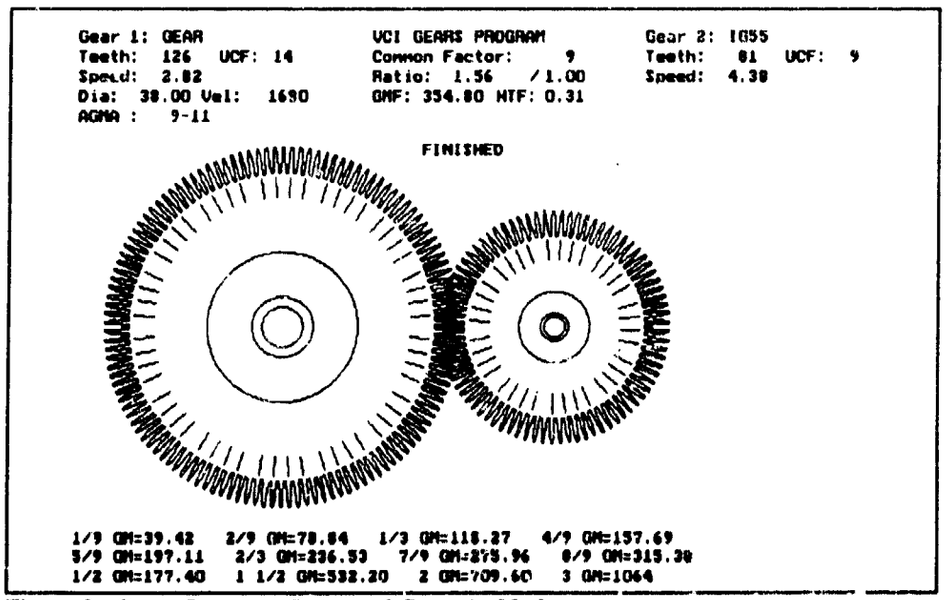


Figure 8. Gears Program Output of Gears in Mesh.

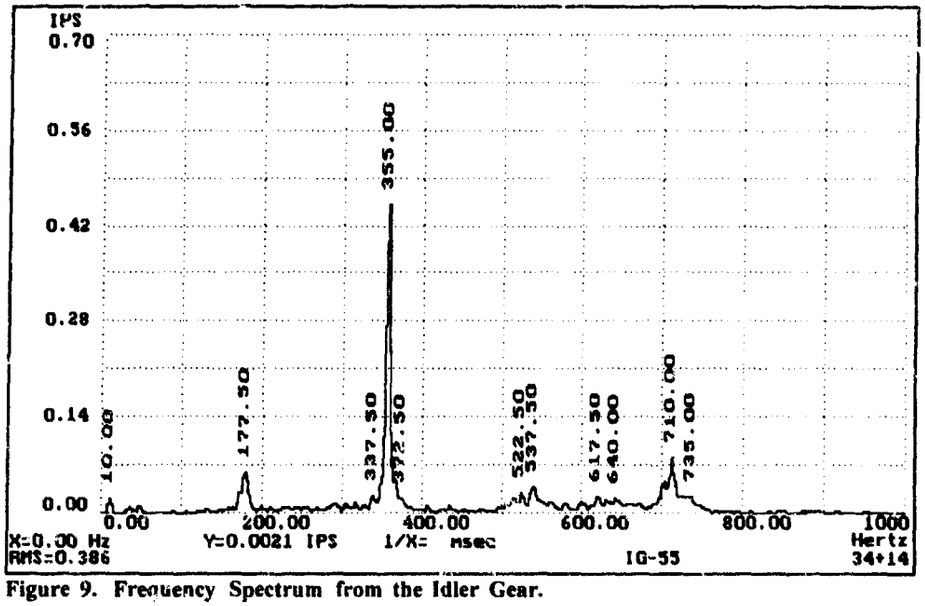


Figure 9. Frequency Spectrum from the Idler Gear.

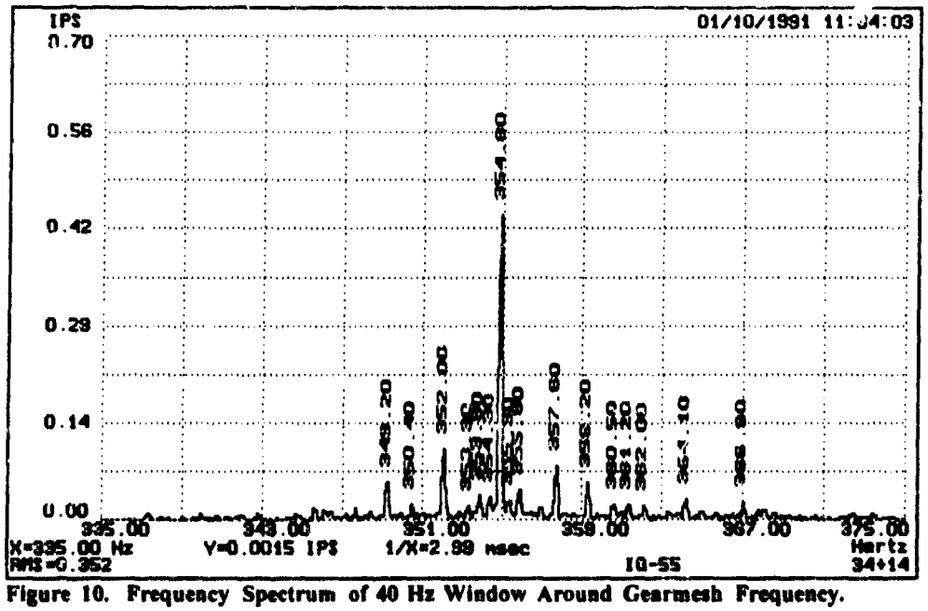


Figure 10. Frequency Spectrum of 40 Hz Window Around Gearmesh Frequency.

A time signal the length of the reciprocal of the hunting tooth frequency, HTF is required for best accuracy. $\left(\frac{1}{0.3}\right)$ = 3.23 *seconds of time* That time signal cannot be presented here because of length.

The next best option is to look at a couple of smaller time periods. Figure 11 contains 178 msec of the time signal from the gears. Remember, this is only 178 milliseconds out of 3.2 seconds. Several points should be made. The zero-to-peak amplitude varies from 0.21 IPS to over 1.2 IPS! The individual cycles are gearmesh frequency. The higher amplitude of every other cycle is half gearmesh frequency. The amplitude of the half gearmesh frequency is 1.2 IPS. The amplitude of this frequency in the above spectra was only 0.06 IPS. This indicates only a few teeth are eccentric and half gearmesh frequency is generated during a small portion of each revolution.

A closer look at the time signal reveals more information. Figure 12 contains about 20 msec of the time signal. Each cycle is gearmesh frequency. When every other cycle is higher in amplitude, that is half gearmesh frequency. Every other cycle of gearmesh frequency is truncated on the negative side. This indicates an in-phase condition. The time period for 532 Hz (1.5 x gearmesh frequency) is 1.87 msec. The time period for 710 Hz (2 x gearmesh frequency) is 1.4 msec. The harmonic appearing near the bottom of each cycle could be 1.5 or 2 times gearmesh frequency or both. The slight phase change can be observed from cycle to cycle. The diagnosis of these gears would be as follows: Gear 1 is loose and has some eccentric teeth. Gear 2 has some eccentric teeth. The eccentric teeth are much more serious than indicated in the frequency spectra.

The eccentricity is probably caused by the manufacturer. The gears could be repaired or fixed by match grinding if the problems were identified early enough. Of course, a better solution would be to specify the proper AGMA gear quality during acquisition. Then if the problem occurs, have the gears replaced under warranty. In either case, a failure is prevented.

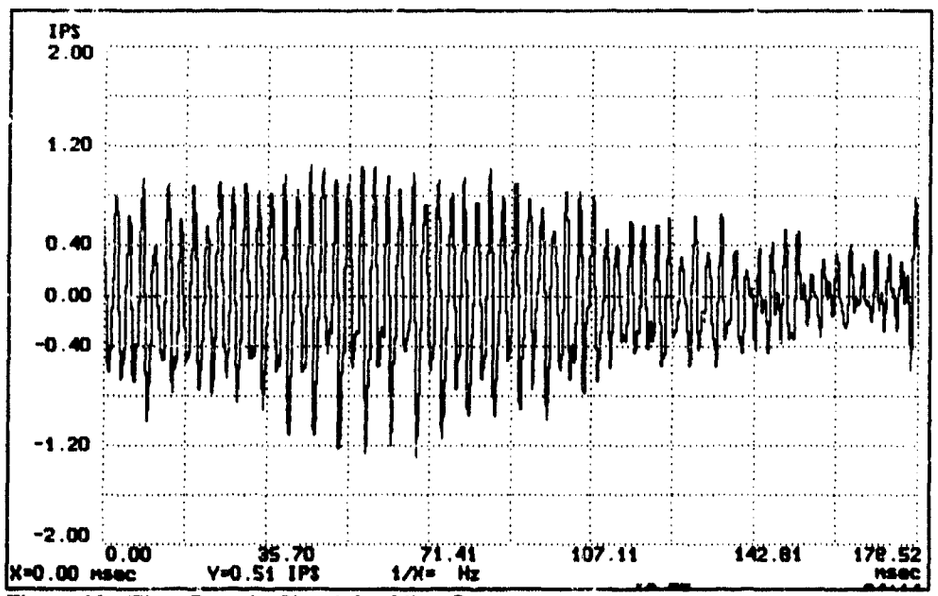


Figure 11. Time Domain Signal for Idler Gear.

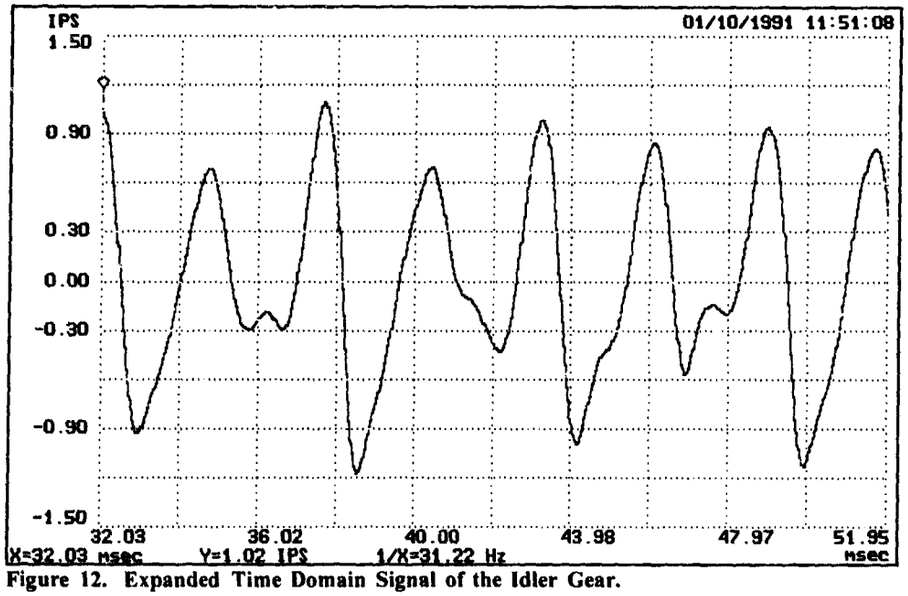


Figure 12. Expanded Time Domain Signal of the Idler Gear.

Conclusion: If the correct hardware, software, and technology are placed in the hands of skilled analysts (engineer or technician), rotating machinery problems can be accurately diagnosed and failures can be prevented. If the time signal and phase is not used, the diagnostics can have an error content of 30% or more.

References:

Drago, R. J. Fundamentals of Gear Design. Butterworth Publishers, 1988.
Taylor, James I. The Vibration Analysis Handbook. VCI, 1994.

CONDITION BASED MAINTENANCE IN POWER GENERATION

T. James Du Bois

Du Bois & Associates
P.O. Box 98
Tujunga, CA 91043-0098

Abstract: The use of Condition Based Maintenance (CBM) is a new concept in the highly regulated environment of the Power Generation industry. The industry is changing, however, and one company, Southern California Edison, has initiated CBM to meet its efforts to lower maintenance costs while maintaining system reliability. The reasons for implementing CBM, the goals of the program, the methods used to implement it, the obstacles that had to be overcome and the successes resulting from its implementation are the topics discussed in this presentation.

Key Words: Attrition; auxiliaries; boilers; benchmarking; breakdown maintenance; condition based maintenance; contractor displacement; decentralization; predictive maintenance; preventive maintenance; root cause analysis; turbine-generators.

Introduction: For decades the Electric Utility Industry experienced steady, at times even rapid, economic growth. The utilities benefitted from having designated service territories and, being regulated by federal and state bodies sympathetic to economic growth and reliable service to customers. With the oil and gas shortages in the 1970's, things began to change.

Predictions of high prices for oil spurred the search for alternative fuels financed in part by federal subsidies. Federal legislation made it mandatory that renewable energy resources be used whenever possible. As the prices of fuel were normally passed on to the consumer, the cost of electricity began to rise. Rate payer groups concerned with these costs began to form and demand new regulation of the utility industry. Deregulation was a definite possibility at both the federal and local levels.

The environmental movement showed that air quality, water quality, and solid waste disposal were national problems that needed to be addressed. The nuclear utilities came to national attention with the accident at Three Mile Island. The costs of building nuclear plants soared. The siting of any new generating station became difficult with the requirement for Environmental Impact Reports and extended regulatory review processes.

New energy producers emerged. Some were large industries desirous of lowering their energy bills and supplying energy for their own use, but able to sell surplus energy to the utility or others. Some were new energy producers who would sell their energy to anyone. Some were renewable energy resources. In other words, the utilities found themselves in a competitive market place for the first time in a long while.

The utilities have been restructuring their organizations and methods to meet these challenges. One of the leaders in these changes is the Southern California Edison (SCE) Company. Within SCE the Power Production Department has taken a bold new direction in an effort to lower their Operating and Maintenance (O&M) cost. Their new direction includes the usage of Condition Based Maintenance (CBM).

Southern California Edison: Southern California Edison is a company within SCECorp. SCE serves a 50,000 square mile territory that extends from the San Diego County boundary on the south to Santa Barbara City and Tulare and Mono Counties on the north. The service territory extends from the Pacific Ocean on the west to the Nevada and Arizona borders on the east. SCE serves over 4,000,000 installed electric meters. Within this service territory there are municipal utilities that serve cities such as Los Angeles, Pasadena, Burbank, and Glendale and the County of Riverside.

The Power Production Department is one of several departments in SCE. It consists of six divisions: Steam Generation; Business Strategy & Policy; Hydro Generation; Transmission/Substation; Power Production Administration; Power Systems Engineering & Construction, and Civil Engineering & Construction. The Power Production Department does not include nuclear facilities. All nuclear generation is administered in a separate department.

The Steam Generation Division consists of seven oil/gas fired conventional steam generating stations, two combined cycle generating stations, and two coal generating stations. These stations contain over 50 individual generating units, capable of producing 12,000 MW of electric power

The Transmission/ Substation Division includes 4800 miles of bulk power (over 220 kV) transmission lines, 200 transmission substations, 704 distribution substations, and 7200 miles of sub transmission (161/33 kV) lines.

The Situation: The oil/gas generating stations had been providing almost 60% of the power generation market in 1982. However, by 1990 renewable and cogenerating Qualifying Facilities (QF's) operating in a baseload mode have succeeded in taking away SCE's market share of Southern California generation. SCE purchases up to 5000 MW of baseload power daily to supply the service territory. As a result SCE generating units have been required to operate in cycling modes that require them to start up in the morning to be ready to supply power during the daily peak but to shutdown or to run at low power generation levels during the night. Capacity factors for the oil/gas units dropped to an average of 17%.

Loss of market share and adverse regulatory outcomes could create a 'death spiral' for service quality. Loss of market to baseload competitors could affect the ability to dispatch power, result in station closures, force layoffs, reduce system reliability, and cause brownouts or other generation related interruptions of service. Obviously these results were unacceptable to SCE.

SCE has assets that are valuable to its customers. The hydroelectric assets provide the lowest cost generation while preserving the environment. The coal units are the low total cost source of base-loaded energy. The conventional gas/oil units provide low-cost capacity to meet intermediate and peak load needs and load following requirements. The other generating assets provide the diversity essential to assuring reliable service.

Structuring for Change: The senior management of SCE recognized that changes were required to compete in the new market of power generation. The changes began in the Steam Generation and Business Strategy and Policy Divisions of Power Production.

The first change was in the business management approach.

During the 1980's the management approach within the Steam Division was control oriented. Central Office management was in command and controlled almost all activities of the field locations (generating stations). To control budget expenditures a five year program of cost containment was instituted which simply set limits to the budget without concern to priorities of replacement of equipment. The general attitude was that the Central Office knew best and the stations should follow instructions. The relationship with the unions was one of 'us' and 'them' and most often of 'stick' not 'carrot'. These methods were all changed in a new management approach for the 1990's.

The 1990's business management approach was one of involvement at all levels. The employees are an integral part of the business and had to understand the issues, goals, strategies, and results of performance. This was a culture change. The employees had to have and know they had the power to influence decisions that affect business performance. Business necessity drives the need for change. Communication became an integral part of the success of the changes.

A focused Business Strategy was developed which was based on competitive success. To develop a higher performance ethic, defined goals were established combined with an incentive award program for achieving those goals.

Starting the Change: The first step was an organizational change. Management responsibilities were decentralized, giving to each of the generating stations the responsibility for administering locally their budgets and operating and maintenance activities.

The second step was to establish strategy action task teams involved in pursuing initiatives for improving cost competitiveness and efficient capital utilization. These task teams were made up of employees from all levels within the Division to obtain the broadest representation and involvement.

The Maintenance Team charter was to recommend an approach to maintenance that would achieve a 35% improvement in maintenance expenses while maintaining reliability and availability of the system. The Management Processes Team charter was to recommend an approach to improve the effectiveness of management processes and business focus.

The Centralized Services Team recommended revised approaches to increase the value and cost-effectiveness of tasks and services centralized resources provide to their customers. The Multiskilling Team investigated approaches to achieving a more multiskilled work force. The Communications Team charter was to recommend an approach to building effective communications throughout the Divisions on strategy and change initiatives.

The third step was to benchmark the performance of the generating stations against the generating stations of other utilities and against independent producers of power.

The Changes: The Maintenance Team identified several methods for improving costs. The methods were associated with the two major categories of overhaul and non overhaul costs. In review of the recommendations it became obvious that a better understanding of the condition of the equipment to be repaired was required.

Prior to 1992, the maintenance philosophy was primarily time-based. Planned maintenance was performed on a preset schedule. Specified time intervals developed from experience, original equipment manufacturer's (OEM) recommendations and judgment, formed the basis for auxiliary overhauls as well as routinely required maintenance. Turbine-generator overhauls were derived from formulas based on operating time. Breakdown maintenance was performed when necessary to repair equipment that failed prior to the scheduled outages. Major boiler maintenance was scheduled to coincide with turbine-generator overhauls.

This resulted in an increasing workscope for major overhauls, an increase in the duration of the outages and rapidly increasing costs. Unfortunately, some equipment was overhauled or subjected to maintenance needlessly. In addition, some equipment suffered from human interference and performed less efficiently after being overhauled or maintained, thus resulting in increased operating costs.

To correct this situation and to achieve the goal of reducing operating and maintenance (O&M) cost by 35%, a Condition Based Maintenance program was recommended and adopted by the generating stations. Condition Based Maintenance (CBM) is a maintenance philosophy that uses the most cost effective methodologies for the performance of equipment maintenance. CBM incorporates a conscious selection process to apply various maintenance philosophies to specific types of equipment, depending on the significance of that equipment to the generation process. It includes the following maintenance philosophies:

Break-down Maintenance: Applied to non-critical, low-cost equipment that would not affect reliability, safety or environmental concerns if it failed.

Preventive Maintenance (PM): Time-based tasks on either critical or non critical equipment, including routine lubrication, instrument checks, or electrical testing. Generally applied where application of available technologies or testing would not effectively detect a developing failure. Machinery assessments are scheduled on a PM basis.

Predictive Maintenance (PdM): The detection of developing equipment failures in their early stages and determining the need for repair or replacement, using various monitoring, testing, analysis, diagnostic and inspection techniques. Failure modes are detected in the early stages, allowing planned maintenance outages.

Proactive or Root Cause Based Maintenance: Application of systematic methods to determine the basic cause of recurring maintenance problems. The objective is to eliminate or minimize the need for maintenance by determining predominant failure modes and causes. Redesign of equipment, or revision of other factors that lead to unnecessary maintenance are addressed.

The condition of major equipment is periodically assessed by the methods and procedures previously selected as being applicable for each specific machine. The results of these assessments form the basis for the decisions on when major maintenance should be performed. The assessments are on-going based on changing conditions of each machine. In many cases an overhaul can be deferred for an additional period of time, in which case another assessment would be scheduled.

It is recognized that even the most sophisticated present-day technologies cannot detect all incipient failures. Thus, there is some additional risk associated with CBM. In many cases, judgments must be made based on historical data or other factors, such as cost considerations.

The cost reduction or savings associated with CBM occur in the following areas:

1. Elimination of unnecessary maintenance.

2. Early detection of developing failures, allowing planned repairs before damage becomes extensive.

3. Reduced maintenance requirements by performing overhauls only when the condition indicates a need. Savings are realized over a long time interval.

After the agreement to implement CBM, presentations on the principles of CBM were made for the management teams at each generating station, and these teams then adapted CBM to their specific location. CBM Coordinators were appointed at each station as the point for implementation.

Initial assessments of critical equipment at each location were made by an Assessment Team that consisted of experienced maintenance and engineering personnel. It was recognized that these assessments would not be complete, but merely a "kick-start", forming the basis for each location to provide more detailed information. Included in these assessments were historical data, known existing conditions, performance data, and the results of inspections and minimal testing. An assessment document was completed for each equipment or device, showing all known conditions and recommendations for future

maintenance or inspections. Recommendations for future overhauls were made. Annual cost projections were made both on the original overhaul schedule, and on the CBM Assessment Team's recommended schedule. Each station then reviewed and added to the assessments as appropriate. Cost savings due to CBM were projected based on the differences between the original and CBM overhaul schedules.

As each station progressed in implementation, various approaches to CBM began to emerge. Emphasis was on differing aspects of CBM at each station. Many innovative approaches were tried and developed. The CBM Coordinators held regular meetings and were able to communicate both successful and failed methods.

The Results: CBM was initiated at the generating stations in the Steam Division in the latter half of 1991. Therefore, 1992 became the first full year in which the effect of CBM could be evaluated. The results were encouraging. A $39.5 million cost improvement was recorded in the Division's Expense Budget over 1991's recorded expenses!

The $39.5 million cost improvement was divided into two categories: $18.5 million in sustainable productivity improvements and $21.0 million in transition year cost reductions. The $18.5 million in productivity improvements were particularly significant since they are projected to continue into the future. CBM accounted for $8.4 million of the cost improvement and was due to the elimination of unnecessary maintenance and improvement of maintenance practices. By reducing the maintenance required by station personnel, these personnel were able to be assigned tasks that would have required the services of outside contractors. This contractor displacement amounted to $4.4 million. Reducing the number of outside contract work resulted in greater competition for the remaining contracts resulting in a further cost improvement of $3.3 million. An additional $2.4 million in cost improvement was attributable to reductions in supplies and improved procurement practices.

Transition year cost reductions of $21.0 million are not expected to be repeated to the same amount in future years. Whenever a new program such as CBM is instituted there will be reevaluations of the schedule of the overhauls. By readjusting the schedule of several overhauls planned for 1992, a total $15 million was saved. This also allowed some stations to perform with station personnel some work that would normally be done by temporary labor resulting in a cost reduction of $2.1 million. By adjusting the operational schedules of the plants some environmental fees were offset for an additional $2.1 million. Other miscellaneous savings amounted to $1.8 million.

To further evaluate the effect of the CBM program, the average cost of turbine-generator overhauls under CBM was forecast for the years 1993 through 1997 and compared to the average cost of overhauls for the

past ten years (not including 1992). The comparison showed a reduction of $6.3 million per year.

The cost improvements in 1992 included the attrition of 151 permanent Power Production Department personnel. The Department expects to further reduce by attrition an additional 520 personnel by the year 2000. Some of these personnel will be in the Steam Generation and Business Strategy and Policy Divisions, thus further reducing the O&M expenses.

The Future: Singular success in one year does not mean that SCE has achieved its goal of being competitive in the power generation market. It does point to the fact that significant improvements in cost reduction can be made in the Power Production Department by:

1. Placing increased attention on making cost-effective capital expenditures that result in reduced operating and maintenance expenses.

2. Becoming more aggressive in cost management.

3. Establishing long-term business planning and tracking.

To achieve these objectives the Steam Generation and Business Strategy and Policy Divisions have established Business Plans for the Divisions and for each of the organizations within the Divisions. These Business Plans set out goals for performance that if achieved will attain the competitive position desired by the Department. These performance areas include: Safety, Environmental Compliance, Operational Performance, Efficient use of Capital, and the Reduction of O&M expense of 35% by 2000 through CBM.

Conclusion: Although the final success of their CBM program will not be seen until the year 2000, Southern California Edison has made a significant step in reducing the operating and maintenance expenses of the Power Production Department through the implementation of a Condition Based Maintenance program.

MACHINERY MONITORING AND DIAGNOSTICS

Cochairmen: Carl Talbott
M&M/Mars

Ruben A. Lebron, Jr.
Naval Air Warfare Center

A NEW APPROACH TO GEAR VIBRATION DEMODULATION AND ITS APPLICATION TO DEFECT DETECTION

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

C. James Li
Department of Mechanical Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180

Abstract: A new scheme for gear vibration demodulation is described. First, a gear vibration model that includes amplitude and phase modulations is proposed. Based on this model, an algorithm which considers all tooth meshing harmonics as modulation carriers is established. This algorithm derives amplitude and phase modulation signals from a gear's vibration signal average obtained by synchronized averaging. This iterative algorithm first finds an approximation to the phase modulation signal from the signal average. Using the phase modulation signal, the amplitude modulation signal is then identified from the signal average. The effectiveness of this scheme is confirmed and compared to the state-of-the-art using a vibration average of a gear which has two defects of different sizes.

Key Words: Condition monitoring; Demodulation of gear vibrations; Gear defect detection; Gear vibration model; Signal analysis; Signal processing; Vibration analysis.

1. INTRODUCTION

For years, monitoring the condition of power transmissions has been deemed imperative. As indicated in an Army report (US Army, 1976), power plants and drive systems contributed to 68% of flight safety incidents related to failures in mechanical systems and to 58% of direct maintenance costs. Thus, gearboxes, the core of power transmission, have received considerable attention in the field of condition monitoring and fault diagnosis. In particular, gear localized defects have been extensively studied, since a large percent (60%) of gearbox damages are due to gear faults, which in turn are mostly initiated by localized defects. For example, a study on gear faults (Allianz, 1978) reported that three types of gear localized defects, namely, forced fracture, fatigue fracture, and incipient cracks, were responsible for about 90% of gear faults found in gearboxes (56% 17% , and 16%, respectively).

Vibrations externally measured on a gearbox have been used to monitor the operating condition of the gearbox and diagnose the fault, if there is any, 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 rotation 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, contained 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.

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, may be employed. 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. 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 of the signal average may be sufficient to detect the damage, as in a case that will be shown shortly (Fig. 2). But the detection of defects at a very early stage requires sophisticated signal detection and processing techniques to enhance the defect information contained in the signal average. Due to the modulation nature of defect related vibration, a number of existing gear defect detection techniques consider amplitude and phase modulations as defect signature. It has been found (McFadden and Smith, 1985) that by bandpass filtering the signal average about the largest meshing harmonic, eliminating that harmonic and then enveloping, a signal considered representing the modulation energy could be obtained. A further study (McFadden, 1986) tried to demodulate the signal average around the largest meshing harmonic to provide both the approximate amplitude and phase signals. However, in all these studies, only a single tooth meshing harmonic and its sidebands were utilized to find the modulation signals, while, in fact, all meshing harmonics carry information about the modulation signal. Consequently, a lot of useful modulation information were ignored or misinterpreted.

In this paper a new gear vibration demodulation scheme is described. As opposed to the narrowband approach mentioned above, the scheme takes all tooth meshing harmonics into consideration as modulation carriers and recovers amplitude and phase modulation signals from the wide band signal. Since gear localized defects produce local modulations to gear vibrations, they can be easily identified from modulation signals by inspection. The effectiveness of the scheme in gear localized defect detection is examined with the vibration average of a gear containing both an advanced defect and a smaller defect.

2. GEAR VIBRATION MODEL: MODULATIONS

This section presents a gear vibration model based on which a demodulation scheme will be derived. To establish such a model, 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 tooth meshing frequency. 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:

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

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

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

Now take into consideration 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 meshing frequency and its harmonics as they go through the meshing. Let these changes be described by the amplitude and phase modulating functions $a_k(\theta)$ and $b_k(\theta)$, then the modulated vibration is

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

The amplitude modulating function changes the envelope of $x(\theta)$, and therefore should be independent of tooth meshing harmonics. On the other hand, since the phase modulating signal at any given instant produces the same time delay to all tooth meshing harmonics in the vibration, it is linearly proportional to the harmonic number k. Thus Eq. (3) is simplified into:

$$y(\theta) = [1 + a(\theta)] \sum_{k=0}^{K} X_k \cos[kN\theta + \phi_k + kb(\theta)] \tag{4}$$

where $a(\theta)$ and $b(\theta)$ themselves can be expanded into Fourier series in harmonics of rotational frequency of the gear:

$$a(\theta) = \sum_{p=1}^{P} A_p \cos(p\theta + \alpha_p) \tag{5}$$

$$b(\theta) = \sum_{q=1}^{Q} B_q \cos(q\theta + \beta_q) \tag{6}$$

This is because, obviously, they repeat themselves from one rotation to the next, as the signal average does.

If the modulating functions $a(\theta)$ and $b(\theta)$ can be extracted from the signal average $y(\theta)$, then we will be able to evaluate the gear condition and detect any defects on the gear because information on tooth variations from average is carried in $a(\theta)$ and $b(\theta)$. In the following section, an algorithm for extracting modulation signals from the signal average is derived and applied to gear localized defect detection.

3. DEMODULATION METHOD

3.1 Separation of Phase and Amplitude Modulations: Suppose that amplitude and phase modulating signals $a(\theta)$ and $b(\theta)$ are narrow-band signals, containing no components with frequency higher than the tooth meshing frequency, i.e., $0<P,Q<N$. Then the analytic signal $y_a(\theta)$ of real measurement $y(\theta)$ can be approximated by (Rihaczek, 1966)

$$y_a(\theta) \approx [1 + a(\theta)] \sum_{k=0}^{K} X_k \exp\{j \cdot [kN\theta + \phi_k + kb(\theta)]\} \tag{7}$$

where j is the complex numeric, and $y_a(\theta)$ has $y(\theta)$ as its real part and the Hilbert transform $\hat{y}(\theta)$ of $y(\theta)$ as its imaginary part:

$$y_a(\theta) = y(\theta) + j \cdot \hat{y}(\theta) \tag{8}$$

$$\hat{y}(\theta) = \frac{1}{\pi} \int \frac{y(\vartheta)}{\theta - \vartheta} d\vartheta \tag{9}$$

Let

$$z(\theta) = \sum_{k=0}^{K} X_k \exp\{j \cdot [kN\theta + \phi_k + kb(\theta)]\} \tag{10}$$

then Eq. (7) becomes:

$$y_a(\theta) \approx [1 + a(\theta)] \cdot z(\theta) \tag{11}$$

Note that the real signal $[1+a(\theta)]$ does not contribute phase angle to the right hand side of Eq. (7). Therefore, complex signals $y_a(\theta)$ and $z(\theta)$ have the same phase angle for all θ. If we know X_k and ϕ_k, we can find $b(\theta)$ by equating phase angles of $y_a(\theta)$ and $z(\theta)$, and then $a(\theta)$ can be solved from Eq. (11) by

$$a(\theta) = \frac{\|y_a(\theta)\|}{\|z(\theta)\|} - 1 \tag{12}$$

However, X_k and ϕ_k are not readily available. In the following, an analysis is given to find the relationship between these quantities and modulation signals. Then a demodulation scheme will be derived.

3.2 Meshing Component Retrieving: Instead of having X_k and ϕ_k , what we have are Fourier magnitudes Y_l and phase angles φ_l of the signal average $y(\vartheta)$ through its Fourier expansion

$$y(\theta) = \sum_{l=0}^{L} Y_l \cos(l\theta + \varphi_l) \tag{13}$$

To obtain X_k and ϕ_k from $y(\theta)$, we compare tooth meshing harmonic in the model (4) and the measurement described in Eq. (13), which are two different expressions of the same signal.

Assume that the influence on a carrier, i.e., one of the tooth meshing harmonics, from modulation sidebands of its two neighboring carrier harmonics is negligible. This means that sidebands of a tooth meshing harmonic caused by modulations are negligible when they are away from that tooth meshing harmonic by more than the tooth meshing frequency. Then the modulation effect on each carrier component may be examined individually. In other words, for a particular pair of X_k and ϕ_k, the modulation on the k-th meshing harmonic, which is the k-th term in model (4)

$$[1 + a(\theta)]X_k \cos[kN\theta + \phi_k + kb(\theta)]$$

will be examined. Noting that amplitude modulation does not change the magnitude and phase of the carrier, i.e., the meshing harmonic, we can consider only the phase modulation for examining the carrier component. Without amplitude modulation, the k-th term in Eq. (4) becomes the following:

$$\begin{aligned} v(\theta) &= X_k \cos[kN\theta + \phi_k + kb(\theta)] \\ &= \sum_{i=0}^{I} V_i \cos(i\theta + \gamma_i) \end{aligned} \tag{14}$$

When $i=kN$, V_i and γ_i give the Fourier magnitude and phase at the k-th meshing harmonic, which should be equal to those given by the measurement $y(\theta)$:

$$V_{kN} = Y_{kN} \tag{15}$$

$$\gamma_{kN} = \varphi_{kN} \tag{16}$$

On the other hand, X_k and ϕ_k can be calculated from V_{kN} and γ_{kN} by the following two equations (see Appendix):

$$X_k = \frac{V_{kN}}{M_k} \tag{17}$$

$$\phi_k = \gamma_{kN} - p_k \tag{18}$$

However M_k and p_k are related with Fourier magnitudes and phases of $b(\theta)$ through

$$M_k \cdot \exp(jp_k) = \sum_c \left[\prod_{q=1}^{Q} \mathrm{J}_{n_q^{(c)}}(kB_q) \right] \exp\left[j\left(\sum_{q=1}^{Q} n_q^{(c)}(\beta_q + \frac{\pi}{2}) \right) \right] \tag{19}$$

where $n_1^{(c)}, n_2^{(c)}, \ldots, n_Q^{(c)}$ are permutations of $n_1, n_2, \ldots, n_Q$ satisfying

$$\sum_{q=1}^{Q} n_q^{(c)} q = 0 \tag{20}$$

and $\mathrm{J}_n(x)$ is the Bessel function of the first type.

Therefore, if the phase modulating signal $b(\theta)$ is known, i.e., B_q and β_q are known, M_k and p_k can be obtained by solving Eq. (19), and then X_k and ϕ_k can be recovered using Eqs. (17) and (18).

3.3 Proposed Scheme: Above analysis provides a means to derive X_k and ϕ_k from the signal average when $b(\theta)$ is given. These X_k and ϕ_k can be used, in turn, to obtain another $b(\theta)$ by equating phase angles of two sides of Eq. (11), i.e., comparing the phase of $y_a(\theta)$ and $z(\theta)$. If the initially given $b(\theta)$ is the actual one, it will be the same as the newly obtained one. Using these two relationships iteratively, a procedure can be established to find the true $b(\theta)$. In the following, this iterative procedure is detailed (see Fig. 1):

(a) Calculate the Fourier magnitudes Y_l and phases φ_l of signal average $y(\theta)$ at tooth meshing harmonics, i.e., $l=kN$;

(b) Using Hilbert transform, construct the analytic signal $y_a(\theta)$ from $y(\theta)$ as described in Eqs. (8) and (9);
(c) Make an initial guess on $b(\theta)$;
(d) Find M_k and p_k for all k's according to Eq. (19);
(e) Calculate X_k and ϕ_k using Eqs. (17) and (18), then construct the complex $z(\theta)$ by Eq. (10);
(f) Compare the phase of $y_a(\theta)$ and $z(\theta)$. If the error is small enough, continue; otherwise, adjust $b(\theta)$ and go back to Step (d);
(g) Calculate $a(\theta)$using Eq. (12).

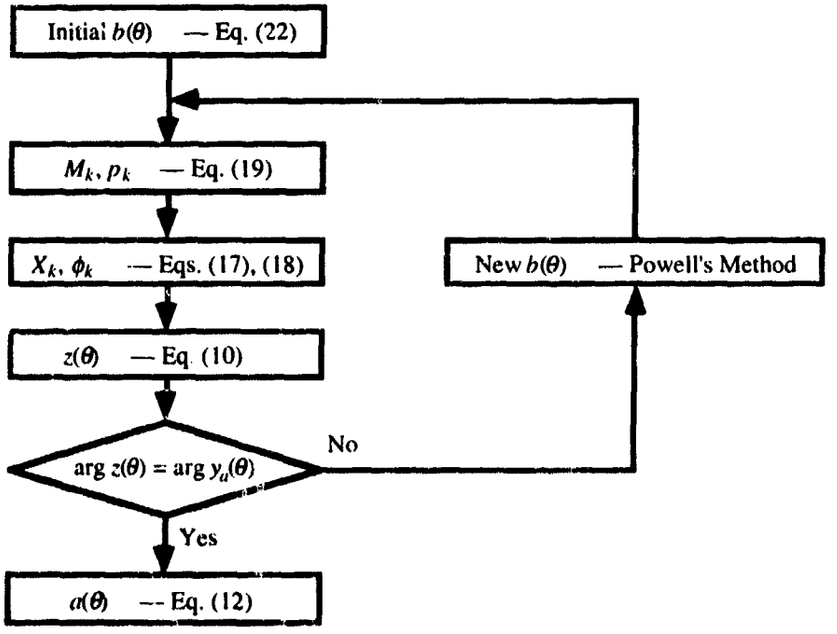


Fig. 1 Diagram of the full-spectrum demodulation procedure

In our implementation of the scheme, Powell's method (e.g. Wismer and Chattergy, 1978) is employed in Step (f) to search for the $b(\theta)$ that minimizes

$$J(b(\theta)) = \int_0^{2\pi} \left[\arg(y_a(\theta)) - \arg(z(\theta))\right]^2 d\theta \qquad (21)$$

To reduce the computational burden, discrete Fourier components of $b(\theta)$, instead of $b(\theta)$ itself, are searched to minimize $J(b(\theta))$. Depending on the initial guess of $b(\theta)$, the search algorithm may converge to different local minimum of the objective function (21) and the convergence time is different, too. The bandpass-demodulation method that considers the largest meshing harmonic only (McFadden, 1986) can be used to obtain a good guess for presumably, the largest carrier harmonic carries the largest fraction of modulation energy. Specifically, the signal average is bandpass

filtered about the largest tooth meshing harmonic, say the k-th. Then an analytic signal $y_{a(bp)}$ is constructed for the bandpassed signal average through Hilbert transform. The initial guess $b_0(\theta)$ can then be obtained by

$$b_0(\theta) = \frac{1}{k}\arg\left[y_{a(bp)}\right] - N\theta \tag{22}$$

4. EXPERIMENTAL INVESTIGATION

The experimental setup, as illustrated in Fig. 2, contains a three-stage reduction gearbox driven by a variable speed DC motor. The load is provided by a disk brake and is adjustable with a screw-spring loading mechanism. The 20-tooth input pinion has two artificially seeded defects of different size at teeth 11 and 18, respectively, simulating fractured teeth at different stage. Since modulation sidebands can be easily distorted by the effect of vibration transmission path (McFadden and Smith, 1986), torsional vibration, instead of translational vibration, is measured using a Hoodwin 34E angular acceleration sensor based on the principle that the eddy current induced in a disk armature is proportional to the rotational speed of the disk (Hoodwin, 1967). To facilitate the synchronous averaging, a 450-line optical encoder is employed to trigger the sampling of the vibration signal.

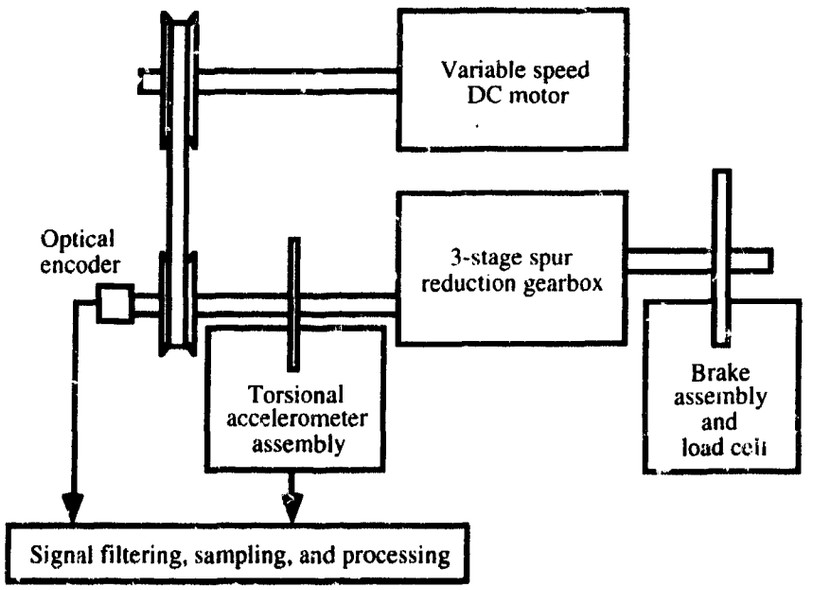


Fig. 2 Experimental setup for gear defect detection

Fig. 3 shows a signal average, obtained from 200 rotations, as a function of angular position (tooth number), and its spectrum indexed by the rotational frequency (order). It is obvious that an advanced fracture, like the one on tooth 11, can be detected immediately by a visual inspection of the signal average. The larger amplitude of vibration around that tooth indicates reduced meshing stiffness. However, the smaller defect on tooth 18 has little noticeable effect on the signal average. Additionally, the band-pass demodulation method (McFadden, 1986), which will be used later to

provide an initial guess about phase modulation signal, also failed to detect the defect (Fig. 4).

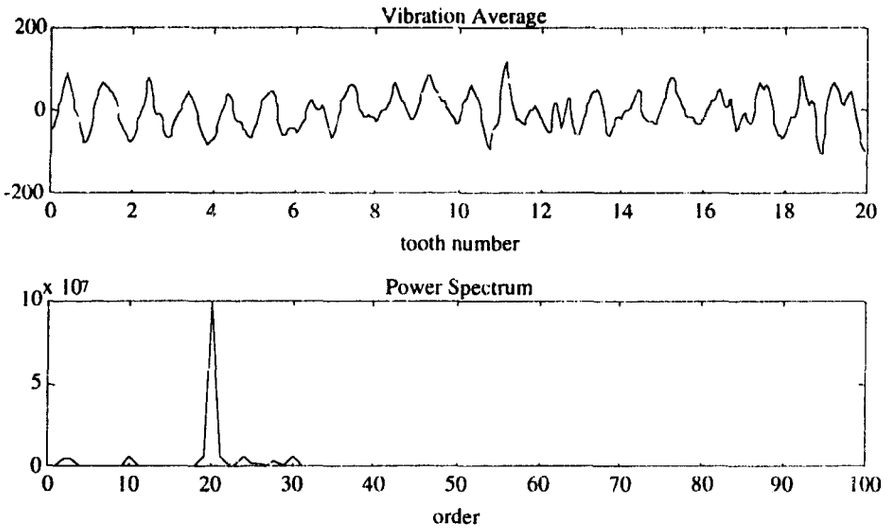


Fig. 3 Vibration signal average and its spectrum

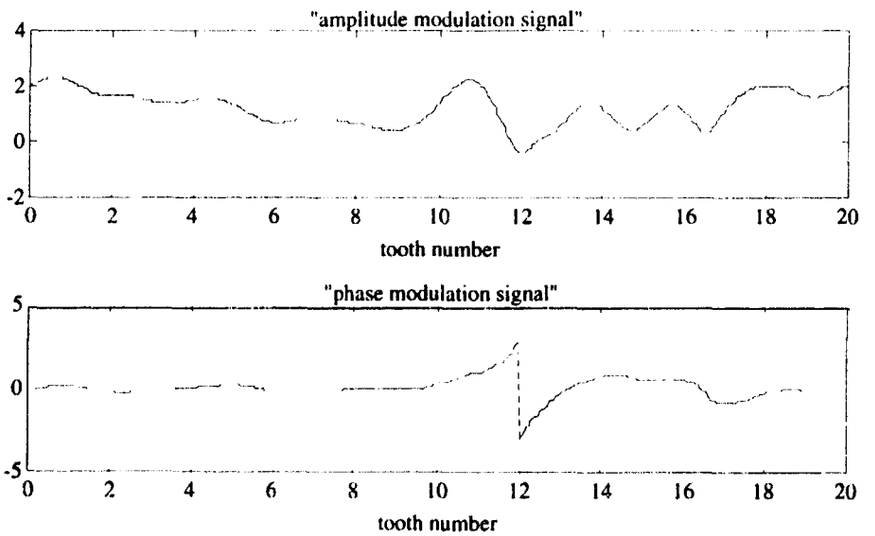


Fig. 4 Results by bandpass-demodulation method

To evaluate the proposed scheme, we applied it to the signal average. The proposed method assumes that a phase modulation signal can be approximated by a Fourier series that contains up to Q=N-1 (here N=20) rotational harmonics. The larger the Q, the better the approximation is to the phase modulation, and the more computation is required. For example, the amount of computing power required for solving Eq. (20) for permutations $n_q^{(c)}$, q=1,2,...,Q, would grow exponentially with respect to Q. To find how sensitive the scheme is to different Q's, we implemented the scheme with

Q=7 and Q=5, respectively. It was found that both phase and amplitude modulation signals were very similar in both cases. Since only a small number (5 or 7, compared to 19 which was assumed maximum number in this case) of rotational harmonics in phase modulation signal were considered, the amplitude modulation signal should give better indication to gear condition than phase modulation signal could. In fact, defects on tooth 11 and tooth 18 were obvious in the amplitude modulation signal solved with Q=7 which is much lower than the N-1(see Fig. 5).

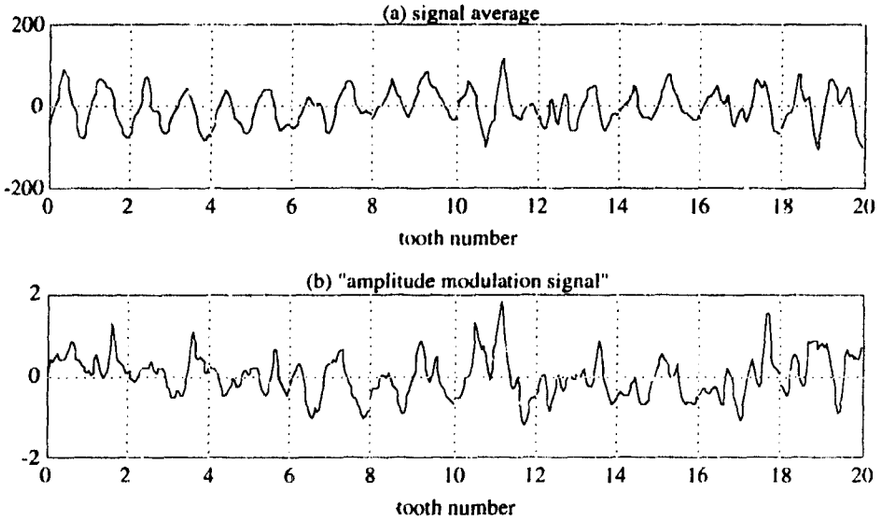


Fig. 5 Vibration signal average and amplitude modulation signal

5. CONCLUSIONS

Torsional vibrations produced by a gear containing localized defects were examined A gear vibration model which includes angular and amplitude modulations was proposed. Means for constructing this model and then recovering the angular and amplitude modulation signals from real signal average was established. It has been found that the amplitude modulation signal recovered by the scheme is very sensitive to defects compared to the bandpass demodulation scheme, although phase modulation signal does provide some indications to defects as well. Why the phase modulation is less sensitive to the defects is due to the fact that we restrained ourselves from approximating it with large number of Fourier components to alleviate the demand on computing power. Nevertheless, the amplitude modulation signal is very effective in indicating local modulations that are associated with localized defects.

The superiority of the proposed full spectrum demodulation method over the state-of-the-art has been demonstrated by experimental study on a gear with two localized defects of different sizes. This method successfully detected the advanced defect as well as the smaller one, which is not detectable to the bandpass demodulation method. It can be used for gear quality evaluation in manufacturing and gear condition assessment and localized defect detection in on-line applications.

References

Allianz Versicherungs-AG, 1978, *Handbook of Loss Prevention*, Springer-Veriag, 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*, T. R. Shives and L. J. Mertaugh ed., 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

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

Hoodwin, L. S., 1967, "Angular Accelerometer: Advantages and Limitations", *Instruments and Control Systems*, April 1967, p. 129

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

McFadden, P. D., and Smith, J. D , 1986, "Effect of Transmission Path on Measured Gear Vibration", *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 108, pp. 377-378

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, NJ, pp. 235-245

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

Wismer, D. A., and Chattergy, R., 1978, *Introduction to Nonlinear Optimization, A Problem Solving Approach*, North-Holland, New York

Appendix CARRIER COMPONENTS IN PHASE MODULATION

When a sinusoidal function of angular position θ, of frequency kN with magnitude X_k and phase ϕ_k, is phase-modulated by another function $kb(\theta)$, the resulted signal can be alternatively expressed in many ways:

$$\begin{aligned} v(\theta) &= X_k \cos\left[kN\theta + \phi_k + kb(\theta)\right] \\ &= X_k \,\mathrm{Re}\left\{\exp\left[j\left(kN\theta + \phi_k + kb(\theta)\right)\right]\right\} \\ &= X_k \,\mathrm{Re}\left\{\exp\left[j\left(kN\theta + \phi_k + k\sum_{q=1}^{Q} B_q \cos(q\theta + \beta_q)\right)\right]\right\} \\ &= X_k \,\mathrm{Re}\left\{\exp\left[j\left(kN\theta + \phi_k + k\sum_{q=1}^{Q} B_q \sin(q\theta + \beta_q + \frac{\pi}{2})\right)\right]\right\} \\ &= X_k \,\mathrm{Re}\left\{\exp\left[j\left(kN\theta + \phi_k\right)\right] \cdot \prod_{q=1}^{Q} \exp\left[j\left(kB_q \sin(q\theta + \beta_q + \frac{\pi}{2})\right)\right]\right\} \end{aligned} \tag{A1}$$

where j is the complex numeric, B_q and β_q are Fourier magnitudes and phases of $b(\theta)$, respectively.

In order to investigate the modulation effect of $b(\theta)$ on the carrier component, let us first consider the modulation effect of only one Fourier component of $b(\theta)$, say the q-th component

$$B_q \cos(q\theta + \beta_q)$$

The modulated signal is

$$v_q(\theta) = X_k \,\mathrm{Re}\left\{\exp\left[j\left(kN\theta + \phi_k\right)\right] \cdot \exp\left[j\left(kB_q \sin(q\theta + \beta_q + \frac{\pi}{2})\right)\right]\right\} \tag{A2}$$

The second exponential function on the right hand side is periodical with period $2\pi/q$, so it has the following Fourier series representation:

$$\exp\left[j\left(kB_q \sin(q\theta + \beta_q + \frac{\pi}{2})\right)\right] = \sum_{n=-\infty}^{\infty} \Theta_n \exp(jnq\theta) \tag{A3}$$

where

$$\begin{aligned} \Theta_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left[j\left(kB_q \sin(q\theta + \beta_q + \frac{\pi}{2})\right)\right] \cdot \exp(-jnq\theta) d(q\theta) \\ &= \exp\left[jn\left(\beta_q + \frac{\pi}{2}\right)\right] \cdot \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left[j\left(kB_q \sin\eta\right)\right] \cdot \exp(-jn\eta) d\eta \\ &= \exp\left[jn\left(\beta_q + \frac{\pi}{2}\right)\right] \cdot \mathrm{J}_n(kB_q) \end{aligned} \tag{A4}$$

where $\mathrm{J}_n(x)$ is the Bessel function of the first type, which has the following properties:

(1) $J_n(x)$ is a real valued function;

(2) $J_n(x) = J_{-n}(x)$ for even integer n and $J_n(x) = -J_{-n}(x)$ for odd integer n;

(3) $\sum_{n=-\infty}^{\infty} J_n^2(x) = 1$.

Substitute (A3), (A4) into (A2),we have:

$$v_q(\theta) = X_k \operatorname{Re}\left\{\exp\left[j(kN\theta + \phi_k)\right] \cdot \sum_{n=-\infty}^{\infty} J_n(kB_q)\exp\left[jn\left(q\theta + \beta_q + \frac{\pi}{2}\right)\right]\right\}$$
$$= X_k \operatorname{Re}\left\{\sum_{n=-\infty}^{\infty} J_n(kB_q) \cdot \exp\left[j\left((kN + nq)\theta + \phi_k + n(\beta_q + \frac{\pi}{2})\right)\right]\right\} \quad \text{(A5)}$$

Therefore, the carrier component at frequency kN (thus n=0) is modulated from X_k and ϕ_k to $J_0(kB_q)X_k$ and ϕ_k. Similarly, for the most general case (A1), it can be obtained by induction that

$$v(\theta) = X_k \operatorname{Re}\left\{\sum_{n_1} J_{n_1}(kB_1)\sum_{n_2} J_{n_2}(kB_2)\cdots\sum_{n_Q} J_{n_Q}(kB_Q)\right.$$
$$\left.\cdot\exp\left[j\left((kN + \sum_{q=1}^{Q} n_q q)\theta + \phi_k + \sum_{q=1}^{Q} n_q(\beta_q + \frac{\pi}{2})\right)\right]\right\} \quad \text{(A6)}$$

This is saying that the carrier component has been affected jointly by all modulating harmonics. Specifically, if we express $v(\theta)$ by

$$v(\theta) = \sum_{i=0}^{I} V_i \cos(i\theta + \gamma_i) \quad \text{(A7)}$$

then

$$V_{kN} = M_k \cdot X_k \quad \text{(A8)}$$

$$\gamma_{kN} = p_k + \phi_k \quad \text{(A9)}$$

respectively, where M_k and p_k satisfy

$$M_k \cdot \exp(jp_k) = \sum_c \left[\prod_{q=1}^{Q} J_{n_q^{(c)}}(kB_q)\right]\exp\left[j\left(\sum_{q=1}^{Q} n_q^{(c)}(\beta_q + \frac{\pi}{2})\right)\right] \quad \text{(A10)}$$

and $n_1^{(c)}, n_2^{(c)}, \ldots, n_Q^{(c)}$ are permutations of $n_1, n_2, \ldots, n_Q$ satisfying

$$\sum_{q=1}^{Q} n_q^{(c)} q = 0. \quad \text{(A11)}$$

OPTICAL OIL DEBRIS MONITOR

J. Reintjes, R. Mahon*, M. D. Duncan, L. L. Tankersley†, A. Schultz**, V. C. Chen**, D. J. Kover††, and P. L. Howard***

Laser Physics Branch, US Naval Research Laboratory
Washington, DC 20375

Abstract: We describe a real time, on-line, optical oil-debris monitor that is expected to provide a cumulative record of the health of engines and gear boxes as well as advanced warning of catastrophic failure. The monitor is based on illumination of the oil lubrication column with a diode laser, followed by imaging in transmission of suspended particles and identification of the particles by analysis of their shape and size using an on-board computerized particle classifier. The optical monitor is capable of recognizing metallic and ferrous particles, such as those from metal gears and bearings, as well as non ferrous particles, such as ingested sand and debris from ceramic or composite bearings. We will describe initial tests of the concept using video detection of images of various types of particles in flowing oil systems and will discuss computational requirements and alternatives for real time particle classification.

Key Words: Bearings; catastrophic failure; gears; lifetime; neural nets; Oil debris; optical; real-time; shape classification

Introduction: Currently operating oil debris monitors detect metallic or ferrous particles based on a combination of their magnetic and electric properties. For example, electric chip detectors based on completion of an electric circuit with magnetically attracted particles can provide real time warnings of catastrophic failure. However, they do not provide advanced warning of gradual degradation of components, and they can be prone to false alarms. Determination of the size and shape of individual debris particles can give information about gradual wear of components as well as an indication of impending catastrophic failure [1]. Such information can currently be obtained off-line from analysis of the debris collected on magnetic plugs, but is not presently available with on-line systems. In addition, none of the existing debris monitors work well with non metallic particles. As such they are not capable of warning of the presence of ingested material such as sand, and they will not operate with advanced ceramic or composite bearings.

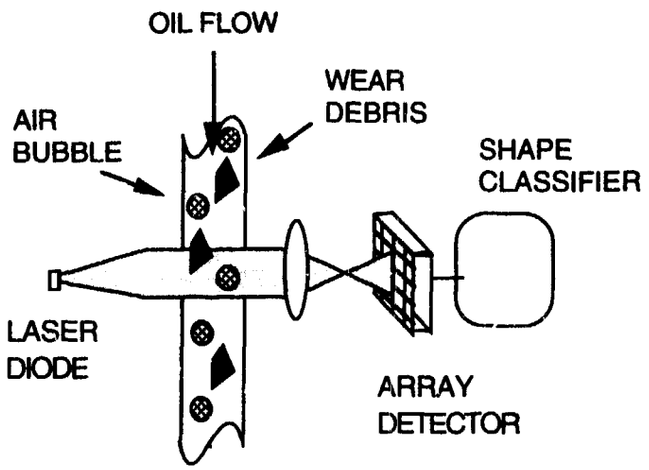


Fig. 1. A schematic diagram of the optical oil debris monitor, showing the illuminating laser diode, the array photodetector and the shape classifier. Objects suspended in the oil flow are imaged in transmission onto the array detector, appearing as dark shadows against a bright background.

Optical oil debris monitor: We are developing an optical oil debris monitor that will operate with all types of debris materials. The monitor is based on a laser illumination system that images in transmission particles that are suspended in the lubrication flow and classifies the particles as to shape and size in real time using on board computer processors. A schematic diagram of the monitor is shown in Fig. 1. The oil flow column is illuminated with a diode laser. The transmitted light is imaged onto an array photodetector such as a CCD or optical transistor array. Any particles that are suspended in the oil column will appear as dark shadows in an otherwise bright field. The shape and size of any such particles are determined by a particle classifier and a record of the particle size and shape is stored in memory. The monitor is expected to provide a cumulative record of the health of the various engine and gearbox components and also can provide a catastrophic warning to the pilot through the use of preset thresholds.

The choice of system components and configuration is driven by a combination of the practical requirements of onboard equipment and various characteristics of the imaging system such as optical transmission of the oil, oil flow speed , optical resolution and field of view. An example of a set of system parameters that provide the

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	1 W peak 1 mW average	OIL TRANSMISSION DETECTOR SENSITIVITY
SIGNAL LEVEL	4×10^5 photons/pixel	LASER POWER ILLUMINATION AREA OIL TRANSMISSION

required performance, and the factors that affect the various system components is given in Table 1.

The use of a laser diode for illumination is driven both by the transmission characteristics of the oil and the practicality of size requirements for an onboard system for helicopters and other aircraft. Optical transmission of used engine and gearbox oil (790 hours of operation) through a 3/4 inch path is shown in Fig. 2. At visible wavelengths the oil is highly absorbing. However, there is a sharp transition in the near infrared, and the oil is relatively transparent for wavelengths longer than about 800 nm. From comparison of a number of oil samples from various sources, the wavelength of the absorption edge, and the transmissivity in the near infrared appear to be dependent on the length of time the oil has been in service and its operating history. However, for all samples tested to date, the transmission is relatively high for wavelengths longer than about 800 nm. This wavelength range is compatible with diode lasers, making them ideal for the illumination source.

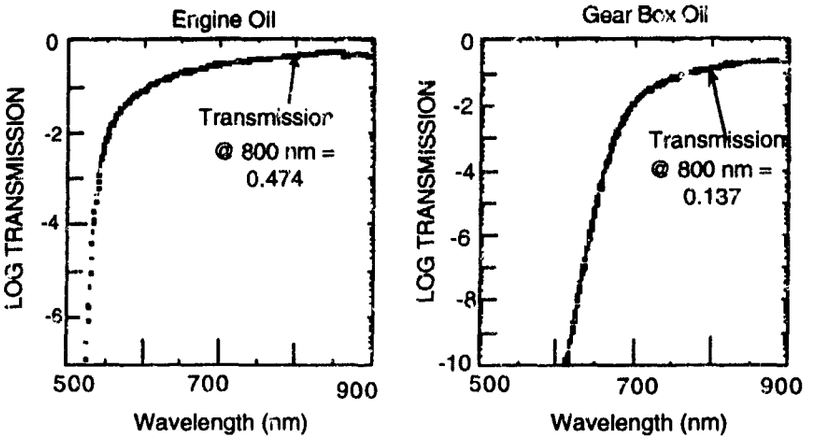


Fig. 2. Optical transmission through a 0.75 in. path as a function of wavelength for engine and gear box oil with 790 operating hours

The operating characteristics of the laser (e. g. pulse duration, repetition rate, illumination area) and those of the optical imaging and detection system are determined by a combination of the expected oil-flow conditions and the desired spatial resolution. Oil flow speeds for various aircraft are expected to be in the range between a few m/sec and 10 m/sec. We have therefore based our preliminary designs on a flow speed of 10 m/sec. The spatial resolution is affected by the illumination area, the number of pixels in the detector array, the flow speed and the need to image particles that are not in a predetermined plane. The optical debris monitor is expected to be able to detect the presence of particles as small as 10-20 μm, and to give size and shape determination for particles larger than about 100 μm. We expect that a reasonably accurate shape identification can be made with about 10 resolution elements on a side, leading to a required resolution of ~10 μm. At 1:1 imaging this requires the detector array to have 1000 10-μm pixels in the direction transverse to the oil flow for a 1 cm field of view. At a flow speed of 10 m/sec, a laser pulse duration of 1 μsec or less is required to freeze the motion to within a single pixel. The dimension of the array in the direction of the oil flow is determined by the need to illuminate the entire flow column, the flow speed and the pulse repetition rate. For example, if a square array of 1000x1000 pixels (1 cm x 1 cm) is used at a flow speed of 10 m/sec, a repetition rate of 1 kHz is needed. If on the other hand an array that is 1 mm in the direction of the oil flow is used, then a repetition rate of 10 kHz is needed. For these illumination areas, currently available diode laser powers between 0.1 and 1 watt are projected to provide ample illumination levels for detection with CCD arrays.

Arbitrary spatial resolution is achievable in principle (down to sizes of the order of 1 wavelength) for particles that are constrained to be in a predetermined focal plane. However, if the debris particles are distributed across the oil flow column, some of them will be out of focus and therefore will produce blurred images. Geometrical blurring at 1:1 imaging restricts the allowable defocusing at the lens apertures required to provide viewing areas of the order of 1 cm. If coherent light, as is produced by a single mode laser, is used the allowable amount of defocusing is increased because of near field diffraction. The outline of out-of-focus particles shows diffraction structure, but retains the original shape for a considerable defocusing distance. For example, we have determined that the shape of 100 μm square particles can be identified at arbitrary positions across a 0.6 in. pipe, and the shape of 50 μm particles can be identified across a pipe about 0.375 in.

Imaging in transmission has been chosen in order to obtain a reliable identification of the shape of the contour of the debris particles without having to restrict their orientation. We have observed that imaging of the particles in reflection (either at 90^{o} or 180^{o}) gives a poor record of the debris size and outline unless the particle is aligned for specular reflection. Imaging in transmission, on the other hand, always gives a reliable record of the major dimensions of the debris particles. Multiple-axis illumination can be used to accommodate chips that are tilted or on edge to give a complete record of the particle shape that is independent of its orientation. In addition, the contrast between the background and the signal associated with the particle is always determined by the one-way transmission of the oil column. As a result the signal strength does not vary with the position of the particle in the oil flow, minimizing variations in the required dynamic range.

It is anticipated that, depending on the specific application, the lubricating oil will generally be mixed with a varying amount of air that is introduced by motion of the aircraft. It is important therefore that the optical oil debris monitor be able to distinguish entrained air from wear debris particles or ingested contaminants. Air in the oil column generally shows up as bubbles of various diameters. Because the index of refraction of the air is less than that of the oil, the bubbles exhibit total external reflection for rays that strike them farther from the center than about 40% of the radius, and behave as a strong negative lens for rays that are closer to the bubble center. For the imaging systems that we have used, most bubbles of the order of 1 mm or less in diameter appear as dark shadows with a round or oval shape. Some transmission is detectable at the center of larger bubbles. Since debris particles from metal bearings and gears are expected to have the form of thin plates with relatively large aspect ratios and irregular contours, they can be distinguished from the air bubbles on the basis of their shape, aspect ratio and contour. Viewing

along multiple axes can be used to confirm the identification of wear debris, as bubbles are expected to be round when viewed from all angles, while wear debris will have a thin dimension in one viewing orientation. Sand particles have been observed to appear as small gray particles. Distinction between these particles and air bubbles can be made on the basis of the gray level of transmission. In addition, multiple axis viewing using both transmission and near 90^{o} reflection can give the air bubbles a characteristic bright pattern that can also make distinction from sand easier.

Imaging experiments: We have investigated several aspects of the optical oil debris monitor on an instrumented oil flow stand using the arrangement shown in Fig. 3. The oil was Mil L 23699 at a pressure of ~ 20 psig. Flow speeds up to 7.5 m/sec were used. Variable amounts of air were injected at a pressure of 30 psig at a point 60 cm upstream of the viewing position. Various types of debris (stamped metal chips,, sand, iron filings, helicopter transmission debris) were injected into the flow at a point 30 cm upstream of the viewing position. The oil column was viewed through a cell with a square flow cross-section about 1 cm across and square windows about 1 cm across. The laser illuminator was a GaAlAs diode that produced 100 mW peak power at a wavelength of 860 nm. The size of the laser beam was adjusted to fill the window of the viewing cell. The pulse duration was varied between 1 and 10 µsec to determine the effect of pulse duration on spatial resolution at high flow speeds. The oil column was 1 cm across in the viewing port to accommodate the age of the oil and the limited power of the laser diode. The oil flow column was imaged with a Nikon f/5 50 mm lens capable of resolving 6 µm. The detection was done with a scientific CCD camera with full frame capability. The detected area was about 480 x 440 pixels and the spatial resolution was about 15-20 µm. Images were recorded on a SVHS VCR that had an equivalent resolution greater than 400 lines.

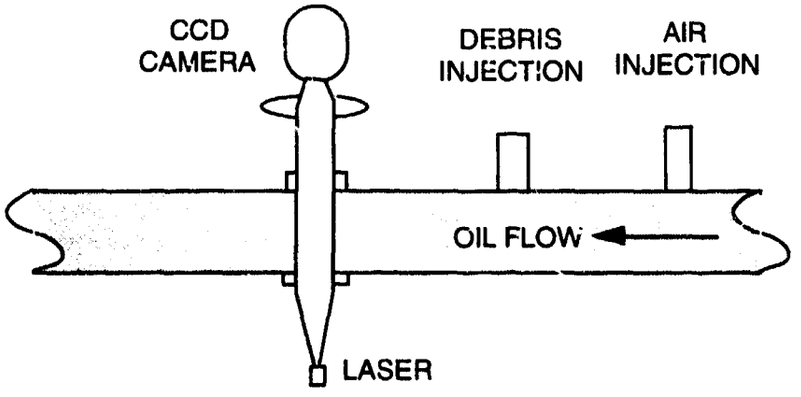


Fig. 3. Schematic diagram of experimental oil flow stand.

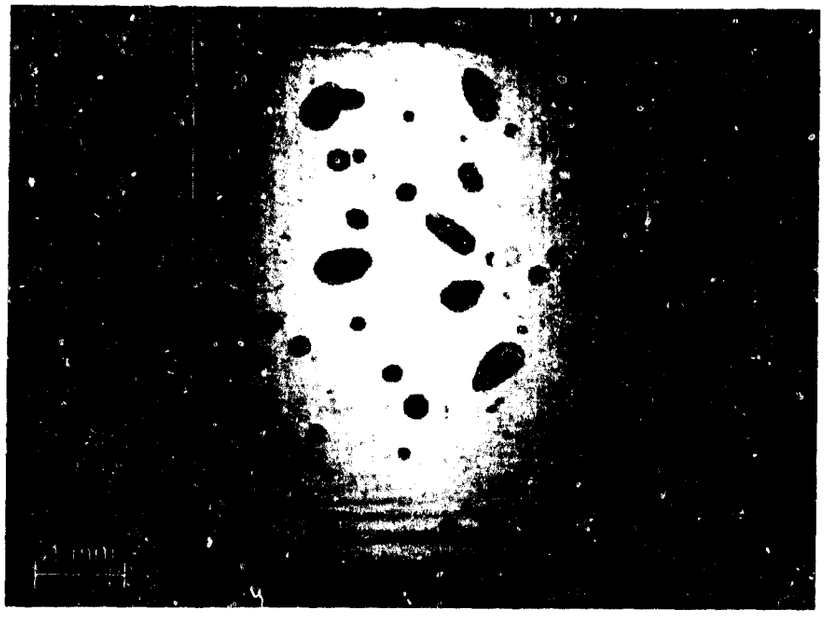


Fig. 4. Image obtained at 5 m/sec flow speed showing air bubbles and three metal chips at bottom.

An example of images obtained with this setup is shown in Fig. 4. This image shows several air bubbles and three metal chips. The metal chips used in this run were stamped from shim stock with a rectangular cross section of 300 x 380 μm. One metal chip is seen at the bottom face on, and two are seen edge on. This picture indicates the typical appearance of debris in the oil column. Chips that are face on are readily distinguished from the surrounding air bubbles on the basis of their shapes. Flakes that are side on are also distinguishable but may require a second viewing angle to identify their size. Other observations using sand, debris from a motor-generator failure, and 40 μm iron filings were made with the setup of Fig. 3 and will be reported at the meeting.

Two axis Illumination: Two-axis illumination can be useful to obtain an additional perspective of the suspended particles, to aid in seeing particles that may be partially obscured by air bubbles, and to provide additional discrimination between air bubbles and solid particulates. We have made observations with a double viewing arrangement as shown in Fig. 5. Here one laser was used to create shadow images of the suspended particles as before. A second laser beam was passed through the cell at 90^{o} to the first beam and was reflected back on

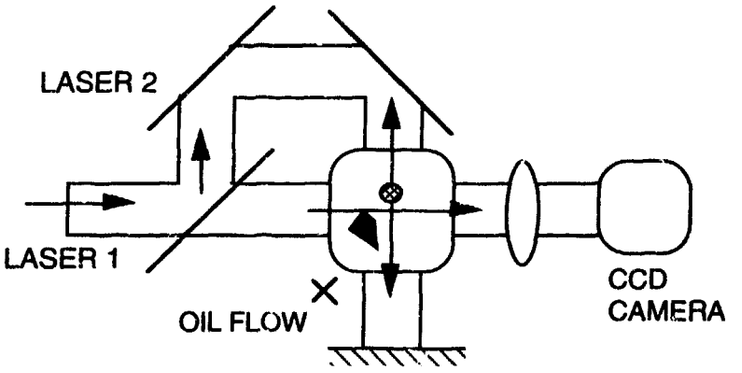


Fig. 5. Schematic diagram of arrangement used for two axis illumination. Light from laser 1 is imaged in transmission, light from laser 2 is imaged in 90^{o} reflection. The direction of oil flow is into the paper.

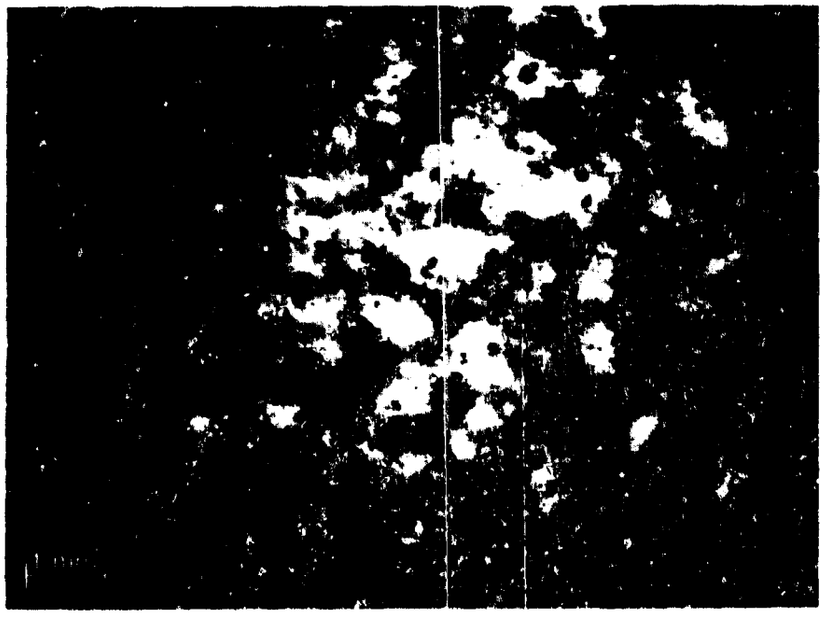


Fig. 6. Image of air bubbles and grit obtained with two-axis illumination

itself. Light reflected from particles near 90^{o} from the second laser beam was also collected and imaged at the detector.

An example of images obtained with this arrangement is shown in Fig. 6. For this picture a combination of air bubbles and coarse grit was used. The air bubbles appear as dark shadows with bright highlights. Larger bubbles are distinguishable on the basis of their round contour as before. The smaller bubbles appear almost completely as bright highlights. The grit appears as small dark shadows that are easily distinguishable from the small bright bubbles. Without the highlights, the grit would be difficult to distinguish from the smaller bubbles.

Image Processing: There are several approaches that are possible for computer processing of the images. In principle, classification of the wear debris based on the particle shape, measurement of particle size, and distinction of wear debris from air bubbles is fairly easy using any of a number of approaches. Some of these are neural networks, Bayesian classifiers or statistical classifiers. One of the major challenges in our system is the high data rate that is required because of the combination of high resolution, large viewing area and high oil flow speeds. Because of the presence of air bubbles, which are distinguishable from the wear debris only on the basis of their shape, it is necessary to examine every frame. Air bubbles are expected to appear more or less constantly, while particles from wear are anticipated to occur only every 10 seconds or less often.

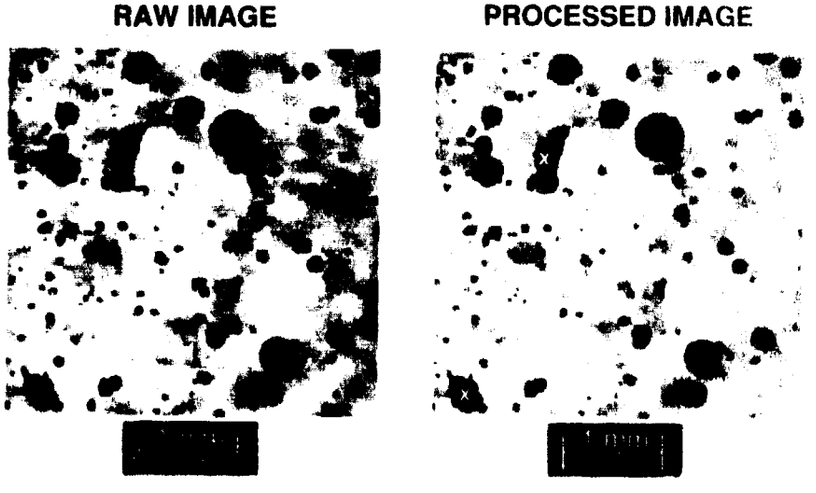


Fig. 7. Unprocessed image of metal chips and air bubbles (left) and processed image (right) showing objects that have been identified as "not round".

One strategy being considered takes advantage of this situation by eliminating air bubbles on the basis of their characteristic round shape in a rapid preprocessing step. Any remaining images of particles can be transferred to a particle classifier for further analysis as to size and shape. The first step, screening out air bubbles, however, must be fast and must be capable of handling the full flow of data. One approach is to use a low resolution convolution to identify the centers of any objects within the field of view and a simple radial spoke comparison to reject all objects that are within a predetermined range of roundness. An example of the use of such an algorithm is shown in Fig. 7. Here the raw image contains both air bubbles and irregular metal chips. In the processed image the centers of the two objects that are not classified as round have been identified. In a complete system, the outlines of these particles would be passed on for further analysis.

References:

1. Albidewi, A., Luxmore, A. R., Roylance, B. J. and Wang, G.; "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.

* 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

WAVELETS FOR HELICOPTER TRANSMISSION FAULT DETECTION AND CLASSIFICATION

Robert R. Tenney
James C. Deckert
Alonso E. Rhenals
ALPHATECH, Inc.
50 Mall Road
Burlington, MA 01803-4562

Abstract: In this effort we demonstrated the technical feasibility of using continuous wavelet transforms to detect and classify failures in a helicopter transmission using single-channel accelerometer data from Navy bench tests. The key technical advance in this effort was the construction of an analyzing wavelet that: 1) provided appropriate resolution in time/frequency space for this class of mechanical systems, 2) allowed features characterizing each fault type to be readily identified, and 3) permitted efficient calculation of feature values from sensor data. Feeding these features into a conventional three-level, feed-forward neural net with deferral (when no clear decision is possible, to a subsequent sample) led to perfect detection, no false alarms, and modest deferral rates (< 4%). We are now extending this approach to more complex systems in realistic environments. The principal issues we are addressing include: 1) use of multichannel amplitude and phase information, 2) feature selection tied to operational mode, and 3) design of a real-time feature extractor employing off-the-shelf electronics technology. These will permit the technique to be applied to transmissions with several dozen moving elements on-board a helicopter with many other vibrating components, operating over a wide range of flight regimes.

Key Words: Condition-based maintenance; detection; diagnosis; gears; helicopters; rotating machinery; signal processing; time-series analysis; transmissions; vibrating systems; wavelets; wavelet transforms

Introduction: To understand the context for this work, consider the basic aspects of any problem of fault detection. First of all, by its very nature, *fault detection involves distinguishing normal behavior in the available data from behavior indicative of a fault.* That is, albeit in varying forms and in varying levels of detail, all fault-detection systems incorporate *models* of what characterizes normal behavior *and* of what characterizes abnormal behavior (although the latter might only be given implicitly through the specification of the allowable range of behavior that should be considered as normal). It is intuitively clear that the better our models are—based either on detailed physical modeling and explicit prior modeling of fault or change modes [1–4] or on adaptive data-based approaches for learning the model [5, 6]—the better job our fault detectors can do.

A second aspect of the fault-detection problem, namely *the presence of uncertainties both in the form of noise and, typically more critical, in the form of uncertainties in the dynamics* of the system under normal and degraded conditions, is discussed in detail in [5, 7]. Specifically, the challenge to the designer is to specify a system that is *maximally sensitive* to certain phenomena, namely faults, and *minimally sensitive* to others, namely unpredictable fluctuations and variations due to noise and model errors. While a number of analytical tools have been developed to deal with this in particular contexts [5, 6], this still remains, in our opinion, the fundamental challenge in any fault-detection problem.

The third distinguishing aspect of fault-detection problems is that *they are fundamentally transient problems*. This is *not* to say that the signals to be detected are necessarily transient events. Rather, the whole point of fault detection is to *detect a change* and to do so in a timely enough fashion so that catastrophic system failure is avoided. Thus not only do we want a fault-detection system to be sensitive to the occurrence of faults, but we also want its transient response to such a fault to be fast, ideally allowing us to isolate potential problems before they become real ones.

Finally, as in any detection problem, there are two principal subfunctions of any fault-detection system: 1) to *focus* the available signals and measurements into *maximally informative statistics*, i.e., into a comparatively small set of features that behave in maximally distinguishable fashions (taking into account noise and uncertainty) under normal and failed conditions; and 2) to *analyze the patterns* in these statistics (perhaps across features and time) in order to detect and, if appropriate, to isolate faults.

Previous Approaches: Over the last 20 years a wide array of fault-detection techniques has been developed and successfully applied. By far the most widely used methods are the generalized likelihood ratio (GLR) and multiple model (MM) methods originally developed in [1] and [2], respectively, and applied successfully to a variety of problems [e.g., 2, 4]. The basic theories behind these methods begin with dynamic state models, under normal and failed conditions, for the system to be monitored, and produce designs that have as their outputs conditional probabilities (or log-likelihoods) that each of the modeled fault conditions has occurred. That is, the GLR and MM focus the available data in order to produce sets of sufficient statistics that make the subsequent steps of optimal detection comparatively simple—e.g., we can compare probabilities to thresholds.

Thus for the problems they address, GLR and MM are *optimal* and, furthermore, they are extremely efficient. Of course, they achieve this optimality by addressing particular classes of problems in which the models, including failure modes, are specified parametrically. While this limitation on the interpretation of GLR and MM as statistically optimal procedures exists, it has not prevented its successful and reliable application in a variety of contexts, such as those mentioned previously, in which model uncertainty is far less structured and well-understood. Furthermore, analytical methods have been developed [5, 6] to produce maximally robust designs for such systems. These considerable successes notwithstanding, however, there still exist fundamental limitations on the domain of applicability of these methods as stand-alone systems for fault detection, necessitating the investigation of more black-box, adaptive methods that begin with far less-structured prior knowledge of the phenomenon being modeled than either GLR or MM assume is available.

In recent years one such methodology has emerged [8–10] that builds directly on black-box, adaptive model-learning methods, such as the celebrated LMS algorithm [11] used in neural network adaptation, by constructing fault-detection systems that monitor the sequence of parameter-adjustment steps made by the adaptive algorithm. Unfortunately, this approach does *not* solve the helicopter transmission fault detection and classification problem. Specifically, the basis for this approach is the detection of subtle changes in the spectral behavior in complex systems. The class of multivariable (to allow for multiple sensors or sensor channels) autoregressive moving-average (ARMA) models represents an extremely rich class of black-box models for estimating spectral characteristics for such systems, and this approach employs powerful mechanisms both for estimating the AR portion in highly nonstationary environments and for detecting changes in it. In some problems, such as offshore platform monitoring, it is reasonable to assume that the model being estimated indeed represents something close to a full, realistic model of the entire system (i.e., only a known, finite number of physically meaningful modes exist, allowing one to identify model order and to explicitly identify different modes with the physical

structure in a simple way). As convincingly discussed in [8, 10], however, in applications such as rotating machines the dynamics are sufficiently complex and uncertain that complete modeling of the entire system is neither feasible nor desirable for fault detection. Rather, one looks for *projections* of the dynamics, i.e., one or more lower-order models of parts of the dynamics that are sufficient for detection and diagnosis of faults. The question, of course, is *where do we get those projections*. In our work we achieved great success using wavelet transforms [12–15] to obtain the projections.

The most attractive characteristics of wavelet transforms are their superior time-frequency resolution capabilities [12, 14, 15]. These are particularly critical characteristics in the present context, since each vibrating component radiates a set of quasi-stationary narrowband signals, often forming a harmonic series. Many external disturbances take the form of impulsive transients that can mask the narrowband elements for short periods of time. Good frequency resolution in the analyzing wavelet allows us to zoom in on critical frequency bands, in order to apply the advanced detection methods described earlier, while rejecting transient disturbances (e.g., data recording/processing artifacts). Good frequency resolution also allows us to track the movement of these critical frequency bands as operating conditions (e.g., drive shaft rotation rate) change.

Thus, the objectives of our effort were: 1) to assess the efficacy of wavelet techniques to select features upon which simple and reliable classifiers could base decisions regarding abnormal changes in system behavior, and 2) to develop and test an algorithm for robust and reliable detection of failures in vibrating systems such as helicopter transmissions.

The Continuous Wavelet Transform: Wavelets offer several different ways to access the structure of a signal in time and frequency. In this project, we found that the *continuous wavelet transform* (CWT) [16] provided an ideal tool to identify significant features for fault detection in vibrating systems such as transmissions, and to provide the basis for extracting those features. Like other image-visualization techniques such as the short-term discrete Fourier transform (DFT), the CWT converts a one-dimensional signal into a two-dimensional image. The CWT eliminates windowing artifacts, however, and provides more flexibility to trade time resolution for frequency resolution. In the CWT each horizontal line of the image corresponds to the time-series response of one filter in a constant-Q (relative bandwidth) filter bank driven by the observed sensor data. This filter bank provides complete coverage of time and frequency behavior. One distinct advantage of the CWT is that it actually represents entire families of transforms—different transforms result from using different analyzing wavelets as the basis of the filter bank design. Here we selected the Kiang wavelet (named after Nelson Kiang of MIT, who derived tuning curves for auditory nerve neurons in the middle 1960s), illustrated in Fig. 1, as the analyzing wavelet due to its superior ability to localize events in frequency.

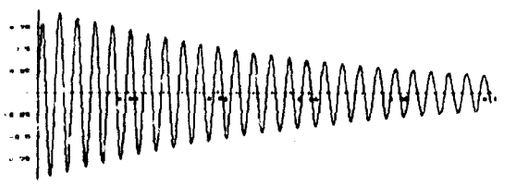


Figure 1. The Kiang Wavelet

Each line corresponding to a single frequency, across time, of a CWT image is constructed by applying the signal to a filter corresponding to the Kiang wavelet, and computing the magnitude of the (complex) output. Because the impulse response of each filter is a wavelet, each filter has a bandpass transfer function. Because the filters differ

only by a scale factor applied to their impulse response (wavelet dilation), higher-frequency filters have proportionally higher bandwidths [17].

A subset of test data originally obtained by Mark Hollins of NATC was provided by NCCOSC/NRaD. The data were for two accelerometers (we used only one) attached to studs near the output end of the intermediate transmission of a TH-1L helicopter, and a shaft tachometer signal (which we did not use) was also provided. The transmission is quite simple, as shown in Fig. 2. Under normal operations, such as the bench tests that generated these data, the 72 Hz shaft rotation rate causes the 27 gear teeth to mesh at just under 2 kHz. The CWT image shows a strong line at this frequency, as well as at integral multiples of this fundamental (although the fourth and sixth harmonics are weak). The CWT image also shows several other horizontal lines, notably one near 1 kHz, that correspond to other vibrating elements such as the bearings in the shaft-support assemblies. Figure 3 shows a frequency slice at one time instant of the CWT for a normal transmission. After examining many representative normal images, we inferred that *robust* features should: 1) persist over long periods of time (> 500 msec), and 2) be quite narrow in frequency (< 0.02 octave) at places in the spectrum where vibrations yield high energy.

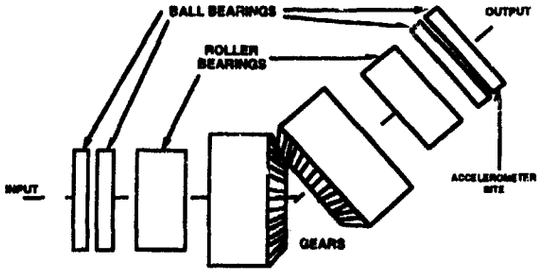


Figure 2. Illustration of Intermediate Transmission for the TH-1L Helicopter

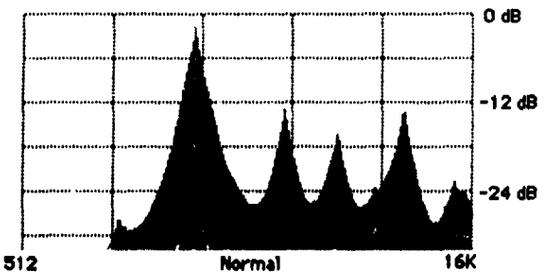


Figure 3. Frequency Slice of CWT at One Time Instant for a Normal Transmission.

We extensively evaluated both the CWT and another wavelet technique, the wave packet transform (WPT) [18, 19], as the basis for selection and extraction of features indicative of failures. Our conclusion was that for vibrating systems, *the CWT provides an extremely useful visualization of the structure of the sensor signals*, whereas computationally realistic versions of the WPT could not provide robust features. (The WPT employs a tree of finite impulse response (FIR) filters. For the same frequency resolution as the CWT, these impulse responses would have to be roughly 1000 points long—far longer than the Daubechies 20 wavelet used in [19]—and thus each filter requires 1000 multiplications per sample. To achieve 0.02 octave resolution around the 2 kHz mesh fundamental, we would need 1024 of these filters. Moreover, the WPT features could not easily be made robust to variations in shaft rate. Nonetheless, we do believe that the WPT may provide very good features for incipient failure detection in systems characterized by short transients during normal operation, such as switches.

By selecting features from high-energy regions of the time/frequency space, as revealed by the CWT, we can obtain *robust* features. By comparing CWT images from varying fault conditions, we can identify a comparatively small number of features that are *maximally informative*.

Fault Conditions: The test data included five fault conditions (bearing inner race, outer race, and rolling element; and gear spall and 1/2 tooth cut) in addition to the normal, unfailed case. These fault conditions had been realized by disassembling the transmission, inserting a specific (non-catastrophic) fault, reassembling the transmission, and then collecting data.

One might assume that classical spectral analysis would provide insight into the spectral locations of maximally informative features. If one thinks of the CWT as a function of two variables, time and frequency, then these spectra represent slices through that function at fixed points in time. CWT spectra differ from classical DFT spectra in both qualitative and quantitative ways. Qualitatively, the constant-Q filter banks provide finer frequency resolution at low frequencies, and coarser resolution at high frequencies, than the constant-bandwidth filters implicit in the DFT. Quantitatively, the wavelet spectra appear on a logarithmic frequency scale.

We first examined the shapes of the CWT spectra for the six fault classes at several time samples, and then examined multiple time segments of CWT images (with color denoting energy). These examinations indicated that significant differences in the CWT images for the six fault classes were fairly stable in time.

Recall that we want fault classification features to be robust. In vibrational data, robustness comes from both temporal persistence and high energy. Of course, "high energy" is a relative term. For example, the subharmonic peak near 1 kHz contains very little energy relative to the harmonics of the mesh frequency, but is quite prominent and obviously useful in discriminating failure cases from the normal case. Therefore, we interpret high energy to be relative to the average energy level nearby in time/frequency space. One way to highlight high-energy features is to estimate this average local energy level, and only consider parts of the CWT image that exceed it. Thus, we examined the CWT images with low-energy regions masked out via a simple morphological filter applied across frequency, masking areas that fall below an estimated noise floor. The technique provided additional highlighting of the significant differences in structure around the third harmonic, and also of the 1 kHz line.

At this stage, then, our CWT analysis of the accelerometer signal in normal and failed operations indicated regions in time/frequency space containing robust features.

Feature Selection: The masked CWT image provides clear insight into the structure of a time signal. It is also extremely computationally intensive to compute. The CWT images used 480 filters, each of which required 10 multiplies per sample. At the 48 kHz sample rate of the helicopter data, real-time display of the masked CWT would require a throughput of about 1/4 billion multiplies per second. While the inherent parallelism of the filter bank structure probably allows one to design special-purpose processors to realize this throughput, our methodology does not need this. Rather, it uses the full masked CWT analysis to select (small) portions of time/frequency space that must be represented as a feature set for a failure detector/classifier. Thus we distinguish the *design phase*, which identifies a set of features on which classification may be based, from the *implementation phase*, in which features extracted from data are presented to a classifier in real time.

Previous attempts to perform adaptive classification on these data used raw DFT energy levels as the feature sets [20, 21]. This is a natural thing to do if one has a capable adaptive pattern recognizer. By "capable" we mean that it can deal with highly non-convex, non-linearly-separable feature sets. Why? Because it turns out that feature clusters formed from raw DFT amplitudes are not convex even in simple cases where one wishes to discriminate an unstable narrowband signal from noise—very much the situation encountered in the region of the third harmonic of the mesh frequency, which forms a clear peak in the normal case but is just part of a broader energy distribution in the other cases. The problem is that variations in the center frequency of the narrowband signal may cause it to move from one DFT bin to another—a situation that becomes exacerbated when large DFTs are used to provide fine frequency resolution. This movement sketches out a region in the vector space of DFT magnitudes that is not convex, and that cannot be separated from other feature regions by simple hyperplanes. Thus a very sophisticated algorithm is needed to construct non-linear separation surfaces between non-convex feature clusters. As most such algorithms are variations of a gradient search (such as many neural net training algorithms), one must take care to avoid locally satisfactory solutions and lack of global convergence.

Because the masked CWT image provides more structural insight to a design engineer than does a large vector of DFT amplitudes, one can formulate alternative feature sets that may offer simplifications. Here the masked CWT images immediately suggested eight frequency regions that would provide maximum information—the lowest six harmonics of the mesh frequency, the sub-harmonic line at about 1050 Hz, and a region at about 2.7 times the mesh frequency (note that the ratios of the diameters of the inner races, bearings, and outer races are not integers, so mechanical systems often give rise to non-integral ratios between spectral lines). To provide robustness against variations in recording amplitude, we normalized the energies in these 8 regions to the energy in the mesh fundamental—i.e., we expressed the other 7 energies in dB relative to the mesh fundamental energy. To provide robustness against variations in shaft rotation rate, we defined all of these frequency regions as multiples of the mesh fundamental. To provide some features of the wavelet spectra obvious to the eye but not captured by raw power levels, we also computed the curvature of the spectrum at each of these locations (recall that curvature around the third harmonic appeared to provide discrimination information for both inner-race and outer-race faults). Calculation of curvature of the fundamental also allowed us to track variations in its frequency.

Thus the masked CWT suggested 15 features of potential use for classification—7 normalized power levels and 8 spectral curvatures. Note that these features represent only a small fraction of the available frequency range. Note also that this is substantially less than the 256 to 16,384 features used in the previous DFT approaches. But how did we know whether or not these 15 features were adequate for fault classification?

Before investing the effort required to train, tune, and evaluate an adaptive classifier, we computed values for these features for representative 500 msec segments of the six data sets. Because the filters required to evaluate the CWT at these frequencies have time constants on the order of 5 msec, we could compute feature values every 10 msec with some assurance that they would be statistically independent—at least with respect to correlations introduced in the feature-extraction process. Thus each 500 msec segment yielded 50 feature vectors of dimension 15. The 50 vectors can conceptually be plotted as six clusters in 15-dimensional space, and the projection of these clusters onto two-dimensional subspaces show the shape and separation of these clusters. We computed the Fisher separation coefficient between each pair of clusters, and for each feature, as:

$$|m_1 - m_2| / \sqrt{(\sigma_1^2 + \sigma_2^2)}$$

where m_1 and m_2 are the sample means of the feature value for the two cases, and σ_1 and σ_2 the sample standard deviations. Intuitively, this statistic captures the separation between two clusters along one axis of the feature space, measured in units of standard deviations. Thus a separation of three or more is virtually perfect. Not all features will separate all pairs of clusters; one would like, however, for there to be at least one feature that does separate each pair. Therefore, we selected the feature for each cluster pair that maximized this statistic, and present the results in Table 1. Note that the only cluster pair not virtually perfectly separable consists of the bearing rolling-element/outer-race fault, both of which were inserted in a bearing assembly near the input side of the transmission, and hence relatively far from the accelerometer supplying the signal. These statistics demonstrate that the features derived from the masked CWT should be adequate for fault detection and identification. In fact, only 8 or 9 of the 15 features contribute significantly to this table, suggesting that we could reduce the size of the feature set by almost half without sacrificing performance. We chose not to do so simply because 15 is already a small number of features for off-the-shelf adaptive classification algorithms to use.

TABLE 1. CLUSTER SEPARATIONS, TH-1L DATA

	ND	IR	RE	OR	SP	TC
Normal	0.00	5.67	10.79	9.79	11.14	11.28
Inner Race	5.67	0.00	4.13	5.25	13.31	6.05
Rolling Element	10.79	4.13	0.00	2.07	5.89	3.45
Outer Race	9.79	5.25	2.07	0.00	7.18	3.29
Gear Spall	11.14	13.31	5.89	7.18	0.00	8.61
Tooth Cut	11.28	6.05	3.45	3.29	8.61	0.00

Before discussing our approach to adaptive classification, we make one additional point. The evaluation of feature-cluster topology not only indicates the separation of feature clusters, and hence suggests the performance to be expected of a classifier, but also allows a design engineer to validate the feature set. Specifically, recall that the TH-1L data came from a single transmission into which faults had been sequentially inserted. How does one know that differences between the cases are due to the fault, and not some artifact of the insertion process? After all, if bolts are tightened in different orders, or lubricant levels are different, the transmission path between the vibrating elements and the accelerometer will be affected. One needs to select features for classification based on not only their statistical separation, but also the physics of the fault mechanism. The risk in not doing this is that data presented to the classifier may contain variations upon which classification may be based, but that bear no causal relationship to fault mechanisms.

Fault Detection and Classification: Having selected a feature set that clearly separates the fault cases from the normal case and from each other, we turned to the problem of numerically evaluating the boundaries between the classification regions centered on each cluster. Because it was handy, we selected a commercial artificial neural network (ANN) software package to deduce these boundaries directly from the feature vector samples. Because we had 15 features, we set up a neural net with 15 input neurons. Because there were 6 classes (unfailed plus five fault types) to discriminate, we set up 6 output neurons. Because the statistical analysis assured us that the clusters were convex and linearly separable, we set up one hidden layer neuron for each output (recall that hidden-layer neurons provide ANNs with the ability to construct non-convex classification regions) [22, 23]. Thus our nets had a total of 27 processing elements (PEs).

Convergence to the specified RMS error of the difference between the desired and the actual outputs occurred relatively fast—after between 5,000 and 10,000 random presentations of the feature vectors included in the training set. *Given the simplicity of the ANNs we used, their small size, and the excellent feature-cluster separation made possible by the judicious utilization of the CWT, no sophisticated training algorithms were required.*

From the test set results, we computed the following measures of effectiveness for each vibrating system: probability of false alarm, probability of missed detection, probability of misclassification, and probability of deferral. *Probability of false alarm* is the probability that a fault is announced when there is no fault present. *Probability of missed detection* is the probability that no fault is announced when there is a fault present. *Probability of misclassification* is the probability that a fault type is announced when a different fault type is present. *Probability of deferral* is the probability that the classifier defers a decision when a case for decision (a feature vector) is presented to it.

For the purpose of this work, a feature vector leads to an ambiguous situation when the ratio between the maximum output value and the next larger output value for a given feature vector is less than the *acceptance threshold.* In such a situation the classifier refuses to announce a decision and considers the next feature vector. The consequence of this deferral is to decrease the probabilities of false alarm and missed detections, and to increase the time delay for a classifier decision—the ANN knows that it is confused, and prefers to wait a bit instead of risking a mistake. For instance, feature vectors for the transmission system are computed every 10 msec, so the price paid in time delay for each deferral is a 10 msec delay in the time to detect a fault.

Performance Results: Given the preceding insight into the derivation of high-energy wavelet features and the convex, separable clusters they form in feature space, it should be no surprise that good classification results are possible. *A feature set that captures the important discriminants between normal operation and faults vastly simplifies the problem of designing an adaptive classifier that achieves good performance.*

Table 2 presents the performance results for the helicopter transmission, and two sets of shipboard pump data we analyzed. The complexity is the total number of PEs in the corresponding ANN. For the transmission and condensate pump systems, the test set was independent from the training set; for the fire pump data these two sets were the same.

TABLE 2. PERFORMANCE RESULTS

	TRANS-MISSION	CONDENSATE PUMP	FIRE PUMP
TRAINING SET SIZE	1125	240	480
TEST SET SIZE	6750	1400	4800
ACCEPTANCE THRESHOLD	1.4	1.2	2.0
PROBABILITY OF FALSE ALARM	0.000	0.000	0.000
PROBABILITY OF MISSED DETECTION	0.000	0.000	0.000
PROBABILITY OF DEFERRAL	0.035	0.020	0.020
PROBABILITY OF MISCLASSIFICATION	0.046	0.000	0.000
COMPLEXITY	27 PEs	28 PEs	45 PEs

The performance results in Table 2 clearly show that *the wavelet feature sets selected above permit perfect detection performance with very low deferral rates.* While these results are pleasing, we feel that an even more important principle has been demonstrated. *We used exactly the same method to find features for the pumps as we used for the transmission data.* There was no trial and error for the pump classifiers—these results are from the very first feature sets we picked. This offers limited but important evidence that our results are not accidental—that *we have a methodology to analyze data from vibrating systems and derive small, focused feature sets that support high-confidence fault detection and classification.*

Conclusions: We have developed an extremely flexible and powerful methodology exploiting the power of wavelet techniques to detect failures in vibrating systems. The essential elements of this methodology are: 1) an off-line set of techniques to identify high-energy, statistically significant features in the CWT; 2) a wavelet-based prepro-

cessor to extract the most useful features from the sensor signals; and 3) simple ANNs (incorporating a decision-deferral mechanism to defer any decision if the current feature sample is determined to be ambiguous) for the subsequent classification task. In the transmission and pump data sets used in this study, the algorithms designed using this method achieved *perfect detection performance* (1.000 detection probability, and 0.000 false alarm probability), with a probability < .04 that a decision would be deferred for a few milliseconds—again based on only 500 milliseconds of training data from each sample case. Based on this effort we have developed a set of guidelines for finding robust feature sets in CWT images. These guidelines can be summarized as follows:

- Features should be *robust to external disturbances:* we seek high-energy-content features to be derived from morphological filtering on the frequency axis of the CWT, with narrow bandwidths to reduce their sensitivity to impulsive disturbances. Features must also be redundant to exploit the correlation among features, and they must be computed as frequently as permitted by the largest time constant in the preprocessor.
- Features should be *diverse:* it is desired to include features across a wide range of frequencies—for example, the first six harmonics of important narrowband vibrations (transmission) or octave samples of broadband components (pumps). In addition it is desired to include one or more low-energy features to support disturbance rejection.
- Features should *distinguish normal from abnormal conditions:* for this we need to compute statistics (mean, standard deviation) on each CWT bin and look for significant differences that will lead to features with high discriminating power.

While the full CWT may represent a significant computational requirement, it is only used in the off-line *design phase* to identify critical features. The final implementation of a fault-detection system consists of a comparatively simple wavelet-based preprocessor, followed by a very simple ANN—a configuration ideally suited to real-time implementation using either digital *or* analog hardware.

The ability to visually identify critical features during the design phase leads to a second, even more important consequence in the *implementation* phase: *it makes the job of the adaptive classifier far easier.* Because so much of the signal structure is obvious, and because the need for robust features leads one to focus only on high-energy areas of the CWT images, the number of features needed for classification can be astonishingly small. In comparison to feature sets based on the DFT [20, 21], which may have from 256 to 16,384 elements in a feature vector for the faulty systems examined here, we found that fewer than 20 features sufficed to obtain reliable separation among classes for the transmission and pump data available to this effort. Small feature sets lead to extremely simple three-layer ANN classifiers—on the order of 50 processing elements—that could then be designed with noteworthy ease and without the need for exotic training algorithms. Also, the amount of data needed to train such simple nets is quite small—we were able to train classifiers using only 500 msec of data from each sample case.

This effort showed the considerable promise of our wavelet-based method for failure detection in vibrating systems [24, 25]. However, more demanding applications, such as machinery (transmissions) mounted on moving platforms (helicopters) that have other sources of high-energy vibration (engines), raise some additional technical issues—including use of multichannel phase information, feature selection tied to operational mode, and design of a real-time feature extractor—that we are now investigating.

References:

1. Willsky, A.S. and H.L. Jones, "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems," *IEEE Trans. Aut. Control*, Vol. AC-21, February 1986, 108-112.

2. Willsky, A.S., J.J. Deyst, and B.S. Crawford, "Two Self-Test Methods Applied to an Inertial System Problem," *J. Spacecraft and Rockets*, 12, July 1975, 434-437.
3. Willsky, A.S., "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, November 1976, 601-611.
4. Willsky, A.S., E.Y. Chow, S.B. Gershwin, C.S. Greene, P.K. Houpt, and A.L. Kurkjian, "Dynamic Model-Based Techniques for the Detection of Incidents on Freeways," *IEEE Trans. Aut. Control*, AC-25, 3, June 1980, 347-360.
5. Chow, E.Y. and A.S. Willsky, "Analytical Redundancy and the Design of Robust Failure Detection Systems," *IEEE Trans. Aut. Control*, AC-29, July 1984, 603-604.
6. Lou, X.-C., A.S. Willsky, and G.C. Verghese, "Optimally Robust Redundancy Relations for Failure Detection in Uncertain Systems," *Automatica*, 22, 1986, 333-344.
7. Willsky, A.S., "Detection of Abrupt Changes in Dynamic Systems," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, M. Basseville and A. Benveniste, Eds., Springer-Verlag Lec. Notes in Control and Information Sciences, 77, 1986.
8. Moustakides, G., M. Basseville, A. Benveniste, and G. Le Vey, "Diagnosing Mechanical Changes in Vibrating Systems," IRISA (Institut de Researche en Informatique et Systèmes Aléatoires) Report 436, Rennes, France, October 1988.
9. Basseville, M., "Detecting Changes in Signals and Systems—A Survey," *Automatica*, 24, 1988, 309-326.
10. Le Vey, G., "Analyse Modale et Surveillance Vibratoire des Machines Tournantes," PhD Thesis, University of Rennes, France, 1988.
11. Widrow, B., R. Winter, and R. Baxter, "Layered Neural Nets for Pattern Recognition," *IEEE Trans. ASSP*, 36, 1988, 1109-1118.
12. Daubechies, I., "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE Trans. Inform. Theory*, 36, 1990, 961-1005
13. Mallat, S.G., "Multifrequency Channel Decompositions of Images and Wavelet Models," *IEEE Trans. ASSP*, 37, December 1989, 2091-2110.
14. Ruskai, M.B., G. Beylkin, et al., (Eds.), *Wavelets and Their Applications*, Jones and Bartlett Publishers, Boston, 1992.
15. Mallat, S.G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. PAMI*, 11, July 1989, 674-693.
16. Grossman, A., R. Kronland-Martinet, and J. Morlet, "Reading and Understanding Continuous Wavelet Transforms," in *Wavelets, Time-Frequency Methods and Phase Space*, J.M. Combes, A. Grossman, and Ph. Tchamitchian (Eds.). Springer-Verlag, 1989.
17. Vetterli, M. and C. Herley, "Wavelets and Filter Banks: Relationships and New Results," *Proc. ICASSP*, Albuquerque, NM, 1990.
18. Coifman, R.R., Y. Meyer, S. Quake, and M.V. Wickerhauser, "Signal Processing and Compression with Wave Packets," preprint, April 1990.
19. Learned, R.E, "Wavelet Packet Based Transient Signal Classification," SM Thesis, MIT Dept. EECS, September 1992.
20. Kuczewski, R. and D.R. Eames, "Helicopter Fault Detection and Classification with Neural Networks," *Proc. Int'l. Joint Conf. on Neural Networks*, June 1992.
21. Dellomo, M.R., "Helicopter Transmission Fault Detection: A Neural Network Based Approach," The MITRE Corporation, preprint, September 1991
22. Lau, C. and B. Widrow, Special Issue on Neural Networks, I: Theory and Modeling, *Proc. IEEE*, September 1990.
23. Lau, C. and B. Widrow, Special Issue on Neural Networks, II: Analysis, Techniques, and Applications, *Proc. IEEE*, October 1990.
24. Tenney, R.R, A.E. Rhenals, J.C. Deckert, and A.S. Willsky, "Incipient Failure Detection Using Wavelets," ALPHATECH TR-565, December 1992.
25. Deckert, J.C., A.E. Rhenals, R.R. Tenney, and A.S. Willsky, "Wavelet Features for Failure Detection in Vibrating Systems," ALPHATECH TR-567-1, January 1993.

Acknowledgments: This work was supported by the Office of Naval Research under contract N00014-92-C-0040 and by the Defense Advanced Research Projects Agency under contract DAAH01-92-C-R300.

SETTLEMENT OF ALARM THRESHOLDS IN VIBRATION MONITORING FOR ROTATING MACHINERY

C.S. Liu

Chung-Cheng Institute of Technology
Ta-Shih,Tao-Yuan
Taiwan, R.O.C.

Y.S. Shin and J.J. Jeon

Department of Mechanical Engineering
Naval Postgraduate School
Monterey, California 93943

Abstract: The setting of alarm levels plays a vital role in a machinery condition monitoring and diagnostic system. Two approaches to set vibration alarm levels using vibration signals produced by shipboard fire pumps are presented for the time and frequency domains. In the time domain, cross peak analysis (CPA) is proposed to extract the dominate peak points. The distribution of these cross peak points is found to have a lognormal distribution and can be normalized to a Normal distribution in the VdB domain. The computed $\mu+2\sigma$ value in the VdB domain is the suggested alarm level. In the frequency domain, 1/1 octave band analysis (OBA) is introduced. Three artificial fault simulations were conducted to compare 1/1 OBA with the broadband method. The results show that 1/1 OBA is more sensitive to changes in VdB level than the broadband method.

Key Words: Alarm level; Condition monitoring diagnostics; Cross peak analysis; Fire pump; 1/1 octave band analysis; Vibration.

Introduction: In recent years, the rising cost of machine maintenance has driven engineers to develop more economical and efficient methods to determine machine health and accordingly plan the required preventive and corrective maintenance. The most popular technique in use today is a predictive maintenance program based on condition monitoring. In naval applications, condition monitoring is commonly achieved utilizing vibration measurement and analysis on-board surface ships and submarines [1-7].

This paper focuses on the use of vibration measurement to monitor machine health and to diagnose system problems which could lead to machine failure. In addition to providing accurate and understandable data on the machine's current condition, a monitoring and diagnostic system must also limit the number of false alarms. Alarm threshold setting is therefore vitally important in machine vibration diagnostics. Alarm thresholds set too high may result in premature machine failure caused by an undetected failure condition. Alarm thresholds set too low may result in frequent false alarms causing unnecessary system interruptions and repairs. False alarms also reduce operator confidence in the monitoring and diagnostic system. Because the optimum setting of alarm thresholds in vibration monitoring and diagnostic systems continues to be problematical, the goal of this paper is to establish alternative ways of setting vibration alarm levels by using time domain and frequency domain data.

Background Data: The fire pump on a naval ship provides more than fire fighting water to damage control systems. The seawater provided by fire pumps is also used by vital air conditioning and chill water systems, main drainage systems and decontamination sprinkler systems. At least one fire pump is running whenever a ship is underway or at anchor. In this paper, twenty fire pumps from three classes of ships are grouped into two types in Table 1. The vibration velocity signals were measured by means of transducers strategically placed at pickup locations on the pumps. Schematic layouts of fire pump and pickup locations are depicted in Figure 1 and Figure 2. A list of abbreviation for pickup locations is shown in Table 2. The transducer pickup placement method is uniaxial with one radial pickup at each bearing and one axial pickup at the thrust bearing. Table 3 and Table 4 summarize the vibration source components. These tables identify exciting components within each machine and lists the vibration frequencies generated by each component. The vibration frequencies were normalized as multiples of the machine's shaft rotation rate (orders). One(1) order is equivalent to rotational speed of shaft (i.e., operation RPM or Hz). The time waveforms are recorded on magnetic tape by a frequency modulated recorder. Each location on the pumps was recorded for a one minute time series record.

Table 1: CLASSIFICATION OF FIRE PUMP DATA

Type	No. of Fire Pumps	No. of Pickup Locations	No. of Data Sets
I	12	5	60
II	8	5	40

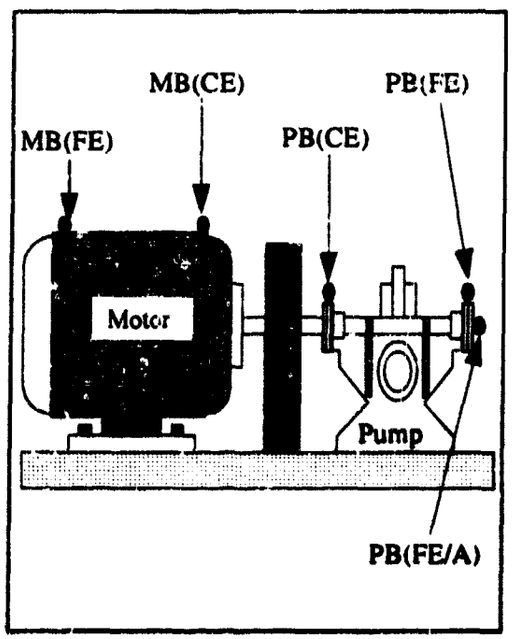


Figure 1 Schematic Layout of Type I Fire Pump

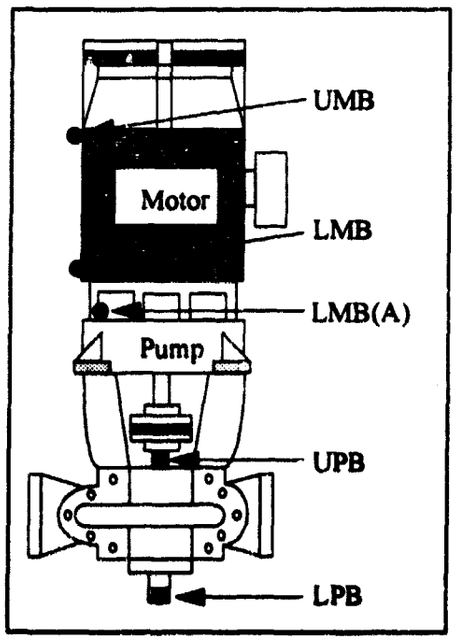


Figure 2 Schematic Layout of Type II Fire Pump

Table 2: ABBREVIATIONS OF PICKUP LOCATION

Abbreviation	Description
MB(FE)	Motor Bearing (Free End)
MB(CE)	Motor Bearing (Coupling End)
PB(FE)	Pump Bearing (Free End)
PB(CE)	Pump Bearing (Coupling End)
PB(CE/A)	Pump Bearing (Coupling End/Axial direction)
UMB	Upper Motor Bearing
LMB	Lower Motor Bearing
LMB(A)	Lower Motor Bearing (Axial direction)
UPB	Upper Pump Bearing
LPB	Lower Pump Bearing

Table 3: Vibration Source Components of Type I Fire Pump

Driver (Motor)			Driven (Pump)		
Description	Element	Order	Description	Element	Order
Motor Shaft (Ref.)		1	Pump Shaft		1
Fan Blading	5	5	Impeller Vanes	6	6
Slots	54	54	Bearing	FAG WT	
Bars	44	44	1 order = 3,595 RPM = 59.9 Hz		
Poles	2	2			
Bearing	MRC 310 & 311				

Table 4: Vibration Source Components of Type II Fire Pump

Driver (Motor)			Driven (Pump)		
Description	Element	Order	Description	Element	Order
Motor Shaft (Ref.)		1	Pump Shaft		1
Poles	2	2	Impeller Vanes	5	5
Bearing	Ball Bearing		Bearing	SKF 6307	

Alarm Level Review: Generally, the vibration of a system can be characterized by a reduced data set in various domains. The criteria to set the alarm level can be considered either in the time domain or the frequency domain as discussed below.

• **Time Domain Criteria**

-**Vibration Severity Criterion Method:** The simplest time domain method is the vibration severity criterion. The root mean square (RMS) value of vibration velocity is usually measured and compared with vibration severity charts [8]. Various companies and national standard organizations published standards for judging vibration severity. For example, International Standards Organization (ISO) standards 2372 and 3945 provide severity guidelines for machinery. This method can be only used for specific types of machinery at a standard operating condition and is insensitive during the early stages of damage [9].

-**Amplitude Probability Criterion Method:** A More sophisticated method was developed by using statistical analysis to examine the distribution of vibration amplitudes. The amplitudes used can be either peak-to-peak, peak or RMS readings of displacement, velocity or acceleration. Both Campel [10] and Murphy [11] use this statistical method to establish alarm levels based on the mean of the reading plus 3 standard deviations. The main drawback of this approach is that it assumes a Normal distribution of the linear readings exists. Vibration readings have a "skewed" rather than "Normal" distribution [12].

• **Frequency Domain Criteria**

-**Broadband Criterion Method:** The broadband criterion method utilizes a vibrometer which can add all the energy dissipated over a wide frequency range (typically 10 to 10,000 Hz). The overall energy is normally calculated by applying the RMS summation method to the spectrum. If the overall energy level exceeds a predetermined level, then an alarm is triggered. This is a simple but inadequate method because it is insensitive to small changes associated with bearing defects, electrical defects and gear tooth defects.

-**Octave Band Criterion Method:** The octave band criterion method is often used in acoustics to determine the energy level changes due to noise and vibration. This method utilizes a constant percentage bandwidth to divide the frequency range of interest into several bands which provide more detailed information than the broadband presentation. A commonly used bandwidth is the one-third octave band. Early researchers used this method to check the change in each band level to determine if the amplitudes exceeded normal values [13-14].

-**Narrowband Criterion Method:** Since the broadband and the octave band criteria methods lack detailed vibration information, a narrowband criterion method is gaining popularity. The bandwidth may up to 10% of the center frequency range (typically 10 to 1,000 Hz). The improved resolution is generally up to 400 or 800 lines over the frequency range of interest. With knowledge of the resonant frequencies of rotating components, the narrowband data is a very useful for diagnosing specific faults.

Data Acquisition And Processing System: A block diagram of the data acquisition and analysis system used is depicted in Figure 3. The tape recorder plays back the machinery vibration data tapes to generate an analog signal, which is then distributed to the data acquisition system and the oscilloscope simultaneously. The oscilloscope controls the quality of the data by monitoring the signal time waveform. EASYEST LX software, developed by Keithley Asyst, was used as the data acquisition system. All of the data

sets were sampled at a sampling frequency of 10 kHz. The sampling duration for each data set was 36.684 seconds. Finally, a 486 PC was used to retrieve these data sequences and perform data processing, with output sent to a laser printer.

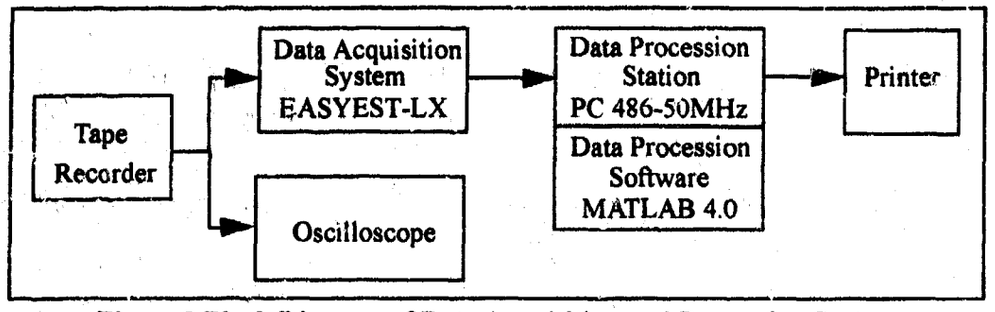


Figure 3 Block Diagram of Data Acquisition and Processing System

Time Domain Approach - Cross Peak Analysis: It is obvious that the peak envelope distribution of vibration time domain signals is relevant to the vibration severity of machinery components. For rotational machinery vibration, since we are interested in the peak envelope distribution of the vibration signal, we use the cross peak data points instead of overall peak data points. The CPA can extract the dominant component from a complicated time waveform data set by selecting the maximum peak value between every zero crossing. Figure 4 compares the difference between sampling points and cross peak points. Using measured RMS values to determine the severity of damage to a component without considering the dispersion of the signal is truly a rough guess. It is better to represent the severity of component damage in terms of a percentage acceptance level. The percentage acceptance level gives the percentage of outcomes which will not exceed this level threshold. For a Normal distribution, the acceptance level is closely related to mean (μ) and standard deviation (σ). For example, $\mu+1.96\sigma$ corresponds to a 97.5% acceptance level which means only a 2.5% probability of the signal amplitude exceeding the $\mu+1.96\sigma$ value. At this point, the statistical moment and central moment of a random variable should be discussed [15]. For the discrete random variable x with probability $Pr(x)$, the nth moment $E(x^n)$ is defined as:

$$E(x^n) = \sum_{k=1}^{\infty} x_k^n \; Pr(x_k) \qquad (1)$$

The first moment is very useful and is given the name mean value (μ):

$$\mu_x = E(x) = \sum_{k=1}^{\infty} x_k \; Pr(x_k) \qquad (2)$$

Of greater significance are the nth central moments which are defined as follows:

$$E\left[(x-\mu_x)^n \right] = \sum_{k=1}^{\infty} (x_k-\mu_x)^n \; Pr(x_k) \qquad (3)$$

The central moment for n=2 is called the variance of the random variable. The standard deviation, which corresponds to the dispersion of the random variable, can be obtained by taking the square root of the variance. Thus,

$$E\left[(x-\mu_x)^2\right] = \sum_{k=1}^{\infty} (x_k-\mu_x)^2 \; Pr(x_k) \qquad (4)$$

$$\sigma = \sqrt{E\left[(x-\mu_x)^2\right]} \qquad (5)$$

For the higher order central moments (n>2), they are often normalized by dividing the nth power of standard deviation. The third and fourth normalized central moments are mostly used to indicate the shape of the probability density function. They can be obtained by

$$Skewness = \frac{E\left[(x-\mu_x)^3\right]}{\sigma^3} \qquad (6)$$

$$Kurtosis = \frac{E\left[(x-\mu_x)^4\right]}{\sigma^4} \qquad (7)$$

The skewness provides information about the position of the peak density relative to the mean value. The kurtosis indicates the spread in distribution. For a perfect Normal distribution, the skewness is zero and the kurtosis is three.

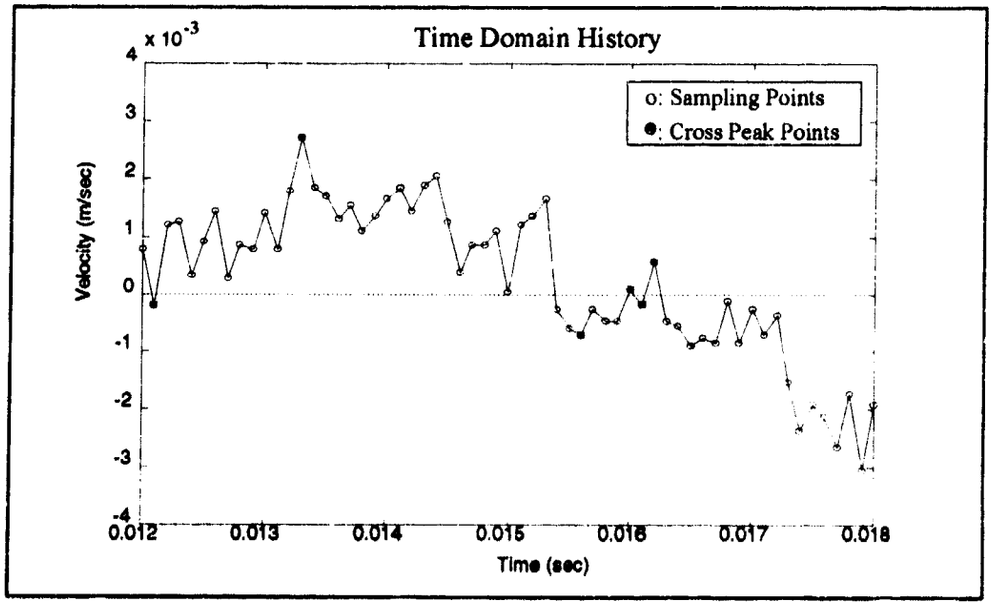


Figure 4 Comparison of Sampling Points and Cross Peak Points

Figure 5 shows the flow chart for our CPA technique. After importing the sampled data, the DC offset was removed by subtracting the mean value. The cross peak points between every zero crossing was then found by a subroutine. Because the dynamic range of the tape record is 40 dB, we set a 1% threshold to eliminate those distorted points. The data sets for the same type of fire pump and same pickup location were then added together to from a combined data set to perform the statistical analysis.

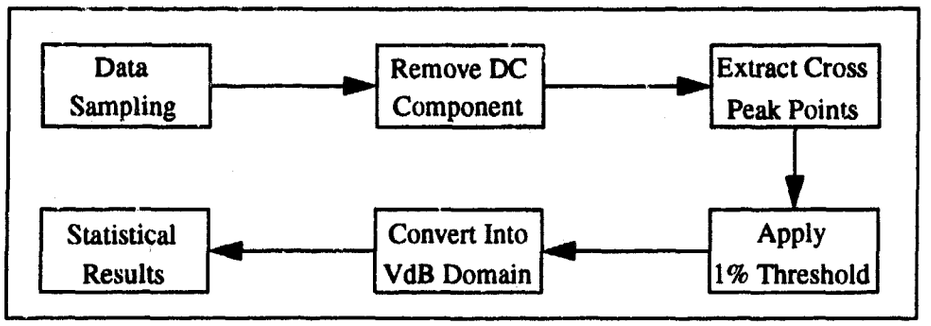


Figure 5 Cross Peak Analysis Flow Chart

In order to understand the distribution of the cross peak points, probability histograms for each pickup location were plotted with a linear velocity scale. Figure 6 shows an example of a probability histogram for a Type I fire pump measured at Motor Bearing (Free End). Inspection of this figure shows an exponential shape distribution which can be approximated as a lognormal distribution. This was true in all cases [16]. The distributions become normal if the data is transformed by a log or natural log algorithm. The method we used to obtain Normal distributions from these lognormal distributions was to convert the linear velocity readings to velocity dB (VdB) readings. The VdB is defined as:

$$Velocity\,(VdB) = 20 \log \frac{V}{V_{ref}} \qquad (8)$$

where V_{ref} is the reference level, normally 0VdB=10^{-8} m/sec. After we computed the mean, standard deviation, skewness and kurtosis, plots of probability density functions and a statistical analysis report were generated.

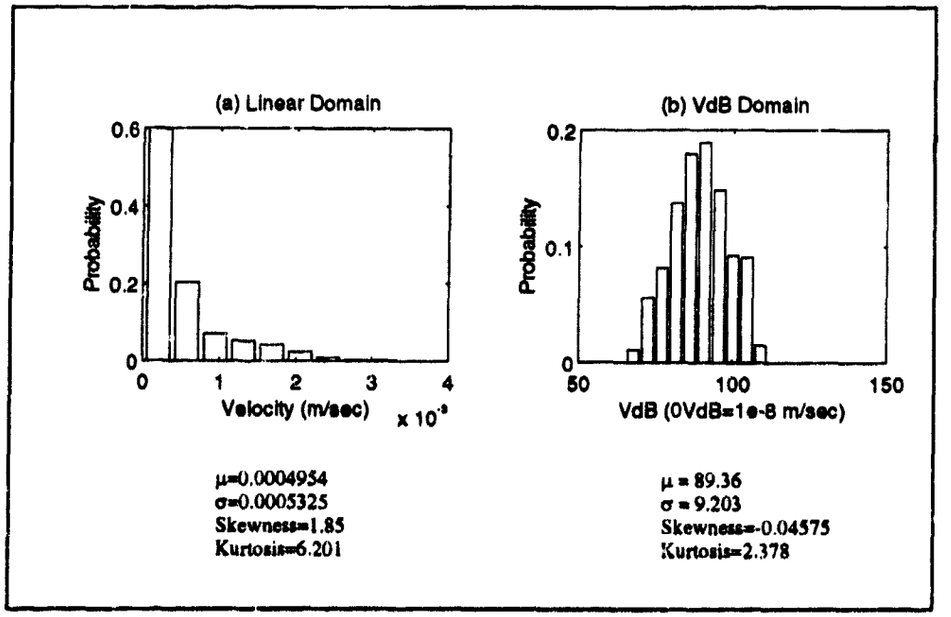


Figure 6 Probability Histograms For Type I Fire Pump Measured at MB(FE)

Table 5 tabulates these results for the linear velocity and VdB domains. A comparison of alarm thresholds between the broadband alarm level (B.B. Alm. Level) and computed $\mu+2\sigma$ level is also shown in Table 6. It should be mentioned that these two alarm thresholds are based on different methods using data from different domains. The broadband alarm level is obtained by using vibrometer readings in the frequency domain. The computed $\mu+2\sigma$ level represent the 97.5% acceptance level of the cross peak envelope in the VdB domain. In general, the B.B. Alm. Level and the computed $\mu+2\sigma$ level in the VdB domain were very close. The relative difference between these two methods for all fire pump were less than 8%. It is also noted from Table 5 that the computed $\mu+2\sigma$ level for distributions with small skewness is closer to the broadband alarm level than for those with high skewness. This is because the computed $\mu+2\sigma$ level for the Normal distribution is very sensitive to skewness.

Table 5: Statistical Analysis Results for Cross Peak Points

Type	Pickup Position	Linear Velocity Domain				VdB Domain			
		μ (mm/s)	σ (mm/s)	Skew.	Kurt.	μ (VdB)	σ (VdB)	Skew.	Kurt.
I	MB(FE)	0.495	0.533	1.850	6.201	89.36	9.20	-0.046	2.378
	MB(CE)	0.539	0.608	2.052	7.195	90.20	9.12	0.049	2.444
	PB(CE)	2.828	2.680	1.517	5.638	104.6	9.78	-0.444	2.476
	PB(FE)	2.717	2.581	1.253	3.959	103.6	10.45	-0.335	2.232
	PB(FE/A)	3.699	3.307	1.086	3.515	106.9	9.90	-0.519	2.472
II	UMB	0.672	0.926	4.985	36.52	92.1	8.69	0.029	2.953
	LMB	0.923	1.046	3.350	19.59	95.3	8.56	-0.077	2.697
	LMB(A)	0.703	0.648	1.881	7.538	93.3	8.46	-0.349	2.663
	UPB	4.141	4.442	1.570	5.446	106.1	11.84	-0.408	2.207
	LPB	3.160	3.246	1.779	6.775	105.1	10.20	-0.373	2.430

Table 6: Comparison of Computed $\mu+2\sigma$ and Broadband Alarm Level

Type	Pickup Position	Skewness	μ (VdB)	σ (VdB)	$\mu+2\sigma$ (VdB)	B.B.Alarm Level (VdB)	Relative Difference (%)
I	MB(FE)	-0.046	89.36	9.20	107.8	107	0.71
	MB(CE)	0.049	90.20	9.12	108.3	109	-0.68
	PB(CE)	-0.444	104.6	9.78	124.2	118	4.95
	PB(FE)	-0.335	103.6	10.45	124.5	119	4.42
	PB(FE/A)	-0.519	106.9	9.90	126.7	121	4.50
II	UMB	0.029	92.1	8.69	109.5	108	1.40
	LMB	-0.077	95.3	8.56	112.4	110	2.11
	LMB(A)	-0.349	93.3	8.46	110.2	108	2.04
	UPB	-0.408	106.1	11.84	129.8	120	7.54
	LPB	-0.373	105.1	10.20	125.5	121	3.59

Frequency Domain - 1/1 Octave Band Analysis: This method divides the frequency spectrum into constant percentage bands having the same ratio of bandwidth to center frequency [17]. Each band has an upper frequency limit (f_2) and lower frequency limit

(f_1). The center frequency (f_c) of any such band is defined as

$$f_c = \sqrt{f_1 f_2} \qquad (9)$$

The ratio of center frequencies of successive proportional bands is the same as $\frac{f_2}{f_1}$ for any one band. i.e.,

$$\frac{f_c}{f_1} = \frac{f_2}{f_c} = \sqrt{\frac{f_2}{f_1}} = 2^{n/2} \qquad (10)$$

where a third of an octave is defined as n=1/3. The method chosen to examine frequency domain alarm levels in this research was 1/1 octave band (n=1). For the frequency range of 10-5,000 Hz of our system, the frequency spectrum can be divided into 9 bands by using the American National Standards Institute (ANSI) preferred center frequencies. The center frequencies and pass bands covering the frequency range 10-5,000 Hz in 1/1 octave bands are given in Table 7.

Table 7: ANSI Preferred Center Frequencies and Pass Bands for 1/1 Octave Band

Center Frequency (Hz)	Pass Band (Hz)	
	Lower Limit	Upper Limit
16	11.2	22.4
31.5	22.4	45
63	45	90
125	90	180
250	180	355
500	355	710
1000	710	1400
2000	1400	2800
4000	2800	5600
8000	5600	11200

Figure 7 shows a flow chart for our 1/1 OBA. In order to have better resolution at lower frequencies, it was necessary to use two different sampling frequencies. For time domain data, the lower sampling frequency spanned 2 kHz and the higher sampling frequency spanned 10 kHz. Twenty blocks of data were taken at each sampling frequency, where a block of data was 2048 data points. After the sampling, a FFT was performed for each block with a Hanning window to obtain the smoothed linear velocity spectrum. The averaging VdB spectrum was then produced by transforming the linear spectrum to the VdB domain and averaging the 20 blocks of data. For the frequency domain data, 10 order and 100 order data sets were used as the lower frequency spectrum and higher frequency spectrum, respectively. Then, the 1st through 6th 1/1 octave band levels were computed using the lower sampling spectrum and the 7th through 9th 1/1 octave band levels were obtained using higher sampling spectrum. Finally, the 1/1 OBA was done after combining these into 9 band levels.

1/1 OBA was performed for all types of fire pumps using time domain data. The mean (μ) and standard deviation (σ) for each octave band level for the same type of fire pump with the same pickup location were also computed. Figure 8 shows an example of the 1/

1 octave band levels for Type I fire pumps. The thin line represents the mean (μ) band level and the thick line represents the mean plus standard deviation ($\mu+\sigma$) band level. The broadband level was also computed by adding all the VdB spectrum using a RMS algorithm. Generally speaking, the dominate levels are located at first six bands (10 to 710 Hz). This implies that the energy of vibration is concentrated at lower frequencies. For fire pumps with a 3555 RPM operation speed, these bands are approximated up to 12 orders.

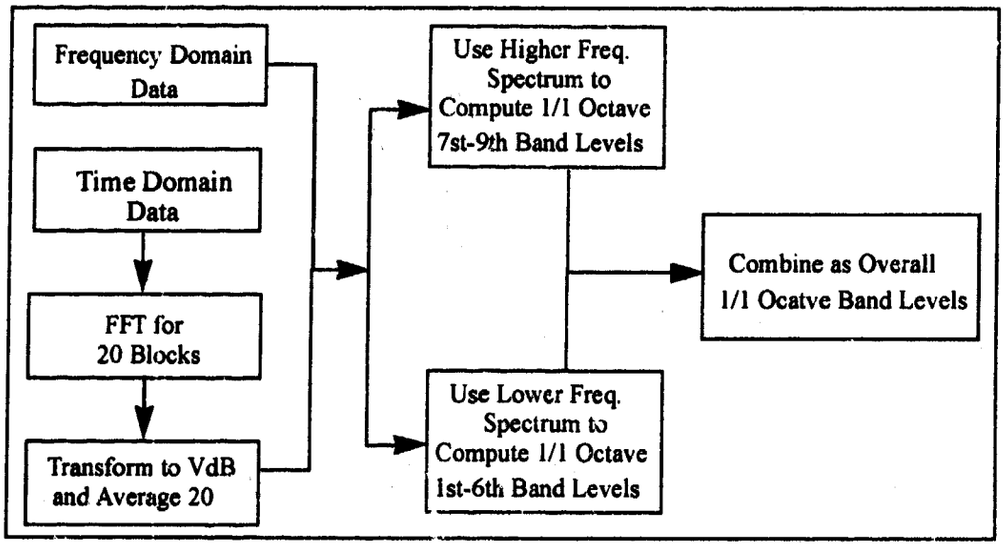


Figure 7 Flow Chart of 1/1 Octave Band Analysis Technique

Artificial Fault Simulations: In order to assess the 1/1 OBA, three cases of artificial fault simulation were performed. Before we simulated these cases, a fault had to be defined. For simplicity, we assumed a fault could be approximated with 5 bars as shown in Figure 9. The bandwidth of the fault is about 4.5 Hz. Based on (Eq. 8), if the amplitude doubles in a linear velocity scale then the VdB level will increase 6VdB at the corresponding frequency. Therefore, we defined a fault with a 6VdB gain at the center bar and use four bars with 3VdB (half the power of the center bar) and 2VdB at the side bands to approximate leakage effect around the corresponding frequency.

› Case I: Simulation of Misalignment Fault

Misalignment occurs when the center lines of two shafts are offset or meet at an angle. The characteristics of misalignment in spectrum include:

- High amplitude axial peaks and radial peaks at 1, 2 and 3 order(s) of shaft RPM.
- Higher harmonics of shaft RPM (greater than 4 orders of shaft RPM) are generally low in amplitude.

Therefore, it is reasonable to add the 6VdB gain fault at the first and second order. Figure 10(a) and 10(b) show the spectrums before and after impose the fault. Figure 9(c) compares the difference of 1/1 octave band levels and 10-5,000 Hz broadband level for these two conditions. As can be seen from Figure 10(c), the 1/1 octave band levels have a 4.46 VdB gain at the third band (corresponds to 1 order of shaft RPM) and a 3.13 VdB gain at the forth band (corresponds to 2 orders of shaft RPM). However, the 10-5,000 Hz broadband level only increased 2.17 VdB. It is obvious that the 1/1 OBA is more sensitive than the broadband method.

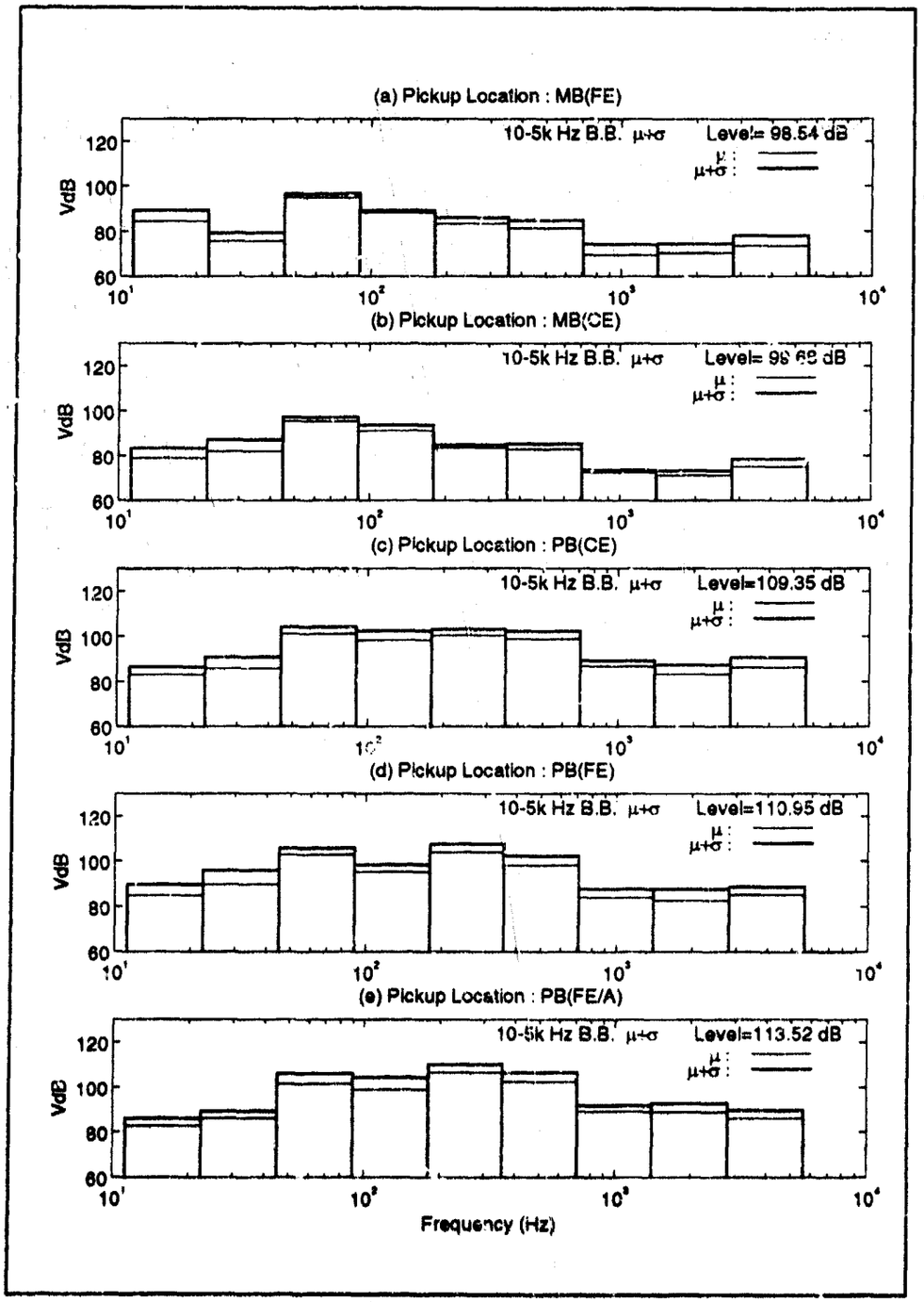


Figure 8 The Summarized 1/1 Octave Band Levels for Type I Fire Pumps (Ref. 0VdB=10^{-8} m/sec)

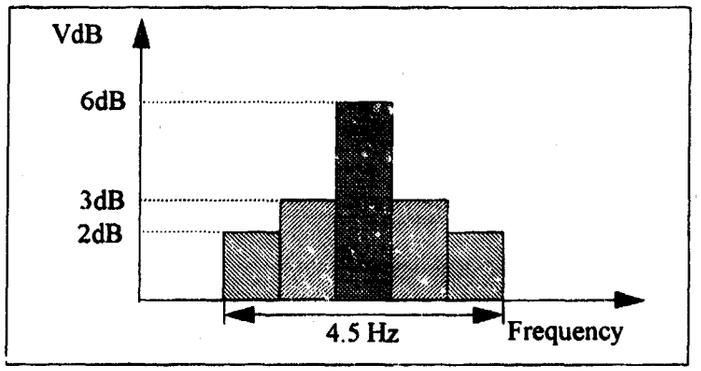


Figure 9 Fault with 6VdB Gain at Center Bar

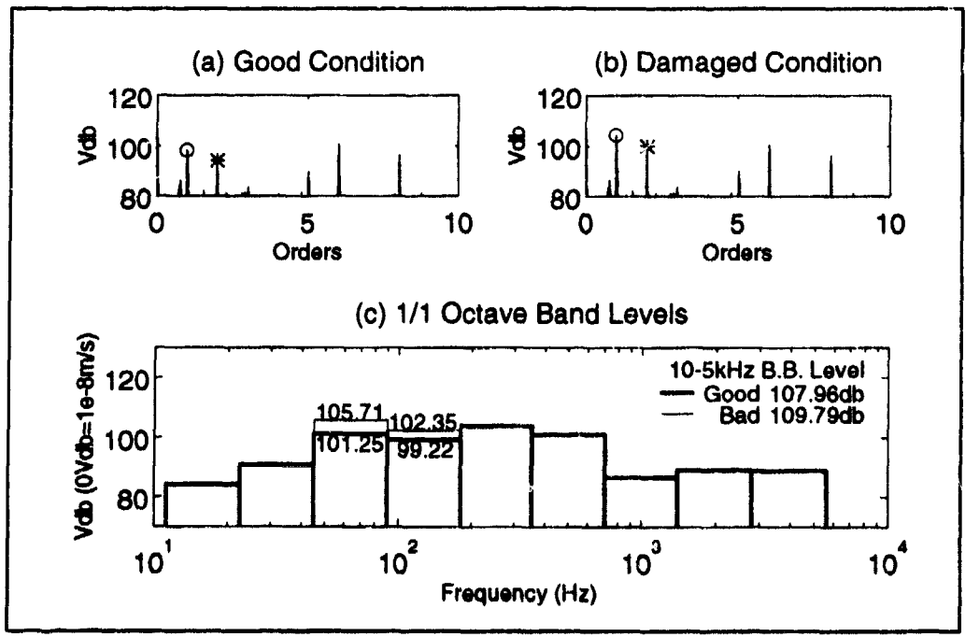


Figure 10 Simulation Result for Artificial Misalignment at Shaft

• Case II: Simulation of Looseness Fault at Impeller

Impeller looseness is a rotating element looseness. The important characteristics of looseness in spectrum are presence of a large number of harmonics and half-harmonics. For example, a fire pump of Type I has six impellers creating a forcing frequency at 6 orders of the shaft RPM, and we expect a high level at 6 orders and its harmonics (such as 12 orders, 24 orders,... etc.). Also, the levels at half-harmonics (such as 3 orders and 9 orders) will increase. Therefore, in this looseness fault simulation, an artificial fault is applied by adding a 6VdB center bar height fault at the 6th order(i.e., 6 x shaft RPM) and adding two faults with 3VdB center bar height at the 3rd order and 9th order. Figure 11(a) and 11(b) shows the resulting spectrum for good and damaged fire pumps. The Figure 11(c) compares the difference of 1/1 octave band levels and 10-5,000 Hz broadband level for these two conditions. At the fifth of the 1/1 octave band levels, there

is a 1.5 VdB gain. However, the 10-5,000 Hz broadband level only increases 0.29 VdB. Again, the 1/1 OBM is more sensitive than the broadband method for fault detection.

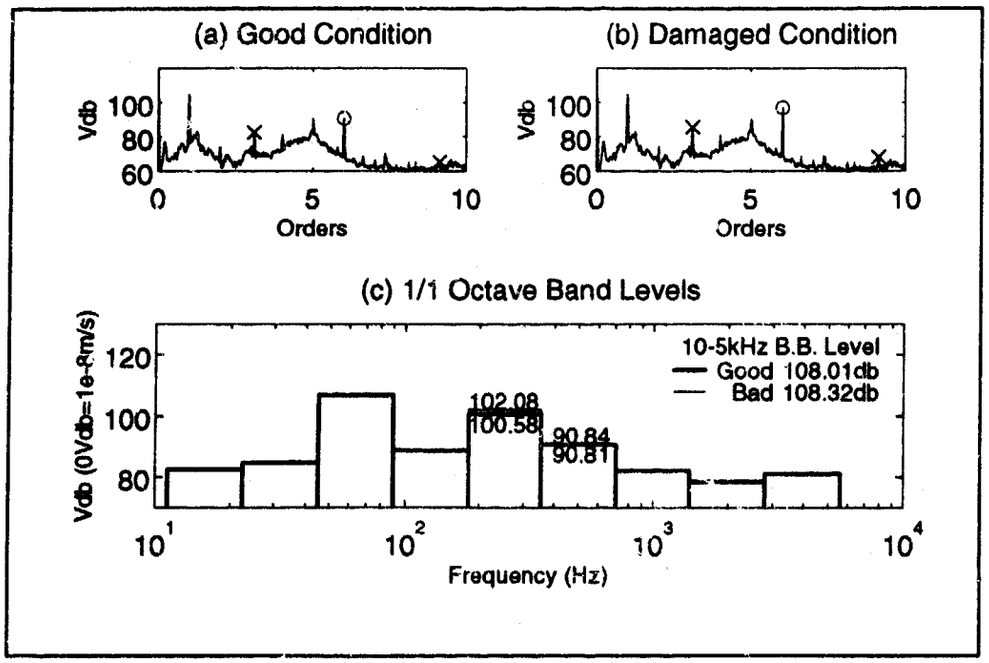


Figure 11 Simulation Result for Artificial Looseness at Impeller

• Case III: Simulation of Bearing Fault

For a steady state condition, some periodic signatures exist which relate to corresponding bearing faults, these are called bearing frequencies. In this section, an artificial fault imposed at the Motor Bearing (Free End) of a Type I fire pump has been simulated. The bearing frequencies of this bearing are tabulated in Table 8.

Table 8: The Bearing Frequencies of MRC310 Bearing

Bearing Frequency	Symbol	Order
Train order of rolling element	f_r	0.38
Relative rotation order of rotating raceway	f_{tl}	0.619
Spin order of rolling elements	f_b	1.98
Irregularity order of rolling element	f_{bs}	3.96
Irregularity order of rotating raceway	f_{ir}	4.95
Irregularity order of stationary raceway	f_{or}	3.04

Suppose, for example, that there is a wear degradation in the inner raceway. Then the presence of a fundamental frequency spike f_{ir} with harmonics would be expected. Figure 12(a) shows the spectrum for a good bearing. The 6VdB fault is imposed at the 4.95 order and is show in Figure 12(b). The Figure 12(c) shows the resulting 1/1 octave band levels. The 4.95 order is located at the fifth band of the 1/1 octave band. The 10-5,000 Hz broadband levels before and after damage are almost unchanged. However, the fifth octave band level has increased 1.57VdB. This is a good illustration as to why a single broadband level cannot be used to accurately determine the condition of a machine.

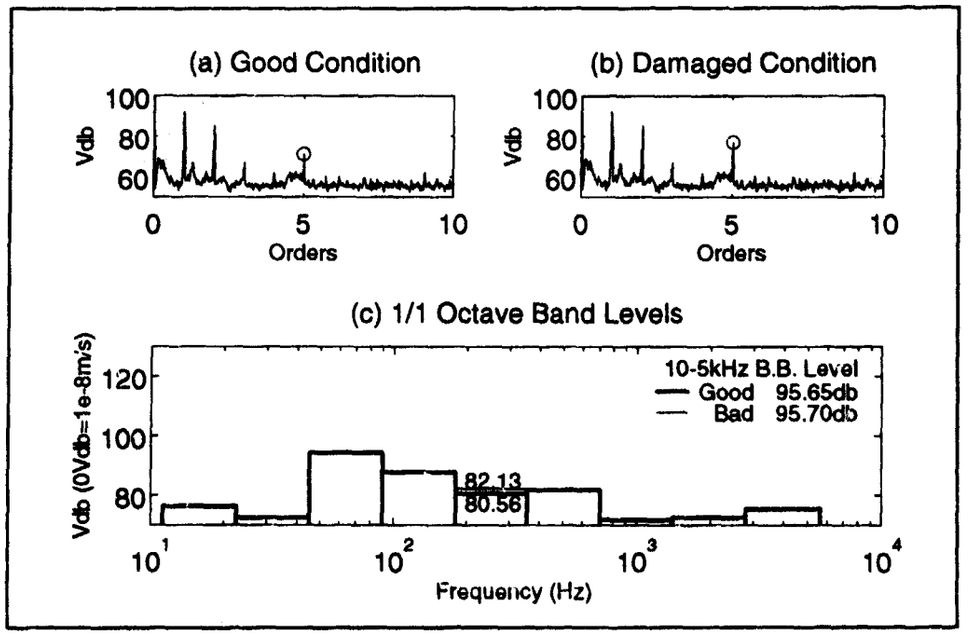


Figure 12 Simulation Result for Artificial Bearing Fault at Motor

Conclusions: Both time and frequency domain analyses were performed using fire pump vibration data to determine an appropriate alert level. Cross Peak Analysis in time domain and 1/1 Octave Band Analysis in frequency domain were used to set the alert threshold levels. The following conclusions are drawn:

Time Domain Analysis (Cross Peak Analysis) :

- The measured peak envelop data at five different pickup locations follows a Gaussian probability distribution in the VdB domain well.
- The $\mu+2\sigma$ value computed using a peak envelop probability density function in the VdB domain gives a broadband peak amplitude alert level rather than the energy content in the vibration signal.
- We used 12 measured data sets (6 fire pumps and 2 different measurement dates) which were available on cassette data tapes. To improve the quality of averaging, more measured data sets would be required.

Frequency Domain Analysis (1/1 Octave Band Analysis) :

- 1/1 Octave Band Analysis (OBA) uses nine frequency band alert levels which provides more detailed information about machine condition than the simple broadband level measured by a vibrometer. When a significant change in VdB occurs in a particular frequency band or bands, narrow band zoom mode analysis can be performed for the selected frequency band(s) to identify the component(s) which may have faults.
- 1/1 OBA divides the frequency range into 10 bins over 10kHz. The VdB level in each frequency bin is quite sensitive to changes in the energy content of the measured vibration signal.

Acknowledgments: We would like to express our sincere appreciation to Mr. Jon Loeliger, Mr. Orlando Taraborrelli and Mr. Authur Cautilli of the Carderock Division,

Naval Surface Warfare Center (formerly Naval Ship Systems Engineering Station), Philadelphia, PA, for their support and funding of this research.

References:

[1]. Xistris, G. D., and Lowe, J. M., "On the Effectiveness of Maintenance Programs and Policies for Shipboard Machinery," Condition Monitoring '84, Proceedings of an International Conference on Condition Monitoring, Swansea, U.K., 10-13 April 1984, pp.55-66.

[2]. Tracor Applied Sciences, "Advanced Vibration Analysis Studies," Tracor Documt No.: T86-0109567-U, Dec. 1986.

[3]. Marshall, B. R., "A Surface Navy Vibration Program Overview: Standardization and State-of-the-Art," Naval Engineers Journal, Vol. 100, May 1988, pp. 90-100.

[4]. Milner, G. M., J., and Smith, R., "Development of an Automatic Machinery Vibration Analysis (AMVA) System," 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, Patuxent River, Maryland, 28-30 October 1986, pp. 237-248.

[5]. Chapman, C. L., and Jolley, R., "Systems and Equipment Maintenance Monitoring for Surface Ships (SEMMSS) Program," Naval Engineers Journal, V. 99, November 1987, pp.63-69.

[6]. Strunk, W. D., "The Evaluation of Accelerometer Mount Transmissibility for U.S. Navy Applications," Proceedings of the 6th International Modal Analysis Conference, January 1988, pp.1384-1389.

[7]. Tracor Applied Sciences, "An Analysis of Submarine Machinery Vibration Monitoring," Tracor Documt No.:T89-01-9546-U, May 1986.

[8]. Mitchell, J. S., "An Introduction to Machinery Analysis and Monitoring," PennWell Books, Tulsa, Oklahoma, 1981, Chapter 9.

[9]. Randall, R. B., "Computer Aided Vibration Spectrum Trend Analysis for Condition Monitoring," Maintenance Management International, v. 5, 1985, pp.161-167.

[10]. Campel, C., "Determination of Vibration Symptom Limit Values in Diagnostics of Machinery," Elserier Science Publications, B.V., 1985.

[11]. Murphy, T. J. and Smiley, R. G., "Set Vibration Alarm Levels without Gauss work," Proceedings of Comadem '89 Conference, pp.386-396.

[12]. Thomas, M. R. and Ball, A. D., "An Algorithm for the Statistical Analysis of Vibration Readings."

[13]. Glew, C. W. A., "The Effectiveness of Vibration Analysis as a Maintenance Tool," Trans. I. Mar. E., Vol. 86, 1974, pp.29-50.

[14]. Carnody, T., "The Measurement of Vibration as a Diagnostic Tool," Trans. I. Mar. E., Vol. 84, 1972, pp.147-159.

[15]. Papoulis, A., "Probability, Random Variables, and Stochastic Processes," McGraw-Hill Book Co., 1991, Chapter 5.

[16]. Liu, C. S., Jeon, J. J. and Shin, Y. S., "Analysis, Approach and Assessment of Vibration Criteria in Shipboard Machinery Condition Monitoring and Diagnostics," NPS Technical Report, NPS-ME-93-005, 1993.

[17]. Pierce, A. D., "Acoustics: An Introduction to Its Physical Principles and Applications," McGraw-Hill Book Co., 1981, Chapter 2.

APPLICATION OF FAULT DETECTION TECHNIQUES TO SPIRAL BEVEL GEAR FATIGUE DATA

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

and

Robert F. Handschuh, and Harry J. Decker
Propulsion Directorate
U.S. Army Aviation Systems Command
Lewis Research Center
Cleveland, Ohio 44135

Abstract: Results of applying a variety of gear fault detection techniques to experimental data is presented. A spiral bevel gear fatigue rig was used to initiate a naturally occurring fault and propagate the fault to a near catastrophic condition of the test gear pair. The spiral bevel gear fatigue test lasted a total of eighteen hours. At approximately five and a half hours into the test, the rig was stopped to inspect the gears for damage, at which time a small pit was identified on a tooth of the pinion. The test was then stopped an additional seven times throughout the rest of the test in order to observe and document the growth and propagation of the fault. The test was ended when a major portion of a pinion tooth broke off. A personal computer based diagnostic system was developed to obtain vibration data from the test rig, and to perform the on-line gear condition monitoring. A number of gear fault detection techniques, which use the signal average in both the time and frequency domain, were applied to the experimental data. Among the techniques investigated, two of the recently developed methods appeared to be the first to react to the start of tooth damage. These methods continued to react to the damage as the pitted area grew in size to cover approximately 75% of the face width of the pinion tooth. In addition, information gathered from one of the newer methods was found to be a good accumulative damage indicator. An unexpected result of the test showed that although the speed of the rig was held to within a band of six percent of the nominal speed, and the load within eighteen percent of nominal, the resulting speed and load variations substantially affected the performance of all of the gear fault detection techniques investigated.

Key Words: Diagnostics; Failure prediction; Fatigue; Gear

Introduction: Drive train diagnostics is becoming 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 turbojet aircraft. This would require a substantial, thirtyfold, increase in helicopter safety statistics. Practi-

cally, 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. Therefore research is necessary to develop and prove various fault detection concepts and methodologies.

A number of methods have been developed to provide early detection of gear tooth damage. McFadden proposed a method to detect gear tooth cracks and spalls using the instantaneous phase of the demodulated time signal [2]. Stewart devised several time domain discriminant methods of which FM0, a coarse fault detection parameter, and FM4, an isolated fault detection parameter, are the most widely referenced [3]. Lyon [4], and Liu [5], proposed using the instantaneous frequency of the demodulated time signal to detect gear surface pitting. Methods NA4 and NB4 were recently developed at NASA Lewis Research Center to provide early detection of gear tooth surface damage, and continue to react to the damage as it spreads and grows in severity [6].

Verification of these detection methods with experimental data along with a comparison of their relative performance is a crucial step in the overall process of developing a highly reliable HUM system.

In view of the aforementioned, 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. Each method is applied to vibration data obtained from a spiral bevel gear fatigue rig at NASA Lewis, where the test gears are run until a fatigue failure occurs. In the test used in this study, a tooth on the pinion developed a pit during the first five and a half hours of the run. The pit was allowed to propagate over a majority of the tooth, resulting in tooth fracture. The performance of each method is discussed, and overall conclusions are presented.

THEORY OF FAULT DETECTION METHODS: All of the methods in this investigation utilized vibration data that was pre-processed as it was collected. To eliminate the noise and vibration that is incoherent with the rotational speed of the spiral bevel test pinion, the raw vibration data was time synchronous averaged immediately after being digitized. During time synchronous averaging, the data was also interpolated to obtain 1024 points over five complete revolutions of the test pinion. Each of the methods below were then applied to the time averaged and interpolated vibration data.

FM0 is formulated to be an indicator of major faults in a gear mesh by detecting major changes in the meshing pattern [3]. FM0 is found by dividing the peak-to-peak level of the signal average by the sum of the amplitudes of the mesh frequency and its harmonics. In major tooth faults, such as breakage, the peak-to-peak level tends to increase, resulting in FM0 increasing. For heavy distributed wear or damage, the peak-to-peak remains somewhat constant but the meshing frequency levels tend to decrease, resulting in FM0 increasing.

FM4 was developed to detect changes in the vibration pattern resulting from damage on a limited number of teeth [3]. 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 original 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, resulting in the normalized kurtosis value to increase beyond the nominal value of 3.

A demodulation technique was developed to detect local gear defects such as fatigue cracks, pits and spalls [2]. The basic theory behind this technique is that a gear tooth defect will produce sidebands that modulate the dominant meshing frequency. In this method the signal is band-passed filtered about a dominant meshing frequency, including as many sidebands as possible. The Hilbert transform is then used to convert the real band-passed signal into a complex time signal, or analytic signal. Using the real and imaginary parts of the analytic signal, the instantaneous phase (I.P.) can be estimated from the filtered sidebands. Teeth with a surface failure, or fatigue crack, will cause a lead or lag in tooth contact during meshing, resulting in transient changes in the gear rotation. These transient changes in rotation will dominate the I.P. function. The standard deviation of the I.P. is then calculated over one complete revolution of the pinion to produce a single number in order to quantify the relative variance of the I.P. at each point in the run.

Another technique was proposed in which the rate of change of the instantaneous phase is calculated [4, 5]. This rate of change, or instantaneous frequency (I.F.), is sensitive to the transient rotational speed changes caused by teeth with surface defects, or root cracks, going through the meshing process. The instantaneous frequency and instantaneous phase are different representations of the same physical phenomenon, however the instantaneous frequency is, by definition, more sensitive [4]. A small change in phase within a very short time would result in a correspondingly large change in the I.F. The I.F. is also calculated from a bandpassed portion of the time signal, using the Hilbert transform. The I.F. is found using equation 1 below:

$$f(t) = \frac{b(t)H'[b(t)] - b'(t)H[b(t)]}{2\pi E^2(t)} - f_c \qquad (1)$$

where

f(t)	instantaneous frequency (Hz)
b(t)	band passed signal
b'(t)	first derivative of band passed signal
H[b(t)]	Hilbert transform of band-passed signal
H'[b(t)]	first derivative of Hilbert transform of bandpassed signal
E(t)	envelope of bandpassed signal (magnitude of complex time signal)
fc	carrier, or center, frequency of band (primary mesh frequency)

The standard deviation of the I.F. is then calculated over one complete revolution of the pinion to produce a single number in order to quantify the relative variance of the I.F. at each point in the run.

NA4 is a method recently developed at NASA Lewis to not only detect the onset of damage, but also to continue to react to the damage as it increases [6]. 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, resulting in the quasi-normalized kurtosis given in equation 2 below:

$$NA4(M) = \frac{N\sum_{i=1}^{N}\left(r_i - \bar{r}\right)^4}{\left\{\frac{1}{M}\sum_{j=1}^{M}\left[\sum_{i=1}^{N}\left(r_{ij} - \bar{r}_j\right)^2\right]\right\}^2} \qquad (2)$$

where

r	residual signal
$\bar{r}$	mean value of residual signal
N	total number of time points in time record
i	data point number in time record
M	current time record number in run ensemble
j	time record number in run ensemble

In NA4, the kurtosis is normalized, however it is normalized using the variance of the residual signal averaged over the run up to the current time record number, where NA4 is being calculated. With this method, the changes in the residual signal are constantly being compared to the running average of the variance of the system, or 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. As with FM4, NA4 is 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 also uses the quasi-normalized kurtosis given in equation 2 above. The major difference is that instead of using a residual signal, NB4 uses the envelope of the signal bandpassed about the mesh frequency. As with the other demodulation techniques, the signal is band-passed filtered about the dominant (primary) meshing frequency. Using the Hilbert transform, a complex time signal is 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 transient variations in the loading. The basic 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 can be calculated using equation 2 above, with the exception of substituting the envelope of the signal in place of the residual signal. NB4* uses the same modification technique as NA4*. NB4 is also dimensionless, with a value of 3 under nominal conditions.

Apparatus and Gear Damage Review: The fatigue damage on the test pinion shown in figures 1 through 8, was obtained using the spiral bevel gear fatigue test rig illustrated in figure 9, at NASA Lewis Research Center. The primary purpose of this rig is to study the effects of gear tooth design, gear materials, and lubrication types on the fatigue strength of aircraft quality gears. Because spiral bevel gears are used extensively in helicopter transmissions to transfer power between nonparallel intersecting shafts, the use of this fatigue rig for diagnostics studies is extremely practical. 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 12-tooth test pinion, and the 36-tooth gear have a 35 degree spiral angle, a 1 in. face width, a 90 degree shaft angle, and a 22.5 degree pressure angle. The pinion transmits 720 hp, at a nominal speed of 14,400 rpm.

Pictures of tooth damage on the pinion at various stages in the test are illustrated in figures 1 through 8. At the first rig shut-down, at about five and a half hours into the test, a small pit was observed on one of the teeth on the test pinion, as illustrated in figure 1. The rig was shut-down an additional seven times to observe and document the damage as it progressed during the run. As seen in figures 2 through 4, the pitted area gradually spread to cover over 75% of the face of pinion tooth. At

approximately twelve hours into the run, pitting started to appear on adjacent teeth, as seen in figure 5. The pitting on the adjacent teeth continued to grow until it covered a majority of the face of three adjacent teeth on the pinion, and part of the face on another adjacent tooth, as seen in figure 7, taken at approximately sixteen hours into the run. The run was stopped when it was found that one of the three heavily pitted pinion teeth experienced a tooth fracture, losing one third of the tooth, as illustrated in figure 8. The break-off occurred sometime between the seventh shut-down (16.16 hours) and the end of the test (17.79 hours).

Discussion of Results: Figures 10 and 11 illustrate the minor speed and load fluctuations present during the run. Figures 12 through 20 illustrate the results of applying the various fault detection methods to the experimentally obtained vibration data. In all of the figures from 10 through 20, the vertical dashed lines numbered 1 through 8 correspond to the eight rig shut-down times, with the resulting damage documented in figures 1 through 8, respectively.

The fluctuations in speed and load over the course of the run were found to have significant effects on the response of the fault detection techniques. As seen in figure 10, the rig speed varied within a band of approximately 6% about the nominal pinion speed of 14,400 rpm. From figure 11, it also can be seen that the gear torque varied within a band of approximately 18%. The sharp change in speed and load that occurred at approximately 8.75 hours into the run, shortly after shut-down #3, affected the response of all of the parameters, in particular NA4. As seen in figure 15, the response from NA4, and NA4*, more than doubled, not due to a major increase in damage, but to the load and speed change at that point. The step changes in speed following shut-downs 4, 5, and 6, and the gradual change in speed in the interims between shut down are clearly evident in a majority of the parameters. The various levels of parameter response for FM0, FM4, NA4, and NB4, as seen in figures 12, 14, 15, and 16, respectively, can be directly linked to the speed changes. The sudden drop in parameters NA4 and NB4 at approximately 3.75 hours, as seen in figure 13, corresponds exactly to the shift in gear torque at that time, as shown in figure 11.

As seen in figure 12, the parameter FM0 shows only moderate changes as the damage starts and progresses. A majority of the variations in FM0 are most probably due to the speed and load variations during the run.

Parameters NA4 and NB4 are the first to react to the pinion tooth damage, as seen in the first 6.5 hours comparison plot of NA4, NB4, and FM4, illustrated in figure 13. The damage observed during shut-down number 1, illustrated in figure 1, occurred sometime between the start of the run and the time of the first shut-down, 5.50 hours. In earlier studies, parameter NA4 was shown to be very reliable and sensitive to the start of pitting damage [6]. At approximately 1.25 hours into the run both NA4 and NB4 increase from the nominal value of 3 to values from 4 to 6, thus indicating the start of tooth damage. Both NA4 and NB4 drop in amplitude at approximately 3.75 hours, coincident with, and thus influenced by, the change in gear torque that remains until the first shut-down (5.5 hours).

Parameter FM4 did show a possible reaction to the pit at approximately 3 hours into the run, as illustrated in figure 14. FM4, however, gives no coherent indication as the pitting grows in severity, even when it is still limited to a single tooth, i.e. up to approximately 10 hours into the run. This is surprising since FM4 was designed to be a single tooth defect parameter. Most of the changes seen in FM4 are due to corresponding load and speed changes.

As seen in figure 15, NA4 not only gives an initial reaction to damage at 1.25 hours, but it also continues to react as the damage increases. When the rig is restarted after shut-down #1, NA4 increases steadily to a value of 7, as the damage progresses from the small pit in figure 1, to the pit

seen in figure 2, (covering 50% of the tooth). NA4 then continues to increase to a value of 17, as the pit grows to cover over 75% of the tooth surface, as illustrated in figure 3. NA4 then progressively drops down, as its run averaged denominator increases. NA4* maintains its sensitivity, due to the denominator being locked, and thus continuing the comparison of current conditions to the denominator representing the system in "good" condition. Again, the speed and torque influences on parameters NA4 and NA4* are clearly seen, especially near shut-downs 3, 4, and 5.

Parameter NB4 shows trends very similar to those seen in NA4, except NB4 gives a more robust reaction to the damage. As seen in figure 16, NB4 increases from a value of 5 for the observed small pit, figure 1, to a value of 25, as the pit covers 50% of the tooth, figure 2. When the pitting covers over 75% of the tooth, figure 3, NB4 increases to a value of 40. As with parameter NA4*, NB4* is used to maintain the response through the end of the test. Again, the speed and load variations clearly affect the response of parameter NB4 and NB4*, as seen in figure 16, near rig shut-downs 3. 4, 5, and 6.

Figures 17 and 18 present the results of computing the standard deviation of the instantaneous frequency (I.F.) of the bandpassed signal using ± 2 sidebands, and ± 1 sideband, respectively. As seen in figure 17, the I.F. gives a robust reaction once the pitting is established, at 5.5 hours. This reaction is coincident with the start-up of the rig, following shut-down #1. However, neither the speed nor load change sufficiently enough at the start-up to cause the I.F.'s initial and sustained reaction at that point. Unfortunately, the ultra-sensitivity of this parameter makes it vulnerable to noise in the signal, even though the noise has been minimized with time synchronous averaging. Because the gear ratio of the test mesh is exactly 3:1, it is nearly impossible to remove the gear vibration from the pinion vibration signal. Some of the noise in the results shown in figure 17 could be due to the combination of gear noise in the signal, and the increased sensitivity of the parameter. Reducing the number of sidebands used to only ± 1, in an effort to reduce noise in the signal, results in eliminating some of the fault information from the signal, as illustrated in figure 18.

The standard deviation of the instantaneous phase (I.P.) is illustrated in figure 19. As seen in this figure, the I.P. gives a robust reaction at 5.5 hours into the run, after the pit has been established. The I.P. increases as the pit grows to cover over 75% of the tooth face, similar to NA4 and NB4, however in a less steady manner. The I.P. continues to react to the end of the test, fluctuating in some cases as a result of the speed changes. It is not known why the I.P. gave no clear reaction to the pitting prior to 5.5 hours into the run.

Figure 20 shows the denominator of parameter NB4, or the run averaged variance of the envelope of the bandpassed signal. Due to the run averaging process, this parameter increases steadily as the pit grows from the initial small pit to the end of the test. It gives a steady indication of accumulative damage without the influence of speed or load fluctuations.

Based on the results just presented, parameters NA4 and NB4 gave the best indication of the start and initial progression of pitting damage. As discussed, NA4 and NB4 detected tooth damage at approximately 1.25 hours into the run, which was the first indication of all the methods investigated. NA4 and NB4 continued to increase as the pitting damage increased to cover over 75% of the pinion tooth.

An unexpected result of this study showed that although the speed of the rig was held to a band within six percent of nominal speed, and the load was held to within a band of eighteen percent of nominal, the resulting speed and load variations present during the test substantially affected the performance of all the gear fault detection techniques investigated. To increase the reliability of the

parameters over speed and load variations, the more promising parameters, NA4 and NB4, should be modified to adapt to the different load and speed baselines to give consistent values based on damage alone, regardless of operating conditions.

Conclusions: Based on the results of applying a number of gear fault detection techniques to experimental data from a spiral bevel gear fatigue rig, the following conclusions can be made;

1) Parameters NA4 and NB4 were the first to react to the gear surface damage, and thus are good indicators of initial pitting.
2) Parameters NA4 and NB4 continued to react as the surface pitting increased to cover over 75% of the face width of the pinion, thus indicating increasing damage severity.
3) The run-normalized variance of the bandpassed signal's envelope (denominator of NB4) was found to be a good accumulative damage indicator.
4) The standard deviation of the instantaneous phase and instantaneous frequency gave robust indications of the pitting damage, once the pitting was established. These parameters, however, are more sensitive to noise in the signal.
5) All of the methods were sensitive to the minor changes in rig speed and load. Additional research is needed to modify methods, such as NA4 and NB4, to give reliable indications, regardless of speed and load.

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) McFadden, P.D.: Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration. J. Vib. Acoust. Stress Reliab. Design, vol. 108, no. 2, Apr. 1986, pp. 165-170.

3) 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.

4) Lyon, R.H.: Using the Hilbert Transform in Machine Diagnostics. Proceedings of the 3rd International Machine Monitoring and Diagnostics Conference, Society for Experimental Mechanics, Bethel, CT, 1991, pp. 204-206.

5) Liu,Y.; Chen, K.X.; Zhang, S.: Early Detection of Gearbox Fault Using Frequency Demodulation. Proceedings of the 3rd International Machine Monitoring and Diagnostics Conference, Society for Experimental Mechanics, Bethel, CT, 1991, pp. 72-76.

6) 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.

Figure 1.—Pinion damage at t = 5.50 hr.

Figure 2.—Pinion damage at t = 6.55 hr.

Figure 3.—Pinion damage at t = 8.55 hr.

Figure 4.—Pinion damage at t = 10.03 hr.

Figure 5.—Pinion damage at t = 12.03 hr.

Figure 6.—Pinion damage at t = 14.53 hr.

Figure 7.—Pinion damage at t = 16.16 hr.

Figure 8.—Pinion damage at t = 17.79 hr (end).

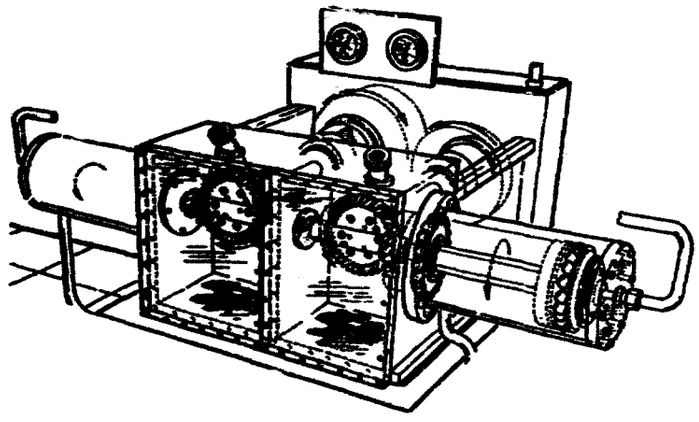


Figure 9.—Spiral bevel gear fatigue test rig at NASA Lewis.

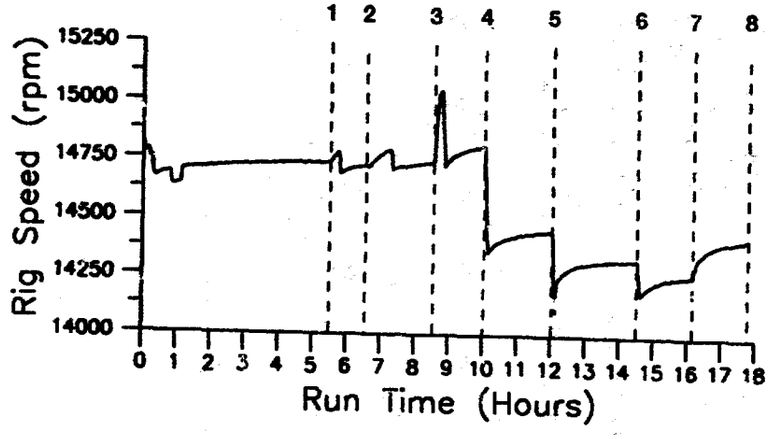


Figure 10.—Rig speed-vs.-run time.

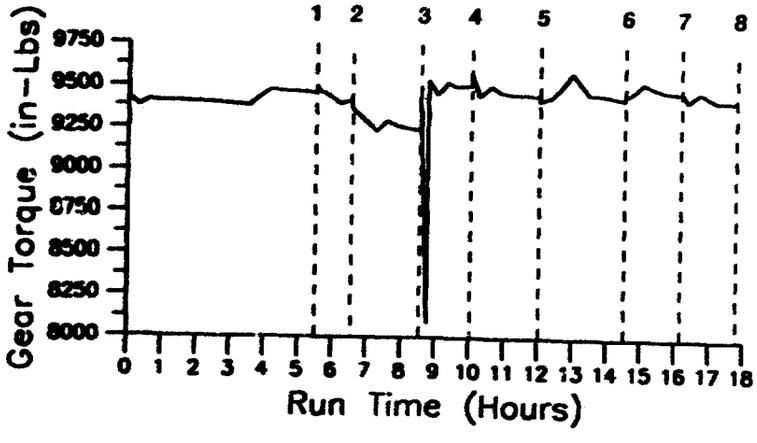


Figure 11.—Torque on gear-vs.-run time.

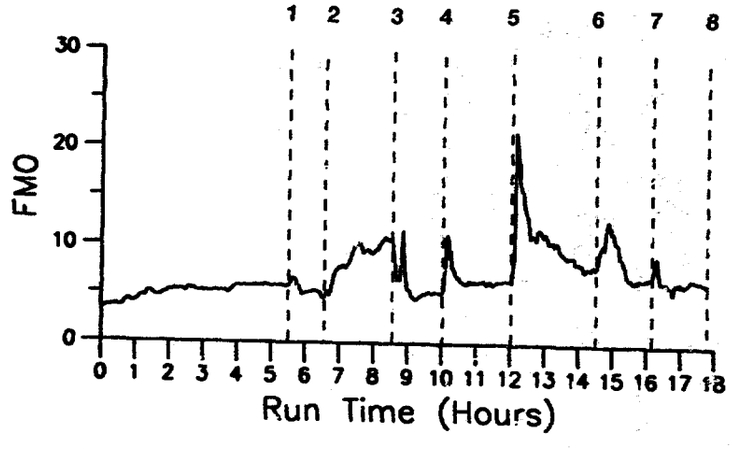


Figure 12.—Parameter FM0-vs.-run time.

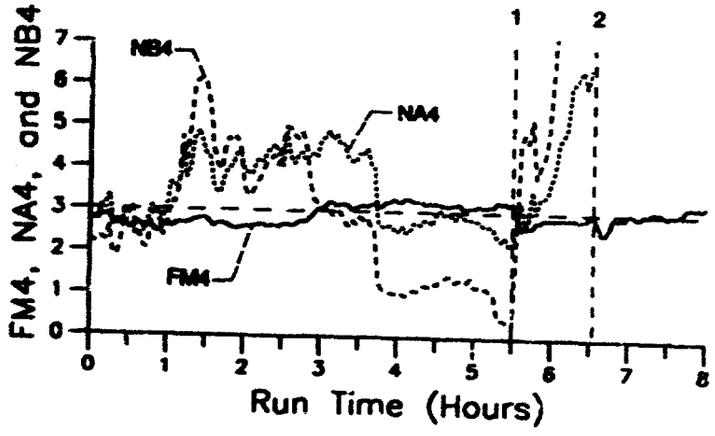


Figure 13.—Comparison of parameters FM4, NA4, and NB4 over first 8 hrs.

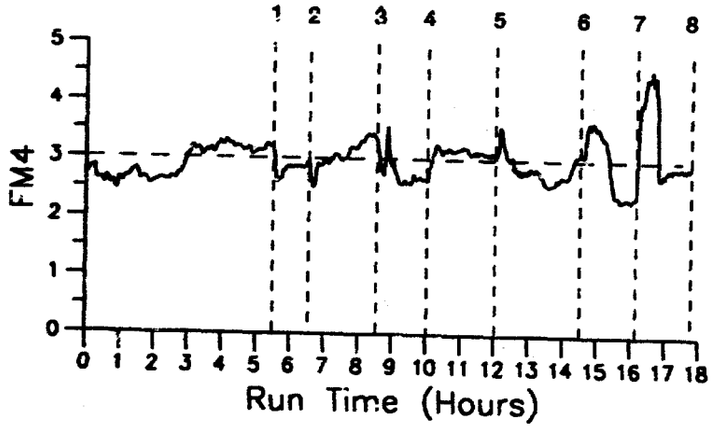


Figure 14.—Parameter FM4-vs.-run time.

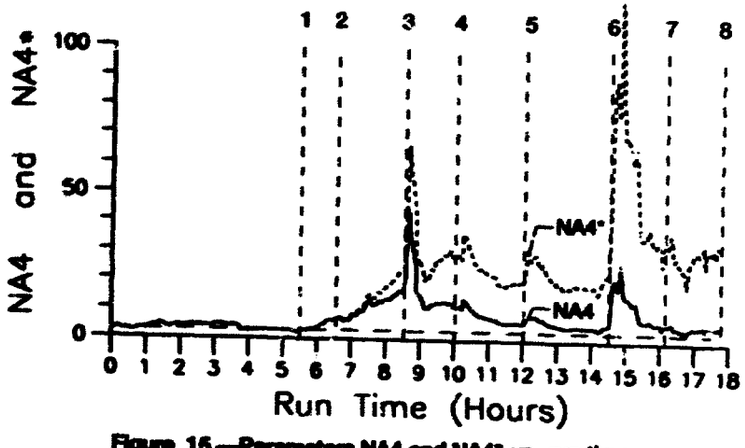


Figure 15.—Parameters NA4 and NA4*-vs.-run time.

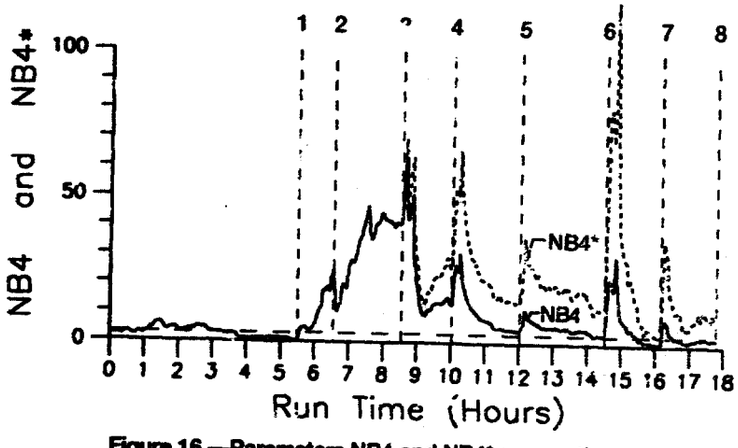


Figure 16.—Parameters NB4 and NB4*-vs.-run time.

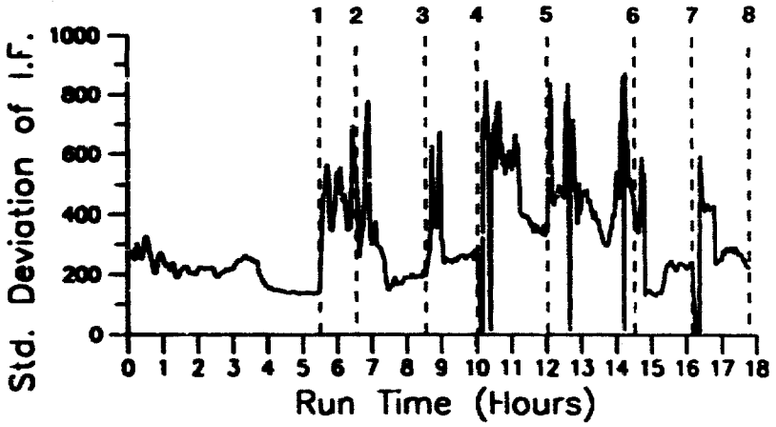


Figure 17.—Standard deviation of instantaneous frequency (with ±2 sidebands).

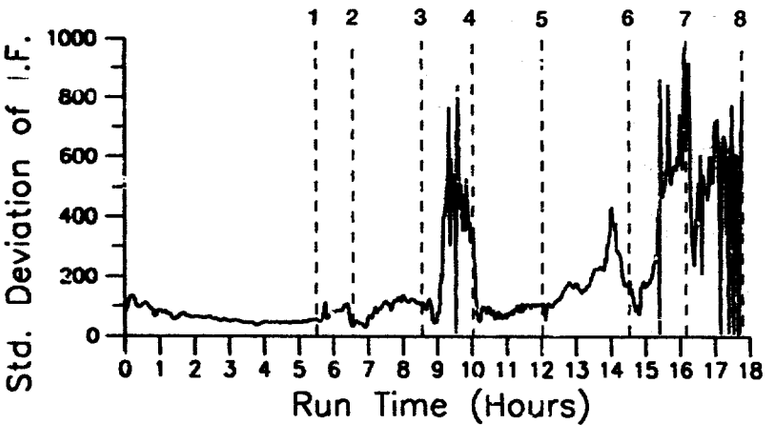


Figure 18.—Standard deviation of instantaneous frequency (with ±1 sidebands).

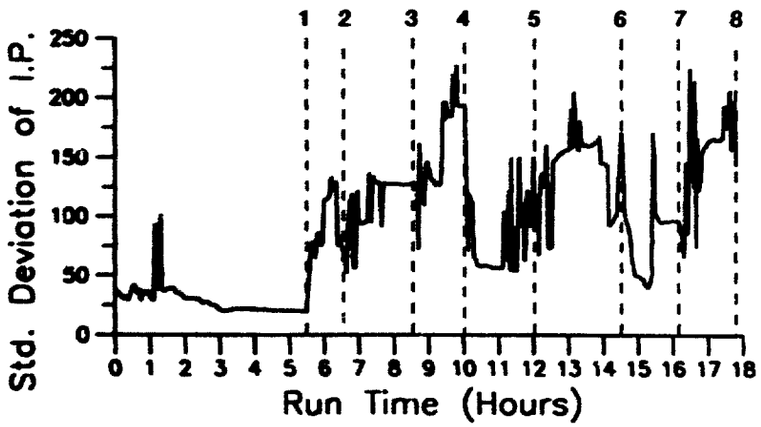


Figure 19.—Standard deviation of instantaneous phase-vs.-run time.

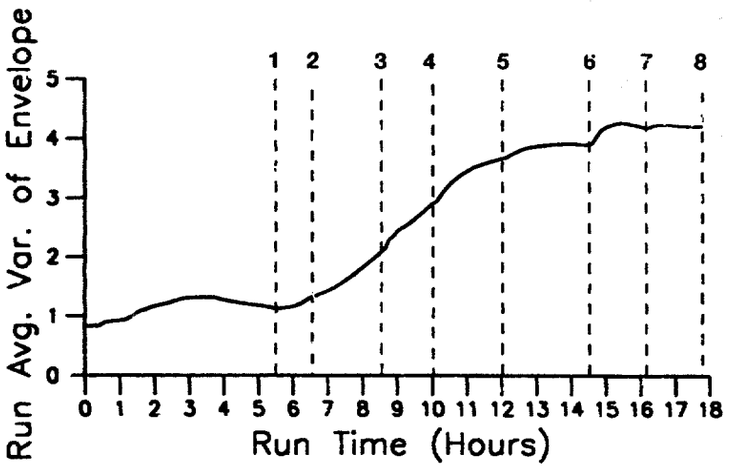


Figure 20.—Denominator of parameter NB4-vs.-run time.

THE AUTOMATED TRANSFORMER OIL MONITOR (ATOM)

Grant John Gorton
NKF Engineering Inc
4200 Wilson Blvd Suite 900
Arlington, VA 22203

Abstract: The Automated Transformer Oil Monitor (ATOM) is designed for condition monitoring and failure prevention for large utility transformers. ATOM was developed by Advanced Optical Controls (AOC) Inc, an NKF Engineering Inc affiliate, under sponsorship of the Consolidated Edison Company of New York. AOC specializes in fiber optic technology and advanced sensor applications. The ATOM project was initiated to provide on-line diagnostics for transformer oil gaseous content to evaluate transformer material condition and predict the onset of transformer failure. The device is designed to operate continuously and provide real-time or near real-time analysis of transformer condition. This is intended to replace the current laboratory analysis of manually removed oil samples. The ATOM extracts gases from the transformer oil without permanently removing any oil from the transformer and determines the concentrations (PPM) of seven different gases of interest. These concentrations can be trended or compared against alarm criteria.

Key Words: Oil analysis; transformers; sensors; detection; diagnostics; gases; fiber optics; neural networks; condition monitoring

Introduction: Large transformers are an integral part of generation and distribution systems for utility power plants of all sizes and types. These transformers represent major investments and their reliability is important to plant power production. Damage or failure of a transformer can mean a major maintenance headache and major dollar expenditures. As large transformers deteriorate, gases build up in the oil in the transformer. Monitoring the concentrations of gases in the oil offers maintenance engineers the ability to assess transformer material condition, detect deterioration prior to failure, and trend life expectancy. The current method of gas sampling requires manual removal of oil from the transformer with shipment to a laboratory for gas chromatography analysis. An automated, on-line monitoring system would save man-hours, save dollar expenditures, increase data availability, and improve material condition assessment.

In 1992, Consolidated Edison Company of New York teamed with Advanced Optical Controls (AOC), an affiliate company of NKF

Engineering Inc, to develop an on line oil monitor for large transformers. AOC has been involved for several years with the Advanced Research Project Agency developing advanced sensing and signal processing technologies for the detection and classification of fires, flooding, and toxic gases for Navy submarine damage control. AOC was contracted to apply this technology to transformer oil monitoring. The Automated Transformer Oil Monitor (ATOM) has four operational goals:

- Detect the seven gases of interest shown in Table 1

- Differentiate the background gases shown in Table 2

- Maintain suitable accuracy for trending and alarm

- Provide real time data

The ATOM has been developed in a four step process:

- Sensing Technology Assessment

- Environmental Performance Demonstration

- Prototype Testing

- Production

Sensing Technology Assessment: There are three basic ways to monitor gaseous content in oil; measure the gases while they are still in the oil, separate the gases from the oil with permeable membrane technology, or separate the gases using a vacuum. While the vacuum separation method is the "lowest tech" method, it has several advantages:

- This extraction process is used in the laboratory so the on line measurements should compare directly with current or future laboratory tests.

- Adjusting the amount of oil from which the gases are extracted offers a simple way to extend the dynamic range of the sensor suite. If gas concentrations are low, suck from a larger oil volume to increase the amount of gases at the sensor. If gas concentrations are high, reduce the volume of oil to decrease the amount of gases at the sensor.

- The extracted gases going to the ATOM sensors are available for collection and manual analysis. This allows the use of gas chromatography for direct calibration of the ATOM.

After extraction, the gases are detected by a series of fiber optic chemical sensors. The sensor signals are analyzed by a neural network that has been trained to use

TABLE 1

DISSOLVED GASES IN OIL USED TO ANALYZE TRANSFORMER HEALTH

	Threshold Levels To Be Detected (Parts Per Million)		
Hydrogen	100	to	2,000
Methane	20	to	1,000
Carbon Monoxide	200	to	5,000
Carbon Dioxide	3,000	to	10,000
Ethylene	10	to	1,000
Ethane	10	to	500
Acetylene	5	to	200

TABLE 2

MAIN BACKGROUND DISSOLVED GASES

Propane

Propylene

Isobutane

N-butane

Isopentane

N-pentane

N-hexane

the sensor input signal to identify the gases present and their concentrations in parts per million (PPM). After development, the sensor suite was tested using sample gas mixtures of varying known concentrations of the gases of interest. During this series of tests, the ATOM demonstrated accuracy's typically in the 95-99 per cent range.

Environmental Performance Demonstration: With the successful testing of the sensor concept, the next step was to develop a full working ATOM unit to test the gas removal and sampling system operation under real life conditions. This unit was named the "environmental vanguard" (EV) unit since its main function was to demonstrate that the monitoring concept would function as well on a transformer in a real life power substation as it did in the laboratory. The main operational requirements for the EV unit were:

- Maximum reliability under installed conditions

- Little or no routine maintenance

- Easy to install and use

- Low cost for both procurement and life cycle

A block diagram of the ATOM is shown in Figure 1. The ATOM is mounted external to and adjacent to the transformer. It is connected by two pipes or hoses for oil supply and return with a flow rate of approximately 3 gallons per hour. The unit is powered by 120 VAC and reboots after power interrupts. Communications can be completed by RS-232 or telephone line. Incoming oil passes through a heat exchanger to reduce temperature. A metering valve restricts oil flow and a gear pump is used to return the oil to the transformer. A solenoid valve controls flow and a check valve prevents back flow in case of malfunction. The circulating oil passes through a partially filled chamber with the head space evacuated by a vacuum pump. The vacuum pump draws the gases through oil traps and then pumps them past the sensors to atmosphere. The signals from each sensor are merged and then processed.

The EV unit was design to be constructed quickly and inexpensively as a demonstration unit and, therefore, differs slightly from the Figure 1 design. The EV unit has a limited sensor suite instead of the complete sensor suite that the prototype and production versions have. It also has an infrared spectrometer installed to sample gases in series with the fiber optic sensors. The unit is controlled by two 386 version PC's with "temporary" electronics allowing AOC and ConEdison to evaluate the EV ATOM's performance. This EV unit was installed on a transformer at ConEdison in November of 1993.

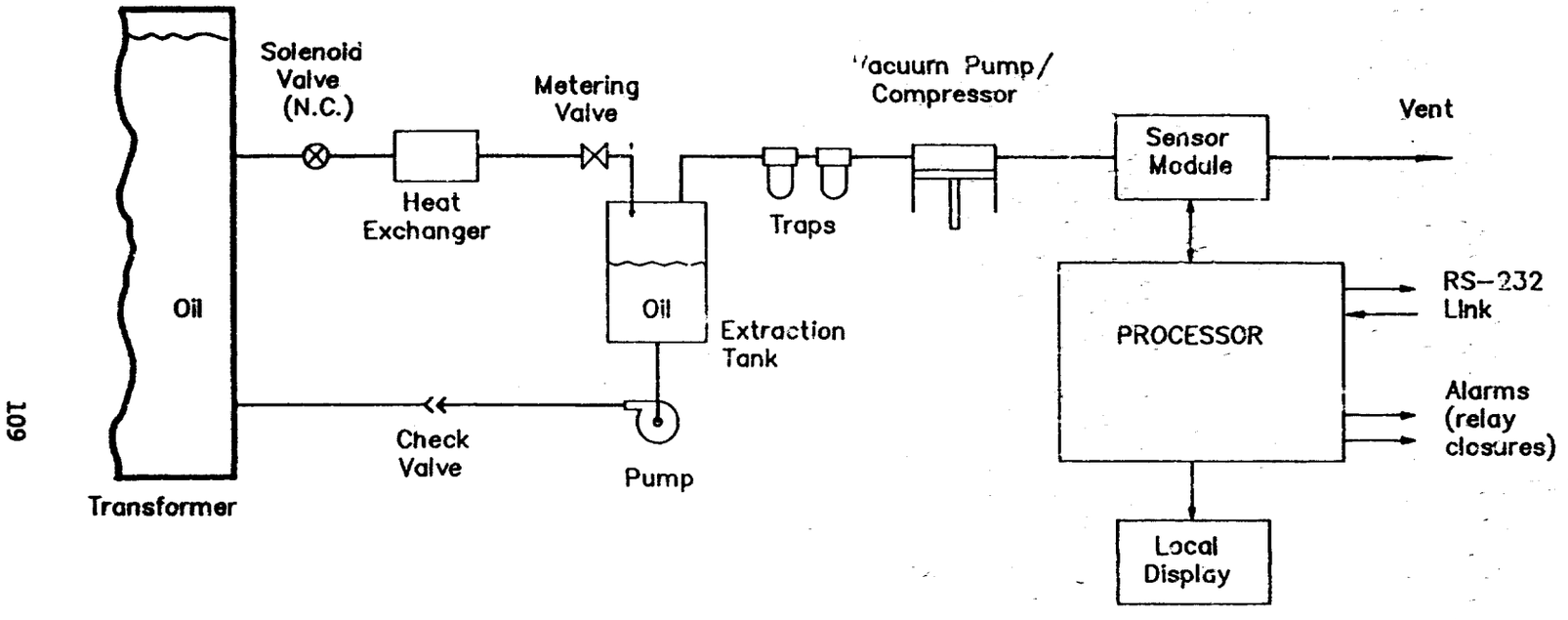


FIGURE 1. ATOM Functional Block Diagram

Prototype Testing: With the EV unit running with no problems identified, the next step was to develop three prototype units to install on transformers at ConEdison. The prototype units have full sensor suites instead of the partial suite if the EV unit. These prototype units do have the IR spectrometer installed, however, and also use a two 386 PC control set up as the EV unit does. The prototype units are sampling their respective transformers every 17 minutes. The data is temporarily stored in the control PC's and downloaded daily to the AOC office for evaluation. The prototype units were installed in January and will operate for a three month test period.

Production Units: The lessons learned and any design modifications which result from the EV and prototype unit operation will be incorporated into a production unit for sale on the open market. It is anticipated that production units will be available in the fall of 1994. AOC and ConEdison are targeting an initial production run of one hundred units for the first year.

Condition Monitoring Application: Setting up a condition monitoring operation for any piece of equipment or system is a four step process:

- Identify Conditions to be Monitored

- Evaluate Technologies to Monitor Desired Parameters

- Develop Performance Criteria

- Prepare a Material Condition Assessment Plan

The identification of transformer oil gases as a desirable condition to monitor was the basis of the ATOM development. The joint research and development effort of AOC and ConEdison to date has concentrated on demonstrating that the ATOM is an effective technology for monitoring transformer oil.

Performance Criteria Development: The data collected by the EV and prototype units, along with historical data from ConEdison will be used to develop Performance Criteria for each transformer to be monitored. Performance criteria development for transformer oil involves the identification of three specific values:

- Initial Performance Value: The concentration of gases which exist in the transformer under normal operating conditions when it has no degradation present. This value may be different for each transformer and may be a function of specific transformer design, transformer load, and/or environmental conditions surrounding the transformer. The

initial performance value represents the baseline for trending and material condition assessment.

- Alert Value: This value represents a warning that either serious degradation is occurring or material condition has reached a serious level. The alert value could be based on concentration of one gas, combinations of gas concentrations, or the rate of change of gas concentrations. The ATOM analysis processor is designed to support evaluation of both concentration and rate of concentration change. The alert value is generally determined to indicate that action is required by the operators or maintenance personnel to prevent a transformer failure. Examples of potential actions include:

-- Decrease power transmission through the transformer

-- Increase the sampling frequency of the ATOM and monitor conditions more closely

-- Plan a repair of the transformer at an appropriate future time.

- Shutdown Value: This value represents an alarm condition where failure of the transformer is eminent. An immediate shutdown of the transformer is required.

The Performance Criteria will be used to set the various alarm functions that are built into the ATOM. The alarm settings may well be different for different transformers.

Material Condition Assessment Planning: Material condition assessment planning involves determining how to trend the variation in concentrations and concentration change rates for the gases being monitored. Some of the issues which have to be resolved include:

- Do concentrations for a given transformer vary based on transformer load or environmental conditions?

- What types of degradations result in concentration changes?

- How predictable is transformer degradation?

The Material Condition Assessment Plan is generally a document that contains all the guidance necessary to monitor the material condition of a system and its components. The document contains the procedures used for monitoring, which would include any ATOM control settings, sampling rates, calibration of the ATOM, and any check or verification samples required (since the ATOM installation supports manual sampling for laboratory analysis). Data storage and retrieval techniques are also covered in the document. For

the EV and prototype units, data is stored locally and downloaded daily to a central facility, the AOC office, for analysis. For production units, the location of the transformer and the current utility maintenance data handling procedures will dictate these decisions. The Plan also documents the analysis techniques used to evaluate all data collected, including any expert systems or artificial intelligence programs. For the ATOM this will mean explanations of how changes in various gas concentrations or changes in rate of concentration build up equate to conditions in the transformer.

Conclusion: The ATOM offers utilities the ability to on-line monitor and trend transformer material condition and provides a cost effective incipient failure detection device for this expensive and critical power distribution component.

FAILURE ANALYSIS

Cochairmen: Robert A. Bayles
Naval Research Laboratory

Gerald Wyen
Wright Laboratories

FAILURE ANALYSIS OF AN INCONEL 718 BARREL NUT FROM AN ARMY ATTACK HELICOPTER

Dr. Kirit J. Bhansali
U.S Army Aviation and Troop Command, St. Louis, MO 63120

Victor K. Champagne
Gary Wechsler
Marc Pepi
U.S. Army Research Laboratory, Materials Directorate,
Watertown, MA 02172

Abstract: During a routine preflight inspection at Ft. Hood, an outboard barrel nut was found to be cracked on an Army helicopter. Subsequent inspections at Ft. Hood and Ft. Rucker revealed an additional seven barrel nuts with large cracks. The components are used in many critical applications. The failures under investigation in this study were relegated to the vertical stabilizer of the aircraft. The failures were all attributed to hydrogen-assisted, stress corrosion cracking. Galling between the unlubricated bolt and the nut threads provided the sustained hoop stress while galvanic corrosion of the carbon steel retaining clip in contact with the barrel nut generated hydrogen as a result of the corrosion process. The microstructure of the material also displayed an acicular grain boundary delta phase which is not as desirable as a spherodized structure. Recommendations were made to utilize a corrosion inhibitive lubricant on the threads of the barrel nut and mating bolt to reduce galling and the consequential high stresses which result from metal to metal contact during torquing. A stress analysis of the part showed that the high strength level of the material could be reduced to increase fracture toughness and resistance to hydrogen cracking. The acicular delta phase should be avoided in accordance with AMS 5662F and the extrusion direction of the material should be parallel to the principal loading direction. Salt fog testing of the proposed barrel nut configuration revealed that the shoulder height base thickness should be increased. Future vendors should qualify their product by conducting a prescribed salt fog test incorporating the prescribed torque requirements. Finally, the material used to fabricate the retaining clip should be changed to prevent galvanic corrosion.

Key Words: Failure analysis, fractography, hydrogen cracking, Inconel 718, stress-corrosion cracking.

Background: During the early 1980's several catastrophic failures of high strength H-11 fasteners were encountered by the aerospace industry. These failures were attributed to service induced hydrogen embrittlement. In response to this problem, the FAA issued an advisory against the use of such fasteners in critical applications. The Army Aviation and Troop Command (ATCOM) in turn issued an engineering order to replace all H-11 fasteners in critical applications on the Apache with those fabricated from Inconel 718, in accordance with an engineering change proposal (ECP) submitted by the primary contractor. The ECP required that the Inconel 718 material be heat treated to 220 ksi minimum ultimate tensile strength and cadmium plated incorporating a vacuum or an electrolytic deposition process. Whereas, the previous design specified the use of a dry film lubricant on both the bolt and

barrel nut, the new ECP did not. Therefore, the torque preload values were recalculated and adjusted from 1,850 to 975 in-lbs. Approximately 8 to 10 months after the ECP was implemented, an outboard barrel nut was found cracked at Ft. Hood, during a routine preflight visual inspection. This incident prompted further inspections at Ft. Hood and Ft. Rucker. An additional seven barrel nuts were found cracked. A total of three cracked barrel nuts were sent to the Army Research Laboratory (ARL) for examination. The first broken part was found during a pre-flight visual inspection at Ft. Hood. The two remaining barrel nuts were detected during visual inspection at Ft. Rucker. All the components were manufactured by the same company (designated within the context of this report as "Contractor A") and were from the same lot number (03). There was also an alternative supplier to these parts (identified as "Contractor B") but no failures were attributed to this company.

In addition to the three failed components examined at ARL, 12 other barrel nut and bolt assemblies were subjected to salt fog testing while loaded. These parts represented three groups. Four parts were fabricated by Contractor A, and four others by Contractor B. The remaining parts represented a newly proposed design which were manufactured by Contractor B. These assemblies were exposed to a salt fog environment over a prescribed period of time or until failure and subsequently examined metallurgically. A series of barrel nuts fabricated by Contractor A were also mechanically loaded to failure over a conical mandrel and the resultant fracture surfaces compared to those of the failed components.

Visual Inspection/Light Optical Microscopy: Figure 1 shows the failed barrel nut assembly from Fort Hood, designated "H" in the as-received condition. The washer, spring clip, cradle and mating bolt have all been identified. A single crack (denoted by arrow) was observed on the barrel nut running perpendicular to the bolt threads. The fracture occurred parallel to the axial loading direction. The cradle was also cracked in two areas. The cadmium plated surfaces of the components displayed typical signs of wear but the shank of the bolt was almost entirely void of the plating from the region located above the barrel nut to the bolt head. The steel spring clip showed a significant amount of corrosion. No other unusual markings were observed.

The crack propagated through the entire thickness of the nut. The company insignia, part number (84209-820) and lot number (03) was still clearly visible. One of the legs of the cradle had actually bent outward from the component, most likely as a result of a post-fracture incident. The two other barrel nut failures from Fort Rucker identified as "R1" and "R2", displayed similar features and were also manufactured by Contractor A, lot number (03). The only significant difference was that both cracks in the cradle of the two Fort Rucker failures were very easily discernable while one of the two cracks in the Fort Hood cradle was barely visible. Corrosion products were found on the surface of the cradle and the nut. The corrosion was reddish-brown in color and most likely originated from the steel spring clip. One of the two raised sections (legs) of the cradle had bent outward and the barrel nut was no longer square because the crack had widened a significant amount.

Examination of the threads on the barrel nuts revealed evidence of metal debris and wear. However, much of the cadmium plating on the threads was still intact. It is believed that some of this damage may have occurred during final fracture. As the crack propagated through the entire thickness of the barrel nut, the bolt is thought to have pulled out during that instant causing the bolt threads to shear. The two Fort Rucker bolt threads experienced shearing while the Fort Hood bolt did not. Some of the galling observed may have also occurred during installation since a lubricant was not used. Wear marks and exposed base material was also observed on the bolt threads that had been inserted into the barrel nut. Frictional forces between the mating threads of the bolt and barrel nut may have been sufficient enough to have caused some excessive wear and stress concentration areas.

Microstructure Analysis: A transverse section (parallel to the fracture plane) was taken from each of the three failed barrel nuts. Figure 2 reveals the extrusion direction of barrel nut R2 which extends parallel to the direction of crack propagation. A transverse section of a Contractor B barrel nut was also examined for comparative purposes and the resultant flow pattern was perpendicular to that of the Contractor A counterpart and would be more desirable under the loading conditions encountered during service. Banding was also observed in both the Contractor A and Contractor B barrel nuts running in the same direction as their respective flow patterns. The Contractor A material displayed heavy banding which consisted primarily of MC carbides and delta phase in the form of a Widmanstatten structure within the grains and an acicular structure at the grain boundaries, as shown in Figure 3. The Contractor B material contained only slight evidence of banding which was comprised primarily of MC carbides as shown in Figure 4. The Contractor A sample contained a large concentration of delta phase in the form of a Widmanstatten structure while the Contractor B material did not. The MC carbides which were aligned preferentially in the extrusion direction seemed to be found in clusters more often in the Contractor A material. The grain boundary delta phase was in the form of an acicular structure in the Contractor A material while the Contractor B material contained a more spherodized structure. Figure 5 is an optical micrograph representative of the Contractor A material at high magnification. The grains are almost completely surrounded by an acicular delta phase whereas the Contractor B microstructure contained a more spherodized grain boundary delta phase which surrounded a significantly smaller percentage of the grains, as depicted in Figure 6. The etchant used during this examination consisted of 100 ml of ethyl alcohol added to 100 ml of hydrochloric acid and 5 grams of cupric chloride.

The grain size was required to be 5 or finer (ASTM E112-63). The grain size of the Contractor A and Contractor B material was measured utilizing this specification. A reference standard illustrating ranges of grain size (1-8) was superimposed on the photographs for ease in comparison. The Contractor A grain size was approximately 8 while the Contractor B material displayed a much finer grain size than that of the reference standard. However, both materials satisfied the specification requirements.

Carbides which form in Alloy 718 are primarily MC type with a nominal NbC composition, although titanium and molybdenum can substitute for niobium in some circumstances. X-ray mapping was performed within the scanning

electron microscope utilizing energy dispersive spectroscopy (EDS) of MC carbides found on both sets of samples. The results indicated a high concentration of Nb and Mo. These carbides being very hard and brittle are usually considered undesirable because they reduce the ductility of the alloy. In addition, they tie up niobium necessary for the formation of the strengthening phases. There was no evidence of Laves phase found in any of the specimens examined. Chemical segregation can result in the formation of large, blocky intermetallic particles known as Laves phase which is also very brittle and adversely affects the properties of the alloy.

A cross-section was taken through the crack origin and prepared metallographically for examination. Figure 7 represents the Fort Hood barrel nut failure. The fracture path appears to be quite intergranular at the origin. The specimen had been polished in relief to highlight the grains.

Chemical Analysis: Material sectioned from a representative failed barrel nut and bolt (both from R1) were subjected to chemical analysis. Atomic absorption and inductively coupled argon plasma emission spectrometry were used to determine the chemical composition of the alloys. The carbon and sulfur content was analyzed by the Leco Combustion Method. The compositional ranges of the material under investigation compared favorable with the published values.

Hardness Testing: A series of microhardness measurements were performed on cross sections of the three failed barrel nuts and a Contractor B barrel nut from inventory, for comparative purposes. Each nut was sectioned in half transversely and metallographically prepared. Knoop microhardness measurements were taken on the surface. The barrel nuts were required to exhibit a hardness of 42 HRC. The failed nut from Ft. Hood displayed an average hardness of 507 HK or approximately 47.8 HRC, while R1 was 515 HK or approximately 48.3 HRC. Specimen R2 was 525 HK or approximately 48.9 HRC and the Contractor B nut from inventory was 510 HK or approximately 48.0 HRC. All values obtained conformed to the governing specification, and were very close to one another.

Tensile Testing: Tensile specimens were sectioned from two failed barrel nuts (R1 and R2), as well as a Contractor B barrel nut from inventory. Substandard specimens were fabricated according to ASTM E8. The specimens were tested in a 20,000-pound capacity Instron universal electromechanical test machine. A 10,000-pound load cell was utilized for the measurement of load. The pull rate was 0.05 inches/minute. Testing was conducted at room temperature at 50% relative humidity. The ultimate tensile strength was required to be 225,000 psi minimum, and 245,000 maximum. The minimum per cent elongation was specified as 8 in 4D. The results of this testing are listed in Table 1.

Table 1
Tensile Test Results

Sample	Diameter inches	Area in^2	Maximum Load pounds	UTS psi	%El
R1	0.066	0.0034	800	233,836	17.6
R2	0.078	0.0048	1,050	219,741	*
Contractor B	0.066	0.0034	705	206,068	16.0

* was not calculated

Fractographic Examination: The three failed barrel nuts were sectioned to separate both fracture halves for examination. Optical and scanning electron microscopy was utilized to identify important features of the fracture surface relating to the failure mechanism. The Fort Hood barrel nut was determined to be representative of all three service failures and will be used within this discussion to describe the manner of crack propagation. Cracking occurred parallel to the axial loading direction or perpendicular to the threads. A discolored region was observed on the fracture surface near the crack origin. Energy dispersive spectroscopy (EDS) was utilized to characterize the chemistry of the surface within this area. The elements associated with the base material were detected as evidenced by Ni, Fe, Cr, Ti, Nb, Mo and Al peaks. A large Cd peak and small O and Cl peaks were found and most likely represented corrosion product from one of the barrel nut assembly components that had been cadmium plated. Corrosion products of this cadmium plated part and the underlying base material could have seeped into the slowly propagating crack in the barrel nut causing a discolored zone. The reddish brown appearance of this stained region suggests that the corrosion products on this surface were rich in iron and may have originated from a steel component in contact with the barrel nut, such as the spring clip. Even though the EDS analysis supports this deduction an argument can be made that the Fe, Cd and O peaks may represent only the barrel nut material.However, the reddish brown appearance of the discolored region indicates corrosion of an iron rich material. The remaining elements are believed to be contaminants, although Si, S and Ca are acceptable within this alloy in very low concentrations. Figure 8 is an optical macrograph highlighting the fractographic features of the surface. The chevron markings and river pattern converge to an area identified as the crack initiation site as shown by the arrow.

Figure 9 is a schematic illustrating a cut-away section of the barrel nut, bolt and cradle showing the location of the fracture origin observed on all three service failures. Cracking had initiated near the first thread of the barrel nut adjacent to the cradle. Figure 10 is a schematic illustrating the four distinct zones noted on the fracture surfaces. The fracture originated within the discolored region identified as Zone 1. This region was relatively free from smearing which can occur if the two fracture faces rub against one another in service, but contained surface contamination and debris. Cracking had occurred in a brittle manner which was discernable, even at relatively low magnification. Zone 1 was an area of slow crack growth showing little ductility and the morphology was intergranular as shown in

Figure 11. Zone 2 was a region where fracture by intergranular decohesion transformed to a more dimpled mode, as depicted in Figure 12. The fracture mode observed in Zone 3 consisted primarily of equiaxed dimples which is associated with ductility. Zone 4 consisted of shear lips adjacent to the outside perimeter which displayed directional dimples. Crack propagation occurred at a faster rate within Zone 3 and final fracture took place within Zone 4.

Further EDS analysis was performed within the various zones of fracture. The spectrum obtained from Zones 3 and 4 revealed primarily those constituents associated with the base material. There was no significant Cd, Cl, or O peaks detected. A very large concentration of Cd, had been detected along with Cl and O near the crack origin site. The corrosion product observed on the surface of the barrel nut cradle was also analyzed by EDS. A significant concentration of Cl was detected. The Cd peak in this case primarily represents the plating on the surface of the cradle. The small Fe and O peak may represent corrosion product from the steel spring clip or the base material of the cradle. Fracture Zones 3 and 4 primarily contained those elements associated with the base material. No significant concentration of Cd or Cl was detected.

Mandrel Testing: A series of Contractor A barrel nuts were subjected to mechanical testing. The barrel nuts were mechanically loaded to failure over a conical mandrel. Subsequent fractographic analysis of the failed barrel nuts revealed surfaces which were entirely dimpled. This type of fracture morphology is indicative of a ductile failure. Since the fracture surfaces of the failed barrel nuts displayed significant intergranular regions at the crack origin, it became evident that initial crack growth may have been environmentally assisted.

Salt Spray Testing:

Procedure

A series of three salt spray tests were conducted on Contractor A , B and newly designed barrel nut assemblies, in an effort to reproduce the failure mechanism and to compare the three barrel nut designs. The differences in design included the height of the conical section , the thickness of the base and the hardness of the material .The assemblies were tested in an actual aluminum fuselage section. Bolts were inserted through a block of 6061-T6 aluminum, which represented the rear vertical stabilizer of the helicopter. Some of the blocks were machined with a 0.6 degree taper, while other blocks were flat and parallel. The taper was designed to induce a bending moment within the nut to replicate the effects of having a bolt inserted at an angle causing a stress concentration and galling when torqued. The bolts were threaded into the barrel nut assemblies and were then torqued and placed in a salt spray chamber. The bolts were incrementally torqued to higher stress levels at regular intervals following a schedule determined by ATCOM. The chamber maintained a salt fog with 5% NaCl by weight, at 118.7 ^{o}F, and 100% relative humidity. During the torquing process, the assemblies were removed from the chamber for no longer than five minutes. Torquing was conducted utilizing a digital torque wrench, having a capacity of 250 ft-lbs.

Results

Two Contractor A barrel nuts broke while being salt spray tested and the resultant fracture surfaces revealed similar features as the failed components (intergranular fracture at the crack origin and ductility in the area of final fracture). However, the region of intergranular fracture within the barrel nuts broken in the salt spray chamber was smaller than that of the three field failures. This was attributed to the higher torque levels prescribed during the laboratory salt fog test as opposed to the lower values specified in service. It is believed that higher stresses would result in faster initial crack growth. The Contractor B and newly designed barrel nuts did not experience failure as did the Contractor A nuts tested under identical conditions. The use of tapered blocks to cause a slight mismatch and galling of the threads on the barrel nut and mating bolt did not contribute to premature failure of loaded barrel nuts during exposure to a salt spray atmosphere.

Discussion: The data obtained from examining and comparing the three Contractor A barrel nuts that failed in service, the two Contractor A barrel nuts that failed in the salt spray chamber test, the Contractor B barrel nuts from inventory and the proposed redesigned barrel nuts that did not fail the salt spray chamber test indicated that part geometry, microstructure, heat treatment, mating materials, lubrication and environment all contributed to premature failure. The smaller cross-sectional area at the conical threaded region (shoulder) of the Contractor A barrel nut contained less volume in which to distribute external loads than the Contractor B and the newly designed barrel nuts. This resulted in higher stresses at this area during service. Another factor contributing to regions of high localized stress would be the installation of the bolt into the barrel nut assembly. The wear and metal debris observed within the threaded section of the failed barrel nuts may have been attributable to improper seating of the bolt while torquing. Misalignment of the bolt caused by an uneven mating surface beneath the bolt head could result in surface galling during torquing. This condition could cause highly stressed areas in the barrel nut. Lubricity is also an important issue when uniform torque is required. The use of a corrosion inhibitive dry film lubricant would have helped to insure a uniform and consistent stress distribution within the barrel nut after torquing. High frictional forces as a result of metal-to-metal contact or parts that have a damaged cadmium plating could be avoided. Therefore, the risk of failure due to stress corrosion cracking would be greater in the Contractor A barrel nuts.

The higher stresses in the Contractor A barrel nuts were compounded further by a higher strength material and an undesirable microstructure. The Contractor A material contained an unacceptable acicular delta phase in the grain boundaries and in the form of a Widmanstatten structure throughout the grains. The ultimate tensile strength of the two Contractor A barrel nuts that failed in service measured higher than the ultimate tensile strength of a Contractor B barrel nut. The higher strength of the Contractor A barrel nuts increased its susceptibility to hydrogen absorption and decreased its toughness in comparison to the Contractor B barrel nut at the lower strength level. The acicular delta phase observed metallographically in the Contractor A barrel nuts is considered a greater stress concentrator than the spheroidal delta phase found in the Contractor B barrel nuts. The delta phase is necessary to inhibit excessive grain growth in Inconel 718, but

it is a brittle phase and decreases the toughness of the material. Metallography also revealed MC carbides in both versions of the barrel nuts. These MC carbides are brittle (many were cracked when examined as a result of prior forming) and are stress concentrators. These carbides were preferentially orientated in the direction of extrusion. The Contractor A material also contained a high concentration of Widmanstatten structure within the regions of banding. Banding in the Contractor A barrel nut occurred parallel to the plane of fracture, which is undesirable. Banding in the Contractor B barrel nut ran in the direction of extrusion perpendicular to the Contractor A fracture plane. The carbides in the Contractor A material seemed to be found more in clusters than in the Contractor B material. Together, the higher tensile strength, the acicular delta phase, and the banding decreased the toughness of the Contractor A barrel nuts.

Inspection of the Contractor A spring clips revealed that the cadmium plating and the steel clip were completely corroded. The cadmium plating of the spring clip was either scratched during installation or corroded during service eventually leaving the steel spring clip unprotected. Corrosion of the spring clip plating and the unprotected steel clip occurred when water was entrapped in the barrel nut receptacle machined in the aluminum fuselage. Oxidation occurred at the steel clip surface while monatomic hydrogen was produced at nearby cathodic sites as part of the electrochemical corrosion process. The high state of stress, grain boundary chemistry (acicular delta phase) in the Contractor A material is thought to have increased the susceptibility of the Contractor A barrel nut to hydrogen charging. In addition, nickel is cathodic to steel in the presence of ionized water. The hydrogen would have preferentially migrated to sites of highest stress and disorder in the grain boundaries. Fractography of the Contractor A barrel nuts that failed in service showed that intergranular decohesion occurred at the crack initiation sites suggesting that cracking in these areas was environmentally assisted. Laboratory testing of Contractor A barrel nuts in a salt spray environment caused intergranular fracture at the crack origin. When the presence of an aggressive environment was removed and the Contractor A barrel nut was forced over a conical mandrel, as in overload testing, fracture occurred in the same plane as the service and salt spray failures, but cracking occurred entirely in a ductile fashion, and there was no initiation region of intergranular decohesion. Instead, the fracture initiated and propagated in a ductile dimpled mode. Therefore, the salt spray chamber and conical mandrel testing showed that fracture of the Contractor A barrel nuts in service would not have occurred in an intergranular mode if an aggressive environment was not present. Further evidence of the presence of an aggressive environment in the service failures of the Contractor A barrel nuts was found by EDS on the surface of the FT. Hood cradle and fracture surface of the barrel nut at the crack origin. Characteristic X-rays of chlorine, calcium and iron were collected that indicated chlorine and calcium ions were most likely present in water that caused the oxidation of iron (most likely the steel spring clip), both of which collected on the cradle surface as the water evaporated or drained out of the barrel nut receptacle in the aluminum fuselage.

According to the original engineering drawing, the only requirement placed on the barrel nut was its ability to withstand a specific load of 37,800 pounds. Each manufacturer had considerable latitude in designing the conical threaded region (shoulder). Contractor A chose to design this region

thinner than Contractor B, yet they both met the original drawing requirements. Based on salt fog tests and microscopic examination of the fracture surfaces, the following hypothesis was developed. Inconel 718 alloy, being a nickel base alloy, is highly prone to galling. Absence of any lubricant during installation causes galling debris to wedge between the threads and in turn provides a source of constant hoop stress in the nut. The presence of a steel clip used for alignment purposes created a galvanic coupling in which the nut became a cathode and the clip the anode. As a result, hydrogen was being charged into the nut during service. Consequently, after a period of time, a service induced hydrogen stress crack developed and grew until the hoop stress overloaded the remaining material. Since the failure was not in the thread, any sustained stress from the preload was ruled out as a potential cause. Since Contractor B nuts did not fail, the magnitude of the hoop stress was suspected to be important. However, the testing conducted to evaluate the effect of misalignment showed no significant detrimental effect.

Conclusion: The failure of the barrel nuts was attributable to hydrogen-assisted stress corrosion cracking. Galling between the unlubricated bolt and the nut thread provided the sustained hoop stress. Galvanic coupling with the carbon steel retaining clip provided the source of hydrogen necessary for this mode of failure in the cold worked and aged microstructure.

Recommendations: 1. A minimum height for the threaded conical section (shoulder) of the barrel nut should be specified and the thickness of the base increased to reflect the new proposed design. 2. A corrosion inhibitive lubricant should be used on the threads of the barrel nut and/or mating bolt to reduce galling and the consequential high tensile stresses which result from metal-to-metal contact during torquing (torque values would have to be re-calculated when using a lubricant). 3. The strength of the barrel nut should be reduced, if deemed feasible by stress analysis to increase the fracture toughness and resistance to stress corrosion cracking of the material. 4. The acicular delta phase banding should be avoided in accordance with AMS-5662F. 5. The carbon steel spring clip should be coated or fabricated from such material as stainless steel to avoid galvanic corrosion with the barrel nut and the resultant mitigation of atomic hydrogen. 6. The extrusion direction should be aligned parallel to the principal stress.

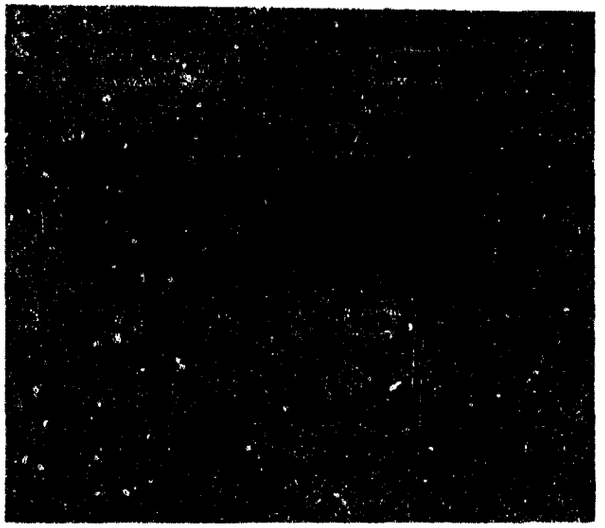


Figure 1 Barrel nut assembly "H" in the as-received condition.

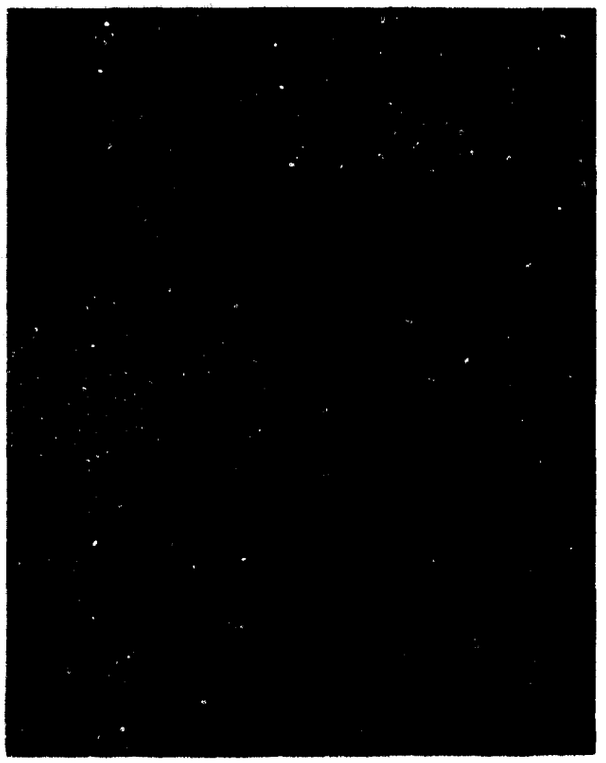


Figure 2 Extrusion direction of nut "R2", parallel to crack. Mag. 20x.

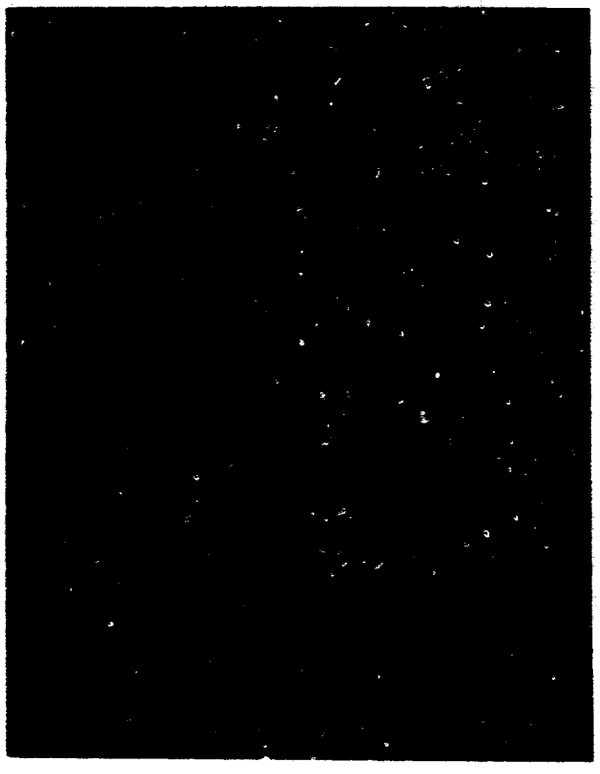


Figure 3 Nut "H" microstructure showing banding and Widmanstatten delta phase in the extrusion direction. Mag. 200x.

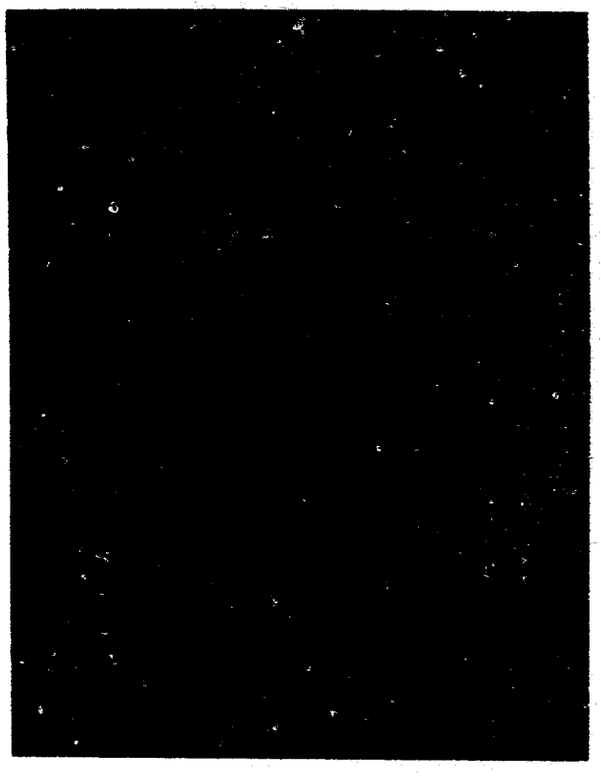


Figure 4 Contractor "B" nut microstructure showing slight banding and MC carbides in the extrusion direction. Mag. 200x

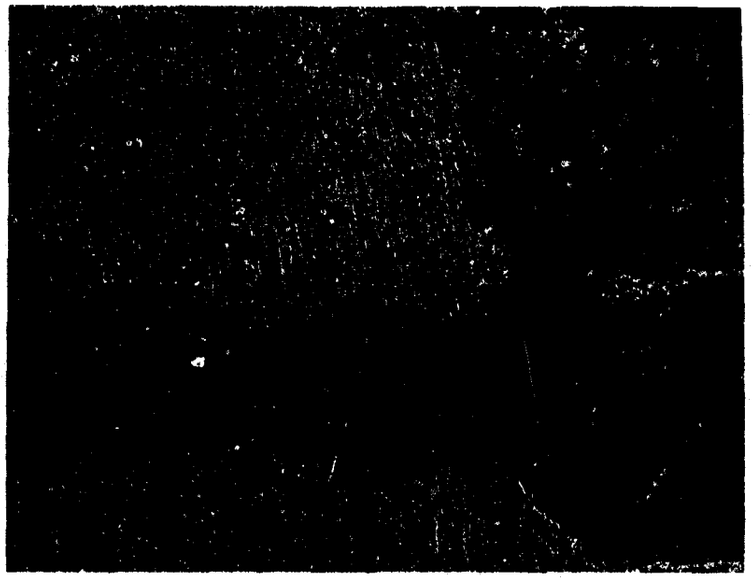


Figure 5 Contractor "A" microstructure displaying acicular delta phase at grain boundaries. Mag. 1,000x.

Figure 6 Contractor "B" microstructure containing a spherodized grain boundary delta phase. Mag. 1,500x.

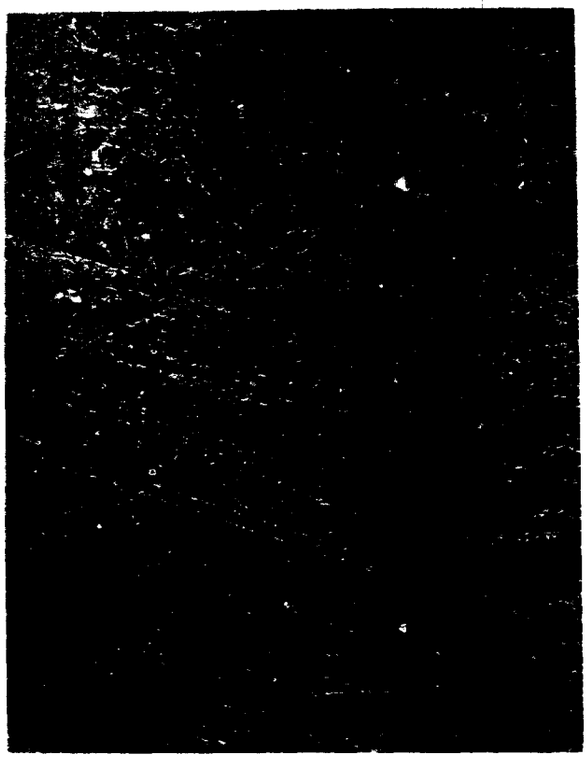


Figure 7 Cross-section through crack origin of nut "H". Note the intergranular fracture path. Mag. 400x.

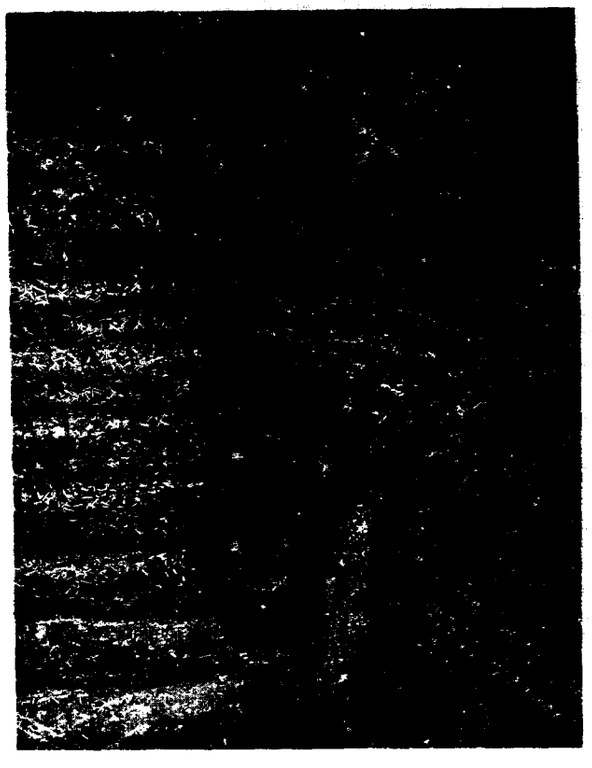


Figure 8 Fractograph of nut "H" showing crack origin and radial markings. Mag. 10x.

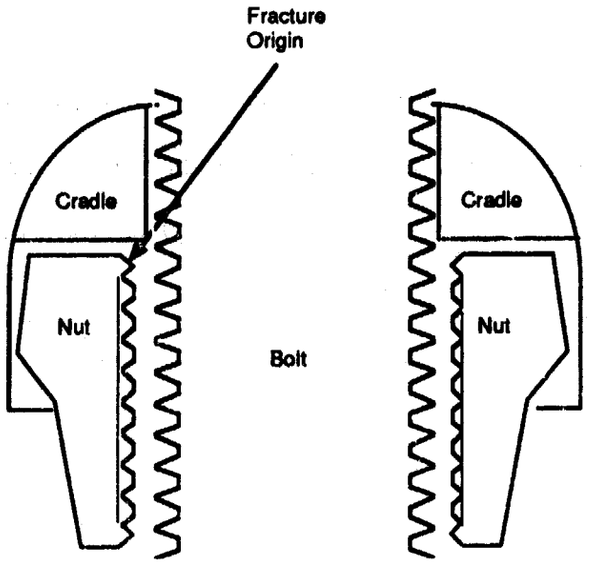


Figure 9 Cut-away section of the barrel nut, bolt and cradle showing the fracture origin observed on all three service failures.

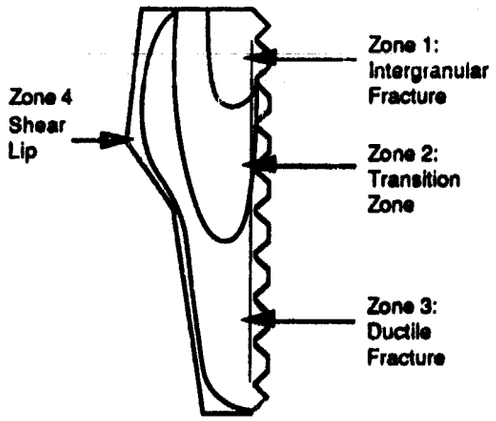


Figure 10 Schematic illustrating fracture zones observed on each failed nut.

Figure 11 Intergranular mode of fracture found in Zone 1. Mag. 1,000x.

Figure 12 Fracture indicative of Zone 2 consisting of intergranular decohesion and ductile dimples. Mag. 1,000x.

ANALYSIS OF CRACKING IN PRESSURE VESSEL STEEL

MarjorieAnn E. Natishan
Robert L. Tregoning
Fatigue and Fracture Branch
Carderock Division of the Naval Surface Warfare Center
Annapolis, MD 21402

Abstract: *Cracking occurred in two pressure vessels during routine flame cutting operations causing concern regarding the integrity of the previous fabrication welds as well as for the safety of the welders during cutting of subsequent pressure vessels. In both cases fracture occurred near the end of the cutting process with only a small ligament of material remaining in front of the cutting torch. Considerable elastic energy was associated with both fractures. Metallurgical analysis of the fractures concluded that there were no weld or other defects present which would have initiated the fractures. Evaluation of the mechanical properties of the welds and base material showed no evidence of deficiencies in mechanical properties which could account for the unexpected fractures. The fracture mode in both instances was identified as Mode I, microvoid coalescence indicating a tensile (or bending) overload failure. Analysis of the stresses present during the cutting process showed that the stresses generated during the cutting process itself, combined with possible residual stresses from the original fabrication of the vessel could be sufficient to drive a crack in this material. Thus the cause of failure was found to be overloading due to the combination of applied stresses during the cutting process and the tensile residual stresses created during the original vessel fabrication process. Variation in the cutting process to eliminate large stress build-up in small remaining ligaments was recommended to eliminate this type of high energy fracture in the future.*

KEY WORDS: Failure analysis; HY-80; Pressure vessel; Weld fracture

Introduction: During routine maintainance of a pressure vessel, in which a large section was being flame cut for removal, a crack initiated ahead of one of the flame cutting torches, propagated several inches and then arrested leaving a ligament several inches in length still connecting the patch to the rest of the pressure vessel. The energy released upon fracture was sufficient to cause the pressure vessel to "shudder." Both the location of the fracture (pressure vessel girth weld) and the energy associated with it caused concern as to the possibility of this type of fracture occurring during service or on other pressure vessels during flame cutting procedures putting personnel at risk for injury. Subsequent to the first weld fracture, another fracture occurred during flame cutting along a girth weld in a second pressure vessel despite the fact that an entirely different cutting procedure was followed. In order to address the safety concerns raised by these failures it was necessary to determine whether the cause of fracture was metallurgical in nature (i.e. a weld defect which would impact service behavior) or mechanical (which would impact future patch

removal operations). The Fatigue and Fracture Branch of the Carderock Division of the Naval Surface Warfare Center was tasked with identifying the mode and cause of fracture of these patch welds. Since the failures are similar, only one will be discussed here.

Background: In November of 1992, a large pressure vessel of HY-80 (high yield (80 ksi)) steel was being flame cut along a previously fabricated HY-100 weld to remove a large access patch. As the patch was being machine flame cut along a circumferential butt weld a four inch section of the weld ahead of the cutting machine fractured. Cutting was continued after the fracture occurred to complete the last 4 to 5 inches of the cut. No pre-existing cracks or defects had been observed by the welders during preparation of the vessel for flame cutting. The cutting sequence used is shown in Figure 1 with the cuts numbered in the order made. The supporting frames were left intact during this procedure. Relief cuts 1 through 4 were made first using a hand held torch. Cut 5 was then started using a machine. When cut 5 was about half complete, cut #6 was started in a direction opposite to that of #5. Upon completion of cuts 5 and 6, cut #7 was started on the bottom end of the pressure vessel. When cut #7 was approximately half complete, cut #8 was started, again in a direction opposite that of #7. When cut #7 was approximately 8 or 9 inches from completion, the weld in front of the cutting machine fractured for a length of about 4 1/2 inches leaving a remaining ligament of about 4 inches. After fracture this ligament was also cut to complete the sequence. When the cut was completed, a triangular piece of the butt weld containing one fracture surface had completely broken off from the remaining patch piece. This piece was removed, photographed, and then given to CDNSWC for analysis. A "D" piece was cut from the patch side of the fracture which contained the other fracture surface, about 3 inches of weld on either side of the fracture and some plate material for analysis by CDNSWC. A section of the weld opposite the fracture on the vessel side of the flame cut was also removed and sent to CDNSWC for analysis. The "D" piece and the weld piece are shown relative to the patch in Figure 1.

The welders were interviewed after the fracture occurred to learn the details of the sequence of events surrounding the fracture and to find out what they observed around the time of the fracture. The observations made by the burners and other personnel included:

1. The patch being removed had moved away from the vessel slightly during cutting (not unusual),
2. The air temperature at the time of cutting was approximately 50° F to 60° F,
3. The crack advanced from the torch towards the weld fusion line and then curved back towards the center of the weld,
4. No pre-existing cracks were observed by the burners preparing the vessel or by the welders hanging the burn boxes along the inside of the weld,
5. The "pop"/crack was of sufficient energy to shake the entire pressure vessel,
6. The bottom piece of the fractured weld metal was missing after flame cutting and is suspected to have burned up during flame cutting,
7. Nothing else unusual was noted about the cuts or the plates.

Procedure: Analysis of this failure included detailed fractographic

characterization of the fracture mode, metallurgical determination of crack path and microstructure in the region of fracture, evaluation of mechanical properties of the weld and base plate and bulk chemical analysis. Careful examination of the patch, prior to removal of the "D" piece, was made to note any deformation in the region of fracture and to look for indications of other defects. Both fracture surfaces were examined in the scanning electron microscope (SEM) to characterize the fracture mode and to identify a fracture initiation site if possible. Energy dispersive x-ray spectroscopy (EDS) was performed on any metallurgical features found on the fracture surface for identification.

Metallographic examination was performed on fracture cross-sections to look at the fracture path relative to the weld and base plate microstructures, and on the weld to characterize any weld defects and to characterize the weld process as much as possible. The metallurgical analysis involved polishing samples down to 1/4 micron diamond polish and then etching lightly with nital for high magnification, microstructural analysis in the optical microscope, or heavily for macroscopic characterization of weld bead geometry, and fracture path relative to the weld microstructure. The top surface of the fracture coupon (the piece that fell out upon completion of the flame cut) was polished, etched and examined along the its length to look for evidence of metallurgical inconsistencies which might define a repair of the original butt weld. Microhardness traverses were performed on the weld fracture surface cross-sections from the weld centerline to the fracture surface to characterize hardness and strength in these areas to compare them to expected values for the HY-100/110 weld and HY-80 base plate material.

Mechanical properties were measured for specimens removed from the weld and base plate material. Testing included longitudinal, all weld metal tensile properties, transverse weld metal tensile and Charpy V-notch impact properties. Tensile testing was performed following the procedures outlined in ASTM E8 [1] and Charpy impact testing followed the procedures outlined in ASTM E23 [2].

Chemical composition was determined for the weld and base plate material using wet chemical spectroscopy techniques.

Results and Discussion:

Visual Inspection of the Pressure Vessel and Patch Regions Adjacent to the Fracture: **Visual inspection of the removed patch, with fracture surface intact, was made prior to removal of the "D" piece. The only deformation observed on the patch was adjacent to the fracture at the toe of the butt weld. At the center of the fracture, where it came closest to the weld toe, there was a buckle (ridge of localized deformation) observed at the HAZ, base plate junction, indicating that some of the energy associated with fracture was absorbed by deformation of the base plate as expected in this weld system. No other defects were observed on the patch surface.**

Deformation of the base plate in the region of fracture indicates that the base plate material deformed plastically to absorb some of the applied stress consistent with design expectations. The philosophy behind use of an overmatched weld (making a weld of higher strength than the base material) is to ensure that any applied plastic stresses are

absorbed in the higher toughness base plate material where damage is more likely to be contained by localized yielding. Weld properties tend to be lower in toughness and less consistent than base metal properties and thus less predictable.

Visual Inspection and Fractography: Visual examination and SEM fractography was performed on both fracture surfaces to determine the fracture mode and identify and defects or features that may have caused or contributed to fracture. The weld coupon that "popped" out upon completion of the flame cut examined first. This piece was triangular in shape, following the profile of the upper half of the butt weld as shown in Figure 2. Visual inspection of this piece showed that fracture initiated in the area of the flame cut at the centerline of the weld, curved out towards the fusion line and traveled along the weld fusion area for several inches before turning back into, and arresting at the center of the weld. The fracture path in relation to the patch cutting sequence is showed in the schematic in Figure 1. Exact initiation and arrest information was obliterated by completion of the flame cut in this region. Weld bead outlines were visible on the fracture surface of this piece suggesting that the fracture path followed the morphology of the weld beads, Figure 2. The weld bead morphology was also apparent at low magnification within the SEM.

At higher magnification the fracture mode was observed to be almost entirely microvoid coalescence, Figure 3. The shape of the microvoids was consistent with mode I loading (in tension) as if the two plates were being pulled apart. This was confirmed when the opposing fracture surface was characterized. Several isolated regions of cleavage were identified, Figure 4, but these were small and not associated with fracture initiation. These small regions of cleavage were most likely due to chemical inhomogeneities typical in the near-weld fusion zone region. Several small lack-of-fusion weld defects were also observed at the base on several of the weld beads at the center of the fractured piece, but again these were not associated with fracture initiation primarily due to their small size and to their location within the fractured piece (at the center).

Energy dispersive x-ray spectroscopy (EDS) of the fracture surfaces showed some manganese sulfide inclusions and carbides at the bottom of dimples which acted as microvoid nucleation sites. But no tramp elements were identified which would have degraded toughness or contributed to fracture initiation. The only areas which were inconsistent with the bulk chemistry were the several regions of weld defects described above. EDS of these areas showed oxides and silicates consistent with "dirty" welds. But these areas were localized and small relative to the weld beads and were not thought to contribute to fracture initiation.

Metallography: In an effort to identify the fracture path relative to the weld, HAZ and base plate, the triangular coupon from the fracture was polished and etched on the top, plate surface and then sectioned perpendicular to the fracture surface, polished and etched. Figures 5 and 6 show the top view and side view of the triangular section of the fracture, respectively. The white etching regions to the left of both photographs were caused by transformation of the microstructure during flame cutting. The weld beads are apparent on both the top and side

surfaces. In Figure 6 it can be seen that the fracture path follows the weld bead morphology as suspected from observing the fracture surface. It is not clear whether the fracture path is along the weld fusion line or following a line between weld beads. To confirm the fracture path relative to the weld, the other half of the fracture (from the "D" piece cut from the patch) was also polished and etched, Figure 7. Another weld bead can be seen to the outside of the fracture on this piece indicating that fracture in this case ran within the weld.

Microhardness of the weld, weld fusion zone, HAZ and base metal in this region showed the HAZ to have the highest hardness, followed by the fusion zone, then the weld and the base plate with the lowest hardness. These hardnesses are typical of overmatched weld systems and would suggest the base plate as the region in which strain would be localized. The fact that the fracture did not propagate into the base material is not totally unexpected and has been observed in other specimens of overmatched weld systems [3].

Mechanical Properties: Tensile properties were measured for the pressure vessel butt weld in both the longitudinal and transverse directions. The transverse weld tensile properties were:

yield stress	108,270 psi
ultimate tensile stress	118,400 psi
%EL	16%
%RA	56%

The longitudinal weld properties were:

yield stress	113,200 psi
ultimate tensile stress	119,200 psi
%EL	19%
%RA	60%

The welds were made according to the specifications described in MIL-E-22200/1D. The properties shown above for the welds are not consistent with those called out in MIL-E-22200/1D for MIL-11018 weld electrode. The transverse yield and ultimate tensile strength falls within the specified yield strength for 11018 weld deposit (98 ksi - 110 ksi) but the elongation measured in this direction is slightly below the required 20%. In the longitudinal direction the yield stress measured from the weld was above the range specified for 11018 and the elongations was slightly belowe the specified 20%. This property variation could be attributed to possible compositional inhomogeniety.

Charpy V-notch impact properties were also measured for the weld in the T-L orientation to coincide with the direction of fracture. These values are presented in Table 1 below. The values at -60° F are above the specified minimum at -60° F in MIL-E-22200/1D for 11018 weld electrode.

Specimen	Test Temperature	Impact Energy (ft-lbs)	Ave. Value
T1	0° F	30	
T2		41	
B1		33	35
B2	-60° F	21	
B3		22	
T3		28	24

Table 1. Charpy impact energies for weld metal specimens tested at 0° F and -60° F.

Chemical Composition: Bulk chemistries were determined for the base plate and weld for the pressure vessel patch. These chemistries are listed in Table 2.

	Pressure Vessel		MIL 11018 Weld Specification	HY-80 Specification
Element	Weld	Base	Max. or Range	MIL-E-16216G
C	0.096	0.153	0.10	0.18
Mn	1.14	0.40	1.30 - 1.80	0.10-0.40
P	0.008	0.012	0.03	0.025
S	0.018	0.014	0.03	0.025
Si	0.30	0.32	0.60	0.15-0.35
Ni	1.96	2.72	1.25 - 2.50	2.00-3.25
Cr	0.79	1.42	0.40	1.00-1.80
Mo	0.38	0.40	0.25 - 0.50	0.20-0.60
Va		0.002	0.05	0.03
Ti	<0.002	<0.002	0.10	0.02
Cu	0.060	0.022		0.25
Sb		0.003		
Arsenic		0.004		
Sn		0.010		
Al	0.009		0.01	
Zr	<0.002		0.01	

Table 2. Chemical compositions of weld and base plate for material adjacent to the fracture surface.

Manganese is slightly below the minimum amount specified for 11018 weld metal in MIL-E-22200/1D. Manganese contributes to solid solution strengthening and toughness in HY steels. The lower than specified amount of this element in the weld metal might explain the reduced CVE measured from the weld specimens.

Stress Analysis: Analysis of the stresses required to cause cracking under these conditions was performed. For the cutting procedure employed it was assumed that the stiffeners acted as a rigid constraint which not only supported the plate's weight, but also prohibited displacement of the plate at the stiffener locations. The patch and vessel surrounding the flame cut was modeled as an edge-cracked panel under tensile loading where the plate height is the distance between plate stiffeners, the plate width is the total length of cut #7. This

modeling also assumed that the flame cut simulated a crack in this plate. Any added stresses due to plate curvature and transverse constraint were neglected and the effect due to weld mismatching was also ignored. The fracture toughness (KIc) of the weld wire was assumed to be 120 ksi(in)$^{1/2}$. For this scenario, the tensile stress needed to drive the crack is less than 500 psi. This low stress indicates that virtually any positive residual stresses which exists would be sufficient to cause fracture of the patch.

The tensile properties measured for the weld was consistent with a higher tensile strength weld as typically used in overmatched welding systems. The philosophy behind the use of overmatched welds is to force any plasticity out into the base plate causing this to be the region of highest stresses. This ensures the integrity of the lower fracture toughness, less homogeneous (less predictable), weld metal. Deformation and fracture occurred primarily in the base metal due to an overload situation caused by the residual stresses across the weld in conjunction with the strain applied by the patch shrinking away from the hull during the cutting process. While deformation occurred in the base metal during fracture of the actual fracture path remained entirely within the weld. Several, small weld defects and regions of cleavage were noted within the weld and the Charpy impact energies were below the minimum specified energies for this type of weld. These factors were probably more influential in determining crack path than in crack initiation site. Fracture initiation occurred just ahead of the cutting torch because this was the point of highest stresses.

Summary: Fracture mode was Mode I MVC which indicates a tensile, overload fracture. Few defects were noted and these were not believed to contribute to the cause of failure. Several, localized regions of cleavage were observed on the fracture surface indicating that this material was in upper transition at the fracture temperature, but no cleavage was observed in the fracture initiation region and cleavage is not believed to be a contributing factor to fracture initiation.

The microstructure for both weld and plate material was martensitic, consistent with HY 80/100. The fracture path was within the weld following the weld bead outline. While this is not typical of overmatched welds, it has been observed in other specimens and is not believed to be cause for concern [3]. The shape of the weld beads indicate that the original welding was performed using a SMAW weld process (stick electrode), but nothing was found to indicate that repair welds had been made previously in the fracture location.

The mechanical properties of the weld was within those specified for 11018 electrodes.

Stress analysis of the loading conditions show that as little as 500 psi tensile residual stress across the weld would be required, in addition to the stresses caused by the cutting process itself, to cause fracture. It is not unexpected to have tensile residual stresses much higher than this across this type of weld.

Conclusions: The fracture in the pressure vessel patch weld is believed to have been caused by the stresses created by the cutting process in conjunction with the tensile residual stresses across the weld.

The weld metal showed below specified amounts of manganese but properties were marginally as specified.

Stress analyses of the loading condition indicates that as little as 500 psi tensile residual stress across the weld, in addition to the stresses created by the cutting process, were required to initiate the type of fracture observed.

REFERENCES

1. "Standard Test Methods of Tension Testing of Metallic Materials," E8-91, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, PA, Vol. 03.01, pp 130-149

2. "Standard Test Methods for Notched Bar Impact Testing of Metallic Materials," E23-92, Annual Book of ASTM Standards, American Society for Testing and Materials, Philadelphia, PA, Vol. 03.01, pp 205-224

3. DeLoach, J.J., Jr. Unpublished work on thick plate welding of HY-80 steel, David Taylor Research Center, March 1991

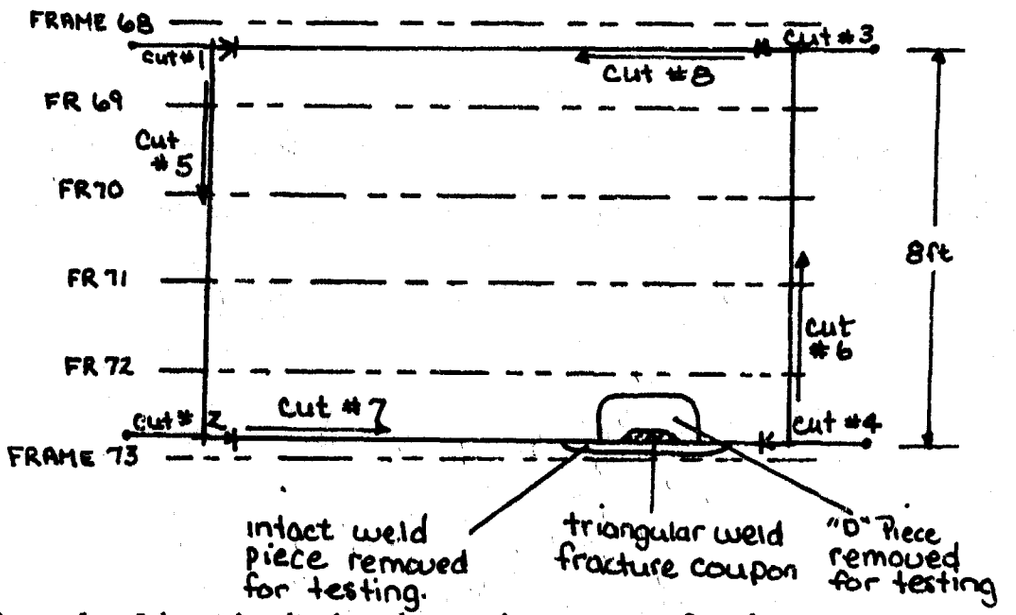


Figure 1. Schematic showing the cutting sequence for the pressure vessel, access patch removal.

Figure 2. Triangular-shaped fracture piece showing the weld bead morphology.

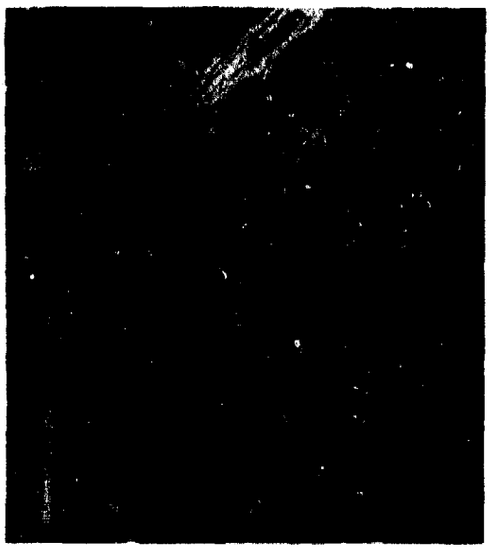


Figure 3. SEM fractograph showing the microvoids typical of Mode I fracture.

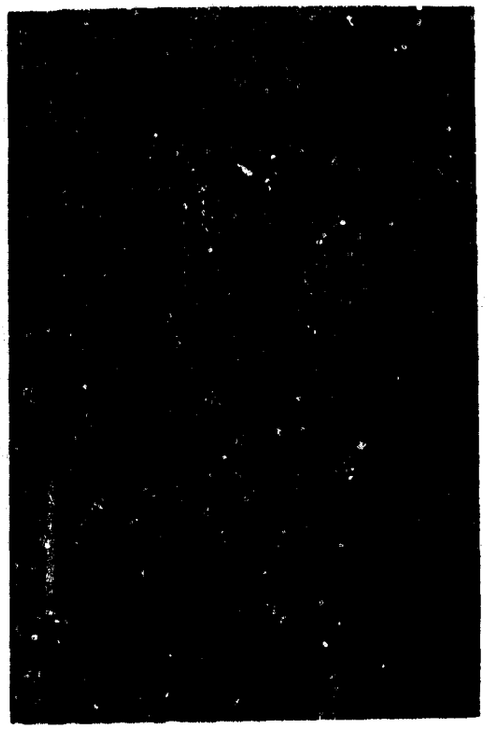


Figure 4. SEM fractograph showing an isolated region of cleavage observed on the fracture surface.

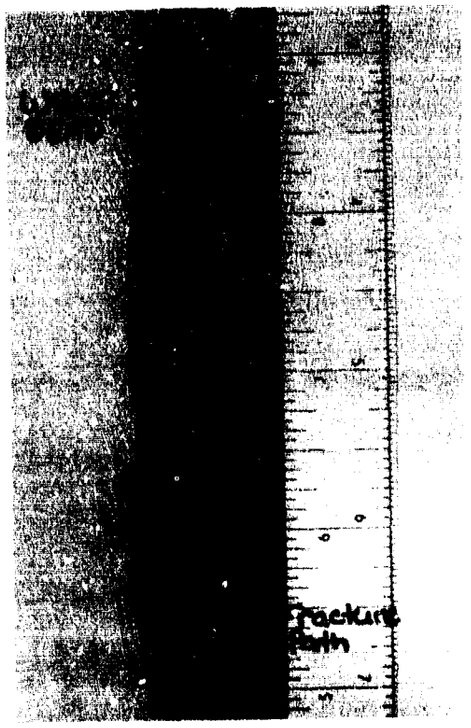


Figure 5. Top view of the triangular fracture piece, polished and etched to show the weld beads.

Figure 6. Side view of the triangular piece, etched to show the fracture path relative to the weld bead morphology.

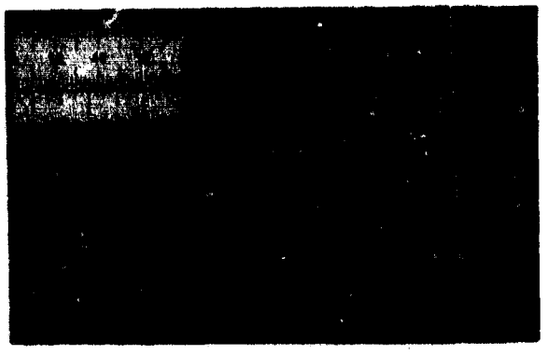


Figure 7. Fracture surface from the patch side of the fracture, etched to show the fracture path relative to the weld and HAZ.

METALLURGICAL EXAMINATION OF A FAILED MIXER PIVOT SUPPORT

Marc Pepi
Victor Champagne
Gary Wechsler
U.S. Army Research Laboratory
405 Arsenal Street
AMSRL-MA-CB
Watertown, MA 02172

Abstract: The U.S. Army Research Laboratory-Materials Directorate (USARL-MD) conducted a failure analysis of a mixer pivot support located on an Army attack helicopter. The mixer pivot support is a flight safety critical component, and is part of the rotor support assembly which fits through the transmission support. The component was found to be broken in two pieces during routine inspection in Saudi Arabia, while the aircraft was being utilized for Operation Desert Storm. Visual inspection of the failed part revealed significant wear on surfaces that contacted the bushing and areas at the machined radius where the cadmium coating had been damaged allowing corrosion pitting to occur. Light optical microscopy showed that the crack origin was located at the machined radius within a region that was severely pitted. Metallographic examination of a cross section taken through the crack initiation site revealed cracks at the bottom of some pits running parallel to the fracture plane. The hardness, chemistry and microstructure of the electroslag remelted (ESR) 4340 steel used to fabricate the component conformed to the required specifications and no apparent manufacturing defects were found. Electron microscopy showed that most of the fracture surface failed in an intergranular fashion with the exception of a shear lip zone which exhibited a dimpled morphology. The failure was set into action by hydrogen charging as a result of corrosion and was aggravated by the stress concentration effects of pitting at the radius and the high notch sensitivity of the material. Energy dispersive spectroscopy (EDS) determined that deposits of sand, corrosion and salts were found within the pits. The failure mechanism was hydrogen-assisted and was most likely a combination of stress corrosion cracking and corrosion fatigue. Recommendations were made to improve the inspection criteria in the field and to lower the hardness of the material.

Key Words: Failure analysis; high strength steel; hydrogen-assisted stress corrosion cracking.

Background: The component was found to be broken in half upon visual inspection after being in service for approximately 1449 days. The original replacement time for the mixer pivot support, as recommended by the Aviation and Troop Command (ATCOM), was 800 days. However, the 800 day service life limitation had been extended to 1440 days due to a lack of available spare parts. A second extension was granted by ATCOM for the same reason and the component was allowed to remain in service for up to six months beyond 1440 days as long as no corrosion was observed

on the surface during inspection. The mixer pivot support was machined from ESR 4340 steel bar stock and hardened to HRC 54-57 as designated on the governing engineering drawing. The component was subsequently cadmium coated by a vacuum deposition process.

Visual Inspection/Light Optical Microscopy: The failure site had been previously analyzed for stress concentrations, and was not identified as the most critical area. Figures 1a and 1b show the failed part in the as-received condition. The upper half (Part A) of the component which contained the bearing was in relatively good condition, as compared to the lower half (Part B). The surface coating of Part A was intact and the serial numbers and manufacturing data were easily distinguishable. There were no obvious signs of corrosion or mechanical damage to Part A. In contrast, Part B showed significant wear on surfaces that contacted the mating bushing. These regions were characterized by dark stains. The cadmium coating appeared to have been almost entirely worn away during service and severe corrosion pitting had occurred in these areas.

Figure 2 is a graphic example of an area approximately 1.5 cm away from the major fracture, but within the fretted region, showing deep pits on the surface of Part B. A few of the pits were large and shallow and may have been formed to some extent by mechanical vibration in addition to corrosion. Deposits of corrosion products and other debris were found in clumps surrounding and filling a number of the pits. Most pits had cracks at the bottom, and Figure 3 shows an interconnecting series of cracks observed in the corrosion layer at the bottom of one pit. It was uncertain from visual inspection whether the cracks extended into the base material. However, metallographic examination performed later in this investigation of cross sections taken through these areas, revealed evidence of cracks originating from the bottom of pits and extending into the steel. Figure 4 is an optical fractograph of the fracture surface. The fracture plane intersected the radius at the crack origin. The radial lines and chevron patterns indicated that the fracture proceeded from the bottom right of the photograph (designated by the arrow) up along the both sides of the central hole. Where the two fractures meet at the top, a ridge is visible. This suggested that the fracture, indeed originated at the bottom radius, as illustrated.

The fracture face of Part B was badly smeared in many areas, caused possibly by improper handling of the fractured component prior to examination (forcing both halves of the fracture face together). Damage may have also occurred when the component was removed from the aircraft or after it had failed. In all cases, the surface smearing was attributable to a post fracture incident. An important feature of the fracture of Part B was a shear lip region where final fracture had occurred. Both faces A and B contained surfaces that were very flat-faced in appearance, displaying no signs of plasticity (often associated with a brittle fracture). The shear lip zone, however, showed evidence of ductility. The existence of a shear lip zone served to further substantiate the location identified as the crack origin, since final fracture would tend to occur in an area opposite the initiation site on this component.

Extensive corrosion pitting was another critical feature found at the crack origin and adjacent to the fracture plane. The pits were concentrated at the crack origin and were considered relatively deep (as confirmed later by metallographic examination) for this material in the hardened condition (HRC 54-57). Another important finding was that the region of most severe fretting was located just below the fracture plane and crack origin. The entire area had experienced severe attack.

Metallographic examination: The part was sectioned such that a longitudinal and transverse specimens could be observed, as well as an area which contained extensive pitting on the exterior surface. Another cross section was taken through the crack origin of fracture Face B. All of these specimens were utilized to characterize the microstructure of the material within specific areas of concern and to measure the depth of pitting.

A fine tempered martensitic structure was observed on each of the specimens examined. The microstructure was consistent with the heat treatment performed on the component. There was no evidence of unusual material defects or large inclusions. The only notable feature was banding which extended vertically, and appeared as alternating light and dark bands. Since the component was machined from bar stock, the banding was the result of rolling during primary processing. The banding runs lengthwise to the mixer pivot support. Figure 5 represents a cross section of a typical corrosion pit. The cracks extend parallel to the fracture plane. The microstructure appeared uniform across the fracture plane and along the radius, with no signs of carburization or decarburization. When the specimen was etched, the cracks were shown to extend in an intergranular fashion. The pitted metallographic specimen was polished such that the pit depth could be measured. The largest pit measured 18 mils in depth.

Tensile Testing: A C-shaped ring of material was subjected to a tensile pull to failure. A rate of 0.05 inches/minute was utilized with a 20 kip electromechanical test machine. The intent was not to record the load and corresponding strain of the specimen until failure occurred, but simply to obtain fracture surfaces which could then be examined under the scanning electron microscope (SEM) and compared to the fracture under investigation. The resulting fracture surfaces displayed a ductile cup-cone surface. A dimpled topography was revealed, indicative of a ductile fracture mode.

Hardness Testing: A series of macrohardness (Rockwell "C" scale with 150 kgf load) measurements were performed circumferentially across a section of the C-ring specimen. Readings were taken on the concave surface of the ring. The average of ten readings was 55.6 HRC which conformed to the required hardness range (54-57 HRC).

Chemical Analysis: The mixer pivot support was specified to be fabricated from electroslag remelted (ESR) 4340 steel bar stock according to the requirements contained within HMS-6-1121. Atomic absorption (AA) and inductively coupled argon plasma spectroscopy (ICAP) were used to determine the chemical composition of the material. The carbon and sulfur content was analyzed by the Leco combustion method. The chemical composition of the material compared favorably to the nominal requirements, as shown in Table 1.

Table 1. Comparison of Chemistries
Weight Percent

Element	C	Mn	Si	P	S	Cr	Ni	Mo	Cu	Al
HMS-6-1121	0.39-0.41	0.60-0.80	0.20-0.35	0.010 max.	0.008 max.	0.70-0.90	1.65-2.00	0.20-0.30	0.35 max.	0.03 max.
				Fe = remainder						
Mixer Pivot	0.39	0.69	0.21	0.008	0.005	0.82	1.99	0.28	0.11	0.012
				Fe = remainder						

Electron Microscopy: The fracture surfaces of the component were examined utilizing the SEM. Extensive corrosion pitting was found along the edge identified as the crack origin, but the exact point of crack initiation was difficult to resolve because of mechanical damage. SEM of the pits showed that many contained cracks. Corrosion did not take place along the part of the radius in which the protective plating was intact. EDS of the coating revealed cadmium (plating material), chromium (sealant) and silicon (surface grease/oils/sand). EDS of a corrosion pit revealed major peaks of iron (base metal), silicon (surface grease/oils/sand), oxygen (corrosion by-product/sand), calcium (surface contamination), and most significantly, chlorine (salt water/sand). The existence of chlorine might be associated with salt water intrusion. Figure 6 is a SEM fractograph of the typical morphology found on approximately 90% of the total fracture surface. The mode of failure that occurred from the crack origin up along both sides of the central hole experienced intergranular decohesion. The only area that fractured differently was within the shear region which displayed a predominantly dimpled topography, as shown in Figure 7. No evidence of fatigue striations were observed but since these features are difficult to resolve in such high strength materials, fatigue could not be entirely ruled out as a failure mechanism.

Conclusions

Crack Initiation: The cracking of the mixer pivot support initiated at the machined radius within a region that was severely pitted. The fracture did not originate in a region where fretting was most severe.

Corrosion Pitting: Deposits of sand, corrosion and salts were found within the examined pits. Some pits were up to 18 mils deep. Metallographic examination of a cross section taken through the crack origin revealed cracks at the bottom of some pits. These cracks ran parallel to the fracture plane.

Materials Characterization of ESR 4340 Steel: The hardness, chemistry and microstructure of the material conformed to the required specifications and no apparent manufacturing defects were found on or within the component.

Mode of Failure: The fracture was brittle in nature, showing little ductility, with the exception of the shear lip region. The morphology of most of the entire fracture surface (approximately 90%) was intergranular while the shear lip region exhibited a predominantly ductile dimpled topography. A section of material (C-ring) from the mixer pivot support was pulled to failure in tension, revealing a resultant morphology of primarily dimples, indicative of a ductile failure.

Failure Scenario: Severe corrosion pitting occurred along the machined radius of the component and served as a crack initiation site. Hydrogen diffused into the high strength material (HRC 56) as a result of the corrosion process and migrated to areas of high stress concentration (crack tip). Evidence substantiating this claim lies in the fact that when a section of material taken from the failed component was pulled to failure in tension, the resultant fracture surface was dimpled, but the failure mode over 90% of the fracture surface under investigation was intergranular. In addition, the final fracture region of the component (shear lip) also displayed a dimpled topography. Both serve as indicators that the material can fracture in a ductile fashion. It has been well documented that hydrogen-assisted cracking occurs in an intergranular fashion in this type of material when heat treated to the hardened condition.

Failure Mechanism: The failure was set into action due to hydrogen charging as a result of corrosion. This condition was aggravated by the stress concentration effects of pitting at the radius and the high notch sensitivity of the material. The failure mechanism was hydrogen-assisted and was most likely a combination of stress cracking and corrosion fatigue.

Recommendations: The pits that served as the crack initiation site occurred over an extended period (probably several months or longer) but definitely did not occur between the last inspection of the component and the time of the failure (9 days). Therefore, it was recommended that the component be removed from service at the first indication of corrosion. Visual inspection with the aid of a magnifying lens can be used to detect corrosion in the field.

The component could continue to be utilized when hardened to HRC 54-57 if the above recommendation is strictly adhered to. In this way, the ballistic properties could be maintained. However, a more conservative approach, which would sacrifice

some of the ballistic properties of the material, would be to heat treat the component to a less hardened condition. This would decrease the notch sensitivity of the material and the inspection intervals could be subsequently increased, since the critical crack size would be increased.

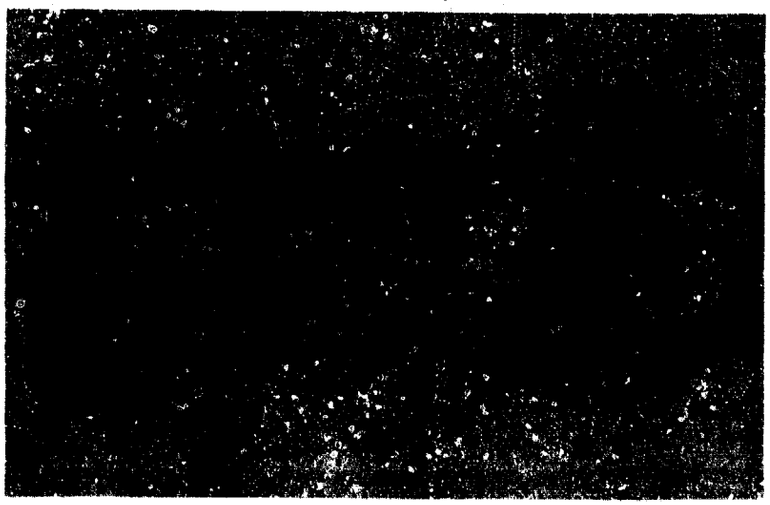


Figure 1a Mixer pivot support (Part A) in the as-received condition. Reduced 40%.

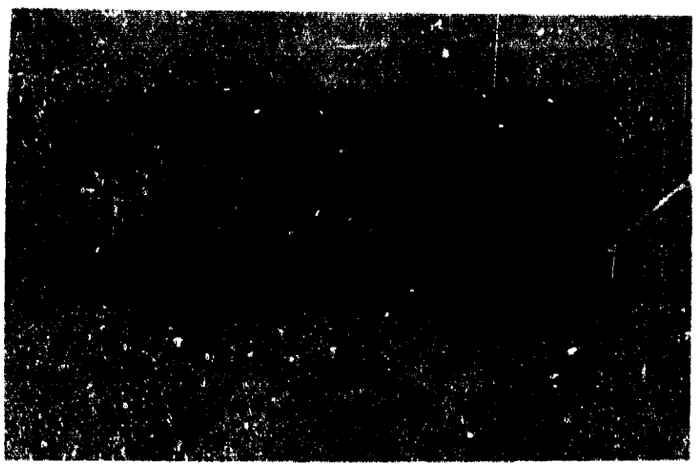


Figure 1b Mixer pivot support (Part B) in the as-received condition. Arrow denotes highly corroded area. Reduced 40%.

Figure 2 Example of severe surface pitting. Mag. 7.5x.

Figure 3 Optical fractograph Part A fracture surface. Arrow identifies crack origin. Mag. 1x.

Figure 4 Corrosion pits with cracks running parallel to the fracture plane. Mag. 500x

Figure 5 SEM of pits and secondary cracking. Mag. 130x.

Figure 6 SEM showing intergranular mode of failure. Mag. 500x.

Figure 7 SEM showing a dimpled/quasi-cleavage fracture. Mag. 2500x.

FAILURE ANALYSIS OF HYDRAULIC VALVE ACTUATORS

W. C. Porr, Jr.[1], R. L. Tregoning[1], M. E. Natishan[1], and H. E. Prince[2]
Metals and Welding Department
Carderock Division, NSWC
3A Leggett Circle
Annapolis, MD 21402-5067

Abstract: *A failure analysis was conducted on two control valve actuator spools that had failed shortly after installation. Metallurgical analysis indicated that the 440C stainless steel spool failed due to environmentally-assisted cracking (EAC) from service stresses and an unknown environment species. 440C stainless steel has a very high susceptibility to EAC in a variety of environments. A fracture mechanics evaluation indicated that less than 0.01" of environmentally-assisted subcritical crack growth could lead to catastrophic failure under normal service stresses due to the low fracture toughness of 440C. Non-destructive evaluation of unjailed spools was inconclusive in determining if these small flaws could be detected prior to failure.*

Keywords: Environmentally-assisted cracking, martensitic stainless steel, valve actuator, fracture mechanics, non-destructive evaluation

Introduction: Two control valve actuator spools failed shortly after installation. Initial analyses indicated that the fracture mode of the failures was brittle, and that corrosion may have played a role in crack initiation. The Carderock Division, Naval Surface Warfare Center (CARDIVNSWC) was requested to analyze these failures and determine if failures in other valves may be expected.

This failure analysis included three main subtasks: Metallurgical evaluation of the two failed valve actuator spools, non-destructive evaluation (NDE) of unfailed actuator spools removed from service, and a fracture mechanics-based critical flaw size evaluation for the actuator spools.

Analysis: CARDIVNSWC received 6 (whole and part) valve actuator spools designated "A" through "F": failed actuator spool A; parts of failed actuator spool B that had previously been destructively analyzed; two spools that had shown magnetic particle NDE crack indications, C&D; and two spools that were

---

[1] Fatigue and Fracture Branch, Code 614

[2] Welding Branch, Code 615

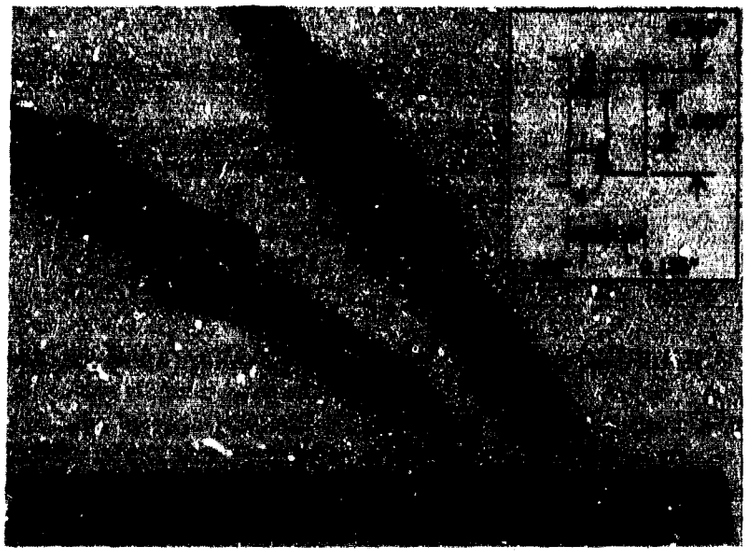


Figure 1 **Photograph of control valve actuator spools: failed spool A alongside an unfailed spool indicating failure site, with "button-head" nominal dimensions indicated.**

removed from service, E&F.

Figure 1 shows an unfailed spool alongside failed spool A, with the nominal dimensions of the reduced neck region indicated. As can be seen in Figure 1, failure in the spool occurred in the reduced "neck" region. The material specified for these spools was 440C stainless steel hardened to a Rockwell C value of 56 to 61. 440C is a high-carbon martensitic stainless steel with nominal composition as listed in Table I. Heat treatment of the actuators was to follow MIL-H-6875H. This includes a two hour temper at 450°F to achieve an approximate minimum hardness of Rockwell C 55.

Metallurgical Evaluation: The failure analysis reported here involved chemical composition analyses, hardness measurements, fracture surface evaluations, and metallography. Chemical compositions of the failed actuator spools A and B were determined with a wet chemistry analysis. Rockwell C hardness was determined for all received spools to verify adherence to specification and give a sense of the tensile strength of the material. Fracture surfaces from the two failed spools were examined and characterized in the scanning electron microscope (SEM) to give insight to the cause of failure. Metallography of the failed spools was conducted to examine the microstructure of the 440C adjacent to the fracture region and characterize features that may be significant to the failure mode.

Hardnesses measured for all the received spools are indicated in Table I. Measured hardnesses were within the specified range of 56 to 61 Rockwell C.

Therefore, hardness measurements give no indication why only two of the six spools examined in this analysis failed.

Table I. **Hardnesses measured for six 440C stainless steel actuator spools. Values indicated are averages of at least eight measurements.**

	A (Failed)	B (Failed)	C	D	E	F
Rockwell C Hardness	56.9	56.7	57.2	57.1	56.7	56.5

The chemical analyses performed on the two failed spools are reported in Table II. The 440C stainless steel used in these parts met specification for all elements, with the exception of Cr. The slightly low Cr content measured for spool A is attributed to the tendency of the analysis laboratory to under-measure Cr, and deemed insignificant in this case.

Table II. **Chemical compositions (weight %) of two failed actuator spools, along with the specified chemistry for 440C stainless steel.**

Element	440C Martensitic Stainless Steel (ASTM A 473-92a)	Actuator Spool A	Actuator Spool B
Cr	16.00 - 18.00	15.2	16.3
C	0.95 - 1.20	.998	1.03
Mn	1.00 max	0.47	0.41
Mo	0.75 max	0.47	0.55
Si	1.00 max	0.43	0.43
P	0.04 max	0.016	0.018
S	0.03 max	0.003	0.002
K	---	<0.001	<0.001
Ti	---	0.002	0.003
Pb	---	<0.002	<0.002
Sn	---	<0.002	<0.002
Fe	Remainder	Remainder	Remainder

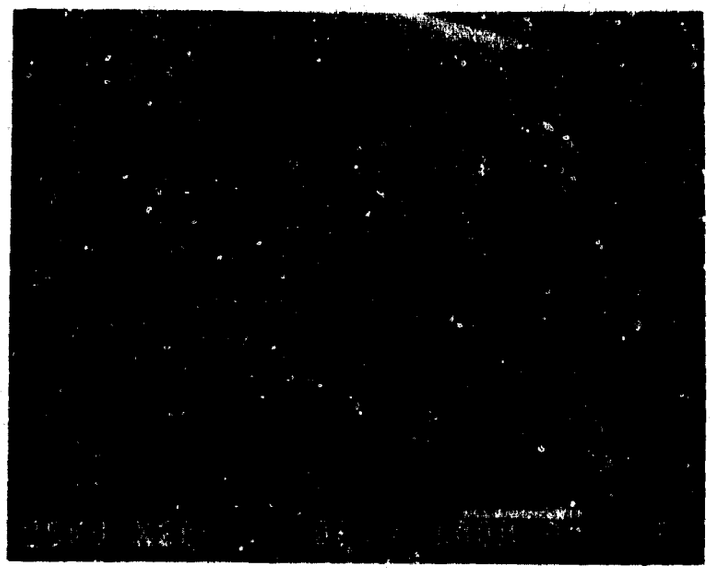


Figure 2 SEM micrograph of the neck and fracture surface of spool A.

SEM examination was performed on failed spool A[3], as illustrated by Figures 2, 3, and 4. Both halves of the fracture surface were examined. Damage to the fracture surfaces prior to their arrival at CARDIVNSWC made their analysis difficult. However, sufficient detail remained to characterize the fracture mode of this spool. The majority of the fracture surface had a ductile and/or quasi-cleavage nature, indicative of overload failure. This is illustrated in Fig. 3. The particles at the base of voids shown are most likely carbides in the tempered martensite microstructure. The most significant observation on the fracture surface was the small region of intergranular fracture on the spool surface located at "a" in Fig. 2, and shown at higher magnification in Fig. 4. This region of intergranular fracture was semi-elliptical in geometry, with surface length of approximately 0.004" and depth of approximately 0.002".

For metallography, spools were sectioned and mounted in epoxy, then polished, and etched with Vilella's reagent, prior to microscopic examination. Metallography of the failed spools A and B showed the spools to have microstructures typical for tempered martensite, with primary and (tempered) secondary carbides. Prior austenite grain boundaries indicated a grain diameter of approximately 20 μm.

Evident in Fig. 5, are regions of prior-austenite intergranular attack adjacent to the fracture surfaces of spools A & B. The depth of this attack in A is approximately 0.002" (50 μm) from the spool surface. This region of

---

[3] The fracture surface of spool B could not be examined by SEM because of a previously performed destructive analysis.

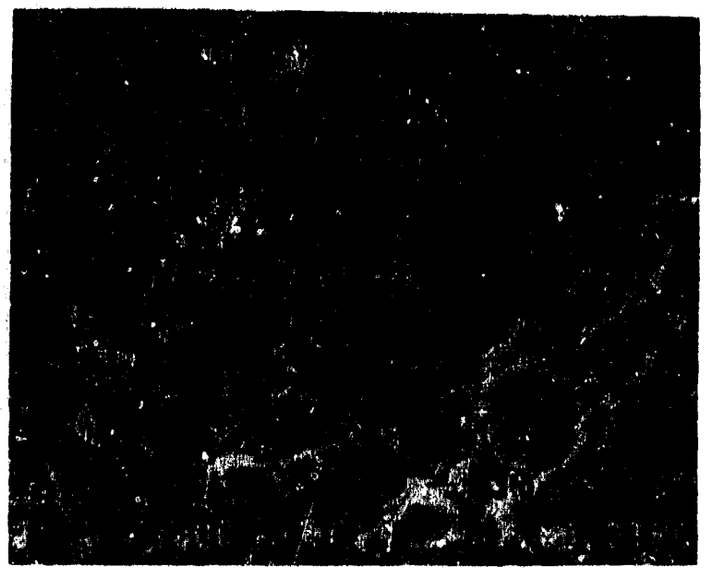


Figure 3 **SEM micrograph illustrating the predominantly ductile fracture mode of the failed spool.**

Figure 4 **Small region of intergranular fracture observed on the fracture surface of failed spool A. (200X)**

intergranular attack is approximately the size and shape of the region observed in fractography of the same piece (Fig. 4). Figure 5 also shows a region of intergranular cracking in a polished, unetched section of spool B. The depth of the region shown is approximately 0.004" (100 μm) and may represent the failure initiation site.

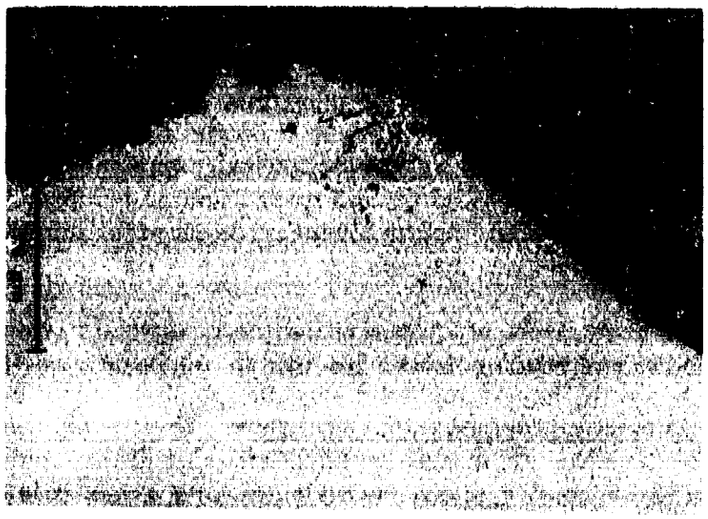
Intergranular attack, like that seen here, is indicative of an environmentally-assisted cracking (EAC) mechanism. Martensitic stainless steels, hardened to high strength levels, are extremely susceptible to EAC, even in mild environments [1]. The spools that failed operated in hydraulic fluid, however the exact environmental / electrochemical service conditions are not known. Environments that can cause intergranular cracking in materials such as this are not specific, contrary to the situation for austenitic stainless steels; cracking can occur in any environment in which hydrogen can be evolved [2]. Therefore, pinpointing the exact environmental species responsible for the cracking in this case may not be critical.

Non-destructive Evaluation: Fluorescent magnetic particle and fluorescent dye penetrant techniques were used to determine if measurable flaws were present in the intact spools C, D, E, and F.

The dye penetrant materials used were Zyglo Cleaner ZC-7, Zyglo Penetrant ZL-22A and Zyglo Developer type ZP-9. All the products were in aerosol cans. The specimens were coated with cleaner and wiped with clean paper towels. A thorough coating of penetrant was applied to the specimens and allowed to dwell for a minimum of 30 minutes. Cleaner was sprayed onto clean paper towels and used to wipe off excess penetrant under black light. After thorough removal of all penetrant under black light, developer was applied.

The equipment utilized for the magnetic particle inspection was a "Magna Flux Black Light" Model ZB 26, Glo-Netic SC-925 Fluorescent Magnetic Particle Oil dispersion and a Magna Flux Alternating Current Yoke, Model Y-6 with flexible contacts. For this procedure the specimen was placed in the yoke, and contacts were flexed inward to hold the specimen firmly in place. The longitudinal magnetic field was applied to the specimen and the wet magnetic particle inspection medium was sprayed onto the specimen under black light. The field was maintained continuously until the oil evaporated and the dry fluorescent particles remained. No attempt was made to de-magnetize the specimens.

The fluorescent dye penetrant inspection produced no indication of any defects. The fluorescent magnetic particle inspection produced a straight line indication 360° around all three specimens. The location of these indications changed as the subject was oriented at various angles to the field direction. Hence, they are considered to be non-relevant indications (artifacts) due to the applied magnetic field. No other indications were observed. The implication of these results is that either flaws do not exist in the spools examined, or the flaws are



A

B

Figure 5 **Optical micrographs of polished (unetched) sections through spools A and B, indicating intergranular attack adjacent to the fracture surface (500X).**

too small to detect by the means employed.

Fracture Mechanics Evaluation: A linear elastic fracture mechanics (LEFM) study was conducted to analyze the valve actuator spool failures. The LEFM determined stress intensity factor, K, allows quantification of crack tip driving force as a function of applied remote loading, flaw size and shape, and component geometry. The critical flaw size necessary to cause failure in these components was first estimated based on conceivable material fracture toughnesses, assumed crack locations and shapes, and design information. Then, the minimum material fracture toughness required for survival of flaws at the NDE inspection limit was determined.

In this fracture analysis, the crack driving force was assumed to result from the combined tensile and bending stresses incurred during proof pressure testing. These stresses are the largest applied during the spool's service life and their magnitude has been ascertained from calculations. These calculations accounted for the reduction in cross sectional area due to the lathe center hole, but neglected stress concentrations caused by the lathe center hole bottom and the shoulder fillet. This fracture mechanics analysis also assumed that a surface breaking semi-elliptical flaw was present in the spool with a surface length to flaw depth ratio of 2. A crack of this shape ratio has the smallest surface flaw length, and the minimum cracked surface area, at failure, and hence, would be the most difficult to detect.

The critical component flaw size range for failure was first estimated by assuming that the material's fracture toughness, K_{Ic}, falls between 10 and 35 ksi-$in^{1/2}$; a typical range for 440C stainless steel [3]. Lack of specific material information for the 440C used in the spools precludes analysis at a definitive toughness within this range. Three potential crack locations were considered in this analysis: in the spool neck near the shoulder, at the lathe center hole bottom, and in the spool neck away from the lathe center. Existing stress intensity factor solutions for simplified geometries representing each of these cases were employed [4,5]. These simplified solutions did not account for the increase in crack driving force near the geometrical stress raisers and assumed plane stress loading conditions. Therefore, the critical flaw sizes determined here should be slightly greater than in the actual component.

The results of this analysis showed that the smallest cracks which can lead to failure occur in the spool neck near the shoulder. This location concurs with the observed failure location in spool A. However, the critical crack size in the spool neck away from the lathe center hole is only between 0.001" to 0.010" larger (depending on exact toughness) than the critical flaw size near the shoulder fillet. Therefore, the spool neck is also highly susceptible to failure, as evidenced by spool B failure at this location. The critical flaw size at the lathe center hole is larger than these two locations and is less likely to be the site of component failure.

The calculated critical flaw depth at fracture was 0.0045" near the shoulder when it was assumed that the K_{Ic} of the spool was 10 ksi-in$^{1/4}$. If K_{Ic} was as high as 35 ksi-in$^{1/4}$, the critical flaw depth was calculated to be 0.044" in this location. This range of critical flaw sizes cannot be unequivocally lowered without definite material toughness information. However, the heat treatment process used for these spools and some limited information linking 440C processing and fracture toughness [3] suggests that the toughness of the 440C used in the spools should be less than 20 ksi-in$^{1/2}$. If the material toughness is in fact below 20 ksi-in$^{1/2}$, the critical flaw depth of these spools is below 0.014". This upper bound estimate of critical flaw depth agrees well with the above described metallographic observations of intergranular cracking depth in the two failed spools inspected by CARDIVNSWC.

The final portion of the fracture analysis determined the necessary material fracture toughness required for this component, assuming the minimum flaw size and shape that could be found by current nondestructive evaluation techniques existed in the component at the worse location (spool neck near the shoulder fillet). The minimum detectable surface flaw length using dye penetrant techniques is approximately 0.0625", with an assumed surface crack length to crack depth ratio of 3.[4] If the earlier fracture analysis is repeated assuming this flaw, the minimum toughness required to avoid fracture of the spool is 24 ksi-in$^{1/2}$. This toughness is higher than the estimated range of toughnesses for the 440C used in these valve spools. This result implies that it is not possible to nondestructively inspect these parts for cracks to the accuracy required.

Failure Assessment Summary: In considering the results of these three separate evaluations, a scenario for failure of the spools consistent with the facts of their service can be considered. The 440C used for this part is highly susceptible to environmental cracking. Whether from residual surface stresses or from applied service stresses, environmentally-assisted intergranular cracking occurred in the spool necks. This cracking progressed until sufficient to constitute a critical flaw which resulted in catastrophic overload failure of the spool. Intergranular crack growth did not need to be extensive prior to failure due to the very low fracture toughness of the 440C material. As the crack sizes leading to failure in the spool were on the order of 0.01", standard NDE crack detection techniques would be insufficient to detect sub-critical flaws. The underlying cause of these failures is inadequate material selection, due to the high susceptibility to environmentally-assisted cracking and low fracture toughness of the 440C martensitic stainless steel used for the valve actuator spools.

---

[4]Personal communication with R. Denale, Head, Welding Branch, Code 615, CARDIVNSWC, July 30, 1993.

Conclusions:

1. Failure of the 440C stainless steel valve actuator spools was caused by environmentally-assisted intergranular crack initiation (EAC). 440C stainless steel has a high susceptibility to EAC, especially when hardened to high strength levels. The environment species that resulted in the intergranular attack of the spool material is not known.

2. The low fracture toughness of the 440C stainless steel used for the spools ensured a low damage tolerance for the component. Critical flaw size for catastrophic failure of the spool was on the order of 0.010".

3. Critical flaw size for this component is too small for NDE detection prior to catastrophic failure under service loads.

References:

1. Uhlig, H. H. and R. W. Revie, Corrosion and Corrosion Control, 3RD Edition, John Wiley and Sons, Inc., New York, 1985, pp. 320-321.

2. Metals Handbook, 8TH Edition, "Failure Analysis and Prevention", Vol. 10, American Society for Metals, Metals Park, Ohio, 1975, pp. 220-221 and 234-235.

3. Lou, B. and B. L. Averbach, "The Effects of Heat Treatment on Fracture Toughness and Fatigue Crack Growth Rates in 440C and BG42 Steels", Metallurgical Transactions A, Vol. 14A, September 1983, pp. 1899-1906.

4. Raju, I. S. and J. C. Newman, Jr., "Stress-Intensity Factors for Circumferential Surface Cracks in Pipes and Rods Under Tension and Bending Loads", NASA Technical Memorandum 87594, August 1985.

5. Mettu, S. R., I. S. Raju, and R. G. Forman, "Stress Intensity Factors for Part-Through Surface Cracks in Hollow Cylinders", JSC Report No. 25685/ LESC Report No. 30124, July 1992.

METALLURGICAL EXAMINATION OF THE LAU-7 NITROGEN RECEIVER FAILURE

Gary Wechsler
Victor K. Champagne
U.S. Army Research Laboratory
Watertown, MA 02172-0001

Dr. David Broek
FractuResearch, Inc.
Galena, Ohio 43021

Abstract: A pressure vessel containing nitrogen gas exploded during a routine filling operation. The pressure vessel was composed of an inner stainless steel liner which was wound with glass fibers in an epoxy matrix. The optical and electron microscopy determined the liner failed in a ductile manner. Metallographic examination and chemical analysis verified the microstructure and composition of the liner met drawing specifications. Optical microscopy of the composite fracture origin revealed dark fibers with debris adjacent to the liner failure origin. Based on this data, the liner failed due to an overload. Since the composite was the load bearing portion of the pressure vessel, the liner was subjected to an overload because the composite had failed initially. The composite may have failed due to damage incurred from handling, proof testing or refilling operations. Recommendations were made to develop a nondestructive inspection plan utilizing acoustic emission testing.

Key Words: Acousitic emission; composite; filament winding; failure analysis; pressure vessel; stainless steel.

Introduction: On January 29, 1991 a LAU-7 Nitrogen Receiver exploded approximately 60 seconds into a routine filling operation at the Miramar Naval Air Station in San Diego, California. The nitrogen receiver was in a constraining cannister when the explosion occurred. Both components fractured into two halves. Each fracture half was propelled through the air in opposite directions, through a wall of the filling station and travelled several hundred feet outside of the building before damaging several vehicles and causing bodily injury to one person. The nitrogen receiver had been in service prior to failure for an undetermined period of time.

Component Description: The LAU-7 Nitrogen Receiver shown schematically in Figure 1 provided compressed nitrogen gas to the Sidewinder Missile. The receiver was manufactured according to the requirements established in MIL-R-81202D and is composed of an inner stainless steel liner (bladder) which was wound with glass fibers within a matrix of epoxy resin (fiberglass). The purpose of the metal liner was to contain the nitrogen gas and prevent leakage. The filament winding provided the structural support needed to withstand the hoop and longitudinal stresses caused by the internal gas pressure.

The metal liner used to contain the nitrogen gas was fabricated from 321 stainless steel. The liner consisted of three major sections; the hemispherical end caps and a cylindrical body to which the end caps were welded. Both end cap to cylinder weldments incorporated backing rings to prevent weld blow through and provide dimensional stability during welding. The finished welds were then subjected to 100% radiographic and liquid penetrant inspection and evaluated to class I of MIL-R-11468 in accordance with MIL-STD-453.

A glass filament composite was wound around the metal liner to prevent the high pressure nitrogen gas from rupturing the liner. The glass fibers were contained within an epoxy resin matrix. During loading the epoxy matrix distributed stresses from the pressurized liner to the glass filaments which provided the structural strength. The glass filament layers were orientated at angles of 90^{o} and $\pm 45^{o}$ from the axial direction. After fabrication the finished component was subjected to leak and pressure tests. The nitrogen receiver was normally charged to 3000

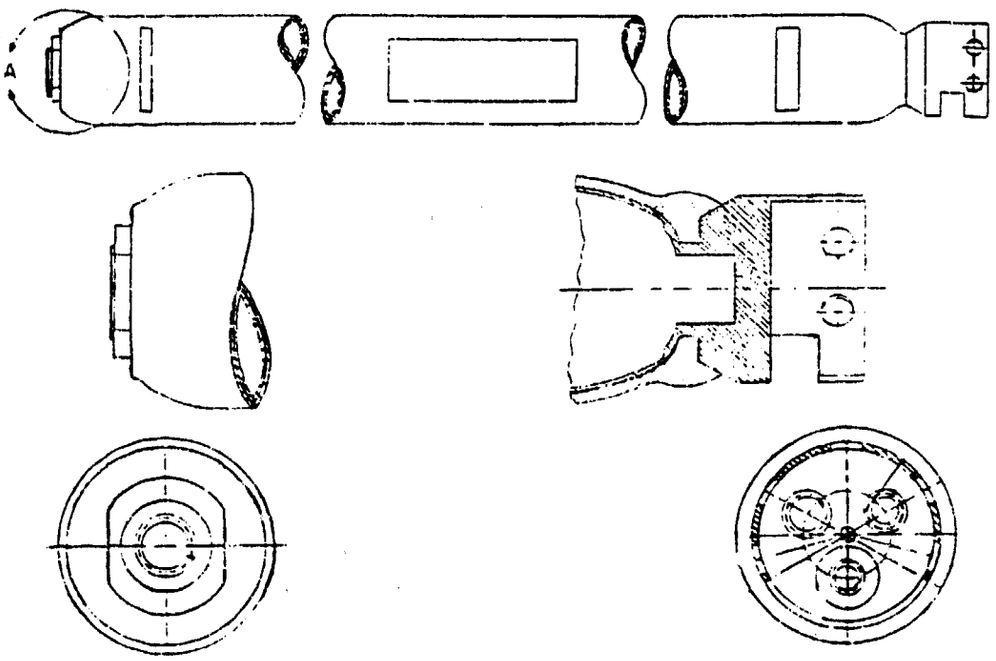


Figure 1. Schematic of the LAU-7 Nitrogen Receiver.

pounds per square inch (psi) during service and hydrostatically pressure tested to 5500 ± 100 PSIG. The filament wound composite was designed to withstand an internal gas pressure of 8500 psi. Completed receivers have been burst tested by the manufacturer to pressures of 10,500 to 12,000 psi.

Visual Examination/Light Optical Microscopy: The nitrogen receiver fractured into two pieces. The largest piece consisted of the cylinder body which was still connected to the end cap containing the skirt as shown in Figure 2. The smaller piece was the locking end cap containing a large hexagonal fitting. Figure 2 also reveals the extensive distortion of the skirt which prevented the identification of the serial number of the nitrogen receiver. The exterior surface of the component was badly damaged by the impact with the wall, after being hurled into the air as a result of the explosion. Exterior surface areas of damage that had darkened with age and weathering were observed on the filament winding. Area A reveals a region in which many of the individual glass fibers had been severed during impact causing the filament winding to unravel and become delaminated. The metal liner beneath this region had been bent inward. Similar damage was observed within Area B. Both areas exhibited damage that occurred after the explosion and was not related to the cause of the failure.

Stainless Steel Liner: The fracture of the stainless steel liner occurred at the circumferential weld joining the cylinder body to the locking cap. One third of the fracture propagated within the heat affected zone (HAZ) while the remaining crack growth occurred outside of the HAZ within the base material of the cylinder body. The resulting fracture surface was predominantly composed of shear lips (fracture occurring at 45 degree angles to the loading direction). This type of fracture is typical of an overload failure in a thin walled pressure vessel. The failure initiation site and direction of crack propagation were determined by examining and interpreting macrofeatures of the fracture surface such as shear lips and the tearing pattern of the metal liner. Another significant feature which aided in identifying the crack origin was an area that had ballooned out due to internal pressure, causing plastic deformation of the metal liner. This region, identified in Figure 3 suggested that the filament winding had a weak spot and lost its ability to constrain the metal liner. Consequently, the metal liner was

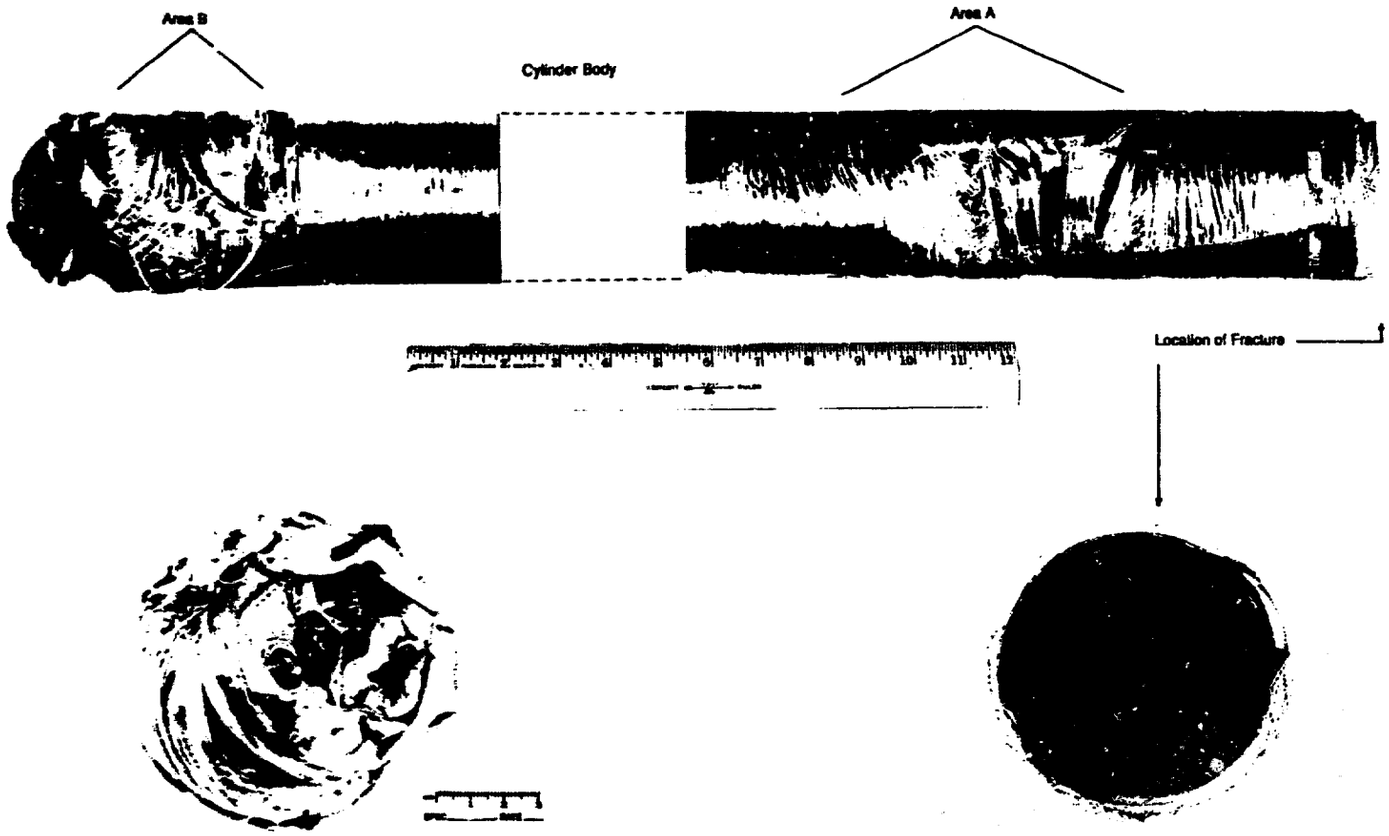


Figure 2. Macrograph of failed nitrogen receiver.

allowed to plastically deform within this weak spot and balloon out until a crack developed and catastrophic failure occurred. The crack origin occurred just beyond the ballooned region within the HAZ and propagated in opposite directions, circumferentially, until the two crack fronts met and final fracture occurred. A cross-sectional view of this area is shown in Figure 3. The ballooned out region could not have occurred due to post-failure damage since there was no evidence of any interior surface damage which would have occurred if the metal liner had been mechanically pushed out as opposed to being formed due to a high concentration of gas pressure. The final fracture zone exhibited ductility as evidenced by the plastic deformation within this region.

Filament Winding: The fracture pattern of the filament winding was also examined and is shown in the schematic of Figure 4. The glass fibers adjacent to the crack origin and ballooned out region of the metal liner were slightly darker than the glass fibers of the final fracture region with small particles of debris in between the fibers. The fibers within the remaining composite fracture were relatively clean. The debris may have been the by products of wear which occurred between the metal liner and composite due to the action of fatigue (refilling, proof testing and/or in-service loading). The fracture surface of the filament winding adjacent to the crack origin of the metal liner was relatively flat (See Figure 5) compared to the final fracture region of the filament winding which failed in shear (See Figure 6). The filament winding fracture region adjacent to the metal crack initiation site was not typical of the global fracture pattern often associated with composite failures and observed within the final fracture region. The composite, initially failed in a very constrained manner as evidenced by the condition of the severed glass fibers, which failed as if they had been cut by a sharp edge. The fibers within the final fracture region, however, had a frayed appearance as expected. This suggests that the composite fibers adjacent to the crack origin may have been damaged prior to failure.

Tears within the composite layers, discoloration within the crack initiation region and the angle of the broken glass fibers indicated the direction of crack propagation coincided with the crack propagation of the metal liner. Voids were observed within the composite adjacent to the fracture as well as other regions. These defects were aligned parallel to the fracture plane.

Locking Cap to Cylinder Body Weld: Inspection of the circumferential weld joining the locking cap to the cylinder body revealed a lack of control of the welding operation. The width of the weld bead varied significantly around the cylinder body from approximately 0.06 inch to 0.133 inch in width. Along the toe of the weld a dark oxide had formed because of insufficient inert gas shielding during welding.

Electron Microscopy: Electron microscopy of the metal liner fracture surfaces confirmed what had been established by visual inspection. The morphology of the shear planes was composed of directional dimples (See Figure 7). A few isolated regions were found that contained equiaxed dimples indicative of failure due to uniaxial loading. Throughout the circumference of the fracture surface an extensive amount of smeared metal was observed obliterating the topography in many regions. Because of this damage the exact fracture origin could not be determined. The initiation site could only be resolved by interpreting macrofeatures observed by optical microscopy as described earlier. The surface of the stainless steel cylinder body adjacent to the weld was rough and exhibited signs of yielding which had taken place during the explosion. Secondary cracking was observed running parallel to the fracture surface.

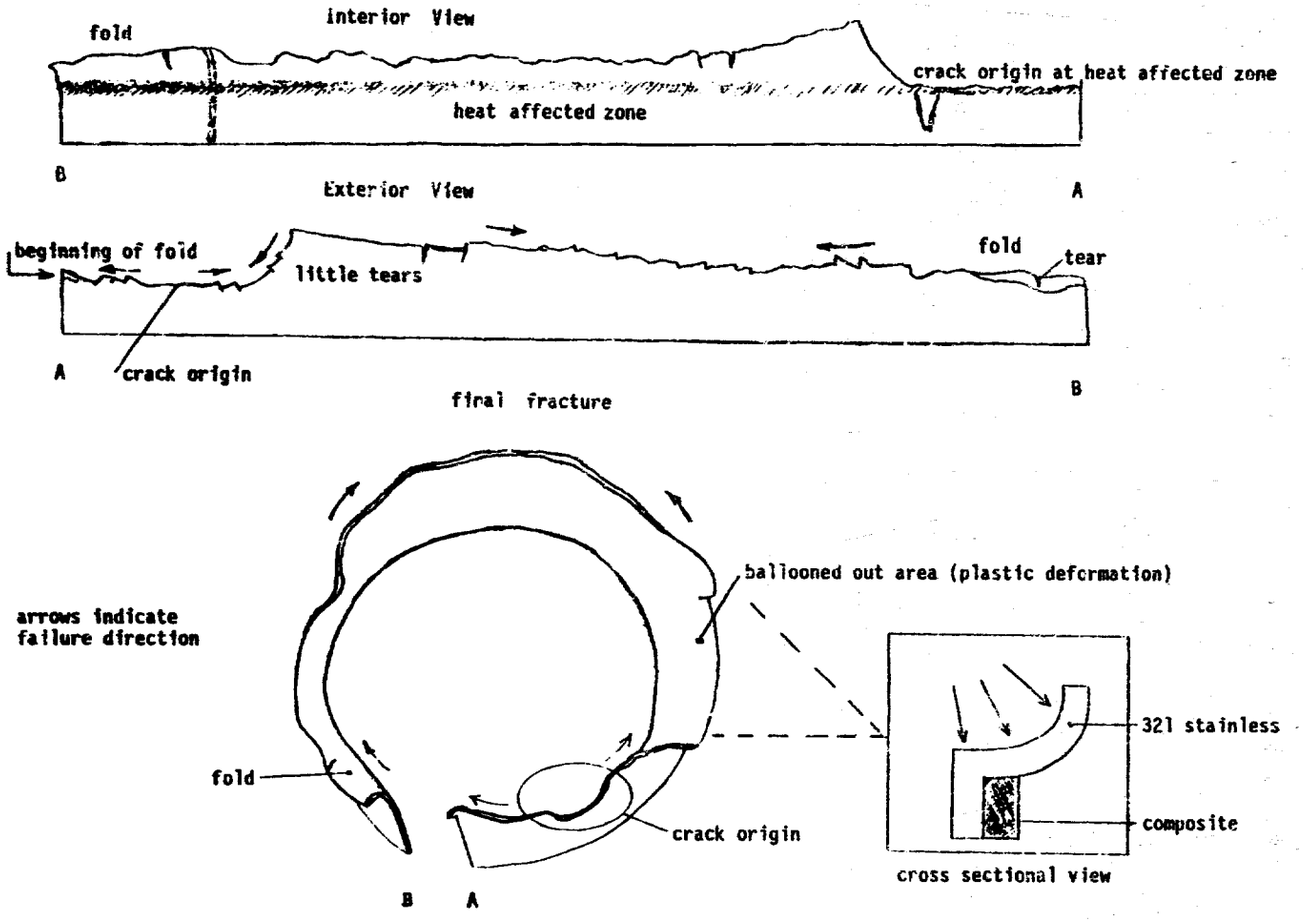


Figure 3. Schematic of the fracture surface of the metal liner of the cylinder body.

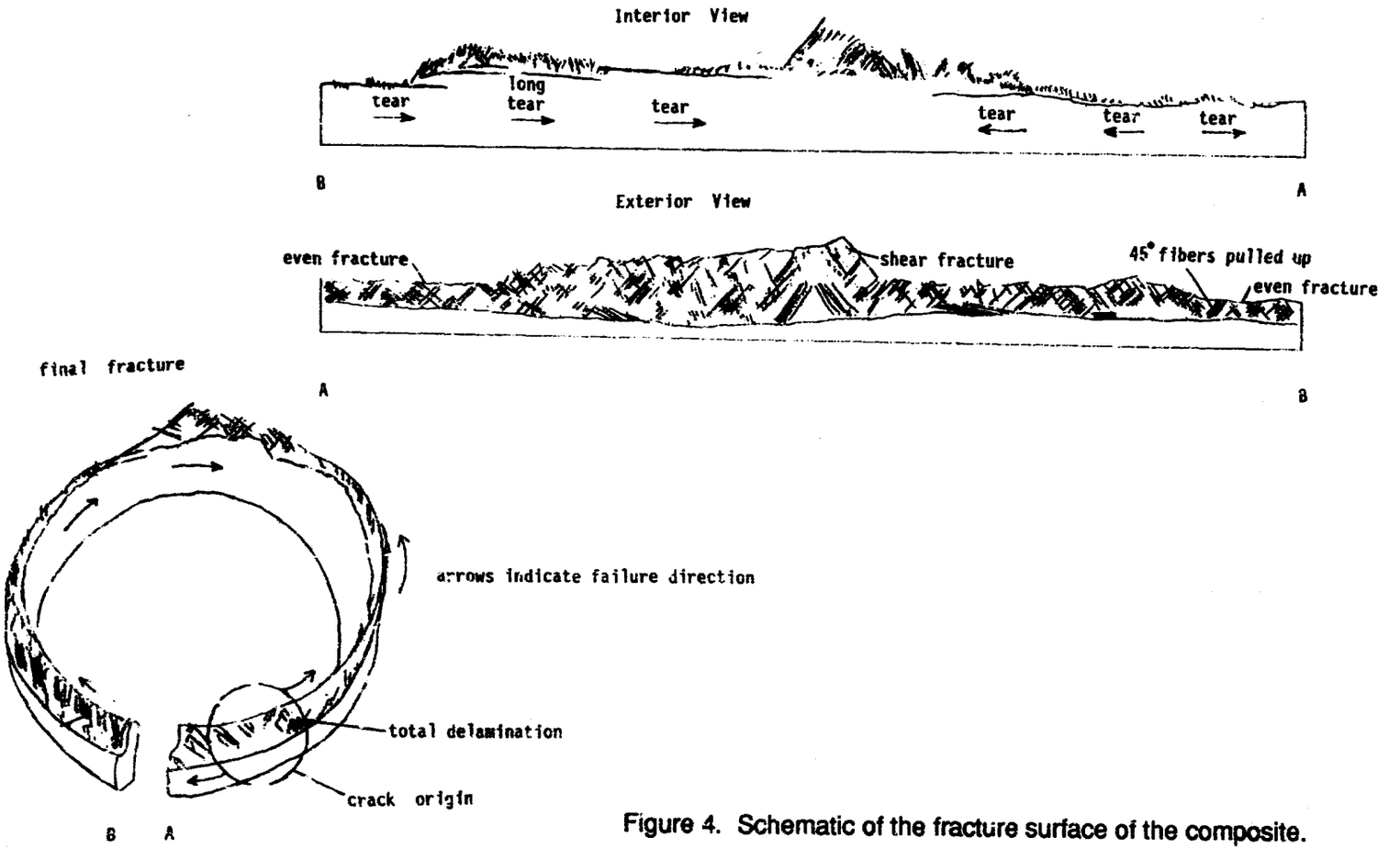


Figure 4. Schematic of the fracture surface of the composite.

Figure 5. Region of crack initiation in the composite. MAG. 7.5X.

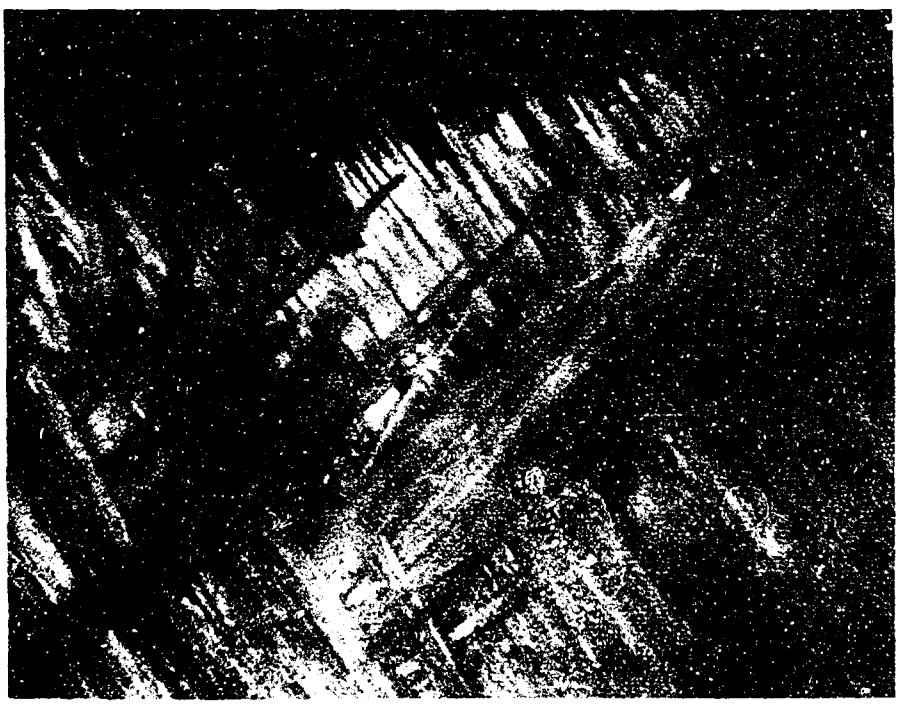


Figure 6. Area of final rupture in the composite. MAG. 7.5X.

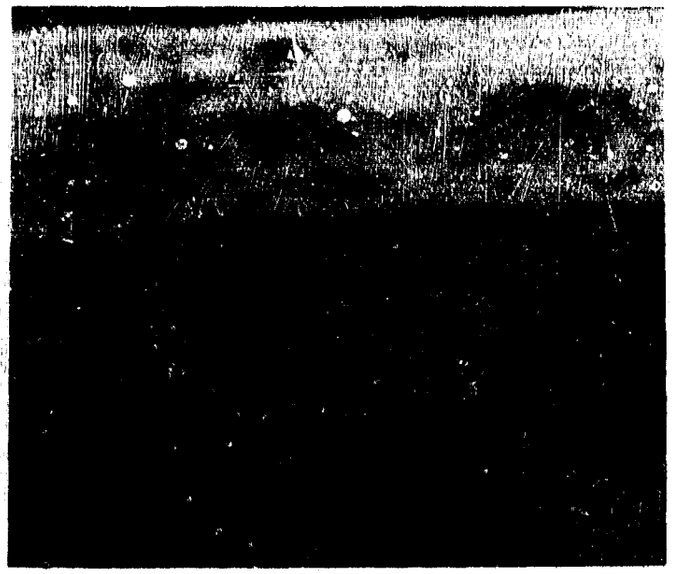


Figure 7. SEM of shear fracture in metal liner. MAG. 120X.

Chemical Analysis: The materials used to fabricate the metal liner of the cylinder body, the locking cap, and the back-up ring along with the weld bead material of the locking cap to cylinder body weld were determined by Inductively-Coupled Argon Plasma Emission Spectroscopy. The carbon and sulfur content were analyzed by the Leco Combustion Method. All of the materials conformed to the compositional range designated for type AISI 321 stainless steel as listed in Table 1. The specified requirements for type AISI 321 stainless steel as listed in the Metals Handbook (published by the American Society for Materials) have been included for comparison.

Table 1. Chemical Composition of the Nitrogen Receiver

	Metal Liner	Locking Cap	Back-up Ring	Weld Bead	AISI 321 Stainless Steel
	%	%	%	%	%
Carbon	0.028	0.055	0.071	0.077	0.08 MAX
Sulfur	0.021	0.005	0.005	*	0.03 MAX
Manganese	0.86	1.37	1.41	1.24	2.00 MAX
Silicon	0.48	0.50	0.44	0.49	1.00 MAX
Chromium	18.3	17.5	17.9	17.4	17.0-19.0
Nickel	10.6	9.93	11.0	10.6	9.0-12.0
Phosphorus	0.031	0.021	0.024	0.022	0.045 MAX
Titanium	0.30	0.57	0.73	0.62	5 X C MIN.

* Insufficient sample

Metallographic Examination: The circumferential weld joining the metal cylinder body to the locking cap which had failed causing the explosion was sectioned and prepared for metallographic examination. Latic acid was used to etch the specimens and reveal the microstructure of the locking cap, cylinder body, back-up ring and weld bead. The weld bead exhibited a typical dendritic structure. The heat affected zones of the three mating pieces can be observed. The microstructure of the cylinder body, the locking cap and the back-up ring was austenitic. Delta ferrite stringers were resolved and appeared as dark horizontal lines aligned with the rolling direction of the material.

Discussion: The metal liner of the cylinder body failed entirely in a ductile fashion indicating that the stainless steel used to fabricate the liner performed as designed and the failure was not environmentally assisted. In addition, the stainless steel exhibited no signs of gross defects or inclusions and although the welding process utilized to join the cylinder body to the locking cap could be improved, there was no evidence collected during this investigation that suggested that either poor welding practices or defective material caused the failure. If the metal liner had failed before the filament winding then leakage of the metal liner would have occurred. The most common location where leakage cracks develop are within or adjacent to the liner welds since the strength of the material in these areas is often less than that of the parent metal.

Cracks in the metal liner can initiate due to the action of fatigue caused by the change in gas pressure that occurs during refilling, proof testing and depressurization. If the adhesive bond between the metal liner and the composite fails, a stress concentration will be created at the perimeter of the delamination. Additional stress can be caused by the difference in elastic limits between the metal liner and filament winding. Expansion of the metal liner can occur during pressurization because the glass fibers of the composite have as much as 4.8% strain before failure (as in the case of E glass). Stainless steel experiences plastic deformation well before this (at approximately 0.2% strain) limit and strain hardens in the process. Buckling of the metal liner occurs when the nitrogen receiver then becomes depressurized. The composite simply returns to its original shape but the metal liner which had undergone strain hardening during pressurization (resulting in a larger diameter) would be crushed by the composite as it is contracting.[1]

When the component failed, an explosion occurred, which strongly indicated that significant damage of the composite had been incurred prior to the refilling operation. It is not unusual for the composite to fail at relatively low gas pressures because of damage accumulated over the service life of the component. Research of the stress rupture characteristics of these types of pressure vessels has shown that significant damage can be incurred during each cycle of proof testing. Microcracks can occur at voids in the epoxy resin and join to form more critical defects. It is not unusual for the void content of such filament wound composites (in the as-formed condition) to range from 5 to 7% by volume. Glass fibers can also be severed due to stress concentration effects. This occurs because of uneven load distribution as a result of defects such as voids or cracks in the resin, delamination between layers of fibers, and debonding between the fibers and epoxy resin.[2,3]

The integrity of the filament winding affects its ability to provide structural support and distribute loads evenly during service. As the component becomes pressurized stress is transferred from the metal liner to the individual fibers of the composite. The ability of the stress to be transferred evenly is hindered by the defects described above and by the misalignment of the glass fibers within the composite, which is typical for these materials. Regions of the composite may contain more fibers per unit area than others and since the resin, which makes up the difference, is weaker than the glass, a stress concentration exists. In addition, static load tests of glass fibers in tension when exposed to water has revealed that moisture acts as a stress corrosion agent. Therefore, if moisture is allowed to seep into tight cracks or crevices due to delamination or debonding in the composite the glass fibers can experience stress corrosion cracking and contribute to premature failure of the component.

An additional concern is the exterior surface condition of the metal liner itself. Examination of the glass fibers near the crack origin indicated that the failure of the composite was very

constrained (almost as if the fibers had been cut by a sharp edge). In instances where the metal liner expands and contracts with changes in gas pressure, any surface asperity (such as weld spatter, burrs, etc.) could cut into the internal surface of the filament winding with each filling or proof testing operation.

Conclusions: The failure of the nitrogen receiver was caused by a weak spot in the filament wound composite material. During the nitrogen gas filling operation, the weak spot was not able to constrain the stainless steel liner (bladder) locally as it expanded. As the weak spot deformed from the increasing local stress field, yielding of the metal liner occurred. Simultaneously, the composite within the stress concentration region fractured as the metal liner continued to expand, locally. The bladder then proceeded to undergo plastic deformation until sudden catastrophic failure of the component occurred in the form of the explosion. The origin of the weak spot in the filament winding may have been the result of any one or a combination of the following:

- exterior damage as a result of impact during service or improper handling;
- interior damage due to expansion and contraction of the component during proof testing;
- propagation of defects during in-service loading and/or refilling operations.

Recommendations: 1. Determine the maximum level of damage the filament wound composite can withstand without affecting service life. Then, perform a visual inspection of Nitrogen Receivers in the field to remove from service those components that exhibit excessive damage.

2. Modify the hydroproof testing of the Nitrogen Receivers. Hill has predicted the burst pressures of pressure vessels from acoustic emission data collected while hydroproof testing to 25% of the nominal burst pressure.[4] This method should be applied to the Nitrogen Receivers to minimize the damage that can be imparted to the filament wound composite during hydroproof testing above the maximum expected operating pressure. First, a sampling of Nitrogen Receivers (the number required will be dependent on the variance of the data collected) would be burst tested while being monitored by acoustic emissions. A correlation would then be established between statistically significant independent variables, dependant variables and the burst pressure. If the correlation is reliable, then subject the remaining Nitrogen Receivers to acoustic emission testing to verify the integrity of the components.

3. Redesign the Nitrogen Receiver based on the knowledge and data that has been developed on stress rupture phenomena in fiberglass. An alternative fiber material could be investigated that would have better performance than glass.

Acknowledgements:

Weld Evaluation, Tom Eager, Tom Melvin; Composite Examination, Yolanda Hinton, Takis Blanas, Gary Foley, Margaret Roylance; Electron Microscopy, Brad Taber, Marc Pepi; Metallography, Andrew Zani, John Mullin; Machining, Dan Snoha.

1. Dr. M. A. Hamstad, University of Denver, excerpt from conversation, April, 1991.

2. Dr. E.v. K. Hill and T. J. Lewis, "Acoustic Emission Monitoring of a Filament-Wound Composite Rocket Motor Case During Hydroproof," Materials Evaluation, June 1985, pp. 859- 863.

3. Mr. D. Spagnuolo, Composites Branch, U.S. Army Research Laboratory, excerpt from conversation, December, 1993.

4. Dr. E.v. K. Hill, "Predicting Burst Pressures in Filament-Wound Composite Pressure Vessels by Using Acoustic Emission Data," Materials Evaluation, December 1992, pp. 1439-1445.

ON THE QUANTUM ELECTRODYNAMIC ORIGINS OF "SCATTER" IN THE MECHANICAL PROPERTIES OF METALS

Minas Ensanian
National Metals Classification and Ranking Center
Eldred, Pennsylvania 16731

Abstract: A pivotal and historic unsolved problem in the quantum mechanical (QM) electron theory of metals (QMETM), has been, an empirical determination of the global thermodynamic state (GTS) of a discrete bulk metal object (DBMO), defined here as any engineering or structural component. In principle, any physical object can be described by a (QM) wave function or state vector. The new science and technology of Electrotopography (ETG) now makes a practical determination of the virtual thermodynamic state (VTS) of a DBMO possible. Therefore, new "criteria" that are "electronic" in nature (as opposed to being crystallographic) can be utilized for the earliest possible warning of impending catastrophic metal failure. The origins of "scatter" in mechanical property test data are examined from a quantum mechanical and (new) thermodynamical perspectives.

Key Words: Electrotopography; ETG; quantum mechanics; free electron theory of metals; catastrophic metal failure; thermodynamic states; electrochemical sensors; equations of state; foundations of physical theory; topological manifolds; symmetry-breaking; energy dissipation processes; scales; the theory of measurement; and, the Bohr-Einstein debate.

Historical Background: As it pertains to its theoretical and or conceptual foundations, a central theme in the new physics is that of "connectedness." In other words, everything is connected to everything else in the universe. On this basis then, the various divisions of science and technology are artificial constructs (regarding natural phenomena) and are not immune to political forces or fortunes.It may therefore be argued that no matter how a given hierarchical entity is structured, knowledge is a two-way street; for example,even a metal physicist may learn something from a factory-floor metallurgical engineer and vice versa. It is interesting to note in passing that, Benjamin Franklin (1706 - 1790) once observed that a drop of oil spreads itself over the surface of a pond; here was a very simple macroscopic phenomena that provided an opportunity to compute the scale of things at the molecular level. In the world of 1994 is such a thing possible on the factory-floor and is it not in everyone's interest for people at all levels to communicate ? What does "interdisciplinary" mean ?

Physicists, like to do physics. That is, they like to make calculations on the basis of so-called "first principles" and as economically as possible, and then see how well the numbers compare with the results of an actual experiment.

If we were to play a game of "definitions" with the fewest possible words, then "science" could be defined in the following manner - structure-property relationships; as it pertained to any physical system. Let's look into this deeper.

Now if we examine the Periodic Table of the Chemical Elements, we will discover that nearly 80% of the chemical elements are metals (which have been defined as a class of solids that have Fermi Surfaces). It can be shown by combinatorial analysis (supported by our empirical knowledge) that the number of potential alloys may exceed the total number of so-called "particles" in the known universe, which is some 81 orders of magnitude. One can only imagine what chemical, physical, and mechanical properties some of these alloys might possess. How could we do a heuristic search ?

The periodic table may be viewed as an attempt to reduce the total physical universe down to a "single" chart. At the moment no further simplification can easily be imagined; that is at the level of quarks and leptons, etc., nor would it be practical in any way. Numerous questions can be raised, for example, how much information does the periodic table contain ? To what extent can we begin with the table and attempt in anyway to reconstruct any part of the known universe ? How would we know that sodium combines with chlorine to form a white crystalline substance essential for life ? How can anything evolve from such a small chart ?

We are presently living in a wonderful world of satellite communications, pocket computers and TV, Fax machines and so on. This is the world of physical electronics and automation (including artificial intelligence (AI), as well as, virtual reality (VR) etc.) with no end in sight. However, this reality associated with the construction of practical devices, electronic or otherwise, cannot be directly viewed as significant progress in our fundamental understanding of the physical universe around us.

Let us now hand a copy of the periodic table (with data) to a modern-day 24 year old Werner Heisenberg or Wolfgang Pauli and ask him to do some physics. On the basis of any set of first principles or rules, we want to synthesize an alloy of arbitrary chemical composition and, to outline its stress-strain curve for a well-defined set of circumstances.

Now with respect to the theoretical foundations there are only two major theories to work with, namely, quantum mechanics, and general relativity ; it must be noted that quantum theory has essentially explained everything (so-to-speak) except gravity. For various reasons the two theories are incompatible. Physics is not a unified science.

If we were to find ourselves in the world's most complete library - where we would begin a search for first principles . A search through current texts and periodicals in solid state physics, physical chemistry, physical metallurgy, statistical, and or, theoretical mechanics, etc..would

be to no avail - there is no place to begin. The required task would be something akin to Bohr's original postulation of the hydrogen atom, or Schrödinger's postulation of his famous wave equation.

In experimental physics and in general we are dealing with natural or other objects that may be looked upon as being made up of a large number of smaller bodies. This is the central problem in the quantum mechanical theory of many-body systems. This means that in the language of classical mechanics, in order to determine the state of such systems at any point in time and space - we would require a knowledge of the position, as well as, the velocity of some 10 to the 23 particles, and which would be impossible to handle.

Now if on the factory floor, a metal physicist and a quality control engineer were examining a bloom of alloy steel - the engineer who is always under time and or economic constraints of some kind, would primarily be concerned with the bloom's "condition" which he could easily determine at times by a simple hardness test. A physicist, on the other hand, would be mostly interested in the thermodynamic state of the system (which obviously is a more fundamental issue). These two notions are of course interrelated.

As a general rule, the condition of a DBMO could be determined by a single measurement, whereas, the virtual thermodynamic state can only be represented by a state vector whose components would all have to be state variables of a single generic type. The paradigm for this discussion is a cubic meter of an alloy steel whose major, minor, and trace chemical constituents total 25 in number. If we were dealing with a cubic meter container of a gas, it would be a very simple matter to determine its thermodynamic state by virtue of the ideal gas law (equation of state), namely, $PV = nRT$. The model cube above refers to a foundry casting.

It must be noted that so-called many-body theory actually refers to a collection of approximate methods for very specific problem domains; and many-body theory is not concerned with any fundamental or complete understanding of natural phenomena. There is but a single general theorem, that of Poincaré, which states, that if you wait long enough the configuration of a system will eventually repeat itself.

Let us examine briefly the "complexities" that are intrinsically associated with our model cube of alloy steel: (1) we are dealing with 25 or more chemical elements (metallic and nonmetallic), (2) they are not uniformly distributed, (3) the same element may be in different oxidation states in different regions of the cube, (4) the chemistry of the metal's surface may vary from area to area, (5) the cube may be slightly distorted, (6) the surface of the cube may have a number of scratch marks, (7) various forms of defect substructure are randomly present, and the relative internal stored energy of the system is not uniform. People who do first principle physics, obviously have to make various assumptions and work with idealized models, and in the absence

of external forces (and or environmental paramaters). As we have previously indicated our ultimate objective is a true first principle synthesis of a practical alloy that has heretofore, been unknown. Let us say, a whole new class of systems has been postulated.

It is immediately evident that we have given our youthful Heisenberg or Pauli, a truly impossible and incomprehensible task; the factory is waiting for a set of instructions that are supposed to be based essentially on first principle considerations, so that they can proceed to produce the new and perhaps exotic material in quantity.

In the field of so-called "naive physics" we are permitted to ask naive questions. For example, why should a quantum physicist have any interest at all in something as seemingly mundane as catastrophic metal failure ? Is there anything there of a fundamental nature (electron theory of metals) to get excited about ? Answer - a metal may be defined as a solid with a Fermi Surface, and, our whole technical civilization we have been told, is based on the skillful exploitation of the "electronic" properties of metals. Now while it is true that quantum mechanics helps us to better understand various chemical, physical, and electrical properties of metals, it has nothing to say about stress-strain curves and the like. In other words, quantum statistical thermodynamics (QST) is structurally inadequate to handle on a first principle basis the mechanical properties of metals.

This situation is glossed-over by some people in the following ways. (1) "we do not touch upon the independent tradition of the study of the mechanical properties of solids, whose roots lie in metallurgy", and (2) engineers have at their disposal a catalogue of various materials which are purely the results of empirical tests and is not something that relates mechanical properties to atomic structure and to the elementary interactions between atoms.

Our interest in a quantum mechanical (that is, a state vector) approach to system characterization and mechanical failure prevention, etc., is based on the premise that in any quasi-first-principle endeavor to forecast the failure (that is, for example, the useful remaining service-life) of a DBMO, must be based on the assessment of its global "electronic state" or configuration since, quantum electronic changes must precede gross macroscopic morphological or structural change. Such changes must be defined in multidimensional linear state spaces;and registered macroscopically.

It is at this juncture that we encounter the conceptual difficulties associated with the many-body problem. The components of a state vector are, first of all, physicochemical entities and the electrical signals by which they are made manifest must be the combined result of two extraordinary phenomena, namely, chemical specificity, and voluminosity; strictly speaking this would mean the following: (1) the signals are global or volumetric in origin, and (2) they are specifically or mainly associated with one of the major,

minor, or trace, chemical constituents of the DBMO, and which would be indicative of a new physicochemical principle and or, cooperative phenomena. We are therefore forced to develop a new "mapping" technique for bulk metals; one that satisfies these two very difficult requirements, simultaneously. The ultimate objective therefore is to represent any DBMO at or near room temperature by a state vector of the form SV = (Fe,C,Cr,Ni,Mn,S,P,Si,B,Mo,Al,W,Zr.....); the components of the vector would numerically represent the average value of either a single waveform or a mapping of a part or the whole of a system by a sensor that was chemically specific. In other words, a mapping of the DBMO would be made using a so-called "iron" sensor, followed by a "carbon" sensor and so on. Average = arithmetic mean.

Some Quantum Mechanical Considerations: The following comments can be made about quantum mechanics - (1) no one can say that they understand it (after Feynman, and Gell-Mann), (2) Einstein spent a great part of the period between 1927 and 1955 attempting to prove that it was incomplete and not a true description of physical reality, (3) In quantum mechanics (QM) a particle can be in any number of different places, all at the same time, (4) two particles that are a billion miles apart can communicate with one another faster than the speed of light, (5) particles evolve out of the vacuum and may also vanish into it, (6) QM only deals with probabilities, yet the Schrödinger equation is deterministic, (7) any physical system (including the universe itself) can be represented (in principle, of course) by a wave function (or a state vector), (8) a wave function represents "potentia" (that is, various possible "outcomes") and operates on the principle of "linear superposition" of all possible quantum states, (9) the force between two elementary entities (such as between an electron and a proton) is the result of an "exchange" of photons, (10) the Sommerfeld fine structure constant (whose origins are unknown - Why did God select this specific number) alpha (1/137)) is a convenient measure of the strength of the electromagnetic interaction. The universe is held together by this number.

Bohr was fond of stating that anyone who is not shocked by quantum mechanics simply does not comprehend it. Some of the reasons are as follows: (1) There is no ultimate reality; there is only our knowledge of something and that can only be gained by an act of measurement (Einstein, " Are you telling me that if I do not look at the moon, it is not there ?), (2) Bohr, "There is no quantum world - There is only an abstract quantum mechanical description", (3) Mathematics is the only reality, according to the great Heisenberg (Note- Kurt Gödel proved that mathematics itself is also inconsistent), (4) quantum mechanics in and of itself, cannot bring about the "collapse" of the wave function; in other words, it cannot provide any mechanism by means of which one can select a single quantum state from amongst an "ensemble of potentia," and (5) quantum mechanics is a "universal" theory and can be applied to all

systems that are composed of atoms or subatomic particles; this refers to a length scale that ranges from 10^{-8} cm to 10^{-16} cm. Bohr takes measurement apparatus for granted.

In its most simplest terms, quantum theory is concerned with the rules governing the gain or loss of energy from an object. It is here that we see the first hint of a direct connection between quantum theory and catastrophic metal failure since the latter is associated with the rate and the various modes of energy dissipation in a DBMO. It must be noted in passing that during its early history, metals served as a proving grounds for the revolutionary new theory. Although he invented solid state physics, Pauli referred to it as "dirt " (See, Rev. Mod. Phys. 59,1,Jan.1987).

The "quantization" of energy is the reason why, for example, all hydrogen or carbon atoms are the same. If energy was not quantized every atom would be different and there would be no uniformity in Nature.

The Theory of Measurement: A quantum measurement can only be made by a set of apparatus that obey the laws of classical physics (essentially, Newton's world). Bohr takes the world of classical physics for granted and draws a line in the sand so-to-speak, and says that there is a "boundary" between the micro and macro worlds,but exactly where it is, no one is sure or knows;Bohr prefers that the boundary be mobile so that someday quantum theory can encompass all of physics (we assume this includes metal failure prevention).

In the quantum mechanical universe there are only two things, (1) measurement or experimental apparatus, and (2) things to be measured, and the only reality is their interaction or relationship; anything else is undefined and absolutely meaningless. It must be noted that no matter how counterintuitive or weird its appearance, quantum mechanics and quantum electrodynamics in particular is, nevertheless, the most successful theory in the history of science and when correctly applied has never been proven wrong. It can provide correct answers to 6 or more decimal places.

According to Bohr, when a measurement is made the infinitely dispersed wave function (an electron, for example, may be found at any place inside a box) must "collapse" to a definite, but unpredictable state at a particular place. Before we pull a mechanical test sample in tension to failure, the potentia (here) refers to an estimated range of values that let us say lie between 63,000 psi and 218,000 psi. The test is completed when a single value within this range has been selected. In the quantum mechanics of metal failure prevention (QMOMFP) it is our principal task or mission to be able to predict this value by means of a state vector with a high degree of confidence. One of the most interesting questions that arises here is whether a better understanding of the mechanism of mechanical failure in metals can provide new insights on the conceptual foundations of quantum mechanics . There are a number of developments that support

this idea and some of whom are indicated during this discussion; there are a number of similarities between wave function collapse, and "scatter" in mechanical property test data. When a mechanical property test specimen (that may be referred to as the macroquantum mechanical system to be measured) suddenly fails , this is operationally equivalent to the reduction of a state vector. The reason is that we have now correlated the state vector and the complete history of the "measurement act" with a single real point or value in the domain of potentia (the "ensemble" of tensile strengths). A dot on a photographic film or a point on the chart of a recording voltmeter are manifestations of wave function collapse in the classical world. A number of problems, however, remain - how are microscopic "events" amplified so as to produce a classical event and how does classical apparatus make up its mind - which value it wants to select ? There is also the question of stability. If the trajectory of a quantum particle is undefined, how then is it possible to construct a solid from first principles ?

The basic quantum mechanical argument is that by the very act of measurement you change the thing that you are trying to measure: we can only study an electron by striking it with a photon. This is the origin of Heisenberg's Uncertainty.

Thermodynamic Considerations: Max Planck's doctoral thesis was concerned with the second law of thermodynamics which gave him an edge over others attempting to solve the problems posed by black-body radiation. Classical thermodynamics is probably the most single powerful general principle of science and phenomenologically it takes no cognizance of the atomic constitution of matter; even at the turn of the century some of the greatest scientists in the world still did not believe in the actual existence of "atoms." It was Einstein who helped to establish this reality. Einstein himself, on the other hand, did not believe that atomic energy would become a practical reality during his lifetime.

The second law without exception applies to all physical processes in the universe; in other words, you cannot gain something for nothing - even the reading of these lines requires an expenditure of energy. The law is statistical in nature and it applies to closed or isolated systems that tend to degrade if left alone.In a truly isolated system and which is impossible, no form of energy can either leave or enter the system. Do not ask a quantum physicist, "what is energy ?"; he does not know.

The following is a reasonable assumption and or argument; namely, that any isolated physical system should or can be represented by a state vector . Energy, has been defined as the capacity to do work, and work has been defined as the product of an applied force and the distance through which it is applied (work is an algebraic quantity and is the product of an intensity factor and a capacity factor and it may be classified as either - gravitational, surface, volume, mechanical, or electrical);chemical work=electron transfer.

For a gas, a state vector can be expressed as SV = (P,V,T) and it must be noted that the variables refer to the global condition of the system, and that they are "emergent" properties. If a physical system can be represented by a state vector and if the system undergoes some thermodynamic change of state (transformation) then, the magnitude of this change can be expressed by a distance . In other words, the argument can be made that thermodynamics is the science that is concerned with distances in N-dimensional linear state spaces when thermodynamic entities undergo changes of state; let us compare this with a definition by F.T. Wall: Thermodynamics is a physical science concerned with the transfer of heat and the appearance or disappearance of work attending various conceivable chemical and physical processes.In classical thermodynamics the definition of energy (First Law) is given only in terms of its changes and not by its absolute magnitude for an isolated system. Absolute values can only be specified by means of Einstein's formula for mass-energy, $E_o = m\ c^2$; postulated in 1900 by Poincaré.

The Notion of Distance in Physical Theory: In the prevention of mechanical failure, for economic, as well as other reasons, the central issue must be concerned with the useful remaining service-life of an engineering component or a structural entity. Intuitively, we are interested in the computation of a distance between the present state of a system and its failed or final state (where a tensile test specimen is broken into two pieces) when the system is under some reasonably defined constraint (regimen).

Operational realities aside for the moment. A state vector for any isolated physical system implies some measure of its energy content or state; which is foreign to classical thermodynamics. Now what happens when "energy" either enters or leaves a system ? For example, if a quantity of wax is brought into contact with a hot body it melts. There is a change in molecular configuration. In a metal the atoms become further apart as the temperature is increased. In other words, there is a "duality" between energy and distance just as there is between mass and energy, space and time, and a particle and a wave. It can be shown that it is impossible for energy to enter or leave a system without a corresponding change in some of the distance parameters of the system; energy and distance are, therefore, two sides of the same coin. They cannot change in value independently.

The Molar Distance of A Pure Metal: The molar distance of a pure metal near 298^0 K is defined by the relation, $D_M = m_A / [\rho \times r^2_{nnd}]$ where m_A, ρ (rho) , and r^2_{nnd} , respectively, refer to the product of the atomic weight in grams of a metal and Avogadro's number, the density of the metallic solid near 298^0 K, and the nearest neighbor distance in centimeters in the standard crystalline material. The atomic volumes of the practical metals range from a low of 4.96 cm^3/ mole for beryllium to 70 cm^3 /mole for cesium. The uranium/lithium atomic weight ratio is 34.303 and their density

ratio is 35.283 . On the other hand, their molar distance ratio is nearly unity at 1.177 . The molar distances of the practical metals, as well as the metalloids silicon, germanium, and carbon, ranges roughly from 1.0 to 2.53 x 10^{16} cm. A quantity that is referred to as the standard molar distance (D_{Ms}) is by definition, made equal to 2.104 x 10^{16} cm.

Einstein's Equation: In the special theory of relativity a mass of 1 gram is equivalent to 9 x 10^{20} ergs of energy. By the simple expedient of dimensional equivalence we can in effect, transform this energy into its distance equivalent.

$$E_o = m c^2 = D_o m_s g \qquad (1)$$

where E_o is equal to 9 x 10^{20} ergs, m_s is equal to 1 gram of mass, and g is the gravitational acceleration at sea level which is 9.8 x 10^2 cm sec^{-2}. The notion of distance has considerable pedagogic, as well as heuristic value in both physic and metallurgical engineering (and which are ultimately connected). As a matter of fact all of the physical constants of physics can be translated into their distance equivalents, that is into centimeters. In equation 1, the quantity D_o is found to be equal to 9.18 x 10^{17}cm.

For nearly eight decades the origins of Sommerfeld's fine structure constant alpha (1/137) has remained one of the greatest mysteries in physics. In the words, of R. Feynman, "all good theoretical physicists put this number up on their walls and worry about it". Alpha represents the ratio of the velocity of the electron in the first Bohr orbit to the velocity of light. Note, it is obtained directly from the ratio of D_o and D_{Ms} and is approximately

$$\frac{\pi \, D_o}{D_{Ms}} = 137.07 \qquad (2)$$

A considerable body of work has recently been developed around the concept of distance translations. For example, if D_o and E_o are divided by Avogadro's number (treated as a dimensionless quantity) we obtained, what are respectively, referred to as, the quanta of distance (1.5244 x 10^{-6}cm) and the quanta of energy (1.4945 x 10^{-3} ergs), which are very useful.[1] Distance, is independent of the path taken.

It is interesting to note that in the context of dimensional units, Einstein's formula and the classical expression for the gravitational potential energy (near sea level) are identical. It is also obvious that arguments in physics must always involve varying degrees of "circularity". A theory prescribes at once both the definitions of its concepts and the conditions in which they are to be related; classical thermodynamics, for example, is related to equilibrium states. In physical science we no longer pretend to have found

the truth but merely consistent descriptions which seem to have predictive value (after, M. Tribus, et al.) . In this same vein, information theory attempts to make the best predictions about a system with nothing more than incomplete information. We now review a new experimental phenomena.

Electrotopography (ETG): It is now a matter of routine to determine the virtual thermodynamic state of a discrete bulk metal object , irrespective of its chemical composition and shape or mass, in real time. Information is transmitted by electrical signals which are related generally to changes in physical state. A digital value, a waveform, and a 3-D or fishnet mapping, respectively, refers to things that are 1,2 and 3 dimensional. A state space on the other hand can have an infinite number of dimensions (Hilbert spaces, for example, in quantum mechanics).

2-9

ETG is a new science and technology that evolved out of an attempt to solve a factory-floor problem in engineering concerned with the classification of mixed populations of critical metallurgical engineering components (CMEC); defined as those objects whose catastrophic failure could result in a loss of human life, valuable property, and or, costly downtime. Classification and Ranking are pivotal to safety .

When two metallic objects are brought into contact they exchange charge but not significance. Due to the statistical variations in the configurations of their atomic constituents, it is impossible for any two physical objects to be exactly the same. Therefore, when they are brought into contact, and due to the differences in their respective Fermi levels, there will be a transfer of charge. Heretofore, it was not possible to measure and to compute the entropy of production in any limited manufacturing cycle. Whenever any two metallic objects are in common contact with an electrolyte we have a galvanic cell and an EMF is generated. In its most simple form (there are 55 general classes of ETG sensors) an ETG mapping device consists of an electrode in the form of a bar upon which has been mounted a doped polymeric electrolyte in the form of a tire, and which very briefly, as it rolls over the surface of a metal, generates an electrical signal (with the units of electrical field strength) that is both volumetric in origin, as well as, chemically specific. Waveform or mapping averages then serve as the components of state vectors or matrices. Experimental details may be found in the references cited. Theoretically, ETG can only be understood on the basis of a quantum mechanical "coupling" between two macroquantum systems.

In quantum mechanics a measurement refers to an interaction between a very large apparatus and quantum entity, such as a single particle. In ETG these same two can be identical and can even exchange their roles: if a bar is cut in half, either section can be the system to be measured or part of the measuring apparatus. Quantum measurements can be done on very large objects (such as, a multi-ton Weber bar or antenna - such measurements are known as quantum nondemolition (see Science, Vol. 209, 1 August 1980, P-547).

Scatter in Mechanical Testing: The word "scatter" is very seldom found in the Index of any textbook in the material or physical sciences. In his description of a reversible process in classical thermodynamics Gibbs defines the notion of a "neighborhood" (which was later taken up by Carathéodory). In other words a system that is taken through a reversible cycle does not have to return to its exact previous position but only to a neighborhood where there is no sensible difference between the two states. Many deep issues can be raised here. For example, in ETG it would be possible to reversibly transform a large solid metal sphere into a cube (experiments of this kind can be done with alloys that melt at 50° C, for example). The state vectors, although they are generically similar are nevertheless, different. One of the most interesting problems in ETG concerns determining the shape of an object, only from the information that is contained in its state vector and a knowledge base for its parent population. Problems are defined only by SV's.

Physics as pointed out by W. Poundstone (The Recursive Universe) is an attempt to define reality recursively. Ideally, it is a set of rules for predicting the situation at time T_1, given the situation at time T_0 . We refer to the latter as the initial state . Industrial quality control in the metalworking industries is strictly based on statistical methods, simply because heretofore there was no way to determine the initial state of a DBMO. Our principle interest is in the description of a stress-strain curve in terms of individual elemental constituents (one at a time) of the metallic body. Tensile tests have N-D chemical perspectives.

Scatter in mechanical testing, chaos in nonlinear dynamical systems, and especially, patterns in cellular automata, all display an extreme sensitivity to so-called "initial conditions." When test specimens are represented by state vectors, then scatter can be linearized, that is, predicted. It is also interesting to note that mutation and other operators (actually algorithms) can be developed, whereupon a single state vector can be enlarged into a neighborhood, that is a population. This is the inverse of wave function collapse or state vector reduction (conceptually speaking).

The Bohr-Einstein Debate: Quantum theory deals with probabilities, and if it takes the classical world for granted, it cannot be universal. In recent years there have been two major international conferences on the Bohr-Einstein debate which is concerned with the ultimate nature of physical reality. In Einstein's viewpoint, "God does not play dice with man." In this author's opinion there is a close fundamental connection between physical reality and catastrophic metal failure. Both subjects are closely associated with the VTS's of physicochemical entities, as well as the geometrization of all natural phenomena (Clifford, Gibbs, Etc.)

Summary and Conclusion: A knowledge of the VTS of a DBMO makes it possible to optimize simultaneously its past history , as well as its future behavior.

The determinations of the past and the future of a physical system may be viewed as another form of complementarity in the theory of quantum mechanics (although, in a limited way). The VTS of the model cube is most descriptive when it is represented by a set of 3-D topological manifolds which are electronic entities resulting from joining a cube's 6 mappings; when properly read , these figures will clearly indicate when the system is in a "failure mode." The volumetric nature of these signals is QM in origin, while the chemical specificity is brought about by chemical biasing.[6] The history of QM measurement is referred to as "state preparation," and must be described in great detail. An ETG sensor screens "potentia" by virtue of the ionic potential. A practical measure of the VTS of a DBMO means that more than a century of modern metallurgy can be revisited by a new thermodynamic tool. Finally, there is preliminary empirical evidence that the ratio of the mechanical property Minkowskian Distances between two members of the same population (that have been tested to tensile failure) and their ETG (that is, physicochemical) Minkowskian Distances may be connected by the fine structure constant alpha. In conclusion, there is no doubt that a deeper understanding of the phenomena associated with "scatter" raises interesting issues in the conceptual foundations of quantum mechanics. There is a connectedness here that must be pursued in an interdisciplinary way; ETG spectra can also be obtained as a function of mechanical vibrational frequencies (electron-phonon scattering phenomena; Ref. 3).

References

1. Ensanian, M., Technical Report ETG-93-1, Experimental FS-EPI, National Metals Classification and Ranking Center, Eldred, Pennsylvania, Part 1, 1993, pp. 1-115.
2. Ensanian, M., "Artificial Intelligence and Deep Knowledge in Metallurgical Testing and Failure Analysis," Proc. Intl. Conf. on Fatigue, Corrosion Cracking, Fracture Mechanics, and Failure Analysis, Salt Lake City, Dec. 2-6, 1985, American Society of Metals, Metals Park, Ohio. Volume on Failure Analysis, Edited by V.S. Goel.
3. Ensanian, M., "Electrotopography: A New Tool for Modeling Dynamical Mechanical Structures," Proc. 1st Intl. Modal Analysis Conference, Orlando, Union College, Schenectady, 1982, pp. 216-222.
4. Ensanian, M., Heat Treating, Vol. XVI, No. 9, Fairchild Publications, New York, Sept. 1984, pp.33-41, (cover).
5. Ensanian, M., Robotics World, Vol.2, No. 3, March 1984, pp. 24-27.
6. Ensanian, M., Technical Symposium Proceedings - Metalform 91', Chicago, March 10-13, Precision Metalforming Association, Richmond Heights, Ohio, Vol. 1, No.1, 1991, pp. 483-585.
7. Ensanian, M., Conference Proceedings, Fastec '87, Atlanta, October 27-29, Society of Manufacturing Engineers, Dearborn, 1987, pp. 7.1-7-36.
8. Ensanian, M., Proceedings 8th Intl. Conf. on Automated Inspection and Product Control, Chicago, 23-25, June 1987, MPC-IITRI, edited by K.E. McKee, IFS, Springer-Verlag, Berlin, 1987, pp. 241-252.
9. Ensanian, M., Proc. of Intl. Joint Conf. on Neural Networks, IEEE, Baltimore, June 7-11, Piscataway, 1992, Vol.III, pp. 438-443.

EVALUATION OF MATERIALS PROPERTIES

Cochairmen: **Kirit J. Bhansali**
U.S. Army ATCOM

Patrick J. Sincebaugh
U.S. Army Research Laboratory

FAILURE MECHANISMS OF TITANIUM ALUMINIDE ALLOYS

Wego Wang
US Army Research Laboratory-Materials Directorate
Watertown, MA 02172-0001

Abstract: This study analyzes the fracture characteristics of three representative titanium aluminide alloys: a super-α_2 Ti-25Al-10Nb-3V-1Mo alloy, a near-γ Ti-48Al-2Nb-2Cr alloy and an XD® Ti-45Al-2Mn-2V alloy containing 7vol% TiB_2. A plate-like α_2 phase in the super-α_2 alloy provides a good combination of ductility and strength. The effects of constituent phases and the controlling mechanisms for tensile and stress rupture failures are discussed. A duplex microstructure of the near-γ alloy is observed in the heat-treated samples. SEM studies show a clearly defined crack initiation point in the tensile specimens. The tensile fractography shows a mixed brittle rocky and river-pattern fracture mechanism at 977K, while switching to a more ductile dimple fracture mechanism at 1033K. It is concluded that a brittle to ductile transition occurs between 977 and 1033K. Interlamellar fracture dominates the tensile failure mechanism of the XD® alloy with fully lamellar or duplex microstructure. The microstructural effects on failure mechanisms are established and the findings will serve as guidelines for future alloy optimization.

Key Words: Fracture mechanism; γ titanium aluminide; super-α_2 titanium aluminide; titanium diboride; XD®

Introduction: The low density, high modulus, and persistent high temperature strength of titanium aluminide alloys make them attractive candidates for applications at high temperatures in advanced gas turbine engines. However, poor room temperature ductility dramatically restricts their otherwise great potential. It is essential to understand the fracture characteristics and the controlling mechanisms to further improve these alloys. This study will compare, analyze and explain the fracture characteristics of three titanium aluminide alloys: a super-α_2 Ti-25Al-10Nb-3V-1Mo (at%) alloy produced by rapid omnidirectional compaction (ROC), a forged near-γ Ti-48Al-2Nb-2Cr (at%) alloy and an XD® Ti-45Al-2Mn-2V (at%) alloy containing 7vol% TiB_2.

Materials and Processing: The super-α_2 alloy used in this study is a titanium aluminide alloy, Ti-25Al-10Nb-3V-1Mo (at%) [Ti-14.1Al-19.5Nb-3.2V-2Mo (wt%)] consolidated from plasma rotating electrode processed powder by ROC. The heat treatment schedules listed in Table I were followed to develop the microstructure that optimized the mechanical properties. Also given in Table I is the average aspect ratio (the ratio between the length and the width) of the transformed α_2 phase (except the as-ROC'ed material where the aspect ratio refers to the primary α_2). Heat treatment 1 produced scattered areas of primary α_2 in a mixed α_2/β matrix with well-defined β grains. Heat treatment 2 gave large amounts of equiaxed primary α_2 along with fine acicular α_2 in a β matrix. Heat treatment 3, cycled above and below the β-transus temperature (1363K), resulted in a transformed colony-type α_2 structure with grain-boundary α_2 present. Heat treatment 4 resulted in a fine Widmanstatten appearance. The microstructures of the super-α_2 alloy are shown in Figure 1.

Figure 1: Microstructures of super-α_2 alloy (a) as-ROC'ed, (b) heat treated by schedule 1, (c) heat treated by schedule 2, (d) heat treated by scheduled 3, and (e) heat treated by schedule 4.

Table I. Heat Treatment Schedule and the Average Aspect Ratio of α_2 Phase

Heat Treatment	Schedule	Average Aspect Ratio
φ	as-received (as-ROC'ed)	3
1	1366K x 2hr + water quenched + 1144K x 2hr + water quenched + 1033K x 4hr + air cooled	6
2	1339K x 2 hr + 1033K x 5 hr + air cooled	9
3	1403K x 1 hr +, 1089K x 1.5 hr + 1403K x 1 hr + air cooled	10
4	1413K x 1 hr + 1089K x 4 hr + air cooled	5

Both the near-γ and the XD® alloys were first cast by a vacuum arc remelting process, followed by hot isostatic pressing (HIP) and isothermally forging, and finally heat treated before testing. The average compositions are Ti-33.4Al-4.38Nb-2.47Cr-0.085Fe-0.012C-0.073O-0.005N+22 ppm H (wt%) for the near-γ alloy, and Ti-29.45Al-2.23V-2.64Mn-2.28B (wt%) for the XD® alloy. The near-γ pancake was supplied by General Electric Aircraft Engines (GEAE) to the Army Research Laboratory-Materials Directorate (ARL-MD) for testing and analysis. The as-received material showed a duplex microstructure of equiaxed primary γ and transformed α_2 and γ laths as shown in Figure 2. The study of the XD® alloy was conducted at Martin Marietta under contract from ARL-MD. The presence of TiB_2 particulates in titanium aluminide alloys could refine the cast grain size, accelerate γ-phase recrystallization and strongly influence morphology transformations.

Figure 2: The duplex microstructure of the near-γ alloy.

Mechanical Properties: Tensile and stress rupture tests were conducted for all three alloys. For a super-α_2 sample heat treated by schedule 4, the high room temperature yield strength (YS) and ultimate tensile strength (UTS), 977.7 and 1174.9 MPa, respectively, were accompanied by poor room temperature ductility, <2%; this is a consequence of the presence of the fine acicular α_2 phase. The tensile strength is compromised when the acicular α_2 coarsens, while room temperature ductility improves as indicated by a sample heat treated by schedule 3 (Sample 3), having a YS and a UTS of 617.8 and 848.8 MPa, respectively, and a reduction of area (RA) of 6.5%. The mixed structure of fine acicular transformed α_2 and equiaxed primary α_2 obtained in the sample heat treated by schedule 2 provides a better tensile strength than Sample 3 but reduced ductility. The tensile properties at elevated temperature follow a trend similar to the room temperature properties. An interesting conclusion can be drawn regarding the tensile properties and the morphology of the α_2 phase. Under otherwise similar conditions, ductility increases as the aspect ratio of the acicular α_2 phase increases but the tensile strength is inversely related to the aspect ratio. The as-ROC'ed material shows a roughly equiaxed primary α_2 structure compared with a much more plate-like structure in Sample 3. Despite the coarse microstructure found in the as-ROC'ed material, it demonstrates better YS and UTS at both room and elevated temperatures, with lower ductility. This further supports the idea that the geometrical shape, i.e., aspect ratio, has a profound effect on the tensile properties and can, at least sometimes, outweigh the effect of refined microstructure. The stress rupture properties of this super-α_2 alloy are remarkable compared with other similar alloys. (1)

The UTS and YS of near-γ alloy remained essentially the same as the test temperature increased from 977 to 1033K while the elongation (EL) and the reduction of area (RA) values dramatically increased from 4.0 to 89.9% and 4.3 to 76.2%, respectively. The UTS abruptly decreased by more than 20%, only a slight decrease in YS, when the temperature increased from 1033 to 1089K. The EL and RA values moderately increased as the temperature increased to 1089K. The moderate change in YS despite the significant deterioration in UTS is partially due to a transition from a brittle to a ductile fracture mode. Stress rupture tests were performed at higher temperatures. The stress rupture time is more than 1,000 hours at 922K/275.8 MPa. (2)

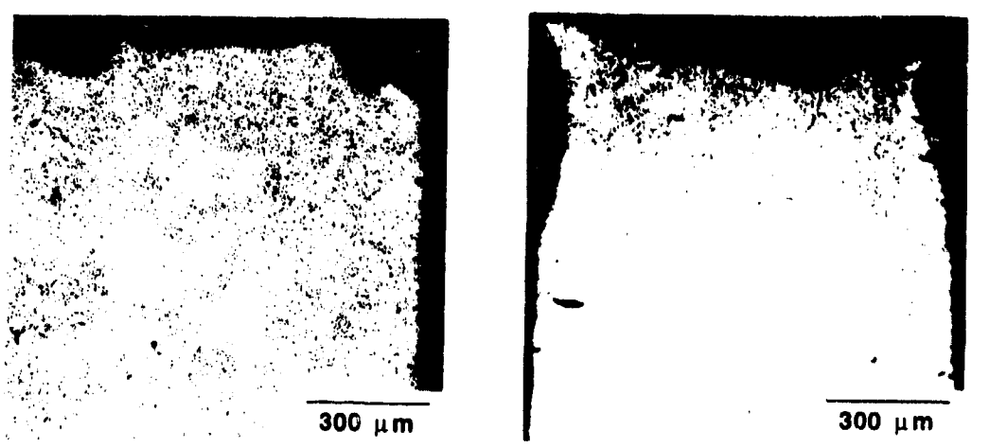
The as-HIP'ed XD™ alloy showed a fully equiaxed single-phase γ microstructure that transformed to a lamellar microstructure after being heated at 1613K for 1 hour followed by air cooling to room temperature. A duplex lamellar and equiaxed (15-20vol%) microstructure was produced in specimens heated at 1613K for 1 hour, air cooled, and subsequently heated at 1473K for 10 hours then air cooled again to room temperature. The lamellar and the duplex microstructures had higher tensile strength compared with their equiaxed counterpart. However, the fully equiaxed γ microstructure produced the highest ductility at ambient temperature. This is opposite to that observed in the super-α_2 alloy: a lower ductility is associated with the equiaxed α_2 structure. For both the lamellar and the duplex microstructures, YS was fairly temperature insensitive at temperatures below 873K, remaining above 550 MPa. The YS would rapidly decrease to 200 MPa as the temperature increased to 1173K. A brittle-to-ductile transition started to occur at 973K. The tensile elongation reached a value higher than 5% at 1173K for both the fully lamellar and the duplex microstructures. The XD® alloy has a higher tensile strength than the near-γ alloy, but an inferior ductility. Compared with the near-γ alloy, the XD® alloy has a similar brittle-to-ductile transition temperature range, however, with more decrease in strength and less increase in ductility during the transition. (3)

Fracture Characteristics: Figures 3(a) and (b) show longitudinal (i.e., parallel to the load axis) cross-sectional microstructures of super-α_2 tensile specimens tested at 700K. They both show a flat transverse fracture surface with very limited secondary cracks immediately underneath the fracture surface indicating a single dominant crack. For the as-ROC'ed sample, as shown in Figure 3(a), a number of transverse cracks were nucleated along the outer specimen surface in the vicinity of the fracture surface. Most of them were short and randomly distributed indicating no preferential nucleation sites. More transverse cracks evenly distributed on the outer specimen surface were observed for samples heat treated by schedule 2. In contrast to the as-ROC'ed specimen, the transverse cracks occurred farther away from the fracture surface indicating a dual deformation process was present during the test. The first is uniformly distributed througnout the specimen in the longitudinal direction, nucleating cracks and contracting the specimen diameter almost up to the shoulders of the specimen; the second is a localized necking process only observed in the vicinity of the final fracture surface as shown in Figure 3(b). In addition, a large number of irregularly shaped cavities, ranging from spherical to elongated, nucleated within the specimen for both as-ROC'ed and heat-treated (by schedule 2) specimens. These crack nucleation sites occurred in several different phase morphologies.

Figure 4(a) shows a longitudinal cross-sectional microstructure of an as-ROC'ed stress rupture specimen tested at 922K/379 MPa that broke after 8.2 hours. During the test multiple transverse cracks nucleated both on the specimen surface (edge) and within the specimen. No longitudinal cracks were detected. Most transverse cracks nucleated at the specimen surface are partially due to high stress concentration resulting from surface roughness and partially due to the brittle oxide surface layer. These surface cracks grow competitively in the transverse direction and some eventually coalesce with outwardly growing interior cracks to produce final failure. Despite the fact that the interior cracks were few and scattered, some are actually fracture-controlling because the "small" surface cracks have to coalesce with interior cracks to reach the critical crack length for propagation leading to failure. No preferential crack nucleation sites were observed. As shown in Figure 4(b), one large crack on the left side was initiated in the fine α_2/β matrix, the middle crack was initiated on the interior boundary (interface) between coarse primary α_2 phase and fine α_2/β matrix, the third major crack on the right side was initiated in a cluster of coarse α_2 phase. The crack propagation rates are seemingly slow as indicated by the short lengths of these surface cracks and are about the same no matter where they are nucleated. The failure is controlled by the interior crack nucleation and the subsequent linkage between the interior and the surface cracks. When multiple cracks happen to

nucleate on the same plane in the specimen, they will soon coalesce and reach the critical length leading to the final fracture. The grain boundaries of prior β grains were not necessarily favorite crack nucleation sites nor preferred paths for crack propagation. Cracks were randomly nucleated throughout a variety of microstructure morphologies. The frequency of crack nucleation and the subsequent crack propagation rate are about the same among coarse primary α_2 phase, fine α_2 /β matrix, prior β grain boundary and the boundaries (interfaces) between these various phases. Crack nucleation and propagation are predominantly affected by the direction of stress (i.e., tension) and stress concentration, and not the different cohesive strength of various metallurgical phases and constituents.

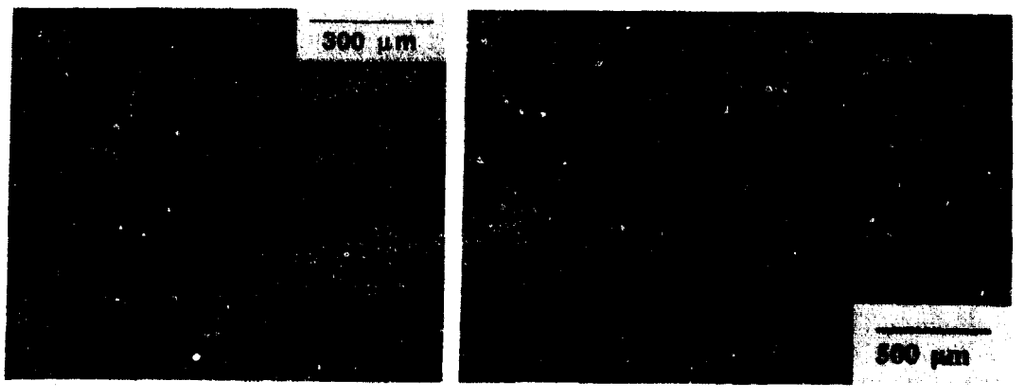
Table II summarizes the fractographic features of super-α_2 specimens tested in various conditions. The prior β grain size for a post-tested specimen is about the same as before the test for both tensile and stress rupture specimens. In the stress rupture specimen, microvoids nucleated during the test at elevated temperature and subsequently coalesced to form the colonies as shown in Figures 5(a) and (b). The colony size observed in the scanning electron microscopy (SEM) fractographs is about 200 μm, about five times larger than the prior β grain size. This indicates that during the stress rupture test multiple cracks nucleate in several prior β grains and then grow simultaneously and eventually bridge together resulting in final fracture. The nucleation and coalescence of microvoids are heavily dependent on temperature, stress and the exposure time at these elevated temperatures and high stresses. The colony-type fracture is less well defined for tensile



(a) (b)

Figure 3: Longitudinal cross-sectional microstructures of failed super-α_2 tensile specimens tested at 700K (a) as-ROC'ed, (b) heat treated by schedule 2.

tests at 700K as shown in Figures 6(a) and (b). Ductile dimples and microvoids are still visible in these tensile specimens, however a shear river-pattern fracture also appears. The fracture surfaces of room temperature tensile specimens as-ROC'ed show predominantly shear river-pattern fracture - no microvoids or colonies can be observed. However, for a sample heat treated by schedule 2, large (about 300 μm) but vaguely defined colonies on the fracture surface were observed as shown in Figures 6(c) and (d). These variations are attributable to the different starting microstructures. The as-ROC'ed specimen starts with a coarse α_2 structure with vaguely defined β grains. On the other hand the microstructure of the specimen heat treated by schedule 2, had a mixture of well defined β grains and primary α_2 phase.



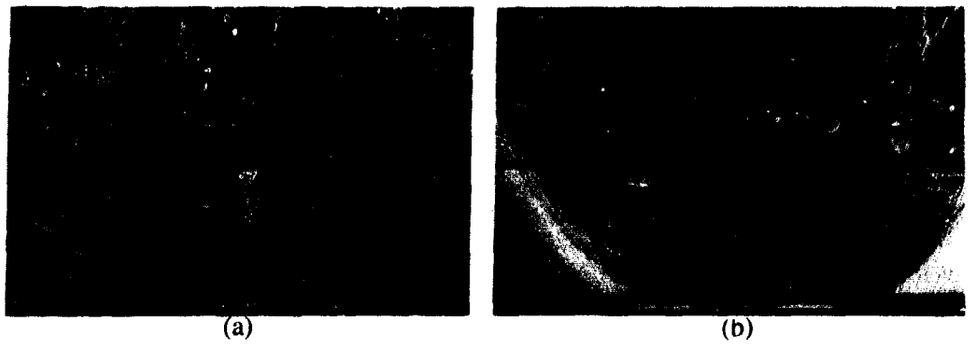
(a) (b)

Figure 4: As-ROC'ed super-α_2 stress rupture fractured specimens (a) longitudinal cross section, (b) transverse surface cracks.

Table II. Fractographic Features of Super-α_2 Alloy

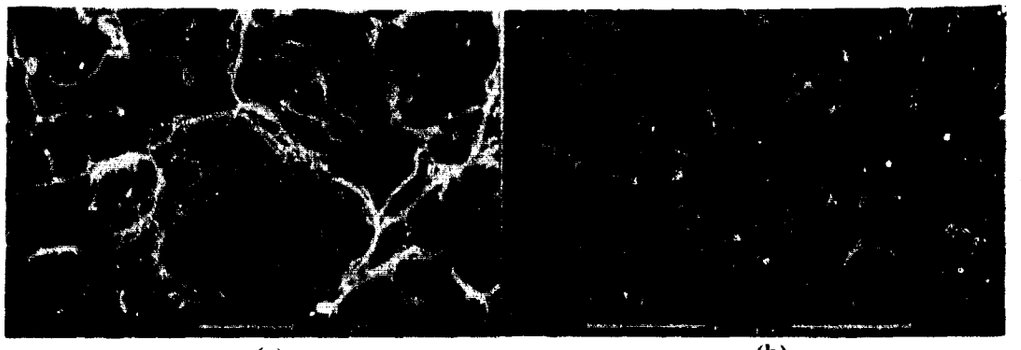
Specimen	SEM	Optical
as-ROC'ed (tensile at 700K)	shear river-pattern, ductile dimples, microvoids	several transverse cracks along the specimen surface near the fracture cross section and irregular cavities within the specimen
as-ROC'ed (stress rupture at 922K/379 MPa)	transgranular colony-type fracture, colony size ≈200 µm	multiple transverse cracks randomly nucleated, prior β grain size ≈ 40 µm
heat treated by schedule 2 (tensile at 293K)	large but vaguely defined transgranular colony-type fracture, colony size ≈ 300 µm	-
heat treated by schedule 2 (tensile at 700K)	mixture of transgranular colony-type & intergranular shear fracture, microvoids	a number of transverse cracks along the surface & irregular cavities within the specimen
heat treated by schedule 2 (stress rupture at 922K/379 MPa)	mixture of transgranular colony-type (interior) and intergranular shear (surface/edge) fracture , microvoid coalescence	-

Both ductile dimple and brittle (mainly intergranular) fracture modes were observed in the near-γ alloy. Figures 7(a) and (b) show the typical tensile-fractured specimen surface tested at 977K. A clearly defined crack origin was observed. The singularity of the crack origin and a river pattern that diverged from this origin in the crack propagation direction led to the conclusion that the specimen failed by a single predominant crack that was initiated on the specimen surface and gradually propagated through the cross section until the remaining cross-sectional area could not sustain the applied load and an abrupt



(a) (b)

Figure 5: SEM fractographs of super-α_2 stress rupture fractured specimens (a) as-ROC'ed, (b) heat treated by schedule 2.



(a) (b)

(c) (d)

Figure 6: SEM fractographs of super-α_2 tensile specimens (a) and (b) tested at 700K, as-ROC'ed and heat treated by schedule 2, respectively; (c) and (d), tested at room temperature, as-ROC'ed and heat treated by schedule 2, respectively.

separation occurred. Figure 7(b) shows a close-up examination at a higher magnification indicating a brittle rocky fracture near the crack origin. Decohesive rupture occurs principally along the grain boundaries. A pronounced intergranular cleavagelike fracture mode with some transgranular contribution can be seen. Clusters of cleavage feathers could also be observed. Figure 7(c) shows the cracks on the external specimen surface resulting from oxide layer on the tensile specimen tested at 977K. The morphology and

distribution of these surface cracks are a function of their respective relative locations to the crack origin and the test temperature. More cracks were observed near the crack origin. They usually appear as linear cracks circling around the external surface. However, the morphology became more complex and the cracks were deeper into the interior for the specimens tested at higher temperatures. These cracks developed due to the propagation of secondary surface cracks. A distinctly different fracture appearance was observed in the tensile specimen tested at 1033K. A mixed ductile/brittle transgranular decohesion with twisted grains dominated the fracture mechanisms as shown in Figure 8(a). Figure 8(b) shows the ductile tubular fracture observed in this specimen. This flutelike tubular fracture results from steps developed during rupture. The tensile specimen tested at 1089K showed a more ductile fracture. The specimens failed by dimple rupture and exhibited transgranular decohesion as shown in Figure 8(c). These findings are consistent with the conclusion obtained from the tensile tests where the ductile to brittle transition temperature is between 977 and 1033K. The EL increased from 4.0 to 89.9% when the test temperature increased from 977 to 1033K.

(a) (b) (c)

Figure 7: The near-γ tensile specimen tested at 977K, (a) overall fracture surface, (b) brittle rocky fracture near the crack origin, and (c) surface cracks.

(a) (b) (c)

Figure 8: Fracture surfaces observed in the near-γ tensile specimens (a) and (b) tested at 1033K, (c)tested at 1089K.

In contrast to the simple tensile fracture, the SEM stress rupture fractographs of the near-γ alloy do not show any crack origin or crack path. Despite the good elongation at fracture the SEM fractographs exhibit a combination of cleavage and transgranular decohesion. Stress rupture tests were performed at three different conditions: 922K/275.8 MPa, 977K/275.8 MPa and 1033K/206.9MPa, respectively. Figure 9 shows the typical stress rupture SEM fractograph for a specimen tested at 977K/275.8 MPa. All stress-ruptured specimens show deformation in different degrees. However, the specimen tested at 1033K/206.9MPa reveals the least deformation and the finest microstructure among these three. It indicates a quasi-ductile fracture mechanism. More brittle fractures and massive shears could be observed in a large area in the specimen tested at 922K/275.8 MPa. The specimen tested at 977K/275.8 MPa has a fracture mechanism in between the two mentioned above. It shows a mixed brittle and ductile fracture mode. The effect of test temperature is particularly significant from 977 to 1033K because this is the brittle-to-ductile transition temperature range of γ (TiAl) alloys. The tensile EL and RA increase significantly in this transition, and these ductility changes were accompanied by distinct fracture appearance changes in both tensile and stress rupture specimens.

SEM fractography of XD® alloy shows that fracture occurs along the interlamellar interface for both the lamellar and the duplex microstructures tensile tested at 293 and 773K, as shown in Fig. 10.

Figure 9: Stress rupture fracture surface of a near-γ specimen tested at 977K/275.8MPa.

Figure 10: SEM fractograpgy of an XD® specimen with lamellar microstructure tensile-tested at 293K. (3)

Discussion and Conclusions: The failure mechanisms of titanium aluminide alloys depend on their morphologies of the constituent phases. The morphological features of these alloys vary with alloy compositions, processing and heat treatment. The tensile fractures of both super-α_2 and near-γ alloys result from a predominant crack propagating through the specimen cross section. Cracks nucleated on the specimen surfaces due to oxidation were observed in both alloys. Secondary cracks were initiated in the specimen interior as well. In the super-α_2 alloy the ductility increases with an increasing aspect ratio of the α_2 phase, however, the strength is inversely related to the aspect ratio. In general, structures with plate-like α_2 phase provide a good combination of ductility and strength.

The room temperature ductility of near-γ alloy is relatively low, and the equiaxed γ microstructure usually has the lowest ductility. A brittle-to-ductile transition is found between 977 and 1033K for the near-γ alloy. Interlamellar fracture is the predominant fracture mode in the XD® alloy for both fully lamellar and duplex microstructures. Similar interlamellar fracture was also observed in the as-ROC'ed super-α_2 alloy that shows an elongated α_2 grain morphology. In the super-α_2 alloy, the colony morphology resulting from the prior β grain plays an important role in the fracture mechanism. It leads to a very different macro-fracture appearance between the super-α_2 and the near-γ alloys despite the observation that both alloys show a heavily deformed micro-structure in the stress-ruptured fracture surface. The stress rupture failure of the super-α_2 alloy is controlled by the coalescence of microvoids and the bridging between simultaneously nucleated cracks on the same plane. Colony-type fracture was observed for stress rupture and tensile specimens with a β phase microstructure. For stress rupture and tensile specimens containing β phase, fracture occurred across colonies. The average colony size was usually five to seven times larger than the prior β grain size and resulted from the linkage of cracks in several β grains. In the near-γ stress rupture specimens, a fractograph with less and smaller ductile dimples was observed .

Acknowledgment: This project was partially sponsored by the U.S. Army Aviation and Troop-support Command, Aviation Applied Technology Directorate, Fort Eustis, VA. The author is indebted to Mrs. Marietta R. Scanlon for her help in experimental work, Mr. Paul Huang for his assistance in SEM. He is also grateful to Dr. Daniel Eylon for supplying the super-α_2 alloy and Mr. James Chesnutt and Mr. Robert Hunt of GEAE for providing the near-γ alloy for this study.

REFERENCES

1. W. Wang, M.R. Scanlon and M.G.H. Wells in High-Temperature Ordered Intermetallic Alloys IV, edited by L.A. Johnson, D.P. Pope, and J.O. Stiegler, Mater. Res. Soc. Proc. **213**, Boston, MA 1991, pp. 931-936.

2. W. Wang, P.J. Huang, J.C. Beck and Martin G.H. Wells, "High Temperature Characteristics of Ti-48Al-2Nb-2Cr," Fatigue and Fracture of Ordered Intermetallic Materials Symposium II, TMS, 1993 Fall Meeting. Pittsburgh, PA (to be published in the proceedings).

3. Pamela Sadler, K. Sharvan Kumar and John A. S. Green, Applicability and Performance Benefits of XD® Titanium Aluminides to Expendable Gas Turbine Engine, Army Research Laboratory Contract Report, ARL-CT-61, 1993.

FATIGUE OF BUTT WELDED JOINTS IN STEEL PIPES

R. J. Scavuzzo, P. C. Lam, T. S. Srivatsan
The University of Akron
Akron Ohio 44325-3909

Abstract: The objectives of this experimental study are to extend the fatigue data base of butt welded pipe from Markl's tests into the low cycle range (100 to 2,000 cycles), to compare the measured fatigue strength with formulas developed by Markl of stress amplitude versus cycles to failure, and to compare results with other available data. Markl's paper covered the range from 2,000 to 4 x 10^7 cycles.

In this investigation, a total of fourteen (14) butt welded pipe carbon steel specimens and seven (7) 304 stainless steel specimens were tested using a four point bending test fixture. All testing was conducted on schedule 80 1¼ inch pipe subjected to a constant cyclic displacement amplitude. Force deflection data were taken so that results could be compared with Markl's stress data. In addition, strain amplitude was measured so that strain-cycle curves could be developed for the butt welded pipes.

Failures in the carbon steel pipe occurred at the end of the weld heat effected zone on the inside of the pipe and not at the outside toe of the weld as previously observed by Markl. An examination of the fatigue fractures by metallography showed the these cracks occurred in the region of change of grain size. In the 304 stainless steel pipe, fatigue cracks started at the toe of the welds at the outside of the pipe.

Keywords: Fatigue; Fatigue of welds; Butt welded piping; Fatigue of piping.

Introduction: Cyclic fatigue life of carbon steel welded pipes was studied about 40 years ago by Markl (1). In 1951, Markl conducted tests on various water filled welded piping components using a cantilever type assembly with fully reversed bending loads. The data on butt welded pipe were obtained from 1935 to 1951 and covered fatigue life above 2000 cycles. Markl's paper published in 1952 covered tests for the following piping components: plain straight pipe, butt-welded joints in straight pipe, short and long radius elbows within-plane and out-of-plane, forged welding tees within-plane and out-of-plane and mitre bends. The results from Markl's test have been used as the basis of piping design in the ASME Boiler and Pressure Vessel Code (2). However, the low cycle regime, less than 2,000 cycles, was not covered by Markl's research.

This paper presents the results of testing on butt welded steel pipes conducted at the University of Akron with the objective of extending Markl's data to the low cycle range (100-2000 cycles). In addition, formulas of stress amplitude versus cycles developed by Markl were compared to this new data in the low cycle regime . Both carbon steel and austenitic stainless steel butt welded pipes were cycled to failure. The welds were based on the ASME Boiler and Pressure Vessels Code (2) using certified welders. All pipes were tested in the as-welded condition without grinding of the weld outside surface or post weld heat treatment.

The microstructure in the weld region and at the fatigue failures was studied and related to the observed failures. Fatigue data are presented on log-log plots of strain amplitude versus cycles and stress amplitude versus cycles. In the stress versus cycles plot, Markl's procedures were followed to determine an elastically calculated stress during plastic bending of the pipe.

Results are compared with Markl's and Langer's equations (1,3). Data from these tests are also compared to other low cycle tests conducted by the General Electric Company (4) and with tests conducted in France (5,6,7).

Test Apparatus: All pipe specimens were tested using the four point bend test fixture shown on Figure 1. The test fixture was mounted in an Instron Material Testing System (Model 1321). Four yokes were used to provide four simple supports for each pipe specimen. The two center yokes are fixed to the load cell of the Instron system. The other two end yokes are fixed to the bottom beam which is attached to the Instron actuator. Both beams are stiff and strong relative to the 1-1/2 inch schedule 80 pipe specimens. The centers of the two end yokes are 48 inches from each other while the other two center yokes are located 15 inches from each end. The center 18 inches of each specimen is subjected to a constant bending moment. As a result, the weld at the center of each pipe specimen was in the middle of the uniform moment region. The pipe specimen was fixed at one end yoke by tightening clamping screws in the yoke. This screw prevented rolling and axial motion of the pipe during testing. Axial slipping was allowed at the other three yoke supports during the large plastic bending of the pipe.

A controlled cyclic displacement was applied by specifying the actuator movement amount a neutral position. The computer controlled cyclic displacement amplitude was constant during each test. The displacement was specified from the initial zero load position of the pipe and the same magnitude was specified in both the up and down direction. The resulting peak load was measured by the Instron's Load cell and recorded during each cycle. Strains in the center portion of each pipe specimen were also recorded until gage fatigue failure occurred. Cyclic strain amplitude was used to determine the magnitude of the cyclic displacement for the desired cycles to failure.

Specimens: Both carbon steel and austenitic seamless stainless steel welded pipes were tested. The carbon steel pipes conformed to ASTM material specification A53 Type F while the stainless steel pipes conformed to ASTM specification A312 grade which is 304 SS. The type F carbon steel pipe is a continuous welded pipe with a minimum yield of 24,000 psi. The 304 pipe is a seamless pipe; the carbon steel is a seam welded pipe. The 4-1/2 foot pipe specimens were butt welded at the center by an ASME code qualified welder. The pipes were 1 1/2 inch schedule 80 having an outside diameter of 1.9 inch and an inside diameter of 1.5 inches. The pipe specimens were 54 inches in length and the weld was located at the center of the specimen. The specimen dimensions and weld location are shown in Figure 2. The weld required preparation is shown in Reference 5. Photographs of the inside and outside of the carbon steel welds are shown on Figure 3; similar photographs for the stainless steel pipes are shown on Figure 4. The welder scribed his initials near the weld in each specimen. The welds passed an X-ray exam and no serious defects were found. Results of the X-ray tests are included in Reference

5. Original plans were to place the worst defect located on the x-rays at the maximum stress location. However, because of the possibilities of fatigue initiation at initials, the initials were placed at the neutral axis.

A total of 21 pipe specimens were tested: 14 CS specimens and 7 SS specimens. The specimens include welded and control (plain) specimens. The control specimens are used to provide base line data to which the welded specimens can be compared. Two CS control specimens and one SS control specimen were tested of the 21 total specimens. A list of the specimens tested is shown in Table 1.

Instrumentation: General purpose electric resistance strain gages were used to measure the resulting bending strains during the initial cycles. The gages were glued to the top surface of the piped specimens near the weld and away from the weld. Signals from each strain gage were amplified with BAM 1, series 14600, bridge amplifiers made by Vishay Instruments. Output from the bridge amplifiers was recorded on a two channel chart recorder. Cyclic strains were recorded until gage failure. The peak positive and negative loads during each cycle were recorded by the Instron digital controller for the entire test. Strain amplitude, which is one half of the total range, was plotted on all graphs.

Test Procedure: Before installing each pipe specimen in the test apparatus, the specimen was filled with water. A water leak was used as a failure indicator. Two rubber plugs were used to close the pipe ends. Load-displacement measurements were made on a number of pipe specimens for use in the Markl's stress analysis method. Two different procedures were followed in this step. In the first procedure the pipe specimen was loaded incrementally and then allowed to relax for a few seconds before recording the resulting load from the Instron control console. Loads decreased during this relaxation period. In the second procedure, a continuous load-deflection record at the test strain rate was obtained directly from an x-y recorder attached to the Instron. Higher loads in the plastic region were obtained (Figures 5 and 6).

A controlled cyclic displacement was applied to the pipe specimen. The specimens were cycled at 4 cycles per minute when the specified displacements were less than 2.0 inches and at 2 cycles per minute when the specified displacements were greater than 2.0 inches. Peak loads and peak displacements were printed on a paper roll every 50 seconds by the Instron control console. The test was stopped when water leaked from a fatigue crack and the number of cycles to failure was recorded.

Results: Results from this test program are shown in Table 1. The results are plotted on a log-log graph in Figure 7. This figure is a plot of the measured cyclic strain amplitude (1/2 of the range) versus number of cycles to failure where the strain amplitude is the strain measured away from the weld region. At the weld, the measured strain was lower than the strain measured away from the weld as shown in References 8 and 9. This result was unexpected. Plastic strains in the heat affected zone were expected to exceed that of the pipe material. At present, this difference cannot be explained. Examination of the test data reveal the following:

(1) Fatigue life of welded pipes is substantially lower than the plain pipes.
(2) Stainless steel pipes have better fatigue life

than the carbon steel pipes.

Comparison of the test results from this test program with Markl's equation ($iS = 245{,}000\ N^{-0.2}$) is shown in Figure 8 where stress amplitude versus cycles to failure is plotted. The stress in this case is calculated from the ordinary beam formula ($S = MC/I$) where the load used to calculate M is obtained from straight line extrapolation of the load displacement calibration curves to the test displacement. The stiffness of the system based on test was about 6,000 lb/in. Considering the pipe only, the calculated stiffness is 7,400 lb/in. Since the fixture does flex, the calculation is felt to verify measured values. Test displacements are well into the nonlinear (plastic) range. The calibration curves are shown in Reference 8. As can be seen from Figure 10, Markl's equation severely under predicts fatigue life in the low cycle range. In addition to being out of range of the equation, this difference is believed to be due to better welding techniques in use today as compared to welds made over 40 years ago when Markl published his paper. Also, there was a large variation in the yield strength of straight pipes used in Markl's data (27.3 KSI to 42.5 KSI). Additional testing of pipe in the high cycle range would provide a more realistic assessment.

Least square fits of the test data from the welded pipes result in the following equations:

(1) All test data points:

$$\delta\epsilon/2 = 102{,}329\ N^{-0.39}$$

(2) CS data points only:

$$\delta\epsilon/2 = 147{,}991\ N^{-0.47}$$

(3) SS data points only:

$$\delta\epsilon/2 = 72{,}444\ N^{-0.32}$$

Langer's equation:

$$S_a = E/4\sqrt{N}\ \mathrm{Log}_{\epsilon}[100/(100 - RA)] + S_e$$

where

S_a = Pseudo alternating stress
E = Elastic modulus
N = Cycles to failure
RA = Reduction in area
S_e = Endurance limit

along with the least square fit of the test data is plotted in Figure 11. Agreement of the test data with Langer's equation is poor. Langer's equation over predicts the fatigue life of welded pipes. The reason for the poor agreement is due in part to the fact that the welded pipes have lower fatigue life than plain pipes. Another source of the poor agreement could be due to incorrect RA values used in Langer's equation. The RA values (SS:RA = 69%, CS:RA = 65%) used were obtained from the non-standard test coupons machined from the pipe specimens.

The fatigue data developed in this investigation was compared to data from two other studies: a study by the General Electric Company in San Jose by Heald and Kiss (4). A study conducted in France by the Commissariat a L'Energy Atomique (5,6,7) and Cyclic strain data developed in France was provided by Dennis Acker and was compared to cyclic strain data developed in this study. This comparison is

presented on Figure 10. As seen on this graph the data agrees well with measurements from this test program.

In the other comparison, cyclic stress data developed by Heald and Kiss (4) was compared to the data of this study. This GE test program included girth butt-welded carbon steel and stainless steel pipes, elbows and tees. Only the pipe data are used in this comparison. In Reference (4), the peak range stress, S_p, and the alternating stress amplitude, S_{alt}, are calculated from the following equations:

$$S_p = K_1 C_1 \delta PTD_o/2t + K_2 C_2 \delta MD_o/2I$$

and

$$S_{alt} = \tfrac{1}{2} K_e S_p$$

For straight pipe without flush welds,

$$C_1 = C_2 = 1$$
$$K_1 = 1.2$$

and

$$K_2 = 1.8$$

The value of the factor, K_e, is stress dependent and listed Table 3 of Reference 8. On Figure 11, data of Heald is compared to the data of this test program which includes the constants K_1 and K_2. On Figure 11, the nominal stresses from the three unpressurized room temperature carbon steel pipe specimens, obtained S_{alt} by dividing by K_2 which is 1.8, are plotted with the nominal stress data from this study. As seen from these two comparisons these data of Heald and Kiss agree more closely with the least squares fit line from data from this investigation than with Markl's equation. It was concluded by Heald and Kiss "that the applicable ASME Section III Code rules for low cycle fatigue are adequately conservative".

From these comparisons, it is concluded that data from this test program agree with that obtained from two other investigations.

Failure Location: All tested pipe specimens failed by developing circumferential cracks in the weld region (Figures 3 and 4). Almost all fatigue cracks in the carbon steel specimens occurred in the junction of the heat affected zone to the base metal where cracks developed on the inside surface and propagated to the outside surface. It was concluded that the cracks started at the inside edge because no crack was observed during testing until failure occurred in the carbon steel pipes. The cracks allowed water to rapidly lead from the pipe. Also from destructive examination of the pipe, the cracks on the inside of the pipe were found to be longer than the outside cracks. On the other hand, cracks in the stainless steel pipe could be observed on the outside of the pipe before leaking occurred. The cracks in the carbon steel specimens were located about 12mm from the center of the weld bead. An exception to this behavior occurred in specimens 10 and 12 where the cracks occurred at the weld edge. Specimen 10 was slightly undercut at the weld joint while misalignment of the welded pipe segments in specimen 12 is believed to have caused the difference in failure behavior. This result is quite different from that observed by Markl. In Markl's tests, "cracks" (in welded pipe) predominantly were located at the edge of the weld were stresses exceeding the yield strength were applied and at the center of the weld when stresses below the yield strength were applied or the weld overlay had been ground off (1)) Fatigue cracks in the stainless steel pipe specimens occurred at the weld edge where the

cracks started on the outside and propagated to the inside. No failures occurred in the weld itself in either the carbon or stainless steel pipe. A summary of pipe failure location is presented in Table 2.

Tensile Tests: Tensile tests were performed on 8 test coupons machined from the tested pipe specimens. The test coupons include 4 unwelded specimens cut from different pipes and four specimens cut from welded pipe. In each case there were two cut from carbon steel pipes and two from 304 stainless steel pipes. The plain test coupons were machined from the pipe ends along the neutral axis. The welded specimens were machined from the weld region and the weld was located in the specimen center. The specimens dimensions are shown in Reference 5. Flat grips were used to hold the test coupons in the tensile testing machine. An extensometer with 1.0 inch gage length and 0.2 inch range was used to obtain stress-strain curves. The extensometer was placed in the center of the test coupon where the weld was located. The one inch extensometer gage length covers a region that includes the weld material and the base material and therefore the measurement reflects the strength of the weld joint and not the weld material alone.

None of the welded test specimens fractured in the weld region where fatigue failures of the pipe specimens occurred. Failure occurred in the base metal away from the heat affected zone of the weld. The stress-strain curves generated are shown in Reference 5. These figures show higher yield and ultimate strengths in the weld region than in the base material in both the carbon steel and stainless steel specimens. However, the increase in yield strength is smaller in the stainless steel specimens than in the carbon steel specimens. A summary of the tensile test results is shown in Table 3. The ultimate strength reported for the weld specimens is that of the base material since failures did not occur in the weld or heat affected zone of the base metal.

Hardness Tests: Rockwell hardness tests using scale A were performed on the failed pipe specimens in the weld region at 0.10 inch increments starting from the weld edge. Scale A is based on the diamond brale and a 60Kg load rather than the 150Kg load used in the "C" scale. the test was an attempt to correlate possible changes in the hardness due to the heat from welding with the fatigue failures. However, no statistically significant change in hardness was observed. Average results from the Rockwell hardness tests using the brale with 60Kg preload were as follows:

Carbon Steel: $R_a = 42$
Stainless Steel: $R_a = 48$

Metallurgical Examination; Carbon Steel Specimens: The macroscopic failures observed in the carbon steel pipes were of two types:

Type A: Cracking at a distance of 0.5% (12mm) from the center of the weld bead at the pipe of the heat affected zone (HAZ) of the CS pipe.
Type B: Cracking immediately adjacent to the weld bead (about 2 mm from the center of the weld bead) in the SS pipe.

The failures can be rationalized by an examination of intrinsic microstructural features of the cyclically deformed pipes. This examination was performed at certain specific locations of the pipe.

Samples for metallographic observation were cut from the two low cycle fatigue (LCF) tested carbon steel pipes. Small sections taken perpendicular to the longitudinal axis of the pipe, about two inches on either side of the weld bead, were cut from the failed pipe. The resultant section was a small cylinder measuring four inches in axial length and containing the weld bead at its center. The cylinder was sectioned again along the longitudinal axis to yield a specimen having a semi-circular cross-section. The circular convex surface of the piece was ground flat along its length with the aid of grinding. Two such samples (A and B) were carefully made from the cyclically deformed carbon steel pipes:

SAMPLE A: had a fatigue crack located 2mm from the weld bead.

SAMPLE B: had a fatigue crack located 12mm from the weld bead.

The samples A and B were mechanically ground using 240, 320, 400 and 600 grit silicon carbide paper with water as a lubrication, and then mechanically polished using 5 micron and 1 micron alumna powder mixed in distilled water. The polished samples were etched using NITAL reagent to reveal the grain boundaries and grain morphology. The etched specimens were observed in an optical microscope and the microstructure photographed using standard bright field technique.

Figures 12 and 13 are typical micrographs taken in the parent metal and at the end of the heat affected zone at the site of cracking, respectively. Inclusions (probably carbide precipitation) at the end of the heat affected zone are large compared to the parent metal or weld metal (Figure 14). The weld region itself had a fine grain sized microstructure. The carbon steel base material was comprised of predominantly large grains in the microstructure (Figure 13).

Inclusions (not identified in this study) were found located at and along the grain boundaries at 12mm from the center of the weld (Figure 13). On account of their intrinsic brittleness the inclusions (particles) crack when the local stress concentration exceeds the strength of the inclusion. During cyclic deformation progressive build-up of dislocations at the inclusion-matrix interface causes localization of strain. The resultant stress or strain concentration ruptures the brittle inclusion thereby initiating a microscopic crack which leads to fatigue failure. The microstructure of the material adjacent to the fatigue crack reveals extensive deformation to have occurred during fatigue testing.

Metallurgical Examination; Stainless Steel Specimen: Failure of the cyclically deformed stainless steel butt welded pipes was consistent with failure occurring at the intersection of the weld bead and the parent material (immediately adjacent to the weld bead). The failed regions comprised of multiple microcracks parallel to the weld bead.

A representative failed specimen was chosen and a sample for metallographic observation chosen such that it contained the weld bead and about 2" on either side of the weld. The sample was mechanically ground using 240, 320, 400 and 600 grit silicon carbide paper and water as lubricant and then mechanically polished using 5 micron and 1 micron alumna

powder dissolved in distilled water. The polished specimens were etched using a solution mixture comprising of 10 ml of nitric acid, 15 ml of hydrochloric acid, 10 ml of acetic acid and 5 ml of glycerol.

Representative micrographs of the stainless steel specimen are shown in Reference 5. The microstructure of the weld metal reveals a dendritically cored structure comprising of large area of ferrite and small quantities of precipitated carbide in an austentic matrix. The observed dendritic coring is characteristic of rapidly frozen austentic metal. The microstructure at the intersection of the weld metal and the parent material is shown in Reference 5. Crack initiation was observed to occur at the interface. During welding the heat generated into the surrounding metal has an influence on the microstructure, that is, the precipitation of particles. The particles have a near-spherical morphology and are most likely the carbides, though no attempt was made in this study to identify their composition. The growth or propagation of the fatigue crack through the heat affected zone was observed in the micrographs.

Conclusion and Recommendations: Fatigue data from 12 A53 Type F carbon steel butt-welded specimens and 6 seamless type 304 SS butt-welded pipe specimens in the low cycle range (150 to 1,200 cycles) were obtain from a 4 point bend test apparatus. In addition, three plain pipe control specimens and one stainless steel pipe specimen . Data were plotted as strain amplitude versus cycles to failure and stress amplitude versus cycles to failure. Stresses were calculated using methods of Markl. The stress amplitude data of the carbon steel pipe consistently fell well above the Markl equation (Figure 10).

Markl's data covered the range from 2,000 cycles to 4 by 10^7 cycles. Equations fitted to this data were extended to the low cycle range for this comparison. Markl's equations were found to be very conservative.

Failures in the butt-welded carbon steel pipe specimens in Markl's data occurred either adjacent to the welds or in the weld metal itself. In this test data, most failures occurred at the end of the heat affect zone of the parent metal about 12 mm from the center of the butt weld. Two failures occurred adjacent to the welds. These failures could be explained by geometric stress concentration effects at the weld. There were no failures in the weld metal.

Strains measured in the heat affected zone of the weld were consistently below strains measured on the unaffected parent metal in both the carbon steel and stainless steel specimens. The yield and ultimate strength of both the weld metal and heat affected zone of the parent metal appeared to exceed the strength of the parent metal away from weld. This trend was also observed in the tensile test specimens.

The fatigue strength of the SS butt welded pipe exceeded that of the CS.

Data obtained in this test series agreed with data from the General Electric Company and with data from Commissariat a L'Energie Atomique, Instut de Recherche Technologique, Department des Etudes Mechaniques et Thermiques, France.

References:
(1) Markl , A.R.C., "Fatigue Testing of Piping Components," Trans. ASME, April 1952, PP287-302.

(2) "ASME Boiler and Pressure Vessel Code," ASME, NY, NY

(3) Harvey, J. F., Theory and Design of Modern Pressure Vessels, D. Van Nostrand Reinhold, 1974 pp 222-32.

(4) Heald, J. D., and Kiss, e., "Low Cycle Fatigue of Nuclear Pipe Components," Jr. of Pressure Vessel Technology, Aug. 1974, pp 171-76.

(5) Autrusson, B. and Aker, D., "Fatigue Analysis of Girth Butt Welded Pipes," Proceedings of the 11th SMiRT Conference, 1991.

(6) Barrou, A. and Autrusson, B., "Experimental Fatigue Analysis of Girth Butt Welded Austentic Stainless Steel Pipes," Proceedings of the 8th SMiRT Conference, 1988.

(7) Autrusson, B. , Brouard and Goulet G., "Fatigue des Soudures Circonferentielles," Commissariat a L'Energie Atomique, Instut de Recherche Technologique, Department des Etudes Mechaniques et Thermiques, Rapport EMT 86/271, 1986.

(8) Scavuzzo, R. J. et al., "Low Cycle Fatigue of Butt Welded Pipe," WRC Project 91-11, Final Report, The University of Akron, 1994.

(9) Scavuzzo, R. J. et al., "Low Cycle Fatigue of Butt Welded Pipe," PVP-Vol. 235, June 1992, pp 157-161.

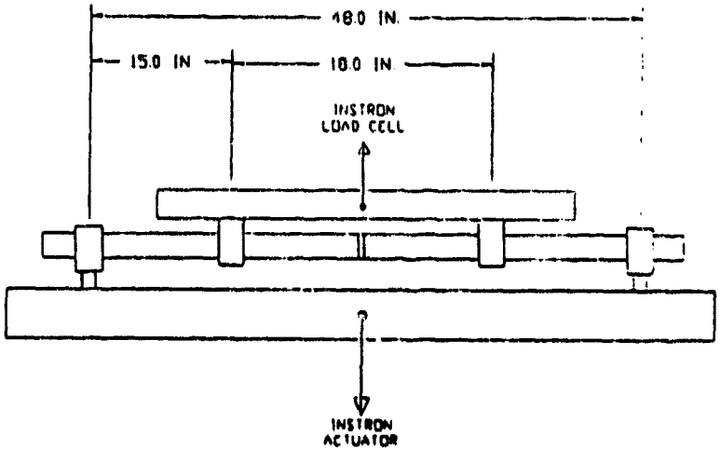


Figure 1. Test Fixture Geometry

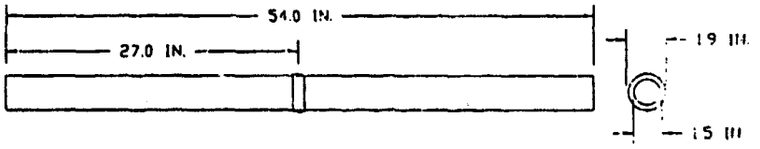
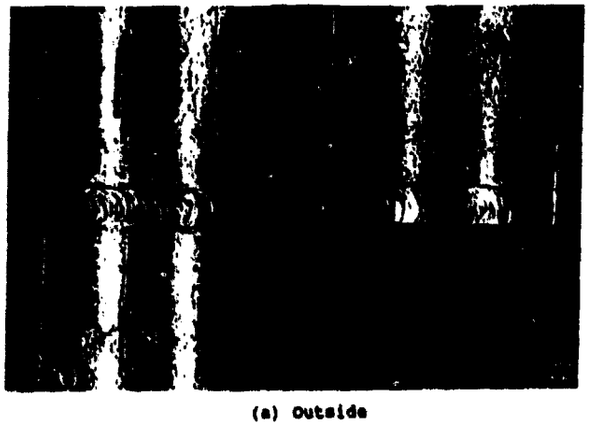


Figure 2. Test Specimen Geometry

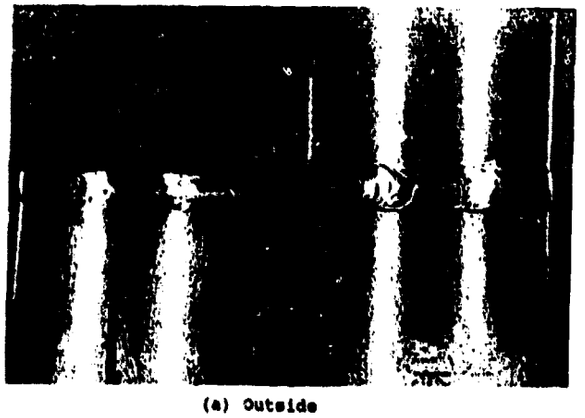
(a) Inside



(a) Outside

Figure 3. Inside and outside photographs of a typical carbon steel pipe butt weld.

(a) Inside



(a) Outside

Figure 4. Inside and outside photographs of a typical stainless steel pipe butt weld.

TABLE 1

FATIGUE RESULTS OF BUTT WELDED PIPES

SPEC. NO.	DISCRIPTION	DISPL. AMP.	STRESS* AMP.	STRAIN AMP.	CYCLES
1	CS - CONTROL	1.70	186318	0.009250	523
2	CS - CONTROL	2.50	282300	0.013400	285
3	CS - WELDED	1.70	191964	0.010270	249
4	CS - WELDED	1.50	169380	0.010100	304
5	CS - WELDED	1.20	135504	0.008300	615
6	CS - WELDED	2.30	259716	0.014720	155
7	CS - WELDED	1.05	118566	0.006530	836
8	CS - WELDED	1.30	146796	0.008480	461
9	CS - WELDED	1.10	124212	0.007800	786
10	CS - WELDED	1.40	158088	0.009170	386
11	CS - WELDED	2.50	282300	0.015640	183
12	CS - WELDED	1.90	214548	0.010680	202
13	CS - WELDED	1.00	112920	0.06230	708
14	CS - WELDED	0.90	101628	0.005800	791
15	SS - CONTROL	1.75	184861	0.009300	751
16	SS - WELDED	1.75	184861	0.009535	596
17	SS - WELDED	2.50	264087	0.013200	195
18	SS - WELDED	1.40	147889	0.007830	1224
19	SS - WELDED	2.00	211269	0.010730	478
20	SS - WELDED	2.25	237678	0.012560	311
21	SS - WELDED	1.60	169015	0.008510	673

* Stress is calculated from the ordinary beam formula (S = MC/I where M is calculated from straight line extrapolation of the linear portion of load-displacement calibration curve.

CS - A53 Type F Carbon Steel
SS - 304 Grade 312 Stainless Steel

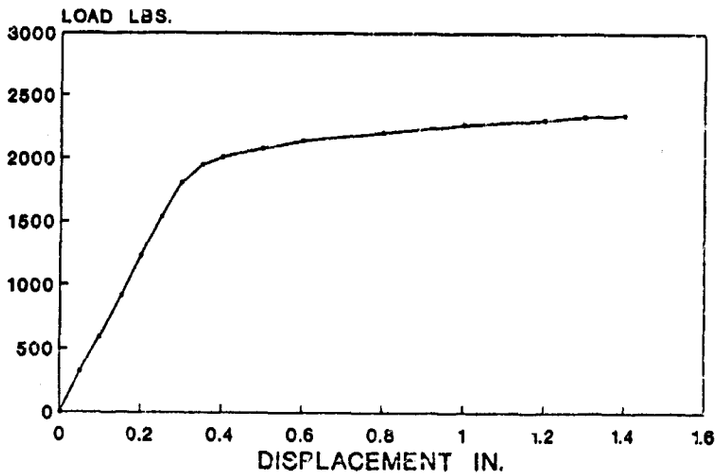


Figure 5. Load versus Displacement for Carbon Steel Specimen 9.

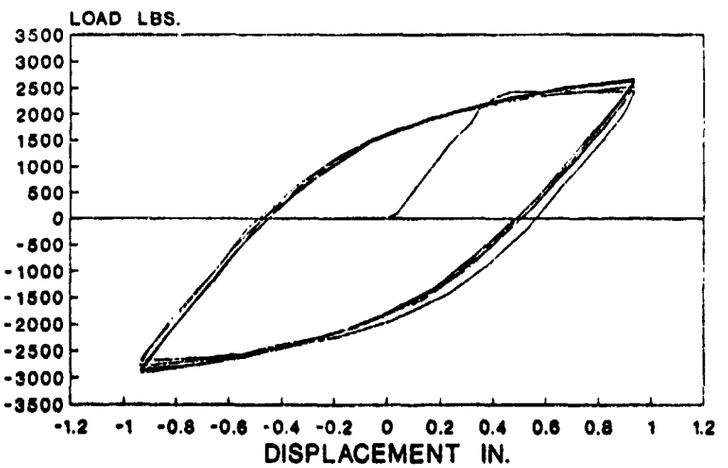


Figure 6. Load versus Displacement for Carbon Steel Specimen 14 with Fast Loading.

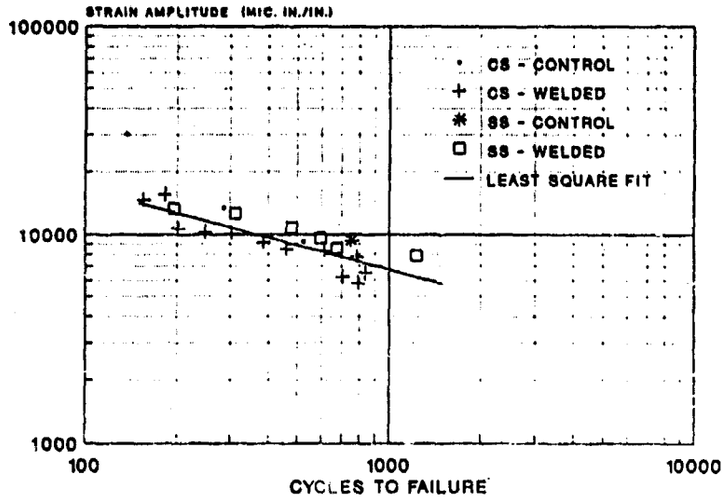


Figure 7. Strain Amplitude versus Cycles to Failure. Strain Measurements were made away from the heat affected zone.

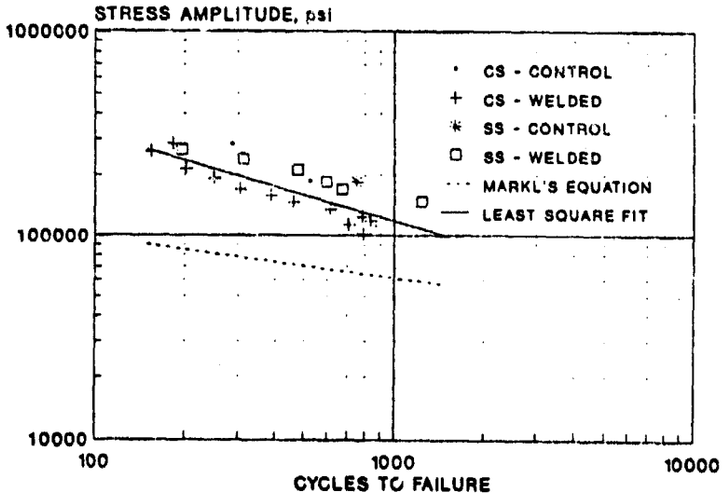


Figure 8. A Comparison of Markl's Formula and Measured Stress Amplitude versus Cycles to Failure.

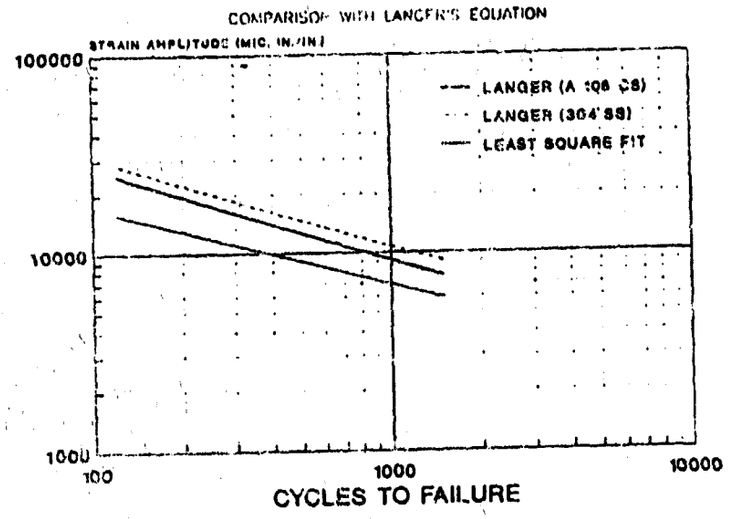


Figure 9. A Comparison of Langer's Formula and Measured Strain Amplitude versus Cycles to Failure.

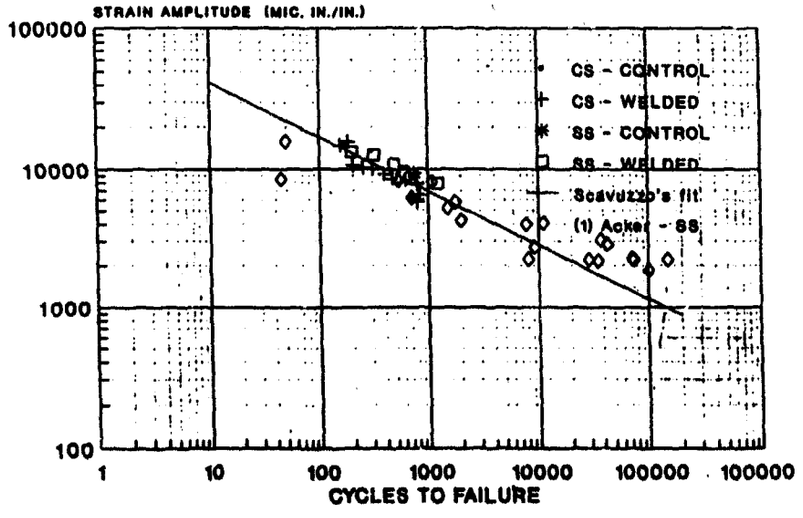


Figure 10. Comparison of Strain Cycling Test Data with data of Acker (5,6,7).

(1) Based on cycles to crack initiation

TABLE 2

LOCATION OF FATIGUE CRACKS

SPEC. NO.	DESCRIPTION	CRACK LOCATION	COMMENTS
1	CS-CONTROL	PIPE CENTER	specimen was notched accedentally in the pipe center at the beginning of test
2	CS-CONTROL	PIPE SUPPORT	
3	CS-WELDED	HAZ	
4	CS-WELDED	HAZ	
5	CS-WELDED	HAZ	
6	CS-WELDED	HAZ	
7	CS-WELDED	HAZ	
8	CS-WELDED	HAZ	
9	CS-WELDED	HAZ	
10	CS-WELDED	WELD EDGE	specimen is undercut at the weld joint
11	CS-WELDED	HAZ	
12	CS-WELDED	WELD EDGE	the two welded segments are misaligned
13	CS-WELDED	HAZ	
14	CS-WELDED	HAZ	
15	SS-WELDED	PIPE SUPPORT	
16	SS-WELDED	WELD EDGE	
17	SS-WELDED	WELD EDGE	
18	SS-WELDED	WELD EDGE	
19	SS-WELDED	WELD EDGE	
20	SS-WELDED	WELD EDGE	
21	SS-WELDED	WELD EDGE	

TABLE 3

TENSILE TEST RESULTS

SPEC. NO.	DISCRIPTION	YIELD STRESS PSI	ULTIMATE STRENGTH PSI	% ELONG IN 4 IN. GAGE	% RED. IN AREA
1	CS-PLAIN	31,700	51,000	20	64
2	CS-PLAIN	29,400	48,500	23	65
3	CS-WELDED	35,000	49,800	18	65
4	CS-WELDED	32,000	49,300	16	65
5	SS-PLAIN	35,000	86,400	80	68
6	SS-PLAIN	35,000	84,700	75	70
7	SS-WELDED	36,400	85,400	69	65
8	SS-WELDED	36,000	83,700	71	60

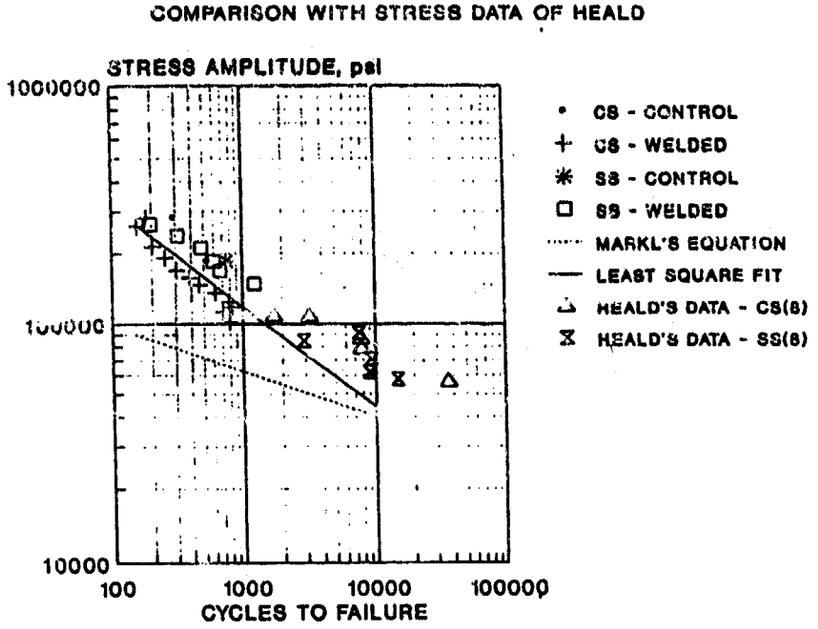


Figure 11. Comparison of Stress Cycling Test Data with data of Heald (8). Some data developed by Heald was done on pressurized pipe. Stress calculations include geometric factors K_1 and K_2.

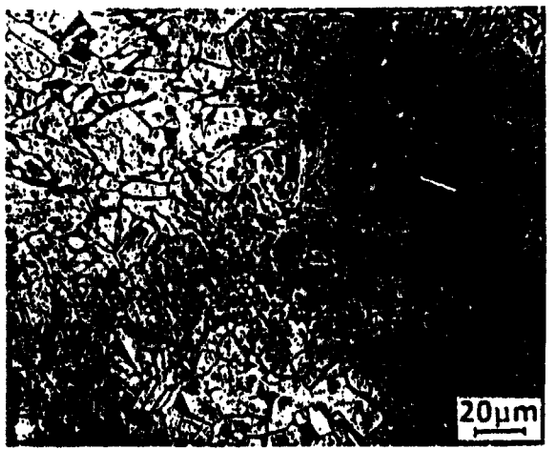


Figure 12. Microstructure of Carbon Steel Pipe Specimen 12 (500x) of the parent metal away from the heat affected zone.

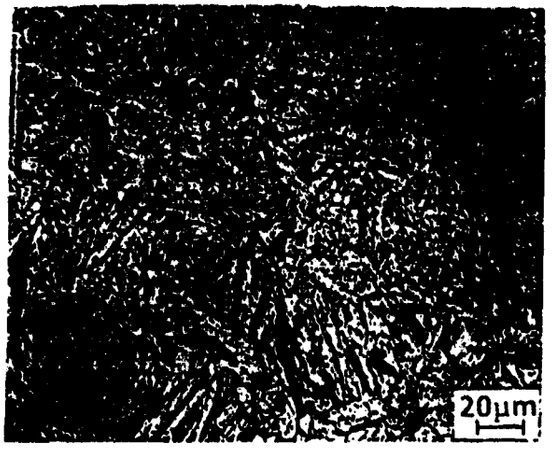


Figure 13. Microstructure of Carbon Steel Pipe Specimen 3 (500x) in the heat affected zone in the region of the fatigue crack.

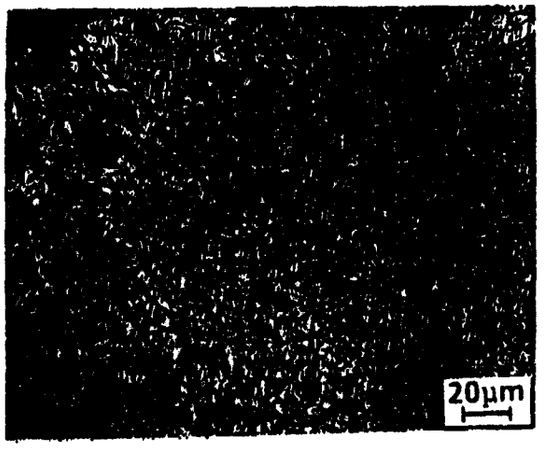


Figure 14. Microstructure of the carbon steel weld metal of Specimen 12 (500x).

INVESTIGATION OF CAVITATION-EROSION OF VARIOUS ALLOYS AS A BASIS FOR CONDITION MONITORING OF HYDRAULIC PUMPS

K. Wolf
Wehrwissenschaftliches Institut fur Materialuntersuchungen
Erding, Germany

R.J. Ferrara
Naval Surface Warfare Center
Annapolis, MD 21402-5067

Abstract: Hydraulic pumps taken from several modern jet engines showed damage of the rotors due to cavitation-erosion. An investigation was conducted to determine the sensitivity of various materials to cavitation-erosion and to characterize and quantify the resulting damage. The materials utilized in this study were E-Cu, CuZn39Pb0.5, AlCuMg2, Ti6Al4V, St37-2k, 90MnV8, and X5CrNi18-9. The primary objective of the study is to characterize the macro and microscopic appearance and degree of cavitation-erosion damage that occurred on aircraft hydraulic pump components in service and on test specimens utilizing an ultra sonic vibration test (USVT) in a fluid containing a corrosion inhibitor. Particles found in the USVT fluid were also examined and characterized. The ultimate goal of this study is to develop a methodology for monitoring the onset and progression of cavitation damage on hydraulic pump components in service from the size, shape and morphology of damage on metallic particles found in hydraulic fluid.
In addition, cavitation-erosion results are compared for two different test methods and fluids; the USVT and a High Velocity Seawater Flow Test (HVSFT). The two tests were conducted totally independent of each other at different times and locations. The USVT is of relatively short duration (several hours), while the HVSFT is conducted over a longer time period (30-days). In this study specimens were exposed in the USVT for various time periods (up to 300 minutes) in water treated with potassium dichromate. In the HVSFT, specimens were exposed for 30 days to high-velocity (120 ft/s) sea water flow.
Samples from the USVT were evaluated utilizing weight loss, surface roughness and optical and scanning electron microscopy. The macroscopic and microscopic appearance of cavitation damage was characterized for each material used in the ultrasound test. In addition, particles from the residue which remained in the test bath of the USVT were examined to characterize the size, shape and damage profile with respect to alloy and test duration.
The samples exposed in the HVSFT were analyzed using weight loss, maximum depth of attack and scanning electron microscopy. In those instances where the same alloy was exposed in both tests, comparisons were made of the appearance of the resulting cavitation damage. It was shown, that the two tests produced significant differences in the appearance and degree of degradation on samples of the same type of alloy.

Key Words: Cavitation-erosion; condition monitoring; high velocity seawater cavitation; hydraulic pump; sem structure; ultra sonic cavitation

Introduction: The time and material dependent degree of degradation due to cavitation-erosion as a result of various sample exposures can be simulated, studied and evaluated through numerous test methods. Usually, the objective of such studies is to find a relationship between degree of cavitation damage and varying exposure times and conditions after giving due consideration to all other possible physical and chemical variables. In this investigation, three damaged hydraulic pump rotors from different jet engines were investigated with respect to cavitation-erosion. Remedial maintenance was deemed essential after a loss of oil

pressure was experienced during a flight mission. After disassembling the pumps it was found that the rotor housings were damaged by fracture at the overflow canals (Figures 1, 2, 3 and 4). Cavitation damage of the hydraulic pump is a serious problem since complete loss of oil pressure would result in failure of all hydraulic actuated systems including ailerons, wings and landing gear. At present the condition of hydraulic systems components of jet engines is not routinely monitored.

The primary objective of this study was to characterize and compare the appearance and degree of cavitation-erosion damage that occurred on the hydraulic pump rotors to that which occurred on specimens tested in an Ultra Sonic Vibration Test (USVT) in a fluid containing a corrosion inhibitor. The macro and microscopic appearance of cavitation damage has been described for several different alloys using the USVT. The appearance of wear particles found in the USVT fluid was also characterized. Comparison of the macro and microscopic size and appearance of cavitation damage on specimens and particles from the USVT to the appearance of cavitation damage on aircraft hydraulic pump components and cavitation particles found in aircraft hydraulic fluid form the basis for a method of monitoring for cavitation damage of hydraulic system components while in-service. The ultimate goal of this study is to develop a methodology for monitoring the initiation and progression of cavitation damage on aircraft hydraulic pump components from the size, shape and damage profile of metallic particles found in aircraft hydraulic fluid.

A secondary objective was to compare the appearance of cavitation-erosion damage resulting from two different test methods and fluids; the USVT and a High Velocity Seawater Flow Test (HVSFT). The two tests were conducted totally independent of each other at different times and locations. The USVT is conducted for a relatively short duration (up to 8 hours), while the HVSFT is conducted over a considerably longer time period. In this study specimens were exposed in the USVT for various time periods in water treated with potassium dichromate. In the HVSFT, specimens were exposed for 30 days to high-velocity (120 ft/s) sea water flow.

Materials and Experimental Conditions: The materials and test conditions for aircraft hydraulic pump components, the USVT and the HVSFT are provided:

Hydraulic Pump Rotors: The damaged hydraulic pumps failed after 11, 96 and 133 hours of operation. The rotors were manufactured from a copper-zinc-lead alloy CuZn39Pb0.5 (equivalent to ASTM B30). The operating pressure of the pumps is up to 270 bar at 6,000 revolutions/min. The influence of other variables such as hydraulic oil pollution, air and water in hydraulic oil, temperature and system oil pressure on cavitation-erosion has already been investigated and therefore was not included in this study [1]. The rotors were disassembled and cut near the cavitation-erosion areas. The surface roughness (R_t) was measured on the rotor with the lowest hours of operation (11 hours); the average value was 10μm.

Ultra Sonic Vibration Test (USVT) Samples: The faces of the USVT cavitation-erosion samples were ground and polished. Their surface roughness (R_t) measured 0.28μm to 0.95μm (as shown in Table 1). Testing of the samples was conducted in accordance with the FVV-procedure, Section R 433, 1986, in which a testing frequency of 20kHz, amplitude of 40μm, atmospheric pressure and a water temperature of 55°C is specified [2]. Depending on the alloy and the measured weight losses, the samples were periodically removed from the testing apparatus and evaluated utilizing the SEM. After completion of the test, surface roughness and weight loss were measured again. Metallographic evaluations on longitudinal sample sections followed.
Cylindrical samples (15 x 15 mm, Figure 5) were used in the USVT and were produced from the following alloys: E-Cu (C1 1000), CuZn39Pb0.5

(ASTM B30), AlCuMg2 (ISO R209), Ti6Al4V (MIL-T-81915), St 37-2K (AISI C1006), 90MnV 8 (AISI 02), and X 5 CrNi 18 9 (AISI 304).

High Velocity Seawater Flow Test (HVSFT) Specimens: Specimens utilized in the HVSFT were flat with a rectangular cross-section having the dimensions 25 x 4 x 139 mm, Figure 6. The surfaces of the flat specimens were prepared with surface roughness of R_t=2.9μm. The seawater flow had a velocity of 117 to 125 ft/s (35.66 to 37.79 m/s), temperature of 0°C and pressure of 93.9 to 94 psig (6.473 to 6.480 bar). After the 720 hour long test, the weight losses and maximum depth of cavitation attack were measured. The alloys used in the HVSFT included: Ti6Al2Cb1Ta0.8Mo, Ti6Al2Cb1Ta0.5Mo, Ti6Al4V, Ti7Al2Cb1Ta, Ti7Al2Cb1Ta0.75Pd, and Ti7Al2.5Mo.
For this study, only the Ti6Al4V and the Ti7Al2Cb1Ta samples were subjected to further evaluation.

Results-General Comments: The USVT and HVSFT were completely independent investigations conducted at different times and locations. While the test methods, fluid mediums and specimen geometries were different, a few of the alloys tested were the same. A detailed examination of the cavitation damage due to test dependent variables were taken into consideration during the evaluation of the damage profiles.

The USVT is a recirculated (closed circuit) system utilizing tap water and a corrosion inhibitor as the fluid medium and therefore there is little or no influence of corrosion on the resulting cavitation-erosion damage. The energy due to the ultrasound and water is directed perpendicular to the surface of specimen. The HVSFT is a once-through (open circuit) system utilizing natural seawater as the fluid medium and thus there is also a corrosion factor involved in the resulting cavitation-erosion damage [3,4,5]. The seawater flow is parallel to the surface of the specimen and cavitating conditions are created due to a hole in the specimen surface. Since both the direction of flow and the fluid medium are substantially different in these two tests, the appearance of the resulting cavitation damage was examined to delineate similarities and/or differences.

Hydraulic Pump Rotors-Visual and SEM Examinations: The damaged areas near the overflow canals of the hydraulic pumps had the same appearance (Figures 4 and 7). The appearance of the damage was characteristic of cavitation-erosion attack. The pump rotors showed cavitated/eroded areas different in size corresponding to the hours of operation (0.2 x 0.2 mm after 11 hours, 0.8 x 0.3 mm after 96 hours and 1.0 x 0.4 mm after 133 hours). The damage near the pump surface was the result of plastic deformation of the ductile metal (Figure 8). As the damage progressed, the stress resulting from the cavitation-erosion caused cracks (Figure 9). There was no evidence of corrosion associated with the cavitation-erosion attack of the pump rotors.

USVT Specimens-Visual and SEM Examinations: A description of the macro and microscopic appearance of the cavitation damage on each alloy follows:

E-Cu (Pure Electrolytic Copper-C1 1000): The sample surface was plastically deformed, cracked and porous after only 5 minutes of exposure. An outline of the grains was distinguishable in the forming tears, but was partially obscured by the build-up of material and voids created by the cavitation attack. After 10 minutes, the damage profile was characterized by deep, hollow cavities and irregular, broad cracks. After 30 minutes, the exposed area was almost entirely destroyed and marked by deep crevices and cavities. After approximately 60 to 90 minutes, the rosette shaped surface regions showed no significant characteristics which could be related to the matrix. After the first few minutes of vibration exposure of the sample, it could not be conclusively determined when and from which location the initial particles "broke free"

CuZn39Pb0.5 (ASTM B30): After 5 minutes, the sample surface displayed the grain structure of the material. The grains were made visible as a result of grain-boundary attack, in the shape of fine ridges and holes. After 10 minutes, the deformation of many alpha-polycrystals was apparent. Less deformed alpha-grains exhibited extensive slip bands and seams of small needle-like holes. The grain-boundaries were "built-up" and exhibited ridges and curvature, as well as small, mostly intergranular cracks. Only a few beta-grains were able to withstand an exposure of 30 minutes. The matrix was cracked and most grains were exposed and massively damaged. After an exposure time of 90 minutes, cracks, holes, and totally deformed alpha- and beta-grains characterized the damage profile. Figure 10 shows the degree of degradation of the CuZn39Pb0.5 sample after an exposure time of 90 minutes.

AlCuMg2 (150 F 209): The age hardenable Al-alloy exhibited a relatively even damage profile after an 80 minute exposure. The grains, due to the type of matrix, were plastically deformed to such a degree that, under visual examination, the damaged region appeared to have a wavy structure. Scanning Electron microscopy revealed continuous, crater-like, and plastically deformed regions, due to the deep cavitation attack. Often, micro-cracks and massively deformed protruding material resembling "ridges" were visible. Specific phases were separated from the Al-matrix and these locations appeared as holes or pores.

Ti6Al4V (MIL-T-81915): The sample prepared from this material proved to be relatively resistant to the cavitation erosion attack. After 80 minutes of exposure, the sample exhibited a fissured surface with numerous protuberances (Figure 11). Both the alpha-grains and beta-grains were "grooved" by the exposure impacts and were occasionally interspersed with fine cracks. The increased resistance to cavitation-erosion damage may be related to the spontaneously formimg and tenacious oxide film that forms on Ti and its alloys.

St 37-2K (AISI C 1006): This sample was removed from the apparatus after 30 minutes. The exposed surface contained intergranular as well as transgranular discontinuities in the metal matrix. After 300 minutes, the grain boundaries exhibited massive damage resulting in exposed grains, intergranular cracks, as well as transgranular discontinuities. Locally, grains exhibited regular flow-patterns. Craters, undulations, and porosity further characterized the damage profile.

90Mn V8 (AISI 02): After 300 minutes, the damaged region could be described as "bumpy", porous, cracked and deformed. In several locations, carbides and channel-like structures could be observed. Nonetheless, the damaged area did seem relatively flat.

X 5 CrNi 18 9 (AISI 304): Cavitation erosion attack was noticeable after only 6 minutes via slip bands, extrusions and depressions. After 10 minutes of exposure the damage was clearly visible. The varying degrees of damage found from one grain to the next leads to the conclusion that the austenitic matrix, depending on grain orientation with respect to the exposed sample surface, has a decisive impact on the propagation of the damage. After 105 minutes, the grain boundaries were broadened, the grains exposed and plastically deformed.

HVSFT Specimens-Visual and SEM Examination: The cavitation erosion damage in the high velocity seawater flow test was unexpectedly high for the Ti6Al4V and Ti7Al2Cb1Ta alloys. The samples exhibited massive damage in the cavitation-erosion region on the parallel faces. The sample ends and side walls of the specimens were also found to contain damage believed to be due to fretting and/or crevice corrosion. This extraneous attack is the result of the combined effects of specimen mounting technique and specimen vibration caused by the cavitating conditions.

Varying types of structures could be observed in the cavitation/erosion damaged region of the Ti6Al4V sample (Figure 12). Exposed grains, micro-

cracks, and plastically deformed matrix components were observed, as were "washed-out" structures and grooves. All of these features are indicative of a cavitation-erosion mechanism, and characterized the damage profile. Similar structures were found in the TiAl7Cb2Ta1 sample. Visual examination of the other Ti alloy samples included in the HVSFT revealed a similar appearance in the cavitation-erosion damaged regions.

Hydraulic Pump Rotors- Microstructure and Hardness Evaluations: Metallographic examination of the brass pump rotors indicated they were manufactured as wrought products by hot forming and machining. The microstructure showed segregated alpha-mixed crystals, spheroidal beta-mixed crystals and intermediate Mn-Si segregation which is considered typical for this type of alloy. Cavitation-erosion attack followed along grain boundaries and lead islands (Figure 13). Brinell hardness (HB) measurements were taken from transverse cross sections. The average of these measurements was 150 HB. The material specification (WL 2.0372 or ASTM B30) requires the hardness of the material to be 133 to 214 HB. The results were all within specification requirements.

USVT Samples-Microstructure and Hardness Evaluations: Hardness measurements and a description of the microstructural features obtained by metallographic examination are provided for each alloy:

E-Cu: The microstructure of the E-Cu sample consisted of alpha-polycrystals. Isolated Cu_2O inclusions were also present. Plastically deformed grains were clearly evident on a longitudinal micro. The cavities caused by the cavitation-erosion exposure sporadically reached a depth of 90μm. There was no apparent relationship between the degradation and grain size, orientation and grain boundaries. The hardness of the sample was measured to be 95 HB.

CuZn39Pb0.5: The brass sample consisted of alpha/beta-polycrystals surrounded by lead islands. The microstructure displayed intergranular and transgranular damage. Matrix decomposition could be observed (Figure 14); and in several locations, the propagation of tears between lead islands was apparent. The hardness of the material was measured to be 130 HB.

AlCu2Mg: The heat treated AlCu2Mg alloy contained Al_2Cu, Mg_2Si, Al_2CuMg and/or Al_5Cu_2 particles in addition to the Al-polycrystals. The cavitation-erosion attack occurred parallel to the direction of grain elongation and reached a depth of 60μm. The propagating material damage could be recognized primarily as fine tears along sub-grain boundaries. The hardness of the alloy was found to be 138 HB.

TiAl6V4: The TiAl6V4 alloy consisted of a microstructure containing alpha- and transformed beta-grains (Figure 15). The areas of the metal sample that were highly deformed due to the cavitation attack displayed discontinuities that were parallel to the surface and often intergranular. Significant differences between the alpha and beta grains were not apparent. The alloy had a measured hardness of 309 HB.

St 37-2K: The microstructure of the St 37-2K sample consisted of ferrite, pearlite and some degenerated pearlite. The discontinuities in the matrix was primarily intergranular in the ferrite. A material removal depth of 170μm was measured in the cavitation damaged region. The alloy hardness was measured at 183 HB.

90 MnV 8: The microstructure of the 90 MnV 8 sample consisted of a ferrite matrix which contained coagulated carbides. The discontinuities (cracking) in the material was found to be both intergranular and transgranular in nature. Anticipated cracking along the carbides contained in the ferrite matrix was not apparent. Maximum depth of attack measured in the cavitation damaged region was 80μm. The alloy had a hardness of 204 HB.

X 5 CrNi 18 9: The hot worked austenitic matrix of the highly alloyed steel X 5 CrNi 18 9 consisted of a gamma matrix (ASTM grain size of 2) containing fine carbide stringers. The relatively rough gamma matrix contained fine cracks. A correlation between the crack direction and the carbide stringers could not be found. Depth of attack in the region of cavitation was very shallow and measured approximately 10 μm. The alloy's hardness was found to be 209 HB.

HVSFT Specimens-Microstructure and Hardness Evaluations: Metallographic examination was performed on transverse and longitudinal sections of the Ti6Al4V and Ti7Al2Cb1Ta HVSFT specimens that contained both undamaged and cavitation damaged surface areas. The Ti6Al4V specimen was examined in detail to facilitate comparison to the Ti6Al4V specimen used in the USVT. The attacked surface areas exhibited grooves shown in Figure 16, with evidence of coring that would be expected in this alloy in this condition. In the damaged areas of the specimen, severe distortion of the α-structure could be observed. The appearance of distortion in the microstructure is typical for a plastically deformed material. The hardness of Ti6Al4V was found to be 276 HB.

Residue Analysis-USVT: The particles which were separated from the USVT samples were removed from the test bath and were identified using scanning electron microscopy and energy dispersive x-ray microanalysis (EDXRM). Figures 17 and 18 show size and shape of representative particles. All particles exhibited an irregular shape with rounded corners and deformed regions. Differences in particle size and shape were clearly related to the specific alloy and exposure time. Attached to the "large" particles were smaller particles of varying sizes, different chemical compositions and structural appearance. Further analysis showed that these smaller attached particles consisted of material which had separated from the other alloys tested. Macroscopic and microscopic characterization of particles produced by cavitation-erosion attack with respect to material type, damage morphology, exposure time and degree of attack provide the basis for establishing a procedure for condition monitoring of hydraulic pumps.

Roughness Measurements: The results obtained from the surface roughness measurements of the USVT samples are shown in Table 1. The establishment of a relationship between roughness measurements and previously discussed results (SEM, metallography and residue analysis) was partly successful. The Ti6Al4V and X 5 CrNi 18 9 USVT samples showed a consistent relationship between roughness and hardness. The relatively high hardness and closely associated microstructure of these materials provided better resistance against cavitation-erosion damage. Homogeneous fine grain microstructural areas displayed more resistance to cavitation attack than heterogeneous areas containing imperfections such as pores, micro-cracks and grain boundary defects. For all samples, the depth of roughness increased with increasing length of exposure; for some samples significantly (see Table 1).

Two Ti6Al4V samples from the HVSFT were selected for roughness measurements and the results are shown in Table 5 along with the roughness measurements on the Ti6Al4V USVT specimens. Although the total roughness measured on Ti6Al4V is much greater in the HVSFT, the calculated rate of roughness increase was actually lower than the calculated rate of roughness increase for the USVT. Assuming a linear rate of increase in roughness, only 0.39μm would have occurred in one hour for the Ti6AL4V HVSF specimens. By comparison, the rate of roughness increase for the Ti6Al4V USVT specimens was almost an order of magnitude higher at 3.46μm/hour. For the CuZn39Pb0.5 alloy, the rate of roughness increase as a pump rotor was 0.9μm/hour, while the rate of increase for this alloy in the USVT was 7μm/hour. The influence of the different flow patterns, angle of impact and other variables on cavitation-erosion results is evident but difficult to quantify.

Weight Loss Measurements: Table 2 lists the weight loss measurements for the USVT samples. It is apparent that, when compared to the remaining samples, alloys E-Cu and CuZn39Pb0.5 experienced significant weight loss after only a short time. Sample St 37-2K exhibited erratic behavior, showing a decreasing rate of weight loss after three hours of exposure followed by an increasing rate once again. This behavior may have been the result of absorption of water into cavities caused by the cavitation damage. The samples of alloys $AlCuMg_2$, Ti6Al4V and 90 MnV 8 exhibited a nearly constant rate of weight loss. The cavitation-erosion attack in the USVT for the CuZn39Pb0.5 and Ti6Al4V specimens versus weight loss and time is shown in Table 3.

The weight loss rates of all of all of the Ti Alloys included in the HVSFT are shown in Table 4. The incubation time and the cavitation-erosion which followed showed differences between the various alloys tested. In the HVSFT, the actual velocity and the angle of impact of the cavitation "bubble" are not known. Consequently the HVSFT results cannot be directly compared to results from the USVT. Nevertheless, at least for Titanium alloys, it appears that a normal component of impact (of the cavitation "bubble") such as occurs in the USVT is a more significant factor than the high seawater flow rate of the HVSFT. Thus in the USVT, the Ti6Al4V specimen showed a weight loss rate of 1.8mg/hour; while in the HVSFT the same alloy showed a weight loss rate of only 0.18 mg/hour.

Conclusions: Based on evaluation of the cavitation damage profiles of the various alloys tested, comparisons of two types of cavitation-erosion tests and the various analytical procedures used in this study; the following conclusions are drawn:

The Ti-based alloys were relatively resistant to cavitation-erosion damage in both USVT and HVSFT exposures.

The rate of weight loss for the Ti6AL4V alloy in the ultra sonic vibration test was higher than that which occurred in the HVSFT.

The USVT is a more controlled and quantifiable test method for evaluating cavitaion-erosion than the HVSFT.

The dissimilar appearance of the damage profiles on the Ti6Al4V alloy tested in both the USVT and HVSFT reflect the substantial differences between the two test methods.

The damage profiles obtained in the USVT were not complicated by corrosion and erosion variables that are inherent in the HVSFT.

The USVT resulted in degradation profiles on samples that can be related to cavitation damage occurring on aircraft hydraulic pump components.

Evaluation of particles from the residue of the closed circuit USV-tests permitted material identification based on energy dispesive x-ray microanalysis and the macroscopic and microscopic appearance of the damage profiles.

The results of this study provide a basis for establishing a procedure for condition monitoring of aircraft hydraulic pump components.

Recommendation: In comparing the two cavitation-erosion tests, analysis of the HVSFT was limited to results obtained on Ti alloys. The cavitation-erosion behavior of other metals such as brasses and bronzes, copper-nickel alloys, Ni-base alloys and stainless steels is also of interest. A joint program between CDNSWC, Annapolis MD, USA and WIM, Erding, Germany to characterize the behavior of a broad range of materials using the two tests would be of mutual benefit.

Acknowledgements: This work has been supported by the Naval Surface Warfare Center (NSWC), Annapolis, MD and the Defense Scientific Institute for Material Testing (WIM), Erding, Germany. The authors would like to thank Mr. J. Patton (NSWC) and Dr. G. Kohlhaas (WIM) for helpful discussions; and Messrs. M. Kollros (WIM) and W. Muellera (WIM) for their professional assistance in conducting tests and SEM-work.

References:

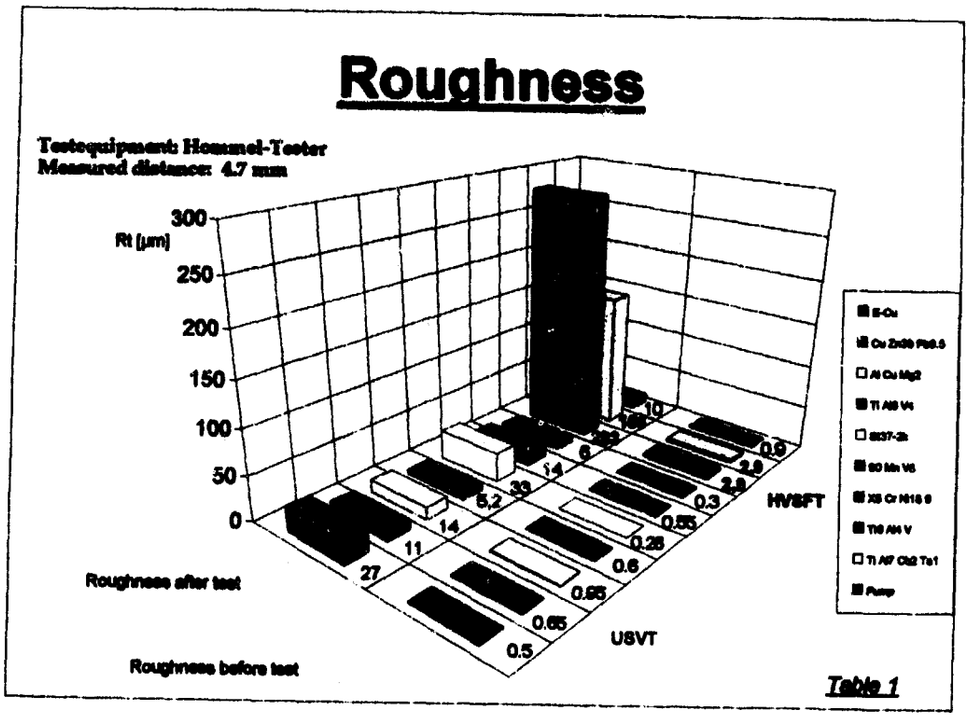
1. W. Hampton, "Das Problem der Kavititionserosion in Flugzeughydrauliksystemen", A.F.R.Ae.S, Hatfield, UK, Luft-u. Raumfahrttechnik 13, Nr 1, Jan. 1967

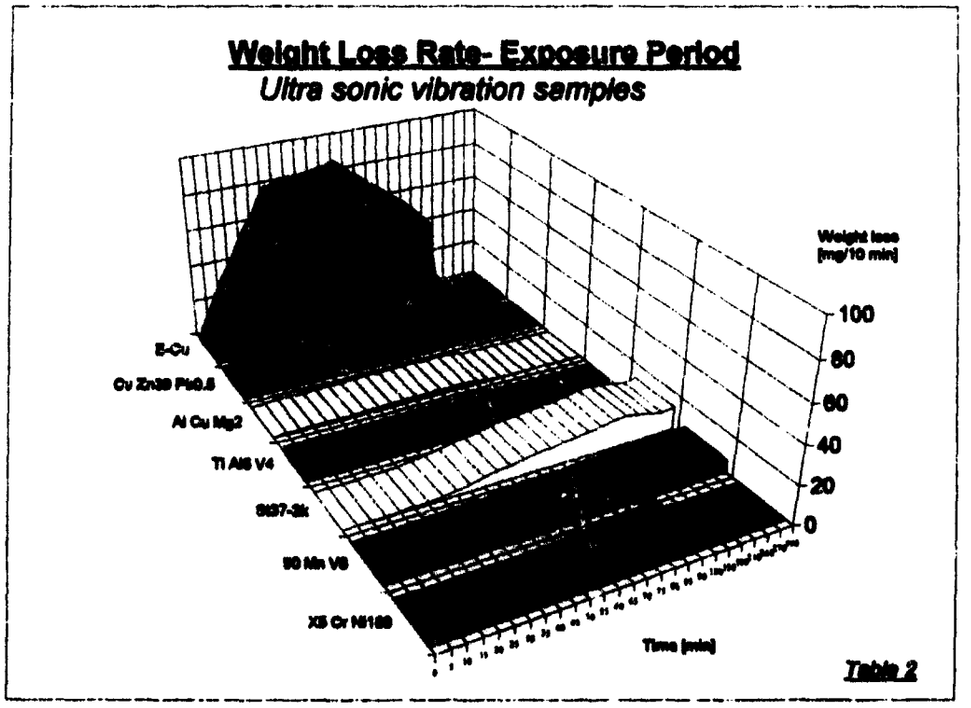
2. FVV-test procedure, section R 443, Frankurt a/M, 1986

3. K.D. Efird, "The Effect of Fluid Dynamics on the Corrosion of Copper-Base Alloys in Sea Water", Corrosion, Vol 33, Jan 1977.

4. G.F. Wilhelmi, "Composite Materials For Ship Machinery Applications", STAR/Symposium/3rd IMSDC, Pittsburgh, June 1988

5. R.J. Ferrara, Gudas J.P., "Corrosion of Copper-Base Alloys with Respect to Seawater Velocity", Proceedings of the Third International Congress on Marine Corrosion and Fouling, 1972


Roughness
Testequipment: Hommel-Tester
Measured distance: 4.7 mm
Rt [µm]
300
250
200
150
100
50
0
Roughness after test
Roughness before test
HVSFT
USVT
27
11
14
5,2
33
14
6
10
0.9
2.8
0.3
0.55
0.28
0.6
0.55
0.65
0.5
E-Cu
Cu Zn39 Pb0.5
Al Cu Mg2
Ti Al6 V4
9137-2k
60 Mn V8
X5 Cr Ni18 9
Ti6 Al4 V
Ti Al7 Cb2 Ta1
Pump
Table 1


Weight Loss Rate- Exposure Period
Ultra sonic vibration samples
Weight loss
[mg/10 min]
100
80
60
40
20
0
E-Cu
Cu Zn39 Pb0.5
Al Cu Mg2
Ti Al6 V4
9137-2k
60 Mn V8
X5 Cr Ni189
Time [min]
Table 2

Correlation of the damage due to cavitation and weight loss rate

Weight loss rate (CuZn39Pb0,5, Ti6Al4V and TiAl6V4)

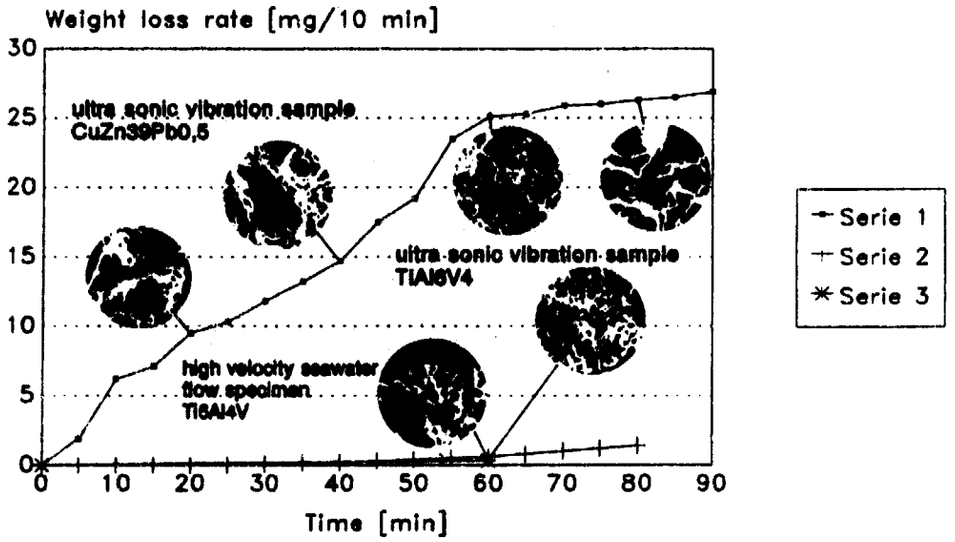


Table 3

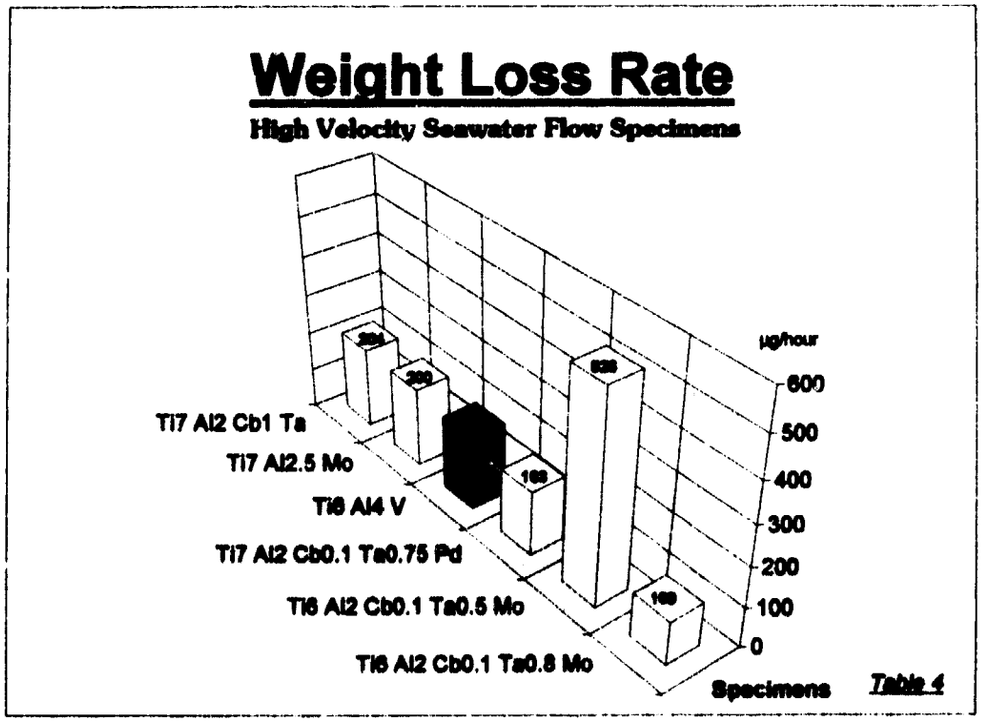


Table 4

Table 5

Hardness, Roughness and Microstructural Features of USVT Samples and HVSF Specimen

Material	Hardness [HB]	Roughness before and after test/ damage [μm]	Microstructural features near the surface due to cavitation-erosion
Hydraulic pump rotor CuZn39Pb0.5	150	0,9/10 [11 hours]	Attack followed grain boundaries and lead islands

USVT-samples

E-Cu	95	0,5/27 [90 min]	deformed grains
CuZn39Pb0.5	130	0,65/11 [90 min]	attacked grain boundaries and lead islands
AlCuMg2	138	0,95/14 [90 min]	damage occurred along subgrain boundaries
TiAl6V4	309	0,6/5,2 [90 min]	deformation discontinuities along highly deformed regions
St 37-2k	183	0,28/33 [90 min]	intergranular attack along ferrite
90 MnV 8	204	0,55/14 [90 min]	cracks along carbides
X5CrNi 18 9	209	0,3/6 [90 min]	cracks along γ - matrix

HVSF-specimen

Ti6Al4V	276	2,9/283 [720h]	cracks along α - grains

*TiAl6V4 = Ti6Al4V

Figure 1 1,6:1
Pump rotor with crack along and between the drilling holes.

Figure 2 1,7:1
The disassembled pump rotor shows the fracture surface and the crack initiation areas near the overflow canals

Figure 3 1:1,2
The pump rotor shows the worn out area and the cracks between the overflow canals

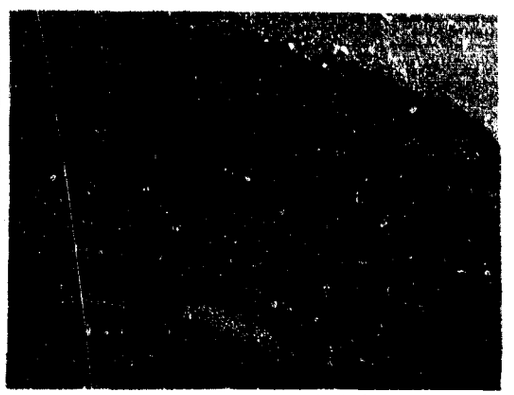


Figure 4 2,5:1
Overflow canals with damaged areas at the overflow canals due to cavitation-erosion

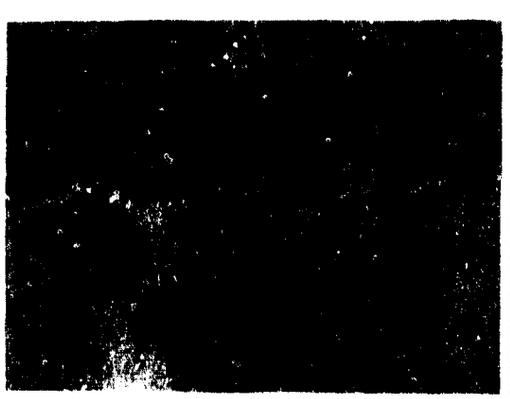


Figure 5 2:1
Specimen used in the ultra sonic vibration test

Figure 6 2:1
Specimen used in the high velocity test

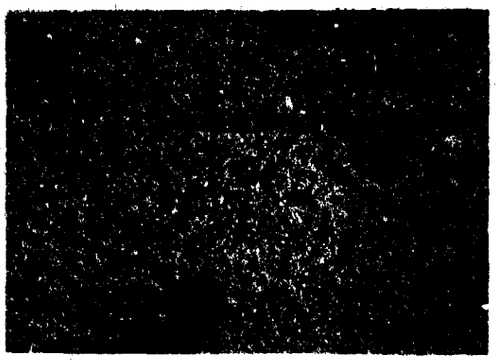


Figure 7 60:1
SEM-image showing an area inside a overflow canal (pump rotor) which was damagedged due to cavitation-erosion

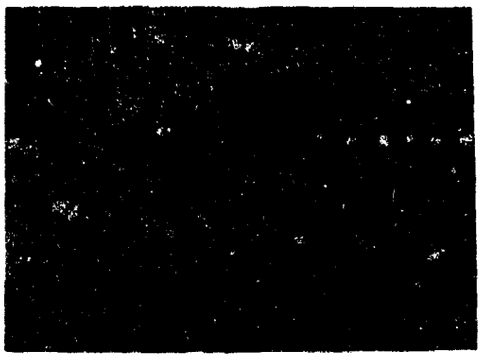


Figure 8 1000:1
Close-up photograph from Fig. 7; the surface shows cracks, holes and deformations

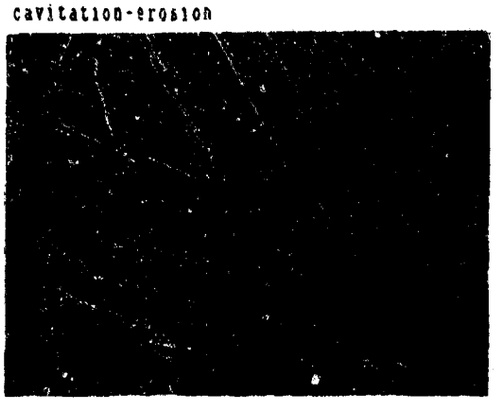


Figure 9 1000:1
Curvature, ridges and mound characterized the surface of the cavitation-erosion damaged area (pump rotor)

Figure 10 1000:1
Cu-alloy after 90 minutes ultra sonic vibration test; cracks, holes and plastically deformed regions characterized the damage profile

Figure 11 1000:1
SEM-micrograph of the TiAl6V4 specimen, exposed for 30 minutes ultra sonic vibration test

Figure 12 1000:1
SEM-micrograph of the TiAl4V specimen, exposed 720 hours high velocity seawater flow

Figure 13 200:1
A photomikrograph of the pump rotor represents locations of cavitation-erosion attack (see arrows)

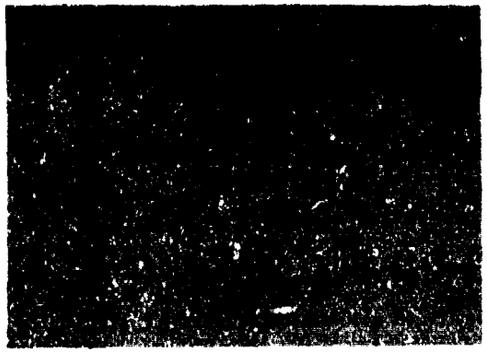


Figure 14 500:1
Transverse section from the Cu-alloy specimen; the surface indicates cavitation-erosion damage due to ultra sonic vibration

Figure 15 500:1
Photomicrograph of the TiAl6V4 specimen which was damaged due to ultra sonic vibration

Figure 16 63:1
Metallographic profile of the Ti6Al4V specimen after 720hours high velocity seawater flow

Figure 17 2000:1
SEM photograph of a Cu-alloy particle which was separated from the ultra sonic test bath

Figure 18 3000:1
SEM photograph showing Ti-alloy particle representative for cavitation-erosion damage

FRACTOGRAPHY OF ADVANCED STRUCTURAL CERAMICS

Jeffrey J. Swab
U.S. Army Research Laboratory - Materials Directorate
405 Arsenal Street
AMSRL-MA-CA
Watertown, MA 02172-0001

George D. Quinn
National Institute of Standards and Technology
Ceramics Division
Building 223, Room A359
Gaithersburg, MD 20899-0001

Abstract: The strength of many brittle ceramic materials reflects the flaws present in the material and the intrinsic fracture toughness. Interpretation of strength data for monolithic or simple ceramic composites which exhibit brittle fracture requires fractographic analysis irrespective of whether the strength data was obtained for quality control, materials research and development, or design purposes. This paper discusses the importance of fractography and recent standardization efforts. It also shows how fractography can be applied to the development of a material data base and the fracture of prototype ceramic components.

Key Words: Advanced ceramics; Flaw characterization; Fractography; Fracture origins; Standardization.

Introduction: In recent years processing improvements have significantly enhanced the mechanical properties and reliability of advanced ceramic materials (i.e., silicon nitride, silicon carbide, zirconia). Additionally tremendous strides have been made to standardize the techniques used to obtain this information. These efforts increase the likelihood that advanced ceramics will find an avenue into applications such as bearings, heat exchangers, and gas turbine engines. Unfortunately the improvement in properties and the development of standardized testing methods alone will not insure their use in these arenas. Other information is also necessary including an understanding of the prevalent failure mechanisms in the ceramic as well standardized methods to characterize these as mechanisms.

In brittle ceramic materials fracture commences from a single location which is termed the fracture origin. The fracture origin consists of some microstructural irregularity which acts as a stress concentrator. In the parlance of the engineer or scientist, these irregularities are termed "flaws" or "defects"; they occur in all materials regardless of how well they are manufactured. This usage **should not** be construed to mean that the material has been prepared improperly or is somehow faulty. These irregularities or flaws can develop

during or after fabrication of the material. Large flaws (relative to the average size of the microstructural features) such as pores, agglomerates and inclusions are typically introduced during processing and can (in one sense) be considered intrinsic to the manufacture. Other flaws can be introduced after fabrication as a result of machining, handling, impact, wear, oxidation, corrosion and high temperature exposure. These flaws can be considered extrinsic flaws. However, machining damage will be considered intrinsic to the manufacture to the extent that machining damage is a natural consequence of producing a finished specimen or component. It is beyond the scope of this paper to discuss the source of flaws or their behavior from a fracture mechanics viewpoint.

Overview of MIL HDBK-790: The release of Military Handbook 790 "Fractography and Characterization of Fracture Origins in Advanced Structural Ceramics" (MIL HDBK-790) by the Department of Defense in July 1992 marked the most comprehensive effort to date to standardize the fractographic analysis of ceramic materials. The objectives of MIL HDBK-790 are two-fold. Firstly it acts as an educational tool for scientist/engineers venturing into fractography of ceramics for the first time and secondly it provides an efficient and consistent methodology to locate and characterize fracture origins in this class of materials for more experienced personnel.

The accurate characterization of fracture origins in advanced structural ceramics is essential for the comprehensive interpretation of the mechanical property data for design. It is also important in quality control and materials research and development. MIL HDBK-790 serves as a bridge between mechanical testing standards and statistical analysis standards to permit comprehensive interpretation of the data for design, Figure 1.

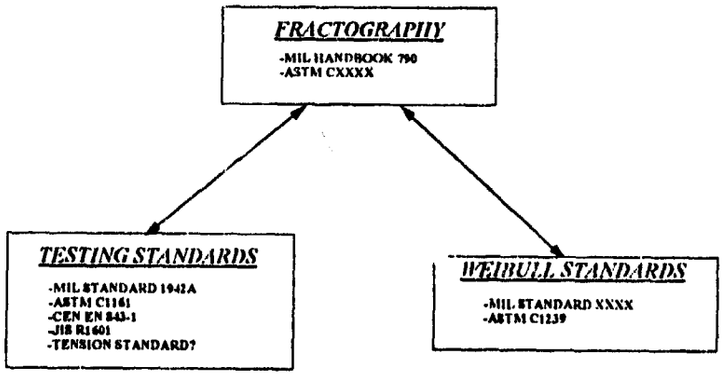


Figure 1. Relationship of Fractographic Analysis to Mechanical Property Testing and Statistical Analysis.

The procedures outlined in MIL HDBK-790 are directed at preserving and finding the primary fracture surface and the strength-limiting flaw. Care must be taken at all times even before the mechanical testing of the specimen or component.

The specimen should be tested to failure in a manner that preserves the primary fracture surface and the specimen for further reference. Care should be taken during handling and storage of the specimen to minimize additional damage and/or contamination of the fracture surface.

A visual inspection (1-10X) is the first step of the examination. This is done to determine crack branching patterns, any evidence of abnormal failure patterns (indicative of testing misalignment), the primary fracture surface, and the location of the mirror and, if possible, the fracture origin. Subsequent optical examination (10-100X) is conducted to locate and, if possible, characterize the flaw. At this time the specimen surface(s) near the origin should be examined for evidence of handling or machining damage. If definitive characterization of the fracture origin can not be made at this point, then the optical examination is done with the purpose of expediting the Scanning Electron Microscope (SEM) examination (10-2000X).

It may not be feasible, practical or even necessary to examine all the fracture surfaces with the SEM. The extent of analysis required will depend upon the following factors; 1) The conduciveness of the material to fractographic analysis. Some coarse-grained or porous ceramics will not leave markings that permit flaw characterization. On the other hand, high strength (≥1 GPa) ceramics will have very small origins which may be difficult to differentiate from the typical microstructural features. Alternatively, a low-to-medium strength ceramic may have flaws which are easily characterized by optical means; 2) The purpose of the fractographic analysis: quality control, research and development, or design. Table 1 gives suggested guidelines for medium-to-high strength ceramics.

Table 1. Sampling Guidelines.

	1 - 10X Visual	10 - 100X Optical	10 - 2000X SEM
Level 1 - Quality Control	Specimens which fail to meet minimum strength requirements	Specimens which fail to meet minimum strength requirements	Optional
Level 2 - Quality Control - Materials Development	All Specimens	All Specimens	Representative specimen -2 of each flaw type -the 5 lowest strength specimens -at least 2 optically unidentifiable flaws
Level 3 - Materials Development - Design	All Specimens	All Specimens	All specimens, or as many specimens such that combined optical and SEM characterizes 90% (100% for design) of all identifiable flaws

During each step, it is important to keep appropriate records and photographs in order to characterize the flaw, show its location and the general features of the fracture. For a new

material, or a new set of processing or exposure conditions, it is highly recommended that a representative polished and/or etched section of the microstructure be photographed to show the typical microstructural features.

An important feature of MIL HDBK-790 is the adoption of a consistent and comprehensive manner of fracture origin characterization including nomenclature. This will enable the construction of efficient computer data bases. Fracture origins will be characterized by the following three attributes: IDENTITY, LOCATION, and SIZE (optional) as summarized in Table 2 and shown in Figures 2-4.

Table 2. Flaw Characterization Scheme.

IDENTITY	LOCATION	SIZE (Optional)
Nomenclature & spatial distribution; volume or surface	Volume (bulk), Surface, Near-Surface, Edge	Mean diameter or estimate of the major and minor axes

Origins are either inherently volume-distributed throughout the bulk of ceramics (e.g. agglomerates, large grains or pores) or inherently surface-distributed on the ceramic (e.g. machining damage, pits from oxidation or corrosion). The volume-distributed origins can, in any test specimen, be located in the bulk, at the surface, near-to-the surface, or at an edge. The variety of locations is a consequence of the random sampling procedure incurred in preparing test specimens.

Figure 2. Flaw in a Y-TZP tested in 4-point flexure at room temperature. Inclusion (I^V), volume, ≈10 x 20 µm.

Figure 3. Flaw in an Al_2O_3/SiC(w) composite tested in 4-point flexure at room temperature. Large grains (LG^V), surface, ≈20 x 35 μm.

Figure 4. Flaw in an injection molded then sintered Al_2O_3 + Y_2O_3-Si_3N_4 tested in 4-point flexure at room temperature. Surface Void (SV^S), surface, ≈100 μm. (Courtesy of A. Pasto, GTE Laboratory, now with Oak Ridge National Laboratory).

IDENTITY: The fracture origins will be characterized by a phenomenological approach that identifies what the flaw is and **not** how it appears under a particular mode of viewing. Descriptions of the mode of viewing may be used as qualifiers, e.g. "pores that appear white when viewed optically,"; but the use of the appearance, i.e. "white spots" should be avoided. This approach is chosen since flaws can appear drastically different in optical versus electron microscopy. The flaw nomenclature has been designed to identify the flaw by name (e.g. pore, inclusion) and is grouped based on the inherent spatial distribution of the flaw type in the bulk ceramic, see Table 3.

Table 3. Fracture Origin Nomenclature.

ORIGIN TYPE	CODE
Inherent Volume	
Pore	P^V
Porous Seam	PS^V
Porous Region	PR^V
Agglomerate	A^V
Inclusion	I^V
Compositional Inhomogeneity	CI^V
Large Grain(s)	LG^V
Crack	C^V
Inherent Surface	
Machining Damage	MD^S
Service Damage	SD^S
Pit	PT^S
Surface Void	SV^S
Miscellaneous	
Other	@
Unidentifiable	?

Note: In the flaw code superscripts "V" and "S" are used to indicate the inherent distribution of the flaw types.

LOCATION: The location of a flaw in a given specimen shall be qualitatively determined The flaw must be characterized as being located in the volume (bulk), at the surface, near the surface, or at an edge (if such exist). The locations shall be used only to specify the location of the flaw in a given specimen. It is not intended to be used to statistically differentiate flaw populations!

SIZE: Since a flaw's true size may not be revealed on the fracture surface, and because fracture mechanics analysis of most flaws is very difficult due to their complex shapes, characterization of the size is required only in a qualitative sense as necessary to identify the general nature of the flaw (e.g. the 20 μm pore versus 1 μm porosity). If the size is measured, the mean diameter should be reported for equiaxed flaws while the major and minor axes should be reported for nonequiaxed flaws.

Application of Fractography to Laboratory Test Specimens: The evolution of a strength database for a ceramic material in the research and development or design stage entails the breakage of hundreds of specimens. This will also involve a significant amount of fractographic analysis. How does one deal with such a large volume of information in a concise and efficient manner? It is suggested that once fractography is completed a fractographic montage, see Figure 5, be created . These montages permit photographs of

the fracture origins to be arranged on a single worksheet around a Weibull graph of the strength results. A mass of optical and SEM photos can be organized in this fashion and notes and observations scrawled on such a worksheet. Patterns can emerge, such as the relationship between the large-grain fracture origins and the low-strength end of the strength data. The montages are relatively easy to store and the ease of retrieval encourages fractographic reassessment or comparisons of one montage to another. The information and relationships obtained from this montage can then be organized in to a labeled Weibull graph, Figure 6. At a glance this graph will permit an assessment of the applicability of the Weibull analysis. In many instances multiple flaw populations can be directly related to irregularities on the strength curve.

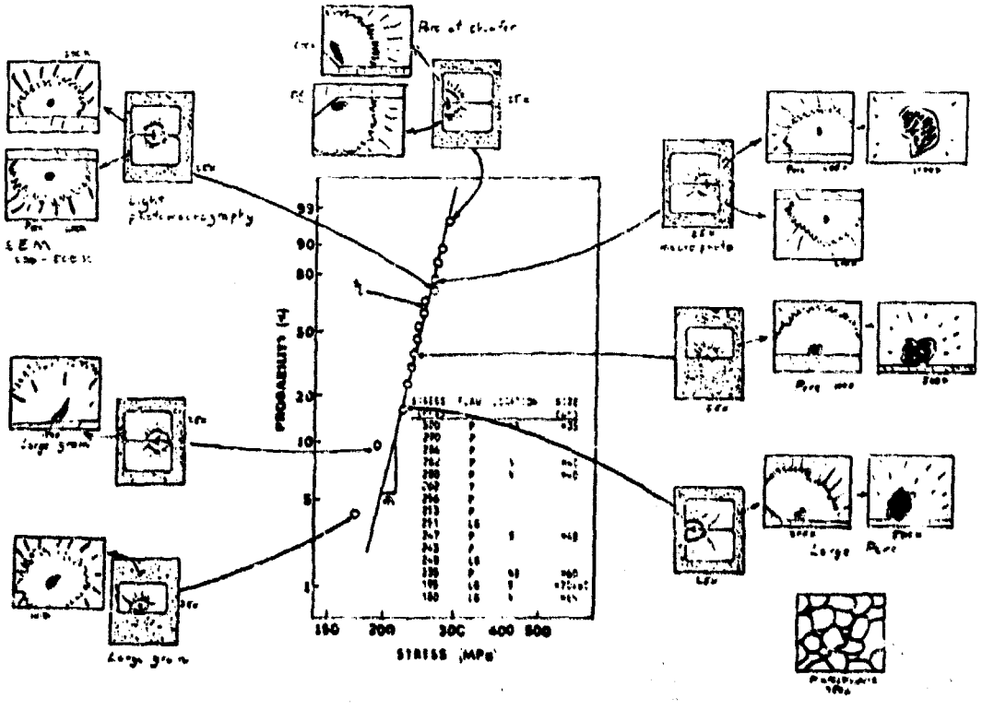


Figure 5. Schematic of a Fractography Montage

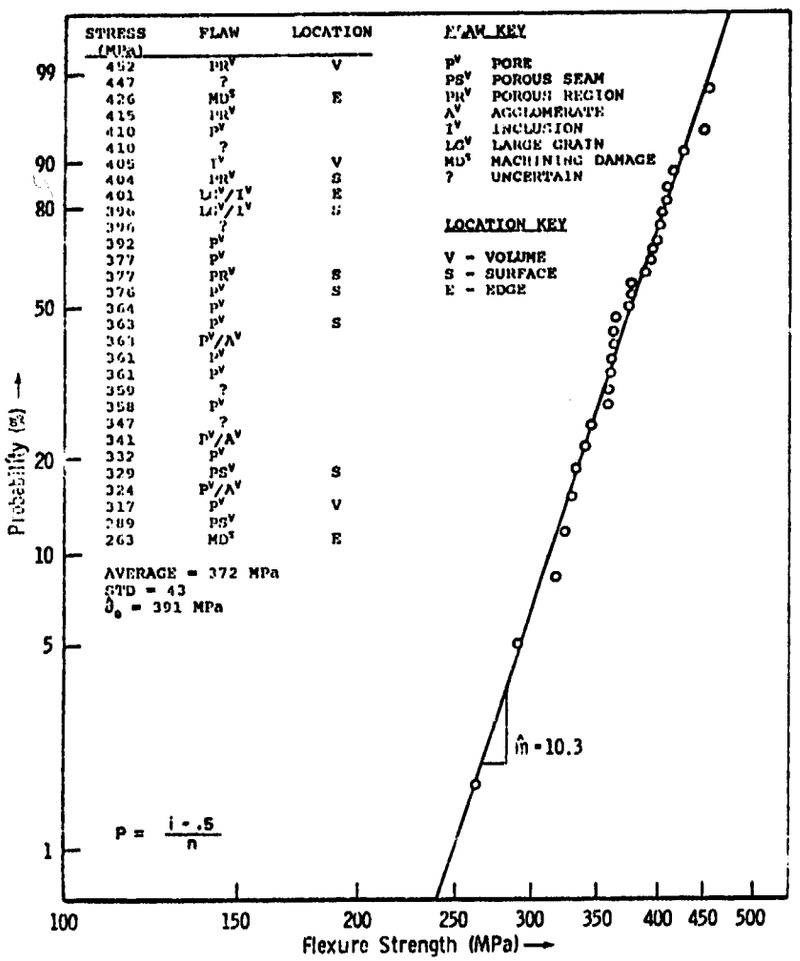


Figure 6. A labeled Weibull graph incorporating the strength and fractographic results.

Application of Fractography to Ceramic Prototypes: Although the procedures described in the handbook are primarily for the analysis of mechanical test specimens loaded in so-called "fast fracture," they can be extended to include other modes of loading and are relevant to component failure analysis. Some examples are given below.

In example 1 a Si_3N_4 gas turbine rotor was spun to failure to determine the failure stress and the strength-limiting flaw. The fragmented pieces shown in Figure 7a were assembled, using the fractographic analysis techniques outlined in MIL HDBK-790, to the configuration in Figure 7b. The assembly is much like completing a jigsaw puzzle except pieces will probably be missing which prohibits complete reconstruction of the component. In this example reconstruction was sufficient to allow for identification of a possible fracture origin at the edge of the inner bore, denoted by A in Figure 7c and the arrow in

Figure 7d. Following the scheme in Table 2 this origin would be labeled as: Machining Damage (MD^S), surface, ?? (Size not measured).

Figure 7a. Fragmented ceramic rotor.

Figure 7b. Reconstructed rotor

Figure 7c. Reconstructed rotor. "A" indicates possible initiation site.

Figure 7d. Possible fracture origin.

Example 2 shows the application of fractography to a ceramic gun barrel. The ceramic liner is a sintered, SiC with a high elastic modulus (≈ 400 GPa) and a low fracture toughness (≈ 3.5 MPa*√m). The liner and steel jacket are assembled through a shrink fit technique. This technique places a residual compressive stress in the ceramic. Failure occurred at the liner/jacket interface. In Figure 8 there is a Hertzian crack at the

liner/jacket interface (large white arrow). A non uniform mechanical fit between the liner and jacket caused point contacts which may have initiated the Hertzian crack. The elastic shock waves produced during firing will reflect at the liner/jacket interface boundary due to the high elastic modulus mismatch. These shock waves could initiate and propagate the Hertzian crack. The path of crack propagation is shown by the small white arrows. The origin would be labeled as Service Damage (SD^S), surface, ?? (Size not measured).

Several recommendations were made to avoid this failure mechanisms: 1) Use a ceramic with a lower elastic modulus. 2) Use a ceramic with a higher fracture toughness. 3) Improve the mechanical fit between the ceramic liner and the steel jacket.

Figure 8. Cross section of a failed ceramic gun barrel.

Summary: The accurate location and characterization of fracture origins in advanced structural ceramics is important whether it is performed as part of the development of a strength data base or during the analysis of a fractured ceramic component. Only with comprehensive fractography can subtle strength differences within a material be accounted for by variations in the flaw type. Fractography of failed components can yield valuable information about the performance of the component in the intended environment. At the present time only the Department of Defense has standardized this characterization technique (MIL HDBK-790). Commercial standardization through ASTM is under way.

BULK PROPERTIES AND FAILURE OF ADVANCED COMPOSITES

M. K. Hinders*, S. D. Bogan and T.-M. Fang

Massachusetts Technological Laboratory, Inc., 330 Pleasant Street, Belmont, MA 02178

Abstract: In this paper we discuss the usefulness of exact and analytic solutions to elastic-wave scattering problems to describe the dynamical behavior of composites. In particular, we consider the scattering of elastic waves from canonical solid inclusions in order to model particle and fiber reinforcements in composites. By investigation of these simple problems we show how to predict bulk properties and failure initiation modes of composites. Results are also directly applicable to ultrasonic nondestructive evaluation of reinforced composite materials.

Key Words: Composites; Dynamic stress concentrations; Elastic wave scattering; Nondestructive evaluation; Ultrasonics.

Introduction: The transfer of advanced composite materials technology from military to civilian applications has created a need for significant extension and further development of the theoretical models which are used for describing bulk mechanical properties, ultrasonic nondestructive evaluation, and prediction of material failure. Advanced composite materials and structural members present unique difficulties for the prediction and monitoring of both their mechanical properties and their probable modes of failure. Although phenomenological and experiment-based models abound, it is most desirable—and in the long run most accurate and least expensive—to develop field-theoretical models which predict bulk properties and failure of advanced composites directly from the known material configuration parameters. A rigorous and analytic theory of the multiple scattering of elastic waves provides the required field-theoretical framework since both the bulk mechanical and wave properties as well as the dynamic failure modes of the advanced composite are derived from the elastic wave scattering analysis.

We discuss the scattering of elastic waves from particles and fibers embedded in an infinite elastic medium of different elastic material parameters in order to predict bulk elastic properties and failure modes of particle- and fiber-reinforced composite materials. The scattering problem is formulated using the exact method of Mie scattering to produce a set of algebraic equations in the unknown modal coefficients of the scattered and transmitted fields. Incident waves are taken to be in turn plane or spherical, longitudinal or transverse; boundary coupling between compressional and shear modes is taken fully into account, and both types of waves are supported in

all media. Inclusions of four types are considered: spheres, spheroids, circular and elliptic cylinders. Differential and total scattering cross sections as well as the extinction cross sections of the four types of scatterers are derived, and in special cases the results are shown to reduce to well-known expressions. Failure modes are predicted by calculation of the dynamic stress concentrations, and an elastic multiple scattering analysis gives the bulk mechanical properties of the composite material. Analysis in this area has been lacking primarily because elastic waves propagating in solid materials involve much more complicated phenomena than are present in wave propagation problems in purely *acoustic* media such as air or water. Unlike inviscid fluids, elastic solids have rigidity–they resist shear or distortional deformations as well as compressional deformations–and as a consequence, shear or transverse waves propagate in solid elastic media in addition to the familiar compressional sound waves. Neglecting shear waves in the ultrasonic nondestructive evaluation of solids, for example, leads to incorrect results, because when purely compressional ultrasonic waves encounter a material discontinuity where reflection, refraction, diffraction, etc. occur, the resulting *scattered* waves are in general always combinations of compressional and shear waves. This mode coupling is a direct consequence of the boundary conditions which must be satisfied at the material discontinuity and cannot be avoided or ignored in a correct description of wave phenomena in solid media.

In advanced composite materials, a rigorous elastic multiple scattering treatment is required to quantitate the propagation constants of the bulk (effective) medium. For particle-reinforced composites a correct quantitative model is the multiple scattering of elastic waves from elastic spheres and spheroids, and for unidirectional fiber-reinforced composites a correct quantitative model is the multiple scattering of elastic waves from elastic cylinders, both circular and elliptic. We here discuss results for the scattering of elastic waves from elastic inclusions in order to develop a rigorous and analytic theory of elastic wave propagation in particle- and fiber- reinforced laminate composites. The field-theoretical solutions then allow prediction of both the bulk properties and dynamic failure modes of composite materials and structural members, and give a rigorous foundation for the quantitative characterization of defects in these materials.

Elastic Wave Scattering: The theory of the scattering of elastic waves dates to the works of Clebsch [1] and Rayleigh [2], with a comprehensive overview of progress in the field given twenty years ago by Pao and Mow [3]. Although a great deal of work on the scattering of elastic waves has appeared since [4], the exact expressions for the fields scattered by elastic spheres and cylinders have been worked out only recently by Hinders [5] – [10]. Extension of these results for the consideration of elastic wave scattering from two-layer elastic cylinders and spheres has also been now completed [11] – [12]. This field of study is often referred to as "Mie Scattering" after G. Mie [13] who is credited with early related developments in optics.

For an isotropic, homogeneous, linearly-elastic medium small elastic displacements

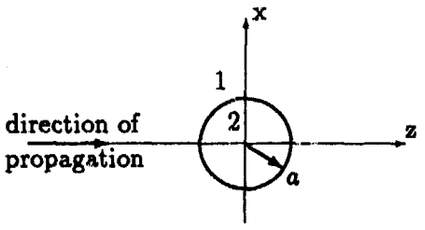


Figure 1: Problem geometry for elastic wave scattering from an elastic sphere.

are governed by the equation

$$\rho\partial_t^2\vec{u} - \mu\nabla^2\vec{u} - (\mu+\lambda)\nabla(\nabla\cdot\vec{u}) = 0 \tag{1}$$

where ρ is the constant medium density, μ and λ are elastic Lamé parameters and $\vec{u}$ is the displacement vector. After some manipulation and assuming harmonic time variation $e^{-i\omega t}$, the equation of motion may be written as

$$(\nabla^2 + K^2)\vec{u} - (1 - \frac{K^2}{k^2})\nabla(\nabla\cdot\vec{u}) = 0 \tag{2}$$

Here, $K = \omega/c_T$ and $k = \omega/c_L$ are the propagation constants for shear and compressional elastic waves with $c_T^2 = \mu/\rho$ and $c_L^2 = (\lambda+2\mu)/\rho$ defining the shear and compressional wave propagation velocities respectively.

It is convenient in the analysis of elastic wave propagation and scattering to consider shear and compressional waves separately by writing $\vec{u} = \vec{u_L} + \vec{u_T}$ where $\nabla\times\vec{u_L} \equiv 0$ and $\nabla\cdot\vec{u_T} \equiv 0$. This Helmholtz decomposition allows us to define three scalar potential generating functions $\pi_L, \pi_{SH}, \pi_{SV}$ where the longitudnial displacement vector is proportional to the gradient of π_L and the two components of the transverse displacement vector are proportional to the curls of π_{SH} and π_{SV}. This is crucial because rather than dealing with the vector partial differential equation (2) for the total elastic displacement field, we can instead consider the familiar scalar wave equations

$$\left(\nabla^2 + k^2\right)\pi_L = 0 \qquad \left(\nabla^2 + K^2\right)\pi_{SH} = 0 \qquad \left(\nabla^2 + K^2\right)\pi_{SV} = 0 \tag{3}$$

whose general solutions separate in a variety of coordinate systems. In an isotropic, homogeneous, linearly-elastic solid we find that the one compressional and two shear wave modes propagate independently. The compressional waves have propagation constant k and both shear wave modes have propagation constant K. However, the modes will be coupled whenever a material discontinuity is encountered, necessitating a full vector treatment of elastic wave scattering phenomena.

Particulate Composites: For a single spherical elastic inclusion embedded in another elastic medium this dynamical boundary value problem have been recently solved [5], and scattering cross sections are given as exact and analytic closed form expressions. The scattering cross section for the sphere represents the ratio of the

average rate at which energy is scattered by the sphere to the average rate at which the energy of the incident plane wave crosses unit area normal to its direction of propagation. Dividing by the geometric cross section of the sphere, πa^2, we write the normalized scattering cross sections for incident longitudinal and transverse plane waves respectively scattering from elastic spheres:

$$\begin{aligned} Q^L_{scat} &= \frac{4}{(k_1 a)^2}\sum_{l=0}^{\infty}(2l+1)\left\{\left|\frac{\Delta_1^L}{\Delta_0}\right|^2 + l(l+1)\left(\frac{k_1}{K_1}\right)^3\left|\frac{\Delta_3^L}{\Delta_0}\right|^2\right\} \\ Q^T_{scat} &= \frac{2}{(K_1 a)^2}\sum_{l=1}^{\infty}\frac{(2l+1)}{l(l+1)}\left\{l(l+1)\left(\frac{K_1}{k_1}\right)^3\left|\frac{\Delta_1^T}{\Delta_0}\right|^2 + \left|\frac{\Delta_3^T}{\Delta_0}\right|^2 + \left|\Delta_5^T\right|^2\right\} \end{aligned} \tag{4}$$

where the quantities Δ_0–Δ_5^T are complicated expressions in spherical Bessel and Hankel functions and their derivatives. For example, Δ_0 is given by the expression:

$$\begin{aligned} \Delta_0 = \left(\frac{\mu_2}{\mu_1}-1\right)^2 &[l(l+1)-2]\left[\frac{k_2 a j_l'(k_2 a)}{j_l(k_2 a)}\frac{K_2 a\psi_l'(K_2 a)}{\psi_l(K_2 a)} - l(l+1)\right] \\ &\times\left[\frac{k_1 a h_l'(k_1 a)}{h_l(k_1 a)}\frac{K_1 a\zeta_l'(K_1 a)}{\zeta_l(K_1 a)} - l(l+1)\right] \\ +\frac{1}{2}(K_1 a)^2\left(\frac{\mu_2}{\mu_1}-1\right)&\left\{\left[\frac{k_2 a j_l'(k_2 a)}{j_l(k_2 a)}\frac{K_2 a\psi_l'(K_2 a)}{\psi_l(K_2 a)} - l(l+1)\right]\right. \\ &\times\left[\frac{K_1 a\zeta_l'(K_1 a)}{\zeta_l(K_1 a)} + 2\frac{k_1 a h_l'(k_1 a)}{h_l(k_1 a)} - 2l(l+1)\right] \\ &-\frac{\rho_2}{\rho_1}\left[\frac{k_1 a h_l'(k_1 a)}{h_l(k_1 a)}\frac{K_1 a\zeta_l'(K_1 a)}{\zeta_l(K_1 a)} - l(l+1)\right] \\ &\left.\times\left[\frac{K_2 a\psi_l'(K_2 a)}{\psi_l(K_2 a)} + 2\frac{k_2 a j_l'(k_2 a)}{j_l(k_2 a)} - 2l(l+1)\right]\right\} \\ +\frac{1}{4}(K_1 a)^4&\left\{l(l+1)\left[1-\frac{\rho_2}{\rho_1}\right]^2 - \left[\frac{k_2 a j_l'(k_2 a)}{j_l(k_2 a)} - \frac{\rho_2}{\rho_1}\frac{k_1 a h_l'(k_1 a)}{h_l(k_1 a)}\right]\right. \\ &\left.\times\left[\frac{K_2 a\psi_l'(K_2 a)}{\psi_l(K_2 a)} - \frac{\rho_2}{\rho_1}\frac{K_1 a\zeta_l'(K_1 a)}{\zeta_l(K_1 a)}\right]\right\} \end{aligned}$$

Note here that we use the subscripts 1 and 2 to indicate the exterior medium ($r > a$) and the spherical scatterer ($r < a$) respectively as shown in Figure 1. The spherical radial functions are related to the half-order cylindrical Bessel and Hankel functions by

$$\psi_l(x) = x j_l(x) = \sqrt{\pi x/2}J_{l+1/2}(x) \qquad \zeta_l(x) = x h_l(x) = \sqrt{\pi x/2}H^{(1)}_{l+1/2}(x)$$

and prime indicates differentiation with respect to argument. Although these expressions look quite complicated, it is important to keep in mind that they are exact, analytic and closed-form. Moreover, the spherical radial functions are well-understood and the series are convergent and well-behaved. In figure 2 we show (5) graphically

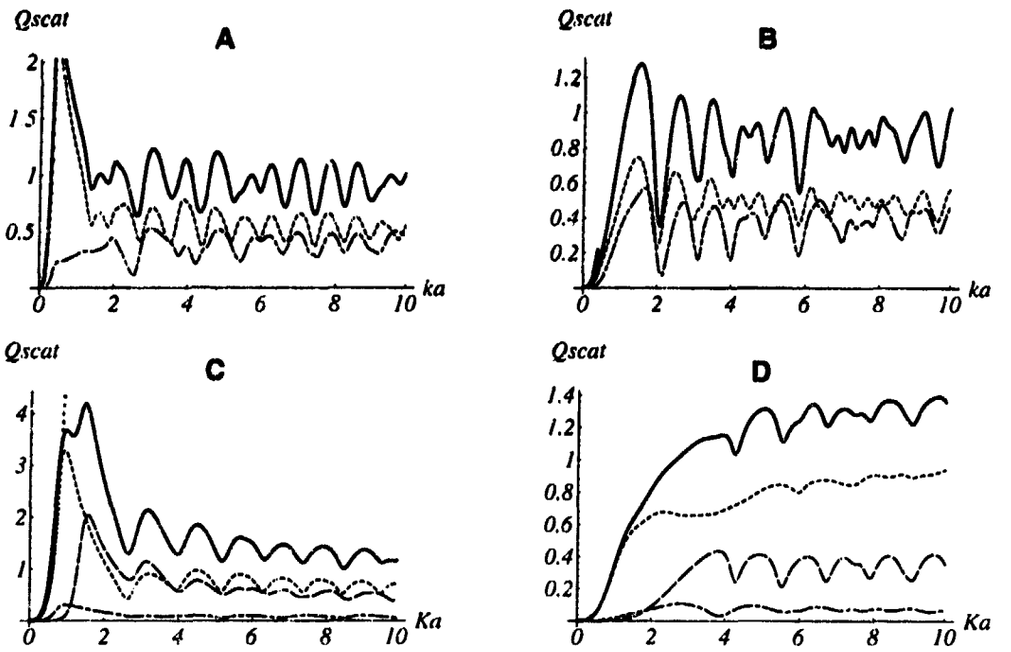


Figure 2: Normalized scattering cross sections for an elastic sphere.

for materials of Truell et al [14]. Plots A,C are for a stainless steel sphere in magnesium (L and T incidence) and plots B,D are for a magnesium sphere in stainless steel. The dark solid line is the scattering cross section and the dotted line is the Rayleigh limit approximation for small scatterers. The dashed lines are the L, SV and SH components that make up the total scattering cross section. Note that the behavior is in general quite complicated, and that the small argument approximation of Truell [15] – [16] is appropriate only for scatterers small compared to wavelength. These scattering cross sections can be used directly to give the bulk properties (e.g. effective wave numbers) of particulate composites by a variety of multiple scattering theories [17]. Moreover, once the scattering problem has been solved exactly, it can be used to predict dynamic stress concentrations of composites under time-varying loads [8]. Particulate composites with flake-like or needle like reinforcements can be described in a similar manner by first solving the corresponding elastic-wave scattering problem in spheroidal coordinates.

Fibrous Composites: For an isolated cylindrical fiber embedded in a solid matrix material we have also solved exactly the dynamical boundary value problem [9] and have extended the analysis to account for an arbitrary interface region between the fiber and matrix [10]. The analysis is exactly analogous to that for the spherical geometry, with the details differing somewhat due to the different general solution of (3) in cylindrical coordinates. Both compressional and shear waves must be accounted for in the scattering problem, and once the exact solution is derived a variety of predictions can be made. Here we concentrate on the dynamic stress concentrations which arise in a fiber-reinforced composite under time-varying loads which set

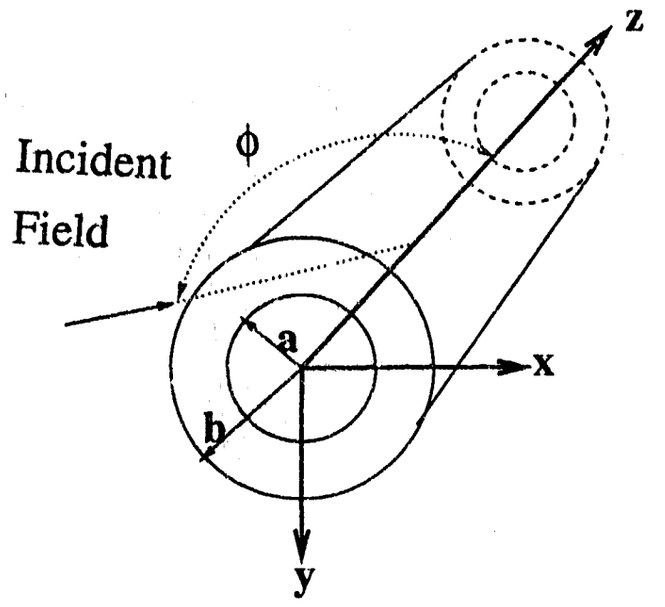


Figure 3: Problem geometry for an infinite cylindrical fiber of radius $r = a$ surrounded by an interface $a < r < b$. The angle ϕ indicates the direction of the incident elastic waves.

up elastic waves in the structure. At the fiber surface we can write the radial dynamic stress concentrations for incident longitudinal waves as:

$$\begin{aligned}
\sigma^*_{rr} &= \frac{\left|\sigma^i_{rr} + \sigma^L_{rr} + \sigma^S_{rr}\right|}{|\sigma^i_{rr}|} \\
\sigma^L_{rr} &= \sum_{n=0}^{\infty} \frac{2\mu_1 i}{k_1 r^2}\left((k_1 r\sin\phi_{1l})\frac{H_n^{(1)\prime}(k_1 r\sin\phi_{1l})}{H_n^{(1)}(k_1 r\sin\phi_{1l})} - n^2 + \frac{(K_1 r)^2}{2} - (k_1 r)^2\cos^2\phi_{1l}\right) \\
&\quad \times \frac{J_n(k_1 b\sin\phi_i)}{H_n^{(1)}(k_1 b\sin\phi_{1l})}\left(\frac{\Delta_1^L}{\Delta_0}\right)\epsilon_n i^n H_n^{(1)}(k_1 r\sin\phi_{1l})e^{in\theta}e^{ik_1 z\cos\phi_{1l}} \\
\sigma^S_{rr} &= \sum_{n=0}^{\infty} -n\frac{2\mu_1}{K_1 r^2}\left(1 - (K_1 r\sin\phi_{1s})\frac{H_n^{(1)\prime}(K_1 r\sin\phi_{1s})}{H_n^{(1)}(K_1 r\sin\phi_{1s})}\right) \\
&\quad \times \left(\frac{K_1}{k_1}\right)\frac{J_n(k_1 b\sin\phi_i)}{H_n^{(1)}(K_1 b\sin\phi_{1s})}\left(\frac{\Delta_2^L}{\Delta_0}\right)\epsilon_n i^n H_n^{(1)}(K_1 r\sin\phi_{1s})e^{in\theta}e^{iK_1 z\cos\phi_{1s}} \quad (5)
\end{aligned}$$

and the shear dynmaic stress concentrations are:

$$\begin{aligned}
\sigma^*_{r\theta} &= \frac{\left|\sigma^L_{r\theta} + \sigma^S_{r\theta}\right|}{|\sigma^i_{r\theta}|} \\
\sigma^L_{r\theta} &= \sum_{n=0}^{\infty} -n\frac{2\mu_1}{k_1 r^2}\left(1 - (k_1 r\sin\phi_{1l})\frac{H_n^{(1)\prime}(k_1 r\sin\phi_{1l})}{H_n^{(1)}(k_1 r\sin\phi_{1l})}\right) \\
&\quad \times \frac{J_n(k_1 b\sin\phi_i)}{H_n^{(1)}(k_1 b\sin\phi_{1l})}\left(\frac{\Delta_1^L}{\Delta_0}\right)\epsilon_n i^n H_n^{(1)}(k_1 r\sin\phi_{1l})e^{in\theta}e^{ik_1 z\cos\phi_{1l}}
\end{aligned}$$

$$\sigma^S_{r\theta} = \sum_{n=0}^{\infty} -\frac{2\mu_1 i}{K_1 r^2}\left((K_1 r\sin\phi_{1s})\frac{H_n^{(1)\prime}(K_1 r\sin\phi_{1s})}{H_n^{(1)}(K_1 r\sin\phi_{1s})} - n^2 + \frac{(K_1 r)^2}{2}\cos^2\phi_{1s}\right)$$
$$\times\left(\frac{K_1}{k_1}\right)\frac{J_n(k_1 b\sin\phi_i)}{H_n^{(1)}(K_1 b\sin\phi_{1s})}\left(\frac{\Delta_2^L}{\Delta_0}\right)\epsilon_n i^n H_n^{(1)}(K_1 r\sin\phi_{1s})e^{in\theta}e^{iK_1 z\cos\phi_{1s}} \quad (6)$$

and finally the hoop dynamic stress concentrations can be written as:

$$\sigma^*_{\theta\theta} = \frac{\left|\sigma^L_{\theta\theta} + \sigma^S_{\theta\theta}\right|}{|\sigma^i_{\theta\theta}|}$$

$$\sigma^L_{\theta\theta} = \sum_{n=0}^{\infty}\frac{2\mu_1}{k_1 r^2}\left(\frac{(K_1 r)^2}{2} - (k_1 r)^2 - n^2 + (k_1 r\sin\phi_{1l})\frac{H_n^{(1)\prime}(k_1 r\sin\phi_{1l})}{H_n^{(1)}(k_1 r\sin\phi_{1l})}\right)$$
$$\times\frac{J_n(k_1 b\sin\phi_i)}{H_n^{(1)}(k_1 b\sin\phi_{1l})}\left(\frac{\Delta_1^L}{\Delta_0}\right)\epsilon_n i^n H_n^{(1)}(k_1 r\sin\phi_{1l})e^{in\theta}e^{ik_1 z\cos\phi_{1l}}$$

$$\sigma^S_{\theta\theta} = \sum_{n=0}^{\infty} in\frac{2\mu_1}{K_1 r^2}\left(\frac{K_1^2}{2k_1^2} - (K_1 r\sin\phi_{1s})\frac{H_n^{(1)\prime}(K_1 r\sin\phi_{1s})}{H_n^{(1)}(K_1 r\sin\phi_{1s})}\right)$$
$$\times\left(\frac{K_1}{k_1}\right)\frac{J_n(k_1 b\sin\phi_i)}{H_n^{(1)}(K_1 b\sin\phi_{1s})}\left(\frac{\Delta_2^L}{\Delta_0}\right)\epsilon_n i^n H_n^{(1)}(K_1 r\sin\phi_{1s})e^{in\theta}e^{iK_1 z\cos\phi_{1s}} \quad (7)$$

where the coefficients Δ_0^L - Δ_2^L are given in [10] for the geometry shown in Figure 3. Note that these normalized stresses are functions of angle, fiber and matrix materials, fiber size, and frequency. Because we have derived exact and analytic expressions for them, we can easily plot them versus any or all of these parameters to explore material behavior. This type of structure is of interest in the chemical vapor deposition of boron fibers, where the core fiber, usually tungsten wire, serves as a substrate for deposition of the shell material. It is the shell of boron that is the actual load bearing part of the composite fiber. In Figure 4 we show, for a boron fiber in epoxy, the dependence of radial, shear, and hoop stresses. The boron is the "shell" region $a < r < b$ surrounding a tungsten "core" of radius $r = a$. The core is small such that $a/b = 0.175$. In each case, we plot the dynamic stress concentrations as a function of $k_1 b$, for the two angles $\theta = \pi$ and $\theta = \frac{\pi}{2}$. These are the directions facing back toward the incident field and facing perpendicular to the incident field. The dynamic stress concentrations are normalized to the incident stress field based on a unit amplitude normally-incident displacement wave of the longitudinal type. The radial dynamic stress concentrations show some definite resonant peaks and nulls. The configuration seems to have a radial stress peak at $k_1 b = 6.6$ in the backward direction and null in the perpendicular direction. This means that at this value of $k_1 b$, the stresses at the fiber-matrix interface are high in the $\theta = \pi$ direction and low in the $\theta = \pi/2$ direction. Since the wave number is proportional to frequency, curves of this type can be used to determine, for a given fiber size, those frequencies which will give rise to potentially damaging stress concentrations. Because we have derived an exact and analytic, closed-form solution we can easily explore all possible configurations to avoid high values of stress concentrations that can locally initiate failure of the fibers or the matrix. As in the spherical geometry, the analysis that we have presented here can be extended to non-circular cylinders, by repeating the analysis in elliptic cylinder coordinates. The resulting expressions allow for modelling of fibers with high eccentricity even to the limit of flat strips.

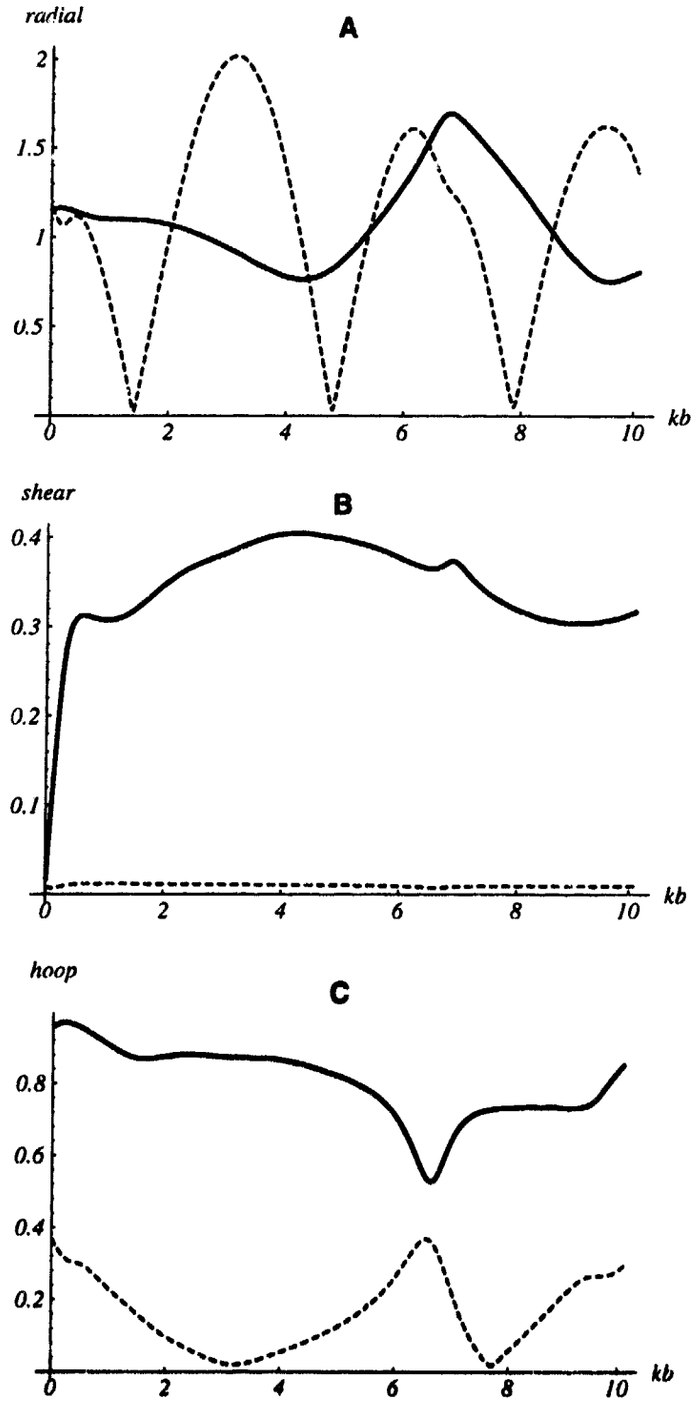


Figure 4: Dynamic stress concentrations for a boron fiber in epoxy. Plot **A** is σ^*_{rr}, plot **B** is $\sigma^*_{r\theta}$, and plot **C** is $\sigma^*_{\theta\theta}$. Solid lines are $\theta = 90°$ and dashed lines are $\theta = 180°$, respectively.

Discussion: In this paper we have outlined a method of analysis which can be used to make predictions about dynamic failure initiation modes in reinforced composites. Travelling mechanical disturbances–elastic waves–scatter from any discontinuity in a material, and this scattering can cause locally high stress levels which may initiate material failure. In composites the reinforcing particles or fibers themselves present discontinuities which scatter the waves, but by modelling those canonical problems where we can derive exact and analytic closed-form solutions we are able to predict the dynamic stress concentrations that will develop between the reinforcement and the matrix under a variety of conditions. With formulæ in hand to describe the complicated dynamical behavior we can explore potential material configurations to avoid those that would be susceptible to failure initiation because of high stress concentrations.

An additional by-product of deriving exact solutions to these scattering problems is the bulk wave properties of composites can be predicted. This is important in ultrasonic nondestructive evaluation of composites, where the heterogeneous composite can be considered as effectively homogeneous, but with scattering that takes place at the fibers or particles represented as an effective loss term. Derivation of the bulk wave properties is done via multiple scattering theory, which hinges on knowing in sufficient detail the scattering behavior for a single isolated particle or fiber.

We have here shown some selected results from our ongoing work, and have given references to more detailed treatments of the background material and solutions. For spherical and cylindrical geometries we have pushed the analysis to its limit and have derived those exact solutions which are tractable. For spheroidal and elliptic cylinder geometries analysis is currently under way to derive the possible exact solutions for elastic wave scattering from solid inclusions. These geometries are of particular interest in modelling the dynamical behavior of reinforced composites, and can also be used directly to model ultrasonic scattering from a variety of defects. Other work in progress is an extension of existing multiple scattering theories to take sufficient advantage of the detailed knowledge of the single scattering problems afforded by our exact solutions.

Acknowledgements: Partial support for this research was provided by the Army Research Laboratory under contract #DAAA15-92-C-0047.

* **Present address:** Program in Applied Science, The College of William and Mary, P.O. Box 8795, Williamsburg, VA 23187.

References:

1. A. Clebsch: *Ueber die Reflexion an einer KugelFläche*, Crelle's Journal **61** 195 (1863).
2. J.W.S. Rayleigh: *On the scattering of light by small particles*, Phil. Mag. **XLI**, 447 (1871).
3. Y. H. Pao and C. C. Mow: Diffraction of Elastic Waves and Dynamic Stress Concentrations (Krane Russak, New York, 1973).
4. V.V. Varadan, A. Lakhtakia and V.K. Varadan, eds.: *Field Representations and Introduction to Scattering* (North-Holland, New York, 1991).
5. M. K. Hinders: *Plane Elastic Wave Scattering from an Elastic Sphere*, Il Nuovo Cimento **106B**, #7, 799, (1991).
6. M. K. Hinders, T. Fang, *Rigorous Solutions for Elastic Wave Scattering*, Fall Meeting of the American Mathematical Society, Philadelphia, (1991).
7. S. D. Bogan, T.-M. Fang and M. K. Hinders: *Spherical Wave Scattering from an Elastic Sphere*, Il Nuovo Cimento B, **107B**, #10, 1215, (1992).
8. M. K. Hinders *et al.*: *Dynamic Stress Concentrations in Particle- and Fiber-Reinforced Composite Materials* in "Topics in Composite Materials and Structures, AMD-vol. 133" V. Birman and A. Nagar, eds. (ASME, New York, 1992).
9. M. K. Hinders: *Plane Elastic Wave Scattering from an Elastic Cylinder*, Il Nuovo Cimento, **108B**,#3, 285, (1993).
10. S. D. Bogan and M. K. Hinders: *Dynamic Stress Concentrations in Fiber-Reinforced Composites with Interface Layers*, J. Composite Materials, **27**, #11, 1272, (1993).
11. S. D. Bogan, T.M. Fang and M. K. Hinders: *Interface Effects in Elastic Wave Scattering* in "20th Reviews of Progress in Quantitative Nondestructive Evaluation" D.O. Thompson and D.E. Chimenti, eds. (Plenum, New York, 1994).
12. S. D. Bogan and M.K. Hinders: *Interface Effects in Elastic Wave Scattering*, Springer-Verlag Lecture Notes in Physics: Monographs, Vol. M19, (1994).
13. G. Mie: *Beiträge zur Optik trüber Medien* Ann. Phys. **25**, 377 (1908).
14. G. Johnson and R. Truell: *Numerical Computations of Elastic Scattering Cross Sections*, J. Appl. Phys. **36**, #11 (1965).
15. C.F. Ying and R. Truell: *Scattering of a Plane Longitudinal Wave by a Spherical Obstacle in an Isotropically Elastic Solid*, J. Appl. Phys. **27**, 1087 (1956).
16. N.G. Einspruch, E.J. Witterholt and R. Truell: *Scattering of a Plane Transverse Wave by a Spherical Obstacle in an Elastic Medium*, J. Appl. Phys. **31**, 806 (1960).
17. L.W. Anson and R.C. Chivers: *Ultrasonic velocity in suspensions of solids in solids–a comparison of theory and experiment*, J. Phys. D: Appl. Phys. **26**, 1566 (1993).

X-RAY DIFFRACTION CHARACTERIZATION OF PROCESS-INDUCED RESIDUAL STRESS

Daniel J. Snoha
U.S. Army Research Laboratory
Materials Directorate
Watertown, MA 02172-0001

Abstract: The U.S. Army Research Laboratory - Materials Directorate (ARL-MD) has utilized the X-ray diffraction (XRD) method of residual stress analysis (RSA) to characterize process-induced residual stress on a variety of polycrystalline metal and ceramic materials. As part of the mechanical failure investigation, modern XRD RSA techiques provide a direct means for quantifying residual stress at the component surface - the location at which most fatigue and stress corrosion cracks originate. Therefore, an understanding of the magnitude and distribution of residual stresses introduced from processing is important when predicting failure modes through fracture mechanics calculations and service loads by finite element modeling. This paper will discuss the procedures for and results from XRD residual stress measurement on the following differently-processed material systems: shot-peened stainless steel, quenched and tempered and welded armor steel, autofrettage gun tube steel, and ground alumina ceramic.

Key Words: Alumina ceramic; armor steel; autofrettage; grinding; mechanical failure; polycrystalline materials; process; residual stress; shot peen; welding; X-ray diffraction

Introduction: The non-destructive, non-contact X-ray RSA technique has found widespread application [1-4] and is generally accepted as being the most accurate of the experimental methodologies for residual stress determination. XRD residual stress measurement is based upon the fact that strain induced in a crystalline material as a consequence of mechanical deformation, phase transformation, thermal expansion, etc. causes a change in the spacing of the atomic planes within the crystal structure from that in the stress-free condition. This change in interatomic, or d-, spacing is evidenced as a shift in the diffracted X-ray peak position. By resolving the angular peak shift and applying the Bragg law $n\lambda = 2d \sin\theta$ (the relation that describes X-ray diffraction) to quantify the d-spacing, the stress on the surface of the specimen can be calculated via linear elastic theory. Assuming that plane stress conditions exist on the surface; i.e., a biaxial system, the relationship of interatomic strain to stress is given by:

$$\varepsilon_{\phi\psi} = [(1 + \nu)/E]\sigma_\phi \sin^2\psi - \nu/E(\sigma_1 + \sigma_2) \qquad (1)$$

where

$\varepsilon_{\phi\psi} = (d_{\phi\psi} - d_0)/d_0$ = strain in the direction defined by angles ϕ and ψ (d_0 is the interatomic spacing in the stress-free condition)

E, ν = material elastic constants
σ_ϕ = surface stress in the direction defined by angle ϕ
ψ = angle between the surface normal and the normal to the crystallographic planes from which an X-ray peak is diffracted
σ_1, σ_2 = principal stresses on the surface.

This equation is used to calculate the stress σ_ϕ in any direction on the surface of the specimen.

Residual stress determined from diffracted X-ray peaks represent strain averaged over a finite measurement volume comprised of the surface area irradiated by the X-ray beam and the depth to which it penetrates (typically, only a few tens of microns). When mechanical deformation processes such as grinding and shot peening produce uniform and continuous plastic strain in the materials surface layers different from that in the bulk, the resultant residual stress is referred to as a macrostress. Plastic deformation nonuniformly distributed from grain to grain in a single-phase material, or between matrix and precipitates with dissimilar yield points in a multiphase system, cause microresidual stresses to form. Macrostress is detected by a shift in X-ray peak position; microstress is observed through peak broadening, and can be described relatively by full-width half-maximum values. The residual stress magnitudes reported herein were generated by the measurement of macrostrain at the as-processed surface except for the armor steel thru-thickness data which were obtained from characterization of sub-surface layers subsequent to electrolytic material removal.

ARL-MD has an advanced X-ray diffraction-based instrument capable of rapid, precise residual stress measurements on polycrystalline metals and ceramics. The Technology for Energy Corporation Model 1610 X-ray Stress Analysis System features a low power X-ray source (100 watts, maximum) and fixed location linear position-sensitive proportional counter (PSPC), and employs the $\sin^2\psi$ stress-measuring technique. The PSPC is a sealed gas detector with a 50 mm (2.0") long carbon-coated quartz fiber wire anode for peak position encoding. At a diffractometer radius (specimen-to-detector distance) of 208.7 mm (8.2"), the PSPC subtends approximately 12° of the instrument's diffraction angle 2θ range of 122° to 166°. Diffraction peaks in the high-back reflection region (2θ values approaching 180°) are much preferred because they show the greatest angular shift sensitivity with a given amount of stress. The Model 1610 allows for the utilization of up to 10 ψ angles per stress measurement.

The $\sin^2\psi$ technique requires a series of peak position measurements for a particular set of hkl planes be made at different tilts (ψ angles) of the crystallographic plane normal referenced to the normal of the specimen surface. The angular position of the diffracted peak is determined by least squares parabolic curve fitting and used to calculate d-spacing from the Bragg relation. A plot is then constructed of d-spacing versus $\sin^2\psi$, and the slope of a least squares line fitted to the experimental data multiplied by the X-ray

elastic constant $E/(1 + \nu)$ is proportional to the stress on the plane of the surface. The slope is found by differentiating Equation 1 with respect to $\sin^2\psi$:

$$\text{slope} = [\partial\ (d\phi\psi - d_0)/d_0]/[\partial \sin^2\psi] \qquad (2)$$

A linear d-spacing versus $\sin^2\psi$ plot indicates that the strain distribution is homogeneous within the irradiated volume and that the assumption of a biaxial stress state is valid. Sin-square-psi plots that split into two branches (ψ-splitting) or exhibit curvature reveal a three-dimensional stress field containing pseudo-macro components of stress (average microstress within a sampled volume of grains). Additional information on the theory of X-ray diffraction and the principles of XRD residual stress measurement is available from references 5 thru 8.

Experimental Procedures: All residual stress data were collected via a divergent beam, four-positive ψ angle arrangement. Other pertinent material- /process-dependent acquisition and calculation parameters are listed in Table 1. The characteristic radiation(s) and lattice planes selected for these investigations are typical for high-back reflection region residual stress characterization. The X-ray elastic constants used in calculating residual stress were bulk values taken from handbooks or literature. Bulk constants represent average elastic properties for all crystallographic directions and may be different from those experimentally determined for a particular set of planes.

Table 1. RESIDUAL STRESS ACQUISITION AND CALCULATION PARAMETERS

	AISI 15-5PH Stainless Steel	MIL-A-46100C High Hardness Armor Steel	AISI 4340 Gun Tube Steel	99.5% Alumina Ceramic
X-Radiation:	$CrK\alpha_1$	$CrK\alpha_1$	$CrK\alpha_1$	$CrK\alpha_1$, $CuK\alpha_1$
Lattice Planes:	(211)	(211)	(211)	(1 0 10), (146)
Diffraction Angle:	156.1°	156.1°	156.1°	135.0°, 136.2°
Irradiated Area:	Rectangular 1 mm x 5 mm	Rectangular 1 mm x 5 mm	Circular 2 mm dia	Rectangular 1.5 mm x 5 mm
Elastic Constant:	180 GPa	169 GPa	154 GPa	246 GPa, 289 GPa

A brief description of each material system and the XRD residual stress analysis application objective follows:

AISI 15-5PH Stainless Steel Tail Rotor Yoke: Taken from a crashed Army attack helicopter and exhibting no physical damage, this component was analyzed for resolving the effects of in-service loading on the magnitude and uniformity of surface residual stresses introduced from shot peening. The tail rotor yoke was sent to ARL-MD from the Corpus Christi Army Depot (CCAD), Corpus Christi, TX. CCAD personnel prepared the yoke for residual stress characterization by plastic bead shot blasting the painted surface to remove the acrylic lacquer and polyamide epoxy primer, then stripping the cadmium plating with ammonium nitrate to expose the martensitic-structured base metal. Measurements were performed in the longitudinal and transverse directions at six locations; four on the nameplate face and two on the opposite face. Figure 1 is a photograph of the tail rotor yoke showing the nameplate face stress measurement locations.

MIL-A-46100C Armor Steel Plate: Residual stresses in light gauge, high hardness (477-534 BHN), high strength (197-209 ksi 0.2% YS, 244-266 ksi UTS) armor steel from quenching and tempering, cutting, grinding and welding have been characterized for comparison to processing parameters as part of ARL-MD's research program. Surface measurements were made on as-produced quenched and tempered plate specimens for establishing baseline residual stress processing data. The measurements concentrated on the stress profiles generated at and near the free-cut edges from underwater plasma arc cutting and edge grinding operations. Magnetic particle inspection confirmed that the cutting and grinding did not create edge discontinuities. Plate weldments were then fabricated per MIL-STD-1941 for the purpose of producing residual stress magnitudes and distributions on a laboratory specimen similar to those found on welded structures. Incremental sub-surface residual stress analysis quantifying interior stress levels was performed on the as-produced plates and the experimental weldment subsequent to electropolishing with a 50-25-25 volumetric solution of phosphoric acid, sulfuric acid and water, respectively.

AISI 4340 Steel Gun Tube Disk: Benet Laboratories, Watervliet, NY requested ARL-MD measure autofrettage-produced residual stresses in a one-inch thick disk specimen sectioned from a 120 mm gun tube for comparison to theoretical stress distribution models. In autofrettage processing a hollow cylinder is deformed into the plastic region by the application of internal pressure causing permanent bore expansion. The resulting beneficial residual stresses increase the elastic strength of the cylinder and retard fatigue crack growth at the bore. Hoop direction measurements were made at one orientation every one-tenth inch along a thru-wall traverse (i.e., I.D. to O.D.) on the electropolished breech face.

Alumina Ceramic Block Specimens: The mechanical properties and overall performance of structural ceramic materials can be influenced by the magnitude, distribution and depth of residual stresses effected by surface finishing techniques. With the objective of using residual stress data as a method for evaluating and optimizing the grinding process, the Department of Industrial Engineering, Lehigh University, Bethlehem, PA submitted 16 differently-ground dual channel 99.5% alumina ceramic block specimens to ARL-MD for residual stress testing services. Table 2 outlines the grinding parameters utilized for preparing the specimens. Longitudinal, or grinding direction, surface

Fig. 1. Army attack helicopter tail rotor yoke with stress measurement locations (1 thru 4) indicated on the nameplate face. Locations 5 and 6 are on the opposite face.

residual stresses were characterized with chromium and copper X-radiations at the center of an arbitrarily chosen channel on the nominal 102 mm x 25 mm x 19 mm block specimens.

Table 2. GRINDING PARAMETERS FOR 99.5% ALUMINA CERAMIC BLOCK SPECIMENS

Specimen	Wheel Bond	Mesh Size	Grit Concentration	Wheel Feed (m/s)	Cut Depth (mm)
1	Resin	80	50	2.9	1.5
2	Resin	80	50	6.9	2.6
3	Resin	80	100	2.9	2.6
4	Resin	80	100	6.9	1.5
5	Resin	180	50	2.9	2.6
6	Resin	180	50	6.9	1.5
7	Resin	180	100	2.9	1.5
8	Resin	180	100	6.9	2.6
9	Vitrified	80	50	2.9	2.6
A	Vitrified	80	50	6.9	1.5
B	Vitrified	80	100	2.9	1.5
C	Vitrified	80	100	6.9	2.6
D	Vitrified	180	50	2.9	1.5
E	Vitrified	180	50	6.9	2.6
F	Vitrified	180	100	2.9	2.6
G	Vitrified	180	100	6.9	1.5

Grinding Wheel: Diamond, 178 mm diameter, 6.35 mm width, 2400 rpm rotational speed, down cut, 22.3 m/s peripheral velocity

Results and Discussion:

AISI 15-5PH Stainless Steel Tail Rotor Yoke: The results of surface residual stress measurements in terms of location and stress-measuring direction are listed in Table 3. The data indicate that this component also may have been subjected to an unusual service overload condition as was proposed as a possible explanation for cause of failure of another tail rotor yoke [9]. Though unlikely, it is not known if the plastic media shot blast used to remove the paint caused any surface deformation. Additionally, the yoke may have been deformed at the time the aircraft crashed. For these reasons the true shot peening-induced stresses may not have been singularly characterized. However, two observations are noteworthy. First, the residual stresses measured at locations 1, 2 and 5, the reduced area where the failed rotor yoke fractured, are significantly lower in magnitude, especially in the longitudinal direction, than those measured at the other locations. The second observation is the uniformity of the measured stresses in both the longitudinal and transverse directions at locations 3, 4 and 6. These values average -707 MPa (-102.6 ksi) and are in good agreement with other reported shot peening stresses [10].

Table 3. RESIDUAL STRESS MEASUREMENT RESULTS FROM SHOT-PEENED ARMY ATTACK HELICOPTER TAIL ROTOR YOKE

Location	Residual Stress, MPa (ksi) Longitudinal Direction	Transverse Direction
1	-223 (-32.4)	-447 (-64.8)
2	-298 (-43.2)	-445 (-64.5)
3	-778 (-112.9)	-694 (-100.6)
4	-674 (-97.8)	-761 (-110.4)
5	-168 (-24.3)	-444 (-64.4)
6	-622 (-90.2)	-714 (-103.5)

MIL-A-46100C Armor Steel Plate: The residual stress data from an as-produced (spray-water roller quenched, tempered at 400°F/~50 m) armor plate are presented in Figure 2. Compressive stresses were measured at all surface locations along a traverse starting at the plasma cut and ground edge. However, at and near the edge the stresses were less compressive in magnitude than those remote to the edge indicating that the cutting and grinding processes altered the as-produced residual stress state. Sub-surface residual stress profiles were compressive to a depth of roughly 0.010" then became tensive and remained so for the balance of the total 0.023" depth thru-thickness characterization. Microstructural examination of the as-produced plates had revealed the existence of an approximately 5 mil-thick decarburized surface layer. Consistent with reported effects of decarburization [11], the measured residual compressive stresses decreased toward the surface. Figure 3 displays residual stresses measured in the transverse direction (perpendicular to the weld line) at the surface of the butt-welded plate specimen along with those

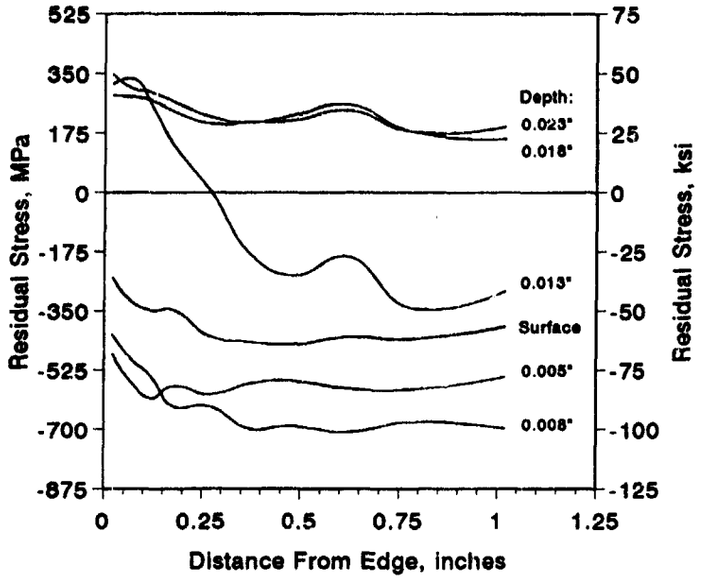


Fig. 2. Surface and thru-thickness residual stress on as-produced high hardness armor steel plate.

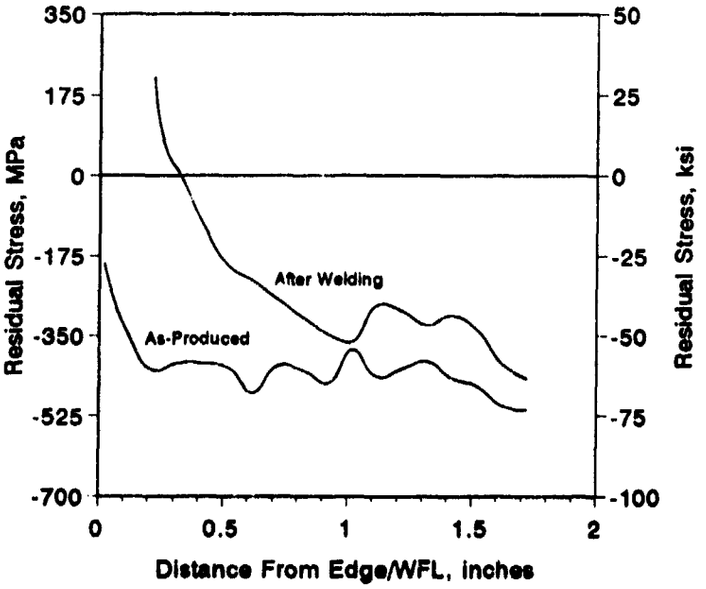


Fig. 3. As-produced and after welding surface residual stress on high hardness armor steel plate.

measured at the same locations prior to welding. The distance offset in the start of the "after welding" trace is equal to half the width of the weld bead. Examination of Figure 3 reveals a steep stress gradient in the heat affected zone (HAZ) with the residual welding stress values and distribution in general agreement with predicted restrained butt-weld stress data [12]. Preliminary results from sub-surface measurements parallel to the weld show tensive stresses in the HAZ approaching 50% of yield strength at 0.005" below the surface. At the 0.018" depth the stress magnitudes increase to 70% of yield.

AISI 4340 Steel Gun Tube Disk: The gun tube disk XRD residual stress investigation is still in progress. However, a graph of percent deviation between predicted (from a two-dimensional, nonlinear elastic-plastic finite element analysis model) and measured residual hoop stresses is shown in Figure 4. Between the 0.14" and 0.94" traverse locations the agreement is excellent; within 4.65%. From 1.04" to 2.04" it is noticed that a slightly increasing percent deviation was obtained indicating a possible relief of residual stress upon cutting the disk from the gun tube. The larger excursions, such as at the 1.04", 1.64" and 2.14" locations, could be attributed to surface preparation irregularities. Whereas, at the I.D. and O.D. (0.04" and 2.24" locations, respectively) the deviations may be due to an error in ψ-position (see Introduction) caused by the "rounding off" of the free edges during electropolishing. These hypotheses are presently being evaluated.

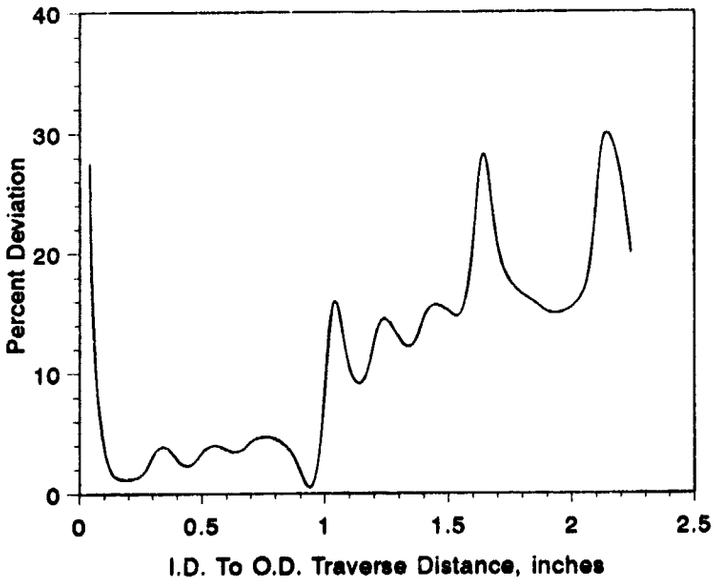


Fig. 4. Percent deviation between gun tube disk predicted and measured residual stress.

Alumina Ceramic Block Specimens: Table 4 lists grinding direction residual stress measurement results. Employing both chromium and copper X-radiations for this study allowed for the determination of residual stress at two surface layer depths. The copper radiation penetrated approximately three times deeper due to the three-fold difference in the linear absorption coefficient of the alumina for these radiations (μ_{Cr} = 794/cm versus μ_{Cu} = 264/cm). Lange et al. [13] have reported the depths of penetration in hot-pressed alumina (1500°C/2 h) for 50% diffracted intensity from chromium and copper radiations as 8 μm and 26 μm, respectively. The variation in the chromium radiation residual stress data, which ranges from -36 MPa to -253 MPa (-5.2 ksi to -36.7 ksi) and averages -138 MPa ± 64 MPa (-20.0 ksi ± 9.3 ksi), suggests that the magnitude of the compressive stress may be a function of the grinding condition. This is not apparent from the copper radiation results as a relatively uniform tensive stress, averaging 240 MPa ± 27 MPa, was measured on all specimens. The crossover from compressive to tensive residual stress occurs at a depth of approximately 5 to 10 microns indicating that the grinding-induced plastic deformation exists in a shallow surface layer.

Table 4. GRINDING DIRECTION RESIDUAL STRESS MEASUREMENT RESULTS FROM 99.5% ALUMINA CERAMIC BLOCK SPECIMENS

Specimen	Residual Stress, MPa (ksi) CrK_α Radiation	CuK_α Radiation
1	-104 (-15.1)	231 (33.5)
2	-253 (-36.7)	187 (27.1)
3	-120 (-17.4)	292 (42.3)
4	-223 (-32.3)	276 (40.1)
5	-70 (-10.1)	261 (37.8)
6	-131 (-19.0)	234 (33.9)
7	-165 (-24.0)	230 (33.4)
8	-94 (-13.7)	214 (31.0)
9	-86 (-12.5)	216 (31.3)
A	-181 (-26.3)	214 (31.1)
B	-191 (-27.7)	260 (37.7)
C	-242 (-35.1)	248 (36.0)
D	-114 (-16.5)	226 (32.8)
E	-109 (-15.8)	237 (34.4)
F	-85 (-12.3)	247 (35.8)
G	-36 (-5.2)	263 (38.2)

Summary: The X-ray diffraction residual stress analysis applications presented herein demonstrate the usefullness and versatility of this technique for characterizing process-induced residual stress in U.S. Army and other material systems.

References:

1. Advances in X-Ray Analysis, Proc. Annual Conf. on Application of X-Ray Analysis., Plenum Press, New York, NY, v.2-, 1958-.
2. Residual Stress and Stress Relaxation, Proc. 28th Sagamore Army Material Research Conf., Plenum Press, New York, NY, 1982.
3. Practical Applications of Residual Stress Technology, Proc. 3rd Intl. Conf., ASM International, Materials Park, OH, 1991.
4. Residual Stresses: Measurement, Calculation, Evaluation, Proc. Conf. on Residual Stresses, DGM Informationsgesellschaft, Verlag, Oberursel, 1991.
5. Klug, H.P. and Alexander, L.E., X-Ray Diffraction Procedures, John Wiley & Sons, New York, NY, 2nd Edition, 1974.
6. Cullity, B.D., Elements of X-Ray Diffraction, Addison-Wesley, Reading, MA, 2nd Edition, 1978.
7. Residual Stress Measurement by X-Ray Diffraction - SAE J784a, Soc. of Auto. Eng., Warrendale, PA, 1971.
8. Noyan, I.C. and Cohen, J.B., Residual Stress: Measurement by Diffraction and Interpretation, Springer-Verlag, New York, NY, 1987.
9. Corpus Christi Army Depot Report No. 88MX022, Corpus Christi, TX, November 1987.
10. Wohlfahrt, H., "Shot Peening and Residual Stresses," Proc. 28th Sagamore Army Material Research Conf., Plenum Press, New York, NY, 1982.
11. SAE J784a, p.29.
12. Masubuchi, K., Analysis of Welded Structures, Pergamon Press, Elmsford, NY, 1st Edition, 1980, p.198.
13. Lange, F.F., James, M.R., and Green, D.J., "Determination of Residual Stresses Caused by Grinding in Polycrystalline Al_2O_3," J. Am. Ceram. Soc., v.66, no.2, 1983, c.16- 17.

SENSORS AND NEURAL NETWORKS

Cochairmen: Henry R. Hegner
ECO, Inc.

David Markley
Radian Corporation

HELICOPTER TRACK AND BALANCE WITH ARTIFICIAL NEURAL NETS [1]

by

Howard J. Taitel, Graduate Research Assistant[2]
Kourosh Danai, Associate Professor[3]
Department of Mechanical Engineering
University of Massachusetts
Amherst, MA 01003
and
David Gauthier, Dynamics Engineer
Sikorsky Aircraft
Stratford, CT 06601

Abstract: Before a helicopter leaves the plant, it needs to be tuned so that its vibrations meet the required specifications. Helicopter *track and balance* is currently performed based on 'sensitivity coefficients' which have been developed statistically after years of production experience. The fundamental problem with using these sensitivity coefficients, however, is that they do not account for the nonlinear coupling between modifications or their effect on high amplitude vibrations. In order to ensure the reliability of these sensitivity coefficients, only a limited number of modifications are simultaneously applied. As such, a number of flights are performed before the aircraft is tuned, resulting in increased production and maintenance cost. In this paper, the application of feedforward neural nets coupled with back-propagation training is demonstrated to learn the nonlinear effect of modifications, so that the appropriate set of modifications can be selected in fewer iterations (flights). The effectiveness of this system of neural nets for *track and balance* is currently being investigated at the Sikorsky production line.

Key Words: Helicopter track and balance; artificial neural networks

Introduction: Helicopter *track and balance* is a tuning procedure for reducing both chassis vibration and the spread of rotor blades about a mean position. *Balance*, which is performed for the reduction of vibration, is the more important of the two since it directly affects the performance of the aircraft. *Track*, on the other hand, is performed mainly for aesthetic purposes as it has been found that well positioned rotor blades increase pilot confidence in the aircraft[4]. Both *track* and

[1] This paper is extracted from (Taitel et al., 1992)

[2] Presently with Mathworks Inc., Natick, Massachusetts

[3] Corresponding author

[4] This research is concentrated mainly on vibrations at the frequency of once per blade revolution (*1 per rev*). Although track does not affect the *1 per rev* vibrations it has the potential to excite other main rotor harmonics such as *4 per rev*.

balance are performed by making modifications to the rotor blades of the aircraft. Thus, the tuning process consists of determining the blade modification, or the set of blade modifications, that will bring the chassis vibration within specification while simultaneously providing suitable rotor track. Since reduction of vibration is the main goal of the track and balance procedure, modifications are generally made in such a way that vibration characteristics are not compromised based on track considerations.

The current method of track and balance assumes that the effect of blade modifications on both track and balance is linear. As such, sensitivity coefficients are utilized to approximate the effect of individual modifications on track and balance. Although this method is generally effective in tuning the aircraft, it suffers from two major limitations. First, since the derived sensitivity coefficients are based on the assumption of linearity, they are valid for only a limited range of vibration and rotor track. Consequently, the effectiveness of the method diminishes for higher ranges of vibration and rotor track, where nonlinear effects may play a role. Second, the coupling effect of modifications (i.e., nonlinear interactions between modifications) is not taken into account by the sensitivity coefficients. As such, only a limited number of modifications are performed for each flight. Both of the above limitations contribute to an increase in the number of flights required to tune each aircraft, and thus result in increased production and maintenance costs.

In this paper, a system of neural networks trained with back-propagation learning is introduced for helicopter track and balance. Neural nets are particularly suited for nonlinear mapping of multiple inputs and multiple outputs. For this problem, a system of networks is designed to both select the set of modifications that will minimize vibration and rotor track, and predict the vibration and rotor track resulting from such modifications. The designed system, which is trained on actual track and balance data, is currently being tested on production aircraft at Sikorsky Aircraft Company.

Track and Balance: Track and balance as applied to Sikorsky's H-60 Black Hawk helicopter is performed as follows. For initial measurements, the aircraft is flown through six different regimes during which measurements of rotor track and vibration are recorded. Rotor track is measured by optical sensors which detect the vertical position of the blades. Vibration is measured at the frequency of once per blade revolution (*1 per rev*) by two accelerometers, 'A' and 'B', attached to the sides of the cockpit (see Figure 1, detail B). The vibration data is vectorially combined into two components: A+B representing the vertical vibration of the aircraft and A-B representing its roll vibration. A sample of peak vibration levels for the six flight regimes, as well as the angular position of a reference blade corresponding to the peak vibration are given in Table 1, along with a sample of track data. The six flight regimes in Table 1 are: ground (*fpm*), hover (*hov*), 80 knots (*80*), 120 knots (*120*), 145 knots (*145*), and maximum horizontal speed (*vh*). The track data indicates the vertical position of each blade relative to a mean position.

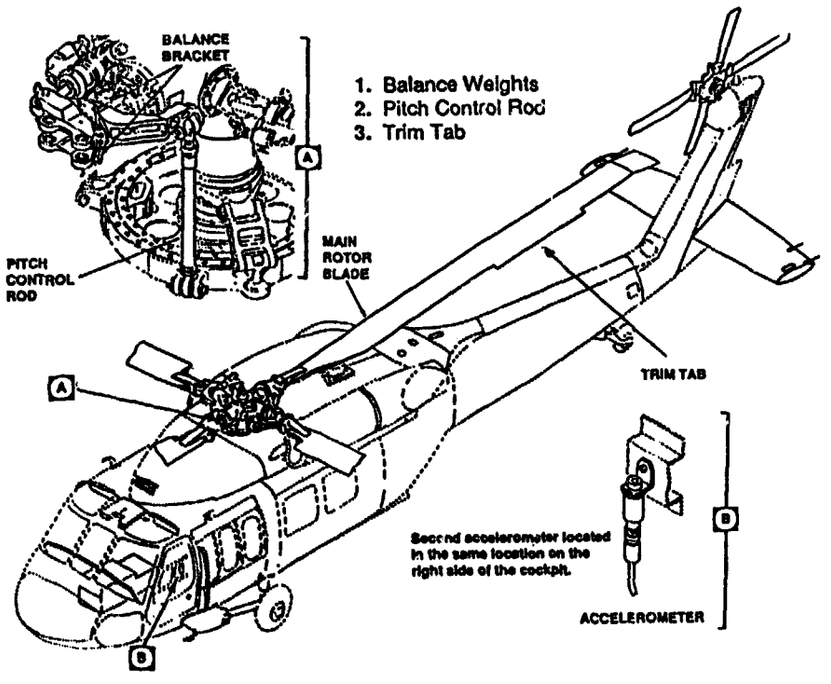


Figure 1: Illustration of the position of accelerometers A and B on the aircraft, and the rotor blade modifications (push rod, tab and hub weights).

In order to bring track and *1 per rev* vibration measurements within specification, three types of modifications can be made to the rotor system: pitch control rod adjustments, trim tab adjustments, and balance weight adjustments (see Figure 1). Pitch control rods can be extended or contracted by a certain number of notches to alter the pitch of the rotor blades. Positive push rod modifications indicate extension. Trim tabs, which are adjustable surfaces on the trailing edge of the rotor blades, affect the aerodynamic pitch moment of the air foils and consequently their vibration characteristics. Tab adjustments are measured in thousandths of an inch, with positive and negative changes representing upward and downward tabbing, respectively. Finally, balance weights can either be added to or removed from the rotor hub to tune vibrations through changes in blade mass. Balance weights are measured in ounces with positive modifications representing the addition of weight. In the case of the Sikorsky H-60 helicopter, which has 4 main rotor blades, a total of twelve modifications can be made to tune the aircraft (i.e., three modifications per blade).

Blade modifications for track and balance can be represented in two forms. The

Flight Regime	Vibration			
	A+B		A-B	
	Mag. (ips)	Phase (deg.)	Mag. (ips)	Phase (deg.)
fpm	0.19	332	0.38	272
hov	0.07	247	0.10	217
80	0.02	86	0.04	236
120	0.04	28	0.04	333
145	0.02	104	0.07	162
vh	0.10	312	0.12	211

Flight Regime	Track (mm)			
	Blade #			
	1	2	3	4
fpm	-2	3	1	-2
hov	-1	3	0	-2
80	1	11	1	-13
120	2	13	-1	-14
145	5	18	-3	-20
vh	2	13	-1	-14

Table 1: Typical track and balance data recorded during a flight.

first form, the *detailed form*, is a list of all twelve blade modifications representing the total track and balance solution. If only vibration (balance) is considered, however, such a detailed specification of the modifications is not necessary. Due to symmetry of the rotor blades in four-bladed aircraft, changes in vibration realized by a modification to a specific blade may alternatively be realized by the negative modification to its opposite blade, or, more generally, by any set of modifications which maintain the relative adjustments to opposite blades. As such, the modifications required for vibration correction can be represented in a *condensed form*, including up to six relative modifications to pairs of opposite blades. The detailed representation, including all of the twelve modifications, is necessary when track is also considered. Two example sets of detailed modifications which have the same condensed representation are presented in Table 2. It can be seen that the condensed modifications are equivalent to the arithmetic difference of detailed modifications on opposite blade pairs 1&3 and 2&4[5].

Neural Net-Based Track and Balance: In practice, track and balance is performed by first specifying a condensed set of modifications to reduce

[5]Condensed modifications on the 1&3 blade pair represent relative modifications of blade 3 with respect to blade 1 and modifications on the 2&4 blade pair represent relative modifications of blade 4 with respect to blade 2.

Type of Modification	Blade Number	*Detailed* Set # 1	*Detailed* Set # 2	Blade Numbers	*Cond.* Set
Push Rod (notches)	1	2	1	Δ13	-2
	2	0	1	Δ24	2
	3	0	-1		
	4	2	3		
Tab (10^{-3} inches)	1	0	0.003	Δ13	-0.003
	2	0.001	0	Δ24	-0.001
	3	-0.003	0		
	4	0	-0.001		
Hub Weight (ounces)	1	5	2	Δ13	-5
	2	3	1	Δ24	-3
	3	0	-3		
	4	0	-2		

Table 2: Two sets of detailed modifications represented by a condensed set of modifications.

vibrations, and then expanding these modifications into a detailed set to satisfy the track requirements. The detailed modifications are constrained by the condensed set so as to maintain the integrity of the vibration solution. This sequential method is in contrast to a simultaneous solution of track and balance, which gives the same priority to both track and balance. In order to preserve the priority of vibration reduction, the proposed neural network-based system of track and balance utilizes the sequential method.

The system of track and balance consists of four neural nets as shown in Fig. 2. The first net in this system, the SELECTION NET, determines the condensed blade modifications (output) that will bring about a given change in vibration (input). To eliminate vibration, the negative of the vibration measurements from the flight are utilized as inputs to this net. The validity of the condensed modifications is then checked by predicting their effect on vibration via the CONDENSED SIMULATION NET. Theoretically, these simulated vibration changes should be the negative of the vibration measurements from the aircraft, so that their summation will generate a resultant vibration equal to zero. However, due to the inexactness of the neural net models and noise, the resultant vibration will most likely not be equal to zero. In cases where the resultant vibration is not within specifications (usually less than 0.20 ips), the condensed modifications may be refined by feeding the resultant vibration back into the SELECTION NET. This feedback is depicted in Fig. 2 by the dashed feedback line. It should be noted that the CONDENSED SIMULATION NET may also serve as a diagnostic tool by indicating out-of-norm behavior. For example, an aircraft with vibrations significantly different from those predicted by the net may suffer from deffective components.

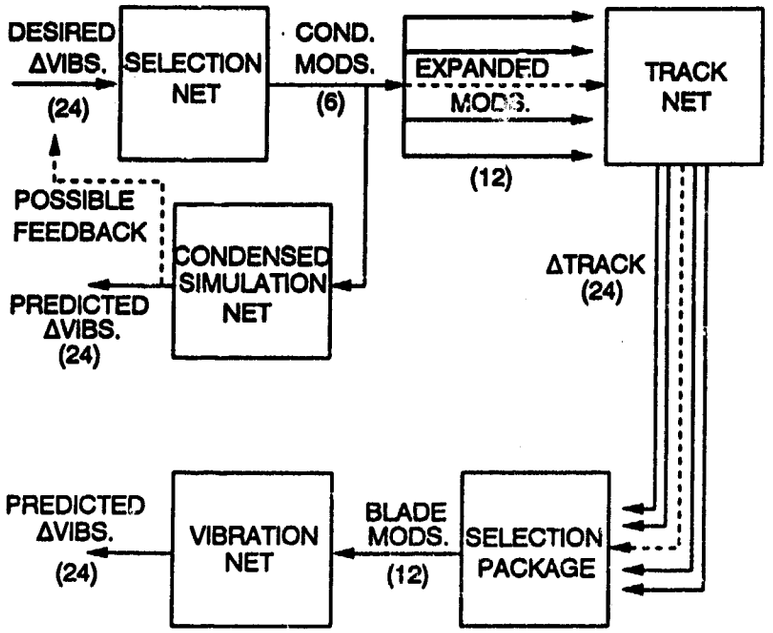


Figure 2: Schematic of the track and balance system. The numbers inside parentheses represent the number of inputs or outputs of individual nets.

Just as with the traditional approach, once the condensed solution has been specified, it needs to be expanded into detailed form to satisfy the rotor track requirements. As previously mentioned, the condensed set of modifications may be viewed as the constraint on detailed modifications so as to ensure that the vibration solution is not compromised for track. This constraint is satisfied by algebraic equations in the expansion program which generates detailed modifications from the condensed set (see Fig. 2):

$$\Delta 3 - \Delta 1 = \Delta 13 \tag{1}$$

$$\Delta 4 - \Delta 2 = \Delta 24 \tag{2}$$

In the above equations, $\Delta 3$, for example, represents the change in push rod, tab, or hub weight of blade #3, and $\Delta 13$ represents the change relative to the change of that parameter in the opposite blade #1. Each one of these detailed sets of modifications is a candidate for the final track and balance solution and it is left to the TRACK NET and the SELECTION PACKAGE to determine which set of detailed modifications provides the best tracking performance. For selection purposes, the TRACK NET simulates the changes in track due to a candidate set of detailed modifications, and then adds these changes to the initial track measurements from the flight to estimate the resultant track. The set of detailed modifications that yields the smallest estimated track (i.e., smallest maximum blade spread) is selected as the solution to the track and balance problem. The selected set of detailed modifications is then checked via the VIBRATION NET, which similar to

the CONDENSED SIMULATION NET, serves as an independent evaluator of the selected modifications. This validation is performed by simulating the effects of the selected modifications and adding them to the aircraft vibrations so as to estimate the resulting aircraft vibrations.

System Training: Track and balance data for approximately one hundred and ten pairs of consecutive flights were used for training and testing of the system. Generally, two types of data were required for training the neural nets. 1) changes in vibration caused by blade modifications and 2) changes in track due to these modifications. Since vibration data are vector quantities (see Table 1), both their magnitude and phase components need to be taken into consideration. In order to keep all of the inputs on the same scale, the vibration vectors were represented to the net in cartesian coordinate form. As a result, the vibration data included 24 elements representing the changes in the x and y components of the A+B and A-B vibration in each of the six flight regimes (see Table 1). To ensure that the nets were properly trained, issues such as noise and generalization needed to be addressed.

Noise: Ideally, identical modifications made to any two helicopters should result in identical changes in vibration. In reality, however, this does not occur due to factors such as small differences between individual aircraft and variances in atmospheric flight conditions (i.e., weather). Prior to neural net training, the severity of the noise problem was investigated by studying the effect of identical modifications on vibration changes (Δvibrations) of different aircraft. One such set of results is shown in Fig. 3 where adjacent pairs of bars represent changes in the x (cos) and y (sin) components of A+B vibration that resulted from an identical set of modifications to two different aircraft. The results indicate that, as expected, large variations exist between the vibrations, and that in some cases the Δvibrations are so inconsistent that they actually differ in direction. This variation is caused by inevitable dissimilarities between various aircraft and rotor blades, and non-uniformity of flight conditions due to variations in weather. (Similar variations are observed in A-B vibration measurements.)

The implication of noise for neural net training is illustrated in Fig. 4 where the target for a set of inputs is not unique, but consists of many targets due to the inconsistency of the data. The fact that one input is mapped to various targets makes learning difficult, and the best that the net can do in such cases is to provide mapping to a target that represents the average of all of the targets.

Generalization and Net Structure: Generalization is defined as the ability of a trained net to correctly classify patterns not included in the training set. Since in general only a finite number of input-output data is available for training, the ability of the net to generalize to patterns it has not 'seen' will ultimately determine its effectiveness. The issue of overfitting is particulary important in the case of helicopter vibration because of the high level of noise in the training data.

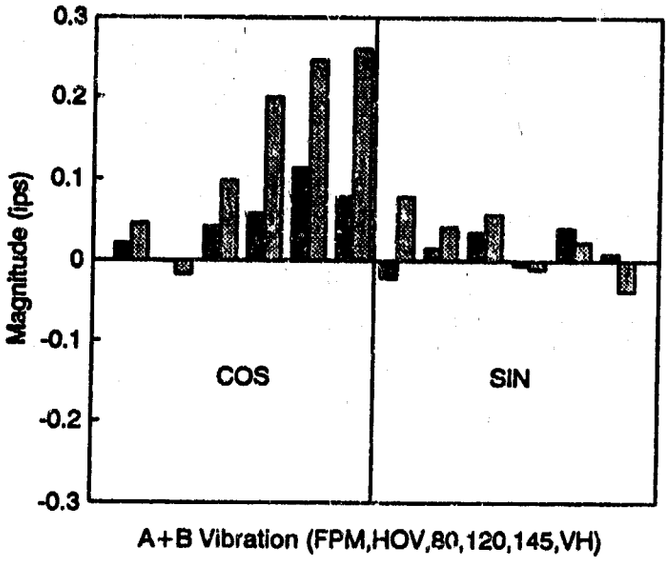


Figure 3: Δvibrations of two aircraft caused by an identical set of modifications.

The complexity of the functions which can be learned by neural net models is directly related to the number of hidden units. Therefore, an important factor in generalization is the number of hidden units of the neural net (Kung and Hu, 1991; Rumelhart et al., 1991; Matsuoka, 1991; Whitley and Karunanithi, 1991). The usual method of trial and error was used to determine the optimal number of hidden units for each net. In these trials, the criteria for 'best network performance' was the sum of the absolute error on a set of ten test patterns (i.e., patterns not used during training). As an example of generalization ability, test results for the CONDENSED SIMULATION NET, in which 14 hidden units were used, are presented in graphical form in Fig. 5. It can be seen that although some of the Δvibrations are accurately predicted by the net, others, the 'FPM' regime for example, are not well predicted. Such absolute comparison, however, is not ap-

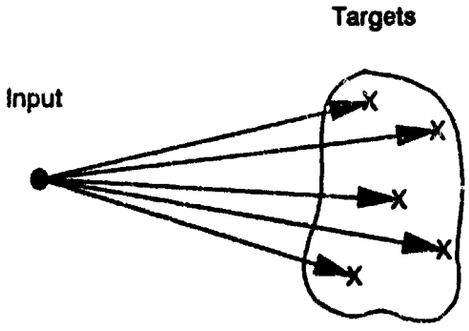


Figure 4: Illustration of one-to-many mapping faced by the CONDENSED SIMULATION NET and VIBRATION NET due to noise.

propriate in view of the large amount of noise present in the data. In order to study these results in the context of the existing noise, predicted Δvibrations from the CONDENSED SIMULATION NET (represented by the 'x') are compared against two target sets of Δvibrations that resulted from the same set of modifications (depicted by adjacent bars) in Fig. 6. The results indicate that the net's output lies within the range of variability of its target values, and given the inconsistancy of the data is in fact providing reasonable generalization. Note that the targets here are from test cases that have not been 'seen' before by the net.

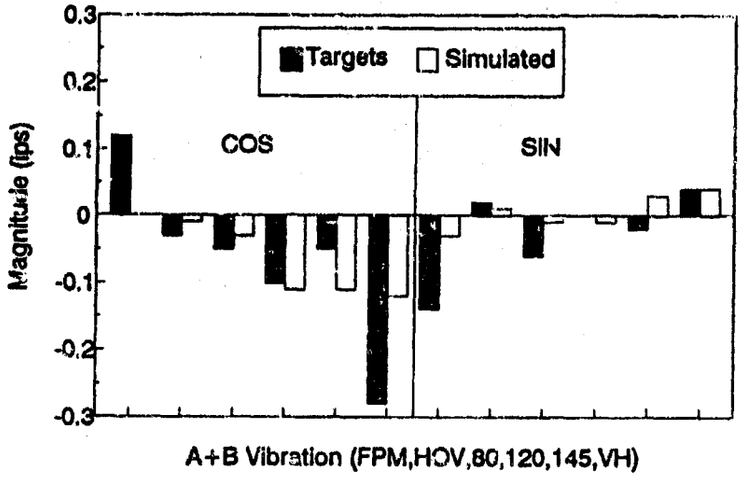


Figure 5: Predicted Δvibrations by the CONDENSED SIMULATION NET compared with their target values.

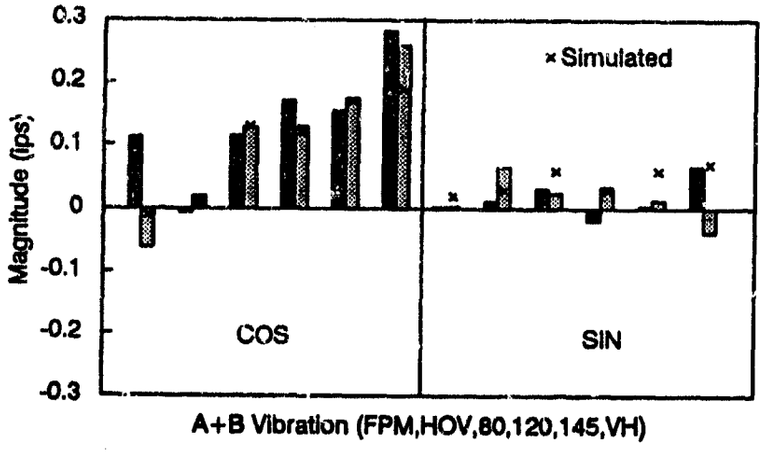


Figure 6: Predicted Δvibrations by the CONDENSED SIMULATION NET (shown by the 'x') compared with their multi-targets (adjacent bars).

Many alternatives to the method of trial and error for determining the number of

hidden units have been proposed. Some notable methods include determining the optimal number of hidden units *a priori* through unsupervised learning (Sanger, 1991), *pruning* procedures (Rumelhart et al., 1991; Kung and Hu, 1991; Hertz et al., 1991), and various cross-validation techniques (Wada and Kawato, 1991; Utans and Moody, 1991). The effect of *pruning* procedures and the cascade correlation algorithm (Fahlman and Labiere, 1990) in predicting Δvibrations by the CONDENSED SIMULATION NET was investigated (Taitel et al., 1992). The results did not indicate much improvement over those obtained by back-propagation. To avoid overtraining, in-training cross-validation was employed, which consists of periodically monitoring the performance of the net on the test set so that training can be stopped once generalization (test set error) begins to deteriorate.

The number of hidden units and learning parameters for the four nets in the Track and Balance system are given in Table 3. Some typical test results for the SELECTION NET are shown in Table 4 indicating that the performance of the net is quite good. Note that the SELECTION NET, contrary to the simulation nets, performs many-to-one mapping of small regions ('clouds') of input space to specific outputs. So, it provides better prediction.

Neural Net	Hidden Units	Learning Rate	Momentum Coefficient	Sigmoid Range
Selection Net	10	0.3	0.6	0 – 1
Cond. Sim. Net	14	0.3	0.6	0 – 1
Track Sim. Net	10	0.3	0.6	0 – 1
Vibration Net	8	0.3	0.6	0 – 1

Table 3: Topology of the nets used in the system of track and balance.

Results: Ideally, the performance of the neural net-based system of track and balance should be evaluated 'side by side' against that of the traditional method. However, such an evaluation would entail flying the aircraft with the modifications from one method, measuring the vibration and track, undoing the changes, and flying the aircraft with the modifications from the other method, so that their vibration and track measurements can be compared. Unfortunately, such an exercise is prohibitively costly and cannot be justified in production environment. A less ideal approach would be to evaluate the performance of the system based on the existing track and balance data, by comparing the system's simulated vibrations and track with those obtained from the modifications of the traditional method. However, that approach would have the problem of being biased towards the neural net-based system by ignoring the inaccuracies of the simulated results. In order to avoid misleading conclusions, the overall performance of the neural net-based system was evaluated independently in production.

Track and balance data from an aircraft before the application of the system are

	Push Rod (notches)		Tab (10^{-3} inches)		Hub Weights (ounces)	
Simulated	0	-2	0	0.005	0	0
Actual	0	-2	0	0.005	0	0
Simulated	0	-2	0	0.004	0	0
Actual	0	-1	0	0.004	0	0
Simulated	-5	5	0.010	-0.010	0	0
Actual	-6	5	0.010	-0.009	0	0
Simulated	0	-4	0.001	0	0	0
Actual	0	-5	0.004	0	0	0

Table 4: Test results comparing the output of the SELECTION NET with its target values.

shown in Table 5 in which the A+B '145' knots and 'VH' vibrations are above the acceptable limit of 0.2 ips. To determine the required blade modifications, the negative of the measured vibrations were specified as the 'desired Δvibration' for the Track and Balance System (see Fig. 2). The modifications determined by the system and subsequently performed on the aircraft are shown in Table 6, and the vibration and track of the aircraft after these modifications are shown in Figs. 7 and 8, respectively. The results indicate that all twelve vibration measurements are within the acceptable range (below the dashed line in Fig. 7), and that the majority of the track measurements have been reduced[6]. The results obtained are indicative of the system's ability to tune the aircraft within one flight, which could be rarely achieved with the traditional method.

Conclusion: A system of feedforward neural nets coupled with back-propagation learning has been designed to deal with noise and nonlinearity of helicopter track and balance. This system uses four neural nets to select blade modifications and to simulate vibration and rotor track resulting from the selected modifications. The track and balance system has shown some promising results and is currently under investigation in production at Sikorsky Aircraft Company.

Acknowledgements: This research was supported in part by the National Science Foundation (Grants # DDM-9015644 and # MSS-9102149).

[6]These results were obtained from one pass through the Selection Net.

Flight Regime	Vibration			
	A+B		A-B	
	Mag. (ips)	Phase (deg.)	Mag. (ips)	Phase (deg.)
fpm	0.06	92	0.15	96
hov	0.05	155	0.03	289
80	0.04	268	0.04	350
120	0.16	346	0.01	344
145	0.35	325	0.06	261
vh	0.54	330	0.16	274

Flight Regime	Track (mm)			
	Blade #			
	1	2	3	4
fpm	-4	6	-7	5
hov	-5	4	-4	5
80	-1	6	-8	4
120	2	11	-17	4
145	5	6	-28	17
vh	12	-2	-33	22

Table 5: Track and balance data recorded during an initial flight.

Push Rod (notches)				Tab (10^{-3} inches)				Hub Weights (ounces)			
0	0	-3	0	0	0	0.008	-0.003	0	0	0	0

Table 6: Modifications selected to zero the vibration and adjust the track of Table 5.

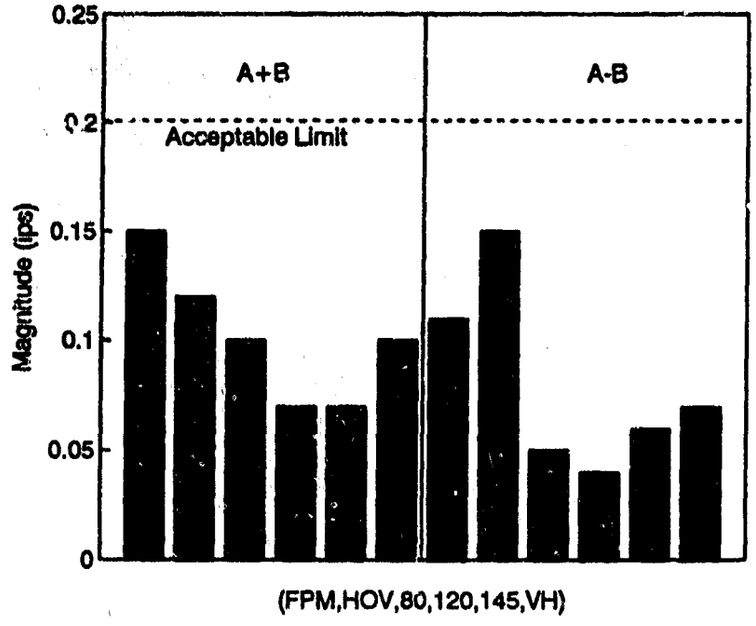


Figure 7: Resultant aircraft vibration (magnitude only) due to modifications in Table 6.

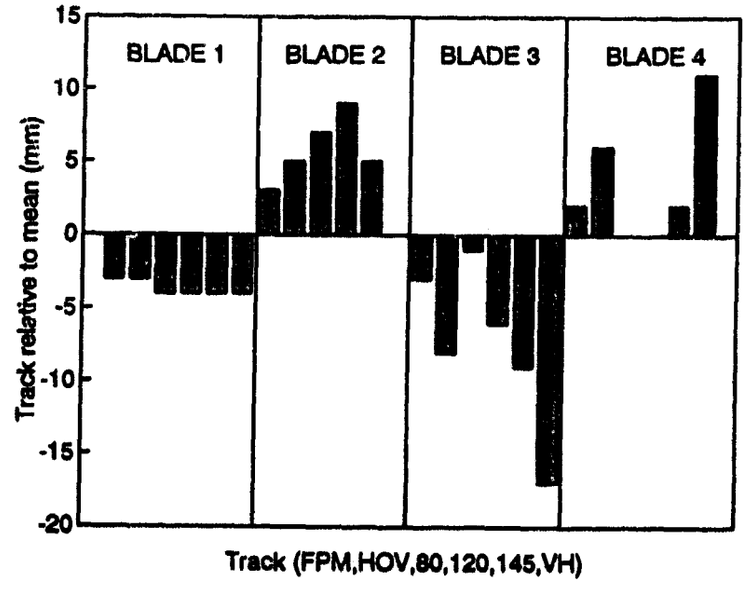


Figure 8: Resultant aircraft track due to modifications in Table 6.

References

Fahlman, S. E., and Lebiere, C., 1990, *The Cascaded-Correlation Learning Architecture*, Technical Report No. CMU-CS-90-100, School of Computer Science, Carnegie Mellon University.

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

Kung, S. Y., and Hu, Y. H., 1991, "A Frobenius Approximation Reduction Method (Farm) for Determining Optimal Number of Hidden Units", *Proc. IJCNN*, IEEE, Seattle, WA, July, pp. 163–168.

Matsuoka, K., 1991, "An Approach to the Generalization Problem in the Backpropagation Method", *Systems and Computers in Japan*, Vol. 22, No. 2, pp. 897–905.

Rumelhart, D. E., Weigend A. S., and Huberman, B. A., 1991, "Generalization by Weight-Elimination applied to Currency Exchange Rate Prediction", *Proc. IJCNN*, IEEE, Seattle, Washington, July, pp. 837–841.

Sanger, T. D., 1991, "Optimal Hidden Units For NonLinear FeedForward Neural Networks", *Pattern Recognition and Artificial Intelligence*, Vol. 5, No. 4, pp. 545–561.

Taitel, H., Danai, K., and Gauthier, D. G., 1992, "Helicopter Track and Balance with Artificial Neural Nets," *Intelligent Control Systems*, ASME DSC-Vol. 45, pp. 47-54.

Utans, J., and Moody, J., 1991, "Selecting Neural Network Architectures via the Prediction Risk: Application to Corporate Bond Rating Prediction", *Proc. First International Conference on Artificial Intelligence Application on Wall Street.*

Wada, Y., and Kawato, M., 1991, "Estimation of Generalization Capability by Combination of New Information Criterion and Cross Validation", *Proc. IJCNN*, IEEE, Seattle, Washington, July, pp. 1–6.

Whitley, D., and Karunanithi, N., 1991, "Generalization in Feed Forward Neural Networks", *Proc. IJCNN*, IEEE, Seattle, Washington, July, pp. 77–82.

PREVENTING MECHANICAL FAILURES IN RESIN TRANSFER MOLDING USING EMBEDDED SENSORS AND NEURAL NETWORKS

Karl Meissner Patrick Sincebaugh

U.S. Army Research Laboratory
Materials Directorate
Watertown, MA 02172-0001
meissner@cns.bu.edu

Abstract: One of the causes of mechanical failures in composite parts manufactured using Resin Transfer Molding (RTM) is the voids or air pockets introduced during the injection stage. This paper discusses an adaptive control technology using multiple injection ports and vents for a more reliable injection process. Reinforcement learning is used to train a Cerebullar Model for Articuled Control (CMAC) type neural network to control multiple injection points.

Key Words: Adaptive Control; CMAC; Embedded Sensors; Reinforcement Learning; Resin Transfer Molding; SMARTWeave

Introduction: The advantages of using composite polymer matrix glass fiber materials instead of traditional materials such as steel and aluminum are well known throughout the materials industry. Composite materials are stronger pe unit weight, and more corrosion resistant than traditional materials. However, the widespread use of composites has been hindered by the lack of a cost effective manufacturing process. Historically, this has limited the extensive use of composites to military and aerospace applications.

Composite parts are currently manufactured using either the Resin Transfer Molding (RTM) or Autoclave Curing process. This paper will focus on a control technology that is being developed to substantially improve the reliability of manufacturing composite parts via the RTM process.

Resin Transfer Molding: The first step in fabricating a composite part via the RTM process is to develop a fiber reinforced preformed mold of the part. The preform is typically produced using fibers such as glass or graphite. Next, the fibrous preform is inserted into a RTM tooling device. This device is a mold, often made of steel or aluminum, that is specifically designed to house the fibrous preform. The mold is designed with a small number of ports and vents. Once the preform has been properly fitted, the mold is sealed shut. A polymeric resin is

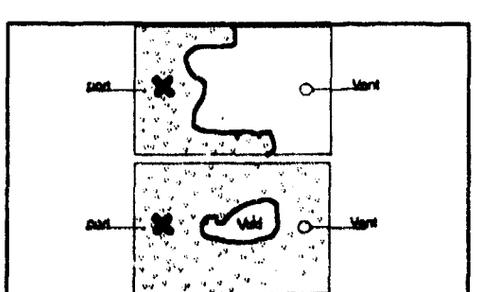


Figure 1: Resin flow at two stages. Void formation due to high permeability edges.

then injected into the mold through the appropriate port(s). The injection of the resin is controlled by equipment that provides low pressure to the designated ports. As resin is being injected through the ports, the appropriate vents are opened. This forces the resin to flow through the preform, and out through the vent. The proper opening and closing of the ports and vents is critical in order to produce a defect-free part. Figure 1 depicts flat plate filling with resin and forming a void.

Once the preform is completely impregnated (wetted) with resin, heat is applied to the to the mold, changing it from liquid to solid. Once the part is cured, it is removed from the mold and is ready for machining.

The wide spread use of the RTM process is currently inhibited by the high scrap rate, and therefore high cost, associated with the process. This can be attributed to the lack of control that is currently available with the RTM process. The proper manipulation of the ports and vents is critical in order to produce a defect free part. Unfortunately, the injection processes is not well understood. Therefore, the fabrication of a composite part using the RTM process is highly dependent upon the experience and skills of the individual operators. [12]

SMART Weave Sensory System: In order to control the RTM process, a sensor system must monitor the flow of resin as it impregnates the preform. The U.S. Army Research Laboratory, Materials Directorate, has completed the first step by developing an economically viable technology to monitor the resin flow. The resin flow monitoring system, known as the SMART (Sensors Mounted As Roving Threads) Weave system, is a sensory grid comprised of an electrically conductive material (ie. graphite fiber). The conductive fibers are embedded within the composite preform by weaving them into the preform cloth during fabrication. The cloth is layered so that a gap exists between the conductive fibers which are aligned in the X and Y directions forming a grid. As the resin flows past a grid junction, it completes a circuit. The position of the resin flow throughout the preform can be obtained by utilizing a multiplexing switching system to acquire data from each junction.

Multiple Point Control: When a new mold is designed the ports and vents are positioned for the most reliable fill rate in a costly trial and error process. Finding the right port positions can take a long time because each part may take several hours or days to complete. It is expensive because a part can cost several hundred dollars in labor and materials.

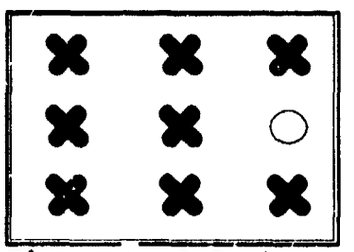


Figure 1: 8 Ports 1 Vent

This problem can be bypassed by using a large number of ports and vents which cover the entire mold. This does two important things. First, the port placement stage can be largely skipped, second, a much higher degree of control over the local flow can be achieved. Additionally, this allows a higher injection rate and shortened fill time.

However, using many ports and vents introduces several new problems. Multiple injection points may cause converging flow fronts which trap air and is likely to cause voids. With tens or perhaps hundreds of ports and vents to control, an operator may have a difficult time keeping up with the changing state of the mold. Ports and vents may need to be turned rapidly on and off at several widespread points. The operator may not know what to do since this technique has not been widely used and RTM dynamics are not well understood.

An automatic and reliable injection controller is needed for a multiple port architecture. The nature of the problem, with tens or hundreds of inputs, tens of control variables and poorly understood dynamics preclude traditional control techniques such as dynamic programming. [2] Using SMARTWeave to sense the position of the resin within the mold and multiple ports to effect the flow path, reinforcement learning enables reliable filling of the mold.

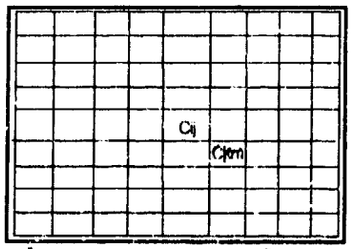


Figure 3: 9x9 Plate

Porous Media Simulation: A simple porous media simulation(PMS) based on Darcy's law[5,6] was written to provide a training set for the learning algorithm.

The simulated mold was a 2 dimensional plate represented by a 9 by 9 grid of cells. The amount of resin at each cell C_{ij} is M_{ij}. At each time step, velocity is calculated based on Darcy's law $V_{ij,km} = u_{km}\ \Delta P_{ij,km}\ f(\Delta P_{ij,km})$ where $V_{ij,km}$ is the velocity of the resin between C_{ij} and C_{km}; u_{km} is the permeability of C_{km}; $\Delta P_{ij,km} = M_{ij} - M_{km}$ is the local pressure gradient between C_{ij} and C_{km}; and $f(P_{ij,km})$ is a sigmoid function to account for the penetration effect. The function f was $1/(1 + e^{((3-3*u_{km}) - (M_{ij}-M_{km}))})$ to reflect the difficulty in entering a cell at low pressure differences. Note that C_{km} is one of the eight neighbor cells of C_{ij}. The flow of the resin is based on the following rule.

If ($M_{ij} > .5$) and ($V_{ij,km} > 0$) then $\Delta M = (M_{ij}/8)V_{ij,km}$. This created realistic flow effects based on local interactions.

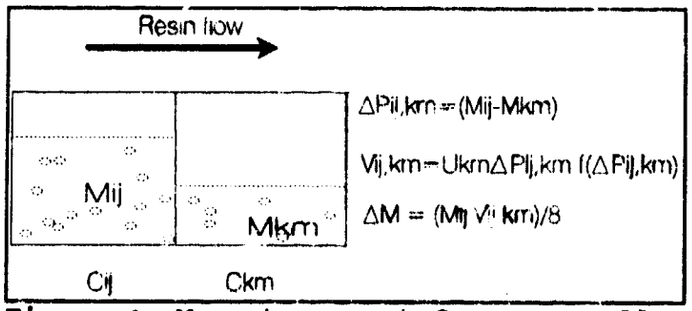


Figure 4: Mass transport from one cell to another

The fiber preform is one of the main sources of variability within the mold. Permeability in a preform can only be inferred by the behavior of the flow. It acts as a hidden variable as the system transitions between states. The permeability around the edges is often 10 times the permeability of central portion of the preform[12]. Variations of the fiber preforms were encoded by varying the value of u_{ij} for each cell. A database of randomly generated permeabilities was used to train the controller. The permeability of cells on the edge of the mold were randomly distributed between 0.85 and 0.95. The permeability of cells on the interior of the mold was uniformly distributed between 0.05 and 0.15. This created edge effects. Resin tends to take the path of least resistance and rush around the edges leaving the interior dry.

The simulation was run until the mold was filled, a void was formed or time ran out. A void

was present if an empty region of cells was cut off from the vents by the resin.

Many real world RTM dynamics such as momentum, temperature and curing were ignored for the sake of simplicity and speed. This simulation captures some important aspects of void formation and demonstrates that reinforcement learning can reliably fill an RTM mold. One of the goals of future work will be to use a more realistic simulation to train the neural network.

Reinforcement Learning: Hard control problems such as autonomous vehicle navigation [8], computer network routing[9], and control of an arc furnace[10] have been implemented using Reinforcement Learning. This technique offers the capability to learn to predict and control tasks which extend over time based on a set of goal states and a function which determines how good or bad a particular goal state is.

The prototypical temporal difference problem consists of a series of transitions between states. For example, as the resin impregnates the mold, more and more cells have resin in them. Every time the amount of resin is changed, a new state is entered. Effectors, the ports or vents, are turned on or off based on the estimated utility of taking various actions. Eventually a terminal state is reached and the system receives a payoff. The molded part is either defective or sound. The system must predict the likelihood of the payoff given a particular state-action in the mold.

Given a series of states, the neural net must solve the temporal credit assignment problem to predict the eventual outcome, z. The temporal credit assignment problem is the question "Which state-action pair was most responsible for the outcome? " At each time step the net has access to the current prediction and the previous one. Learning is based on the difference of successive predictions. If there is a large difference, a large amount of learning takes place. If the expectations of the system are met, the predictions are close and very little learning takes place. At each state S_1, S_2...S_n the network makes a prediction of the utility O_1, O_2...O_n. Learning at each time step is proportional to the difference $O_{k+1} - O_k$ weighted by the reinforcement function. Over time the net converges to good estimates of the utility.

This pseudo-algorithm outlines training regime

1) Initialize the simulation with a randomly chosen preform
2) Sample the state of the mold and present the input to the net
3) Calculate the output of the net
4) Turn the effectors on and off based on output and the exploration algorithm
5) Update the PMS mold one time step.
6) if (the mold is not finished)
   goto 2
   else
   7) reinforcement = : the part is sound 1; the part is defective 0
   8) Train MPC based on the difference of successive estimations
   9) if (the net has not converged)
      start a new trial - goto 1

Multiple Point Controller (MPC): When building a learning system to solve a practical task, problem dependent knowledge is built into the system. This restricts the generality of the learning algorithm but is often needed to make the problem soluble.

The MPC needs to solve two difficult problems. First it needs to find a robust enough strategy to be able to account for random variations in permeability which affect the flow. Second, the MPC must coordinate multiple ports and vents to prevent converging flow fronts and voids.

The neural net could be trained by hooking it directly to a mold and allowing it to train on a real mold. However, real resin is slow, toxic, hard to clean up, and expensive so the MPC was trained with the PMS computer program which can fill a simulated mold in a few seconds. A trained neural network could then control actual hardware. Figure 5 shows one possible hardware configuration.

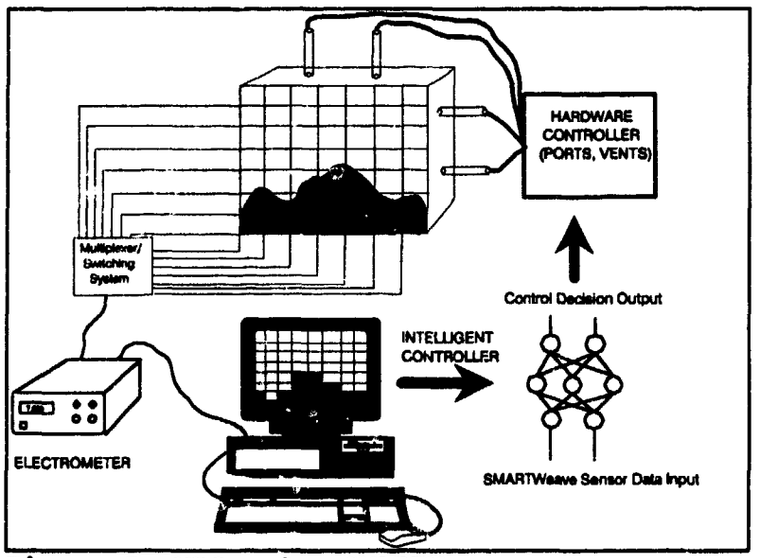


Figure 5: MPC and a SMARTWeave sensor

The inputs for the MPC are restricted to the information available to a system using SMARTWeave. MPC only knows whether or not there is resin at each cell and the state of each effector. An effector refers to either a port or a vent. The effectors can be either on or off. This gives 81 inputs and 9 effectors for the 9x9 plate simulation with PMS.

The MPC uses a CMAC type neural network. A good reference on CMAC is [13] by Miller using CMAC for real-time control in robotics. This neural acts much like a lookup table with generalization. The input space is tesselated into boxes. Given an input, the output is the entry in the corresponding boxes. Generalization allows neighbor boxes to be trained and to affect the output.

The MPC consists of 9 pairs of CMAC, one pair per effector. Each pair of CMAC represent the predicted utility of turning the effector either on or off.

Each CMAC has two inputs. The first input ranging 0 to 19, is a local measurement of how much the resin has covered the area around the effector. This tells the CMAC if it is

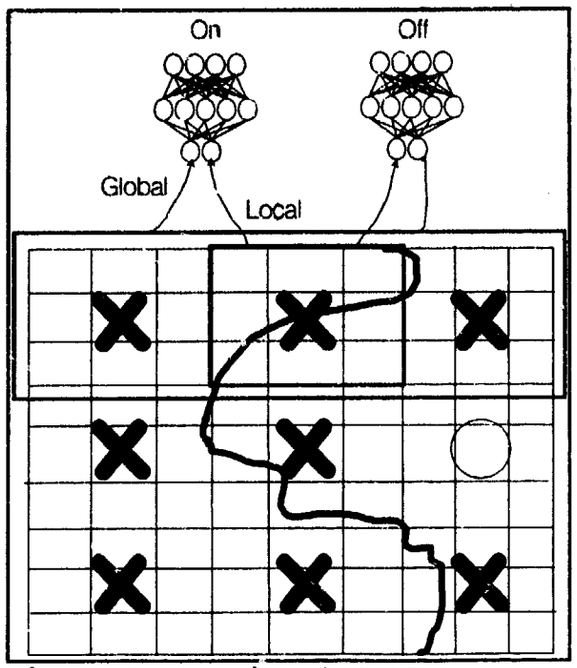


Figure 6: A pair of CMAC connected to one effector. The 3x3 and 3x9 input fields are shown.

covered with resin. Each cell containing resin in the 3x3 area around the effector is counted. If the central cell with the valve in it is covered, then 10 is added to the count.

The second input is a global measure of the flow front. The vent is on the right in PMS so the resin should flow from left to right. The second input ranges from 0 to 9 and is a count of how far to the right the resin has flowed in the 3x9 row that the effector is in vs. the rest of the resin. This encodes information about the way the resin is supposed to flow. It reduces the generality of PMS because it is specific to this type of mold.

The output of each pair of CMAC is an estimate of the utility for turning the valve on or off at each of the nine respective control points. If MPC is trained, then the control strategy is simply to take the action at each effector with the maximum predicted reinforcement.

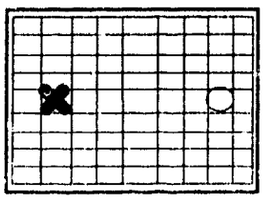
However, initially the MPC is untrained and the estimate is very poor. During training it is necessary to occasionally ignore the estimate and explore the effects of novel control actions. When the output of the on/off pair was close, the network was judged to be uncertain in its preference and one of the actions was chosen at random.

Once an action was made, that effector continued that action until the inputs transition to a new state. Then the CMAC estimated the utility for turning that effector on or off.

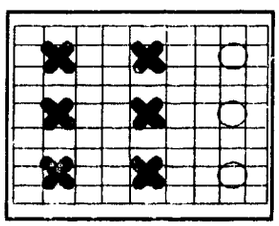
There was a small but finite chance that all the ports would turn off. If this happened the central left port was turned on.

The reinforcement function was very simple for this task. If the mold was completely full then R = 1 otherwise it was 0. This punished the formation of voids or very slow injection strategies.

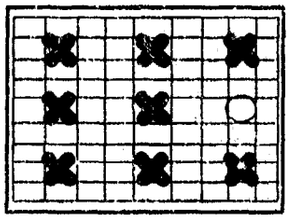
Results: There were three different sets of experiments done.



Simulation 1: 1 Port 1 Vent. The port was located at cell 1,4. The vent was located at 8,4. Several different hardwired control strategies were tried to fill this mold as well as training with the neural network. They all failed. Voids always formed in the low permeability interior while the resin flowed around the edges. There were insufficient ports to control the local flow behavior.



Simulation 2: 6 Ports 3 Vents. The ports were at (1,1), (1,4), (1,7), (4,1), (4,4), (4,7) and the vents were at (8,1) (8,4), (8,7). This mold could usually be filled by turning on all the ports. The resin filled the left half and the flowed right. The two corner vents prevented the resin from sweeping around the edge and forming a void. MPC would converge to a control strategy which could reliably fill the mold after an average of only 62.3 training trials over 10 runs. The time to fill the mold by turning on all the ports was 188.8 time steps while the trained controller took about 198.3 time steps on average.



Simulation 3: 8 Ports and 1 Vent. The two corner vents of the previous configuration were now ports. The resin tended to rush around the edges unopposed and form dry spots in the center. A simple strategy of just turning on all the ports would not work in this case and never filled the mold. MPC was able to converge to a strategy which would reliably fill the mold after an average of 1227.2 training trials over 10 runs. The time to fill the mold was on average 178.2 time steps. Qualitatively it seemed to converge to the same stable strategy each time. It would fill the corners quickly but then turn off the corner ports. It would then carefully fill in the central portion moving left to right, finally turning on every thing when there was a contained bubble around the vent. Figure 10 shows MPC converging after 680 trials. The histogram shows how many trials ended in failure per block of twenty trials. For a long period of time the controller always fails. Around trial 650 it stabilizes a strategy that works and quickly stabilizes to it.

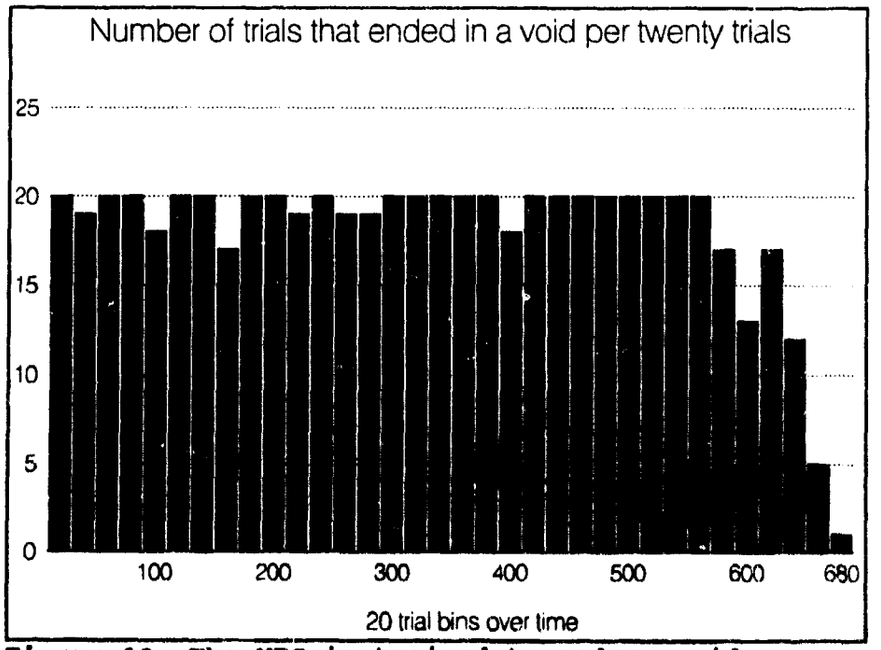


Figure 10: The MPC is trained to reduce voids

Conclusions and Future work: The MPC control was able to converge to a reliable control strategy which prevented void formation. This project is the first pass at the problem of filling the mold to minimize mechanical failure. Much work remains to be done. As already mentioned the realism of the mold simulation will be improved. Coordinated efforts to improve the simulation are ongoing with the Army Computing and Information Center.

This study was focused on demonstrating that an adaptive control strategy was capable of reliably filling the mold with multiple injection points. Additional improvements could be made by attempting to optimize the controller for filling speed. The could be done by making reinforcement a function of speed.

Finally an effort will be made to make the controller generic enough to work on a variety of different mold topologies. Presently it will only work on molds that it has already been trained on.

Appendix A: Formal Definitions: This appendix is meant to formally define Reinforcement Learning. Let s(t) be the *state* of the system at time t. Let S be the set of states s. Let $s(t+1) = T(s(t))$, $\{s(t), s(t+1)\} \in S$ be the *transition* from s(t) to s(t+1). Note that T may be based on some hidden probability such as the permeability. Let $r=R(s)$ $s \in S$, be the *reinforcement function*. By convention R will be non-zero only for *goal states* $g \in G \subseteq S$. The larger the reinforcement value r, the better the goal state is, the lower r is the less desirable g is. The *utility* of a state U(s) is the total discounted r of the future path through the state space. The path is the tuple $\{s(t+1), s(t+2),...s(t+\infty)\}$ resulting from the iteration of T on S. The utility is typically discounted with an exponentially decaying function over the path

$$U(s) = \Sigma^{\infty}_{i=t} \lambda^{i-t} R(s(i))$$

where $0 \leq \lambda \leq 1$. *Reinforcement Learning* is the estimation of the utility U(s) for all S based on a training set of paths and their reinforcement values [1, 2, 11].

This learning law was

$$\Delta W_{i,j,\ldots} = \begin{cases} \alpha[R(t)+O_i(t+1)-O_j(t)] & \text{where i and j are the action at time t+1 and t respectively} \\ 0 & \text{otherwise} \end{cases}$$

where $\Delta W_{i,j,\ldots}$ is the entries in the activated hypercube of the CMAC, α is the learning rate, and O_t and O_{t+1} are the outputs at time t and t+1 respectively. This rule is similar to Watkins' Q-Learning[11], especially when the generalization radius of the hypercube is 0.

CMAC used a number of internal parameters. $\alpha = 0.86/$(volume of the hyper-cube). When the radius was 0, the algorithm was very similar to Q-Learning. The weights of the receptive fields are initialized to 1/(volume of the hypercube) This biases the net to try novel actions.

Dyan has shown that similar laws will converge to U(s(t)) the estimated utility for state action pairs [3,4]. Given even an approximate estimate, a fairly good controller can be constructed.

Bibliography

[1] Barto, A. Sutton, R. Watkins, C. "Learning and Sequential Decision Making", COINS Technical Report 89-95, September 1989

[2] Barto, A. G. Bradtke, S. Singh, S. P. "Real-Time Learning and Control Using Asynchronous Dynamic Programming", COINS Technical Report 91-57, University of Massachusetts: Amherst

[3] Dayan, P. "The Convergence of TD(λ) for General λ", Machine Learning vol 8, 1992, Kluwer Academic Publishers, Boston, Mass.

[4] Dayan, P. Sejnowski, T. "TD(λ) Converges with Probability 1", CNL, The Salk Institute, ftp site: ftp.gmd.de

[5] Dullien, F. Porous Media Fluid Transport and Pore Structure Academic Press 1979

[6] Harlow, F. Anderson, A. Fluid Dynamics: An Introductory Text Los Alamos Scientific Laboratories of the University of California, 1970

[7] Lin, L. J "Self-Improving Reactive Agents Based on Reinforcement Learning, Planning and Teaching", Machine Learning, 8, 1992.

[8] Meeden, L., McGraw, G. Blank, D. "Emergent Control and Planning in a Autonomous vehicle", Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society

[9] Mitchell, L. Boyan, J. "A Distributed Reinforcement Learning Scheme for Network Routing", Technical report CMU-CS-93-165, School of Computer Science. Carnegie-Mellon University

[10] Sheppard, M. et al "Reinforcement Learning in Control" ftp site: ftp.gmd.de

[11] Watkins, C., "Learning from Delayed Reinforcement", PhD thesis, Cambridge University. Cambridge, England 1989.

[12] Walsh, S. Charmchi, M. "Free Surface Flow Surface Flow Observations in the Resin Transfer Molding Process"

[13] Miller, W.T. "Real Time Control Applications of Neural Networks for Sensor Based Control of Robots with Vision" IEEE Transactions of System, Man & Cybernetics vol. 13 No. 4. pp.825-831, 1989

PATTERN CLASSIFICATION OF VIBRATION SIGNATURES USING UNSUPERVISED ARTIFICIAL NEURAL NETWORK

Joung K. Kim and Young S. Shin

Department of Mechanical Engineering
Naval Postgraduate School
Monterey, CA 93943

Abstract: Unsupervised artificial neural networks play an increasingly important role in the pattern classification of vibration signals as a tool for machinery condition monitoring and diagnostics. A major advantage associated with neural networks lies in the fact that the method does not require any explicit apriori statistical information. This paper describes the use of unsupervised artificial neural network for the pattern classification of vibration signatures. The Adaptive Resonance Theory (ART2-A type) paradigm of artificial neural network is applied in this study. Unsupervised learning is a far more plausible model of learning for the pattern classification problem which does not require fault condition data. ART2-A type algorithm is selected due to its desired design principles such as unsupervised learning, stability-plasticity and search direct access.

Key Words: Adaptive resonance theory; Artificial neural network; node; Pattern; Supervised learning; Unsupervised learning; Vibration.

Introduction: Artificial neural network, simply neural network, is a parallel processing system consisting of nonlinear processing elements (nodes) interconnected by fixed or variable weights. The system can be designed to generate arbitrarily complex decision region for specific mappings. Hence the system is well suited for use as detectors and pattern classifiers. Neural network is data-based rather than rule based, thereby possessing the potential of being able to operate where analytical solutions are inadequate. It is reputed to be robust and highly tolerant of noisy data.

Neural network paradigm falls into two classes: supervised and unsupervised learning. Supervised learning [1] requires a pair of each input vector with a target vector which is associated with the desired output. We call this a training pair. Training and adjusting neural network weights is accomplished by sequentially applying the input pairs. When an input vector is applied, the output of the network is calculated and compared to the

corresponding target vector, then the difference is fed back through the neural network and weights are changed according to a learning algorithm that tends to minimize the error. The vectors of the training set are applied sequentially, and errors are calculated and weights adjusted for each vector. The neural network weights gradually converge to values such that each vector produces the desired output vector. Future inputs which are similar to those in the training sample will be classified appropriately.

Despite many successful application [2,3,4,5], supervised learning has been criticized as being biologically implausible. Unsupervised learning is a far more plausible model of learning in the biological system. Human being are able to learn many new things without necessarily forgetting things learned in the past. It is one of the nice features of human memory. Supervised neural network will tend to forget old information if we attempt to add new informations. Supervised learning counts on a target vector for each input of the training pair. This technique is very useful, and in some ways related to the human learning process. However in many applications, it would be more beneficial if we could ask the neural network to form its own classifications of the training data. Developed by Kohonen [6] and Carpenter and Grossberg [7,8,9], it requires no target for the outputs, and hence no comparisons to predetermined desired output. The training set consists of input vectors only. The unsupervised learning algorithm modifies the internal state of the neural network to model the features found in the training data. Unsupervised learning uses the neural network as a self-organizing classifier. Decision regions are constructed in light of similarity of the input exemplars. It requires no target vector for the outputs, and hence, no comparisons to predetermined ideal responses.

Most neural network paradigms are plagued by a problem known as the *stability–plasticity dilemma*, faced by all intelligent systems capable of autonomously adapting in real time to unexpected changes in their world. This dilemma is stated by a series of questions [10]: How can learning systems be designed to remain plastic (the learning state where the internal parameters of the network can be modified) or adaptive, in response to significant inputs, and yet remain stable (a fixed classification set) in response to irrelevant inputs? How does the system know how to switch between its stable and its plastic modes? How can the system preserve its previously learned knowledge while continuing to learn new things? In a backpropagation neural network, for example, the training vectors are applied sequentially until the network has learned the entire set. If, however, we are trying to add a new training vector to an already trained network, it may have the catastrophic side-effect of destroying all the previous learning by interfering with the weight values. With training times for large networks requiring considerable amounts of computer time (hours, even days) this is a serious limitation. A developing theory called Adaptive Resonance Theory (ART), suggests a solution to this problem. The major feature of ART, proposed by Grossberg and Carpenter [7,8,9], is the ability to switch modes between plastic or adaptive and stable, without detriment to any previous learning.

To characterize the vibration signatures, time, frequency, and time-frequency domains are commonly used. Both stationary and non-stationary signals of the time domain are used as an input feature set for the neural network. A simulation program was written and the results of the study shows that ART2-A type algorithm is not only able to learn

and distinguish the patterns from each other but also fast despite the high dimensional input pattern. This paper addresses the pattern classification of vibration signatures using ART neural net model. A complete step by step algorithm of ART2-A type neural network is presented and examples are demonstrated.

Adaptive Resonance Theory: Adaptive Resonance Theory (ART) architectures are neural networks that carry out stable self-organization of recognition codes for arbitrary sequences of input patterns. According to ART theory, adaptive resonance occurs when the input to a network and the feedback expectancies match. It accepts an input vector and classifies it into one of a number of categories depending on which of a number of stored patterns it resembles. The ART neural network compares a given input with previously encountered patterns. If the input pattern is similar to a stored pattern within a specified tolerance ρ (the vigilance), it will be classified into the same category with that pattern. And the stored pattern is adjusted to make it still more like the input pattern depending on the learning rate β. On the other hand, if the input vector does not match any of the previously presented patterns, a new category will be assigned to the given input by storing a pattern which is the input vector. The sensitivity of the neural network is adjusted with the vigilance threshold. High vigilance threshold will increase sensitivity, reduce errors and generate a large number of categories. The number of assigned categories, on the other hand, can be reduced by selecting low vigilance threshold. However, errors will increase An optimal vigilance threshold creates a reasonable number of categories with minimal error.

Two distinct network models based on ART are ART-1 and ART-2. ART-1 [7] is designed to accept only binary input patterns, whereas ART-2 [8], a later development that generalized ART-1, can process both binary and analog inputs. The ART2-A [9] is an efficient algorithm which models the essential dynamics of the ART-2 analog pattern recognition neural network. The ART2-A system accurately reproducing the behavior of ART-2 in the fast learn limit, and sharply delineates the essential computations performed by ART-2. ART2-A ran approximately three orders of magnitude faster than ART-2 in simulations on conventional computers thereby making it easier to use in solving large problems. The ART2-A algorithm also suggests efficient parallel implementations.

The ART network relies on details of architecture far more than most other neural network paradigms. The layers of the ART network have different functions unlike the fairly homogeneous layers of the multilayer perceptron or Kohonen networks. And there are external parts to the layers that control the data flow through the network. Because of this, it is probably worth explaining the ART-2 architecture first, before going on to describe the ART2-A type algorithm in detail.

ART-2 Architecture: The ART neural network is shown schematically in Figure 1. It has three layers: the first is the preprocessing layer F_0, the second is the input/comparison layer F_1 and the third is the output/category layer F_2. The F_1 and F_2 layers are connected together by both bottom-up from F_1 to F_2 and unlike most other networks, top-down from F_2 to F_1. The bottom-up and top-down weights between F_1 and F_2 layers are fully connected. ART-2 neural network has bottom-up weight vectors

from F_1 layer to F_2 layer and top-down weight vectors from F_2 layer to F_1 layer.

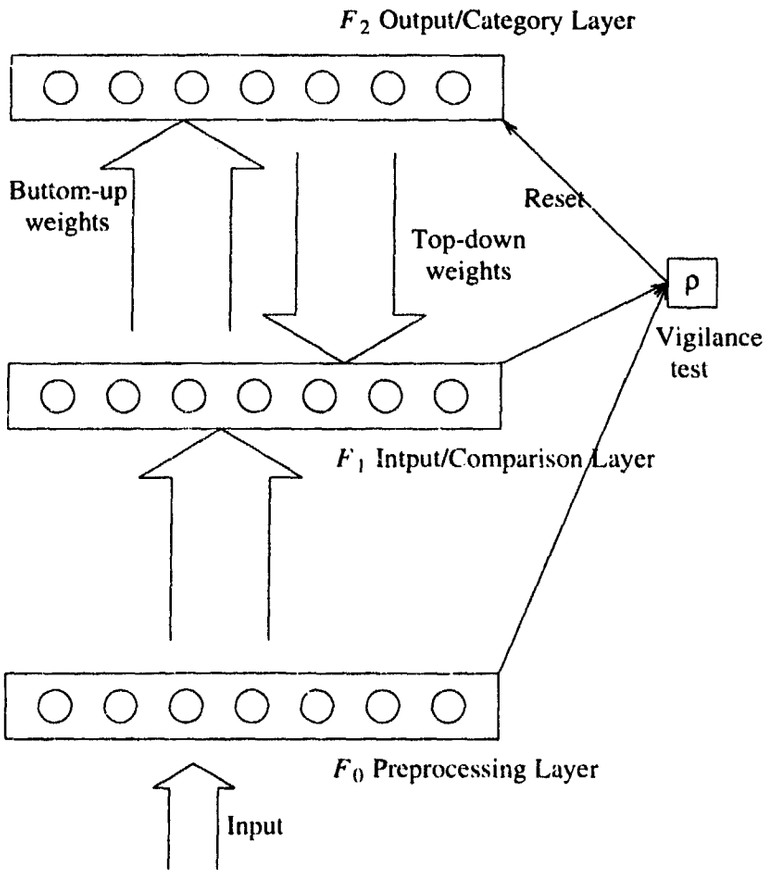


Figure 1. ART-2 Architecture

The functions of preprocessing layer F_0 are noise suppression and normalization. F_0 layer receives and normalizes either binary or analog input vector. Next noise suppression and contrast enhancement is done by a piecewise linear signal function defined by

$$f \equiv \begin{cases} x & \text{if } x \geq \theta \\ 0 & \textit{otherwise} \end{cases}$$

The threshold θ is assumed to satisfy the constraints $0 < \theta \leq 1/\sqrt{N}$ so that N-dimensional input vector is always nonzero if input vector is nonuniform. Next this vector is normalized again and send to input/comparison layer field F_1. The functions of the F_1 field are contrast enhancement, noise suppresion, matching, and normalization. F_1 is composed of three sublevels, for receiving bottom-up input pattern, for receiving top-down expectation and for matching the bottom-up and top-down data. The F_1 layer feeds the bottom-up weights, reads, normalizes top-down input, and matches patterns.

The output/category layer F_2 serves to classify the input vector. Each node in F_2 layer computes an inner product between its weight and output vector from F_1. The node which has the largest output will win the competition and it inhibits all other nodes in the layer.

There are also a reset module among F_0, F_1, and F_2 layers. This plays a vital role in the neural network: it performs more than simple a reset function for output nodes. The reset module measures the similarity between input vector and best matching examplar vector.

When the j-th F_2 node is chosen as a good pattern match to the input pattern, the top-down and bottom-up weights need to be adjusted. Two kinds of learning could be distinguished in ART-2: slow and fast learning. In slow learning, an input vector may be applied so briefly that network weights do not have enough time to reach their asypmtotic value during single presentaion. On the other hand, in the fast learning, the input vectors are applied for a long enough time that the adaptive weights can approach their final values. If they differ by more than the vigilance threshold, a reset signal is sent to disable the firing node in the output/category layer. Thus, the vigilance parameter determines whether a new class pattern should be created for an input pattern. This is the basic architecture of the ART-2 network.

ART2-A Type Algorithm: ART2-A algorithm is a wide range of implementation that adheres to the spirit of ART-2 neural network while they differ greatly in detail. The implementation that follows is based on ART2-A algorithm, with certain aspects changed for compatibility with Capenter, Grossberg and Rosen [3]. The following is the ART2-A type algorithm. The algorithm selects the first input as the exemplar for the first cluster. The next input is compared with the first cluster exemplar. It is clustered with the first if the cosine of the angle between input and the first cluster exemplar is greater than a vigiliance threshold ρ. Otherwise it is the exemplar for a new cluster. This process is repeated for all following inputs. The number of clusters depends on both the vigiliance threshold ρ and the Euclidean distance metric used to compare inputs to cluster exemplars.

Step 1. Initialization:

Set the noise compression ratio θ

$$0 < \theta \leq 1/\sqrt{N}$$

where N is the order of input vector I^0.

Set the vigilance threshold ρ

$$0 \leq \rho \leq 1$$

Set the learning rate β

$$0 \leq \beta \leq 1.$$

The fraction ρ is the vigilance threshold which indicates how close an input must be to a stored exemplar to match. High value of ρ leads to finer categories. On the other hand, low vigilance causes the grouping of input patterns to coarser

categories. Learning rate β ranges from 0 to 1. Setting β equal to 1 gives ART2-A into fast learn limit. Setting β equal to 0 remains the weight vector constant once a node is winning.

Step 2. Apply New Input Pattern Vector:

$$\mathbf{I}^0 = [I_0, I_1, \cdots, I_{N-1}]^T$$

Step 3. Preprocessing: Normalization and noises suppression

$$\mathbf{x} = \frac{\mathbf{I}^0}{\|\mathbf{I}^0\|}$$

$$\mathbf{v} = F_\theta \mathbf{x} \quad \textit{where} \quad (F_\theta \mathbf{x})_i \equiv \begin{cases} x_i & \text{if } x_i > \theta \\ 0 & \textit{otherwise} \end{cases}$$

$$\mathbf{I} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

If the input vector is the first input vector of neural network, then go to step 7.

Step 4. Compute Matching Scores

$$\mu_j = \mathbf{I}\,\mathbf{z}_j = \sum_{i=0}^{N-1} I_i z_{ij} \qquad 0 \le j \le M-1$$

In this equation μ_j is the output of output node j and $\mathbf{z}_j$ is weight vector. M is the number of the cluster exemplar in the output/category layer which is initially equal to zero.

Step 5. Select Best Matching Exemplar

$$\mu_j^* = \max_{0 \le j < M} (\mu_j)$$

If more than one node is maximal, choose one at random.

Step 6. Vigilance Test

If $\mu_j^* \ge \rho$ where ρ is constrained so that $0 \le \rho \le 1$, then the best matching node j selected in Step 5 remains constant. Then go to Step 8.

Step 7. Reset

j is reset to the new cluster node with index M and set

$$\mathbf{z}_j^{(new)} = \mathbf{I}.$$

$$M = M+1$$

Go to Step 9.

Step 8. Adapt Best Matching Exemplar

If j is one of cluster exemplar, then

$$\Psi = \frac{\phi}{\|\phi\|} \quad \textit{where} \quad \phi_i \equiv \begin{cases} \mathbf{I}_i & \text{if } \mathbf{z}_{ij}^{(old)} > \theta \\ 0 & \textit{otherwise} \end{cases}$$

$$\mathbf{w} = \beta\Psi + (1-\beta)\mathbf{z}_j^{(old)}$$

$$\mathbf{z}_j^{(new)} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

where $\mathbf{z}_j^{(old)}$ denotes the weight vector of $\mathbf{z}_j$ at the start of the input representation, and $0 \leq \beta \leq 1$.

Step 9. Repeat by Going back to Step 2

ART2-A type Processing Example: A numerical example will be given for ART2-A type algorithm that is to be categorized as follows. We consider three simple input vectors: [1.0 0.0 1.0], [0.9 0.2 0.8] and [1.0 1.0 0.1]. For this example, the initial parameters will be set at $\theta = 1/\sqrt{3} = 0.5774$, $\rho = 0.8$, and $\beta = 0.01$ respectively.

I			z			Category
0.7071	0.0	0.7071	0.7071	0.0	0.7071	0
0.7474	0.0	0.6644	0.7075	0.0	0.7067	0
0.7071	0.7071	0.0	0.7071	0.7071	0.0	1

Table 1. ART2-A type Classification and Adaption

To follow this table, first input vector [1.0 0.0 1.0] is normalized and suppressed noise from Step 3 and then go to Step 7.

$$\mathbf{x} = \frac{\mathbf{I}^0}{\|\mathbf{I}^0\|} = \frac{[1.0 \;\; 0.0 \;\; 1.0]}{1.4142} = [0.7071 \;\; 0.0 \;\; 0.7071]$$

$$\mathbf{v} = F_\theta \mathbf{x} = [0.7071 \;\; 0.0 \;\; 0.7071]$$

$$\mathbf{I} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = [0.7071 \;\; 0.0 \;\; 0.7071]$$

By step 7, the node 0 is selected as the exemplar for the first cluster 0. So $\mathbf{z}_0$ has weights [0.7071 0.0 0.7071]. Now the second input vector is applied to the neural network. From Step 3,

$$\mathbf{x} = \frac{\mathbf{I}^0}{\|\mathbf{I}^0\|} = \frac{[0.9 \;\; 0.2 \;\; 0.8]}{1.2207} = [0.7373 \;\; 0.1638 \;\; 0.6554]$$

$$\mathbf{v} = F_\theta \mathbf{x} = [0.7373 \;\; 0.0 \;\; 0.6554]$$

since second element is less than θ. Next, the noise suppressed input vector is normalized again using

$$\mathbf{I} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{[0.7373 \;\; 0.0 \;\; 0.6554]}{0.9865} = [0.7474 \;\; 0.0 \;\; 0.6644]$$

Since we have only one cluster exemplar, we compute

$$\mu_0 = \mathbf{I}\mathbf{z}_0 = [0.7071 \;\; 0.0 \;\; 0.7071][0.7474 \;\; 0.0 \;\; 0.6644]^T = 0.9983$$

and best matching exemplar becomes μ_0^*. From Step 6, since $\mu_0^* = 0.9983 > \rho = 0.8$, best matching node 0 remains constant. From Step 8, compute $\Psi = \frac{\phi}{\|\phi\|}$ using

$$\phi_i \equiv \begin{cases} \mathbf{I}_i & \text{if } \mathbf{z}_{ij}^{(old)} > \theta \\ 0 & otherwise \end{cases}$$

So Ψ = [0.7373 0.0 0.6554]. And adaption is accomplished by

$$\begin{aligned} \mathbf{w} &= \beta\Psi + (1-\beta)\mathbf{z}_j^{(old)} \\ &= 0.01\times[0.7474\ \ 0.0\ \ 0.6644] + 0.99\times[0.7071\ \ 0.0\ \ 0.7071] \\ &= [0.7075\ \ 0.0\ \ 0.7067] \end{aligned}$$

Finally, the new $\mathbf{z}_0$ = [0.7075 0.0 0.7067] is obtained by normalizing the vector $\mathbf{w}$. Now apply the third input pattern [1.0 1.0 0.1] to the neural network. From Step 3,

$$\mathbf{x} = \frac{\mathbf{I}^0}{\|\mathbf{I}^0\|} = \frac{[1.0\ \ 1.0\ \ 0.1]}{1.4177} = [0.7053\ \ 0.7053\ \ 0.0705]$$

$$\mathbf{v} = F_\theta \mathbf{x} = [0.7053\ \ 0.7053\ \ 0.0]$$

since the third element is less than θ. Next, the noise suppressed input vector is normalized again using

$$\mathbf{I} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{[0.7053\ 0.7053\ 0.0]}{0.9975} = [0.7071\ 0.7071\ 0.0]$$

Since we have only one cluster exemplar, we compute

$$\mu_0 = \mathbf{I}\mathbf{z}_0 = [0.7071\ \ 0.7071\ \ 0.0][0.7075\ \ 0.0\ \ 0.7067]^T = 0.5003$$

and best matching exemplar becomes μ_0^*. From Step 6, since $\mu_0^* = 0.5003 < \rho = 0.8$, the neural network reset to the new cluster node with index M = 1. From Step 7, the weight vector $\mathbf{z}_1$ becomes [0.7071 0.7071 0.0].

Examples and Discussions: In order to explore the behavior of ART2-A type neural network algorithm in a machinery diagnostics environment, we use three groups of stationary and non-stationary signals; (1) Harmonic waves, (2) Composite signals, (3) Harmonic wave with some glitches. Any pattern classification attempt should be preceded by a feature analysis phase. This is true even for ANN based pattern classification. In the present research, the time domain data of signals is proposed. Each data of signals was sampled at 1000 *Hz*.

1. Harmonic Waves: In this section we will present classification of harmonic data in time domain. We used pure sine waves with frequency component (A) 30 *Hz*, (B) 60 *Hz*, (C) composite signal with two frequency components 30 *Hz* and 60 *Hz*, and (D) shifted signal with 30 *Hz* frequency component. We sampled by forming a time axis for our data, running from $t = 0$ until $t = 0.255$ in steps of 1 millisecond.

To evaluate the performance of the ART2-A type neural network, the inputs, indexed in the left column of Figure 2, were presented in the order A, B, C, D. I^0 is the input vector, I is the normalized and noises suppressed vector of I^0, z_J is adapted weight vector of winning cluster node J at the end of each input presentations. Second column in Figure 2 shows the four patterns after preprocessing for the input patterns respectively. Each input pattern I^0 contrasts with the each pattern I in Figure 2 due to the fact that the pattern

troughs are below the noise level defined by signal threshold θ. In fact, θ is set equal to $1/\sqrt{N}$ where $N = 256$ is dimension of input vector. The ART2-A type simulation categorized the four input patterns into 2 categories. The pattern A and D are in category 0 and the pattern B and C are in category 1. The two subplots at the bottom of third column shows how the weight adaptions are accomplished after the best matching examplar is selected.

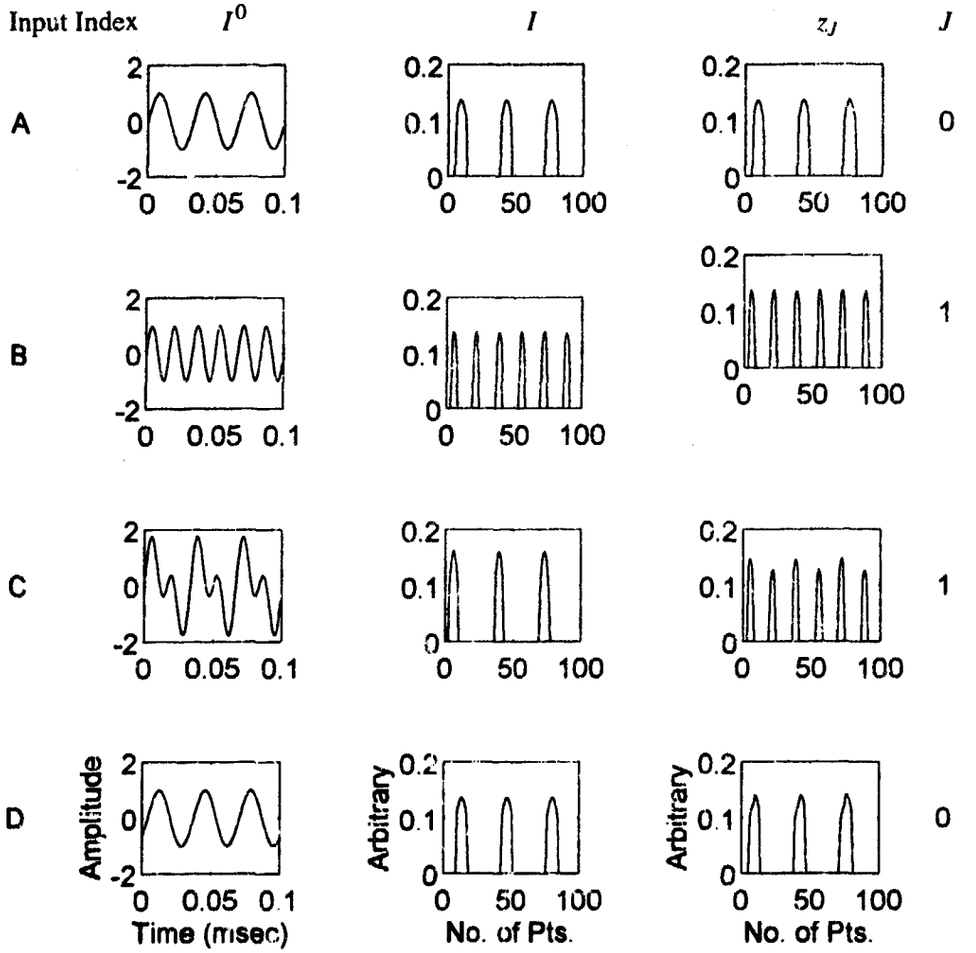


Figure 2. ART2-A type simulation for harmonics ($\rho = 0.5$, $\beta = 0.1$)

Table 2 shows the result of the impact of the vigilance parameter on the sensitivity of the pattern matching of the ART2-A type neural network. The neural network is capable of distinguishing these patterns apart properly. As shown in table 2, the pattern A is assigned to category 0, the pattern B and C are in category 1, and pattern D is assigned to category 2 for vigilance parameter $\rho < 0.6$. For $\rho \geq 0.6$, all the patterns assinged to different category. On the other hand, Pattern A and D are assigned to different category for any vigilance parameter even though the two signals are same except shifted. From this results shows that when we use the time domain data as an input pattern of neural

Vigilance	Category
0.30	0 1 1 0
0.40	0 1 1 0
0.50	0 1 1 0
0.60	0 1 2 3
0.70	0 1 2 3
0.80	0 1 2 3
0.85	0 1 2 3
0.90	0 1 2 3
0.95	0 1 2 3

Table 2. Result of classification for different vigilance threshold for harmonics

network, synchronized samplin ; must be performed.

2. Composite Harmonic Wave: We tested the three sets of signals: (A) the sine wave with stepwise frequency changes, 50 *Hz*, 25 *Hz* and 10 *Hz*, (B) composite signal with three frequency components 50 *Hz*, 25 *Hz* and 10 *Hz*, and (C) composite signal with three frequency components 35 *Hz*, 25 *Hz* and 10 *Hz*. We sampled by forming a time axis for our data, running from $t = 0$ until $t = 0.511$ in steps of 1 milisecond.

To evaluate the performance of the network for learning and differentiating patterns A, B, and C, we tested the three sets of data of 512 dimensionality, each in order "ABC" as one input file. Figure 3 shows the three patterns after learning for the input patterns respectively. Each input patterns I^0 in Figure 3 contrasts with each pattern I due to the fact that the pattern troughs are below the noise level defined by signal threshold θ. The simulation shows the neural network differentiates the pattern A as a different pattern from the pattern B and C.

Table 3 shows the result of the impact of the vigilance parameter on the sensitivity of the pattern matching of the network. The network is capable of distinguishing these patterns apart properly. As shown in table 3, the pattern B and C are in category 1, and pattern A is assigned to category 0 for medium vigilance. For vigilance $\rho \geq 0.7$, all the patterns assinged to different category.

3. Harmonic Waves with Some Glitches: The interesting phenomena on the signal with abnormal components as a fault were investigated. The general rotating machinery has a periodic signal pattern in time domain. Figure 4 shows the plot of the signals with abnormal glitchs at a small region of the time record: (A) is the time signal with one glitch component, (B) is the time signal with two glitch components, (c) is the time signal with three glitch components, and (D) is the pure sine wave with 10 *Hz* frequency components. In this section we will present classification of the signal with an abnormal components as a fault in time domain. Each signal was sampled with 256 time data

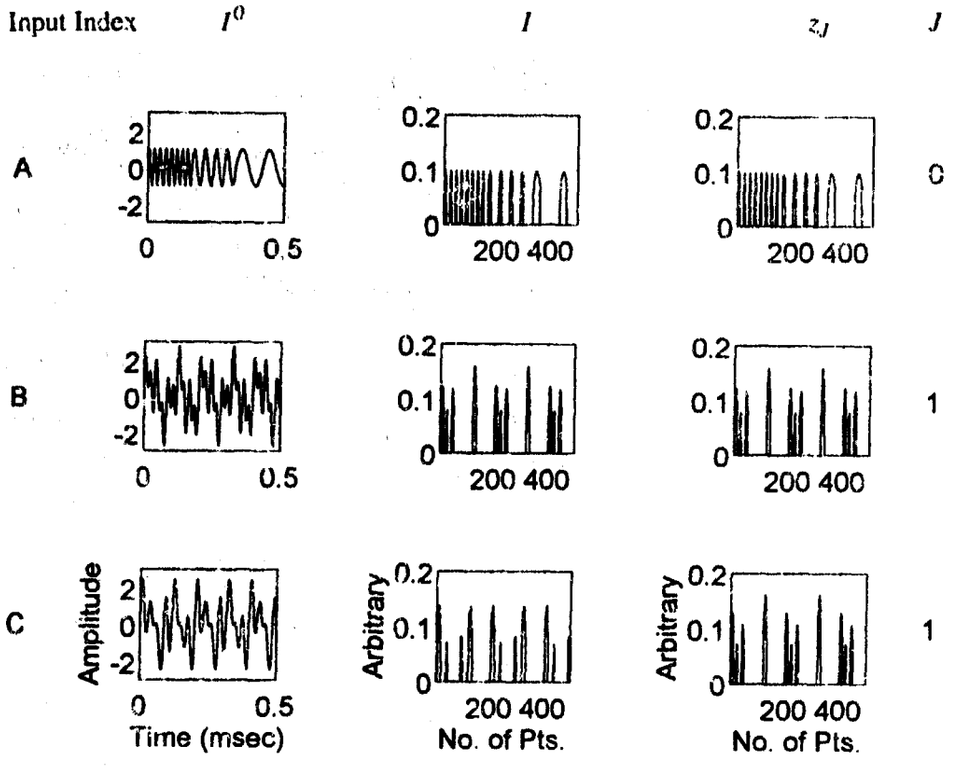


Figure 3. ART2-A type simulation for composite harmonics ($\rho = 0.8$, $\beta = 0.1$)

Vigilance	Category
0.30	1 0 0
0.40	0 0 1
0.50	0 1 1
0.60	0 1 1
0.70	0 1 2
0.80	0 1 2
0.85	0 1 2
0.90	0 1 2
0.95	0 1 2

Table 3. Result of classification for different vigilance parameter for composite harmonics

points by forming a time axis for our data, running from $t = 0$ until $t = 0.255$ in steps of 1 millisecond.

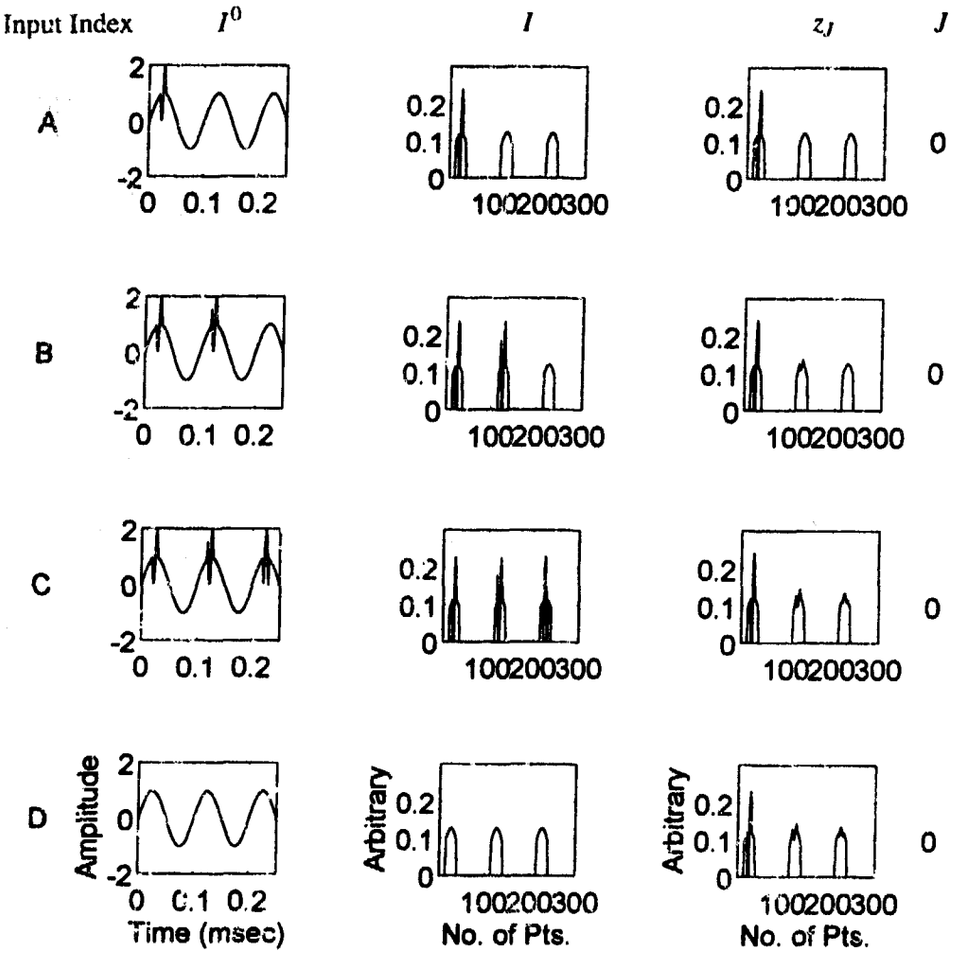


Figure 4. ART2-A type simulation for harmonics with glitches
($\rho = 0.5$, $\beta = 0.1$)

We tested the four sets of data of 256 dimensionality, each in order "ABCD" as one input file. Figure 4 shows the plot of the four input patterns I^0 and I after learning for the input patterns and z_J adapted weights of winning node J respectively.

As shown in table 4, the pattern C, which has three glitch components, is differentiating as a different pattern compared with patterns A, B, D for the vigilance $\rho = 0.9$. For the vigilance $\rho = 0.95$, the patterns A and B are assigned to category 0, pattern C and D are assigned to category 1. These results shows that the feature representation in time domain of the signals with small energy changes can be detected as a fault from the neural network.

Vigilance	Category
0.40	0 0 0 0
0.50	0 0 0 0
0.60	0 0 0 0
0.70	0 0 0 0
0.80	0 0 0 0
0.85	0 0 0 0
0.90	0 0 1 0
0.95	0 1 1 0

Table 4. Result of classification for different vigilance parameter for harmonics with glitches

Conclusions: In this paper, the behavior of ART2-A type neural network in a machinery diagnostics environment has been discussed and demonstrated. A simulation program was built and tested. The synthesis of this ART network do not require any apriori statistical information. The result of this study shows that ART2-A type algorithm is not only able to learn and distinguish the patterns from each other also fast despite the high dimensional input pattern. The training process is fast and the actual classification can be performed in real time. The analysis of results indicates that the vigiliance threshold, plays a critical role in ensuring satisfactory performance. The ART2-A type neural network is indeed very promising in pattern recognition, detection of fault and abnormality for stationary and non-stationary signals in real world problem which has no fault condition data.

References:

[1] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning Internal Representations by Error Propagation," in D. E. Rumelhart and J. L. McClelland, eds., Parallel Distributed Processing, MIT press, Cambridge, Mass., 1986.

[2] Daniel R. Dorrance, "Application of an Artificial Neural Network For F-15 RADAR LRU Fault Isolation", Intelligent Engineering Systems Through Artificial Neural Networks, pp 657-662, 1991.

[3] McDuff, R., Simpson, P., Gunning, D., "An Investigation of Neural Networks for F-16 Fault Diagnosis: I. System Description", Proceedings from IEEE AutoTestCon, pp 351-357, 1989.

[4] Wasserman P. D., Unal A., and Haddad S., " Neural Network On-Line Machine Condition Monitoring Systems," Intelligent Engineering Systems Through Artificial Neural Networks, pp 693-700, 1991.

[5] Kim, D. S., Shin, Y. S., and Carlsen, D. K., "Machinery Diagnostics for Rotating Machinery Using Backpropagation Neural Network," Proceedings of the 3rd International Machinery Monitoring & Diagnostics Conference, pp 309-320, Las Vegas, NV, Dec. 9-12, 1991.

[6] Kohonen, T. "Self-organization and Associative Memory". Series in Information Science, Vol. 8, Springer Verlag, Berlin, 1984.

[7] Carpenter. G. A. & Grossberg. S., "A Massively Parallel Architecture for Self-organizing Neural Pattern Recognition Machine, Computer Vision, Graphics, and Image Processing, 37, pp 54-115, 1987.

[8] Carpenter. G. A. & Grossberg. S., "ART-2: Self-organization of Stable Category Recognition Codes for Analog Input Patterns, Applied Optics. 26, pp 4919-4930, 1987.

[9] Carpenter. G. A., Grossberg. S., & Rosen D. B. "ART 2-A: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition", Neural Networks, Vol 4, pp 493-504, 1991.

[10] Carpenter. G. A., Grossberg. S., "The ART of Adaptive Pattern Recognition by Self-organizing Neural Network", Computer, 21 (3) 77 - 88, March 1988.

MAGNETOSTRICTIVE SENSORS FOR STRUCTURAL HEALTH MONITORING SYSTEMS

Cecil M. Teller and Hegeon Kwun
Southwest Research Institute
P.O. Drawer 28510
San Antonio, Texas 78228-0510

Abstract: A new and very promising sensor technology—magnetostrictive sensors*—for use in structural health monitoring systems is under development by Southwest Research Institute for a variety of applications. The magnetostrictive sensors comprise a network of ferromagnetic "nerves" embedded within or bonded to the component and one or more surface-mounted detectors (e.g., inductive pickup coils). Only simple electronics, including low-noise amplifiers, filters, cables, and connectors, are needed to provide useful signals for a computerized data acquisition, analysis, and display system for evaluation of the service life/integrity of the structure. Potentially, this ***single*** sensor technology is useful for monitoring delaminations, dynamic strain (stress), fatigue cracking, fast fracture, and other damage from cyclic stresses, structural overloads, and ballistic/foreign-object impacts. As an example of the power and diversity of the magnetostrictive sensor technology, this paper presents new concepts and test results particularly relevant to structural health monitoring systems.

Key Words: Delaminations; fatigue; fracture; impact; magnetostriction; sensors; strain; stress; vibration.

Introduction: To ensure safety and structural reliability of structural components, they typically are subjected to periodic inspection. Inspection of critical components is time-consuming and costly because it involves (1) almost completely disassembling the system; (2) inspecting the parts for cracks, corrosion, or other types of damage; (3) repairing that damage or replacing the part; and (4) then reassembling the system. When the critical crack size in a component such as a helicopter rotor is small, these inspections are undertaken at short intervals with a considerable number of man-hours and downtime of the system.

A more promising approach to tracking and assuring the safe operation of complex mechanical systems that does not require disassembly is based on the structural health monitoring system concept (1). This concept comprises innovative sensors, artificial intelligence (AI), and advanced analytical techniques to provide realtime, continuous assessment of structural integrity. With this concept, critical structural components are continuously monitored for detection of inservice dynamic loads, crack propagation, or vibration, and the data are analyzed by an expert system to automatically produce a "bill of health." Such a health monitoring system would allow early detection of problem areas, optimized repair scheduling, reduction in inspection man-hours and unsubstantiated component removals, and improved structural safety and reliability. It is important, however, to be able to nondestructively determine the quality and integrity of structures and to continuously monitor their health. Moreover, as a result of the economic

*Patent pending.

climate of the world today, it is also important to use effective, low-cost, rugged, and reliable sensors for the job.

Crucial to the success of an on-line structural health monitoring system is development of advanced sensors that can feel, like *nerves* in the human body, the pains in structural components and send distress signals to a central unit where, like in the *brain*, the signals are processed and analyzed and the condition of structural health and actions to be taken are determined. Distress symptoms these sensors should be able to detect include:

- Fatigue cracking
- Fracture
- Inservice dynamic loads
- Other damage from structural overloads and ballistic/foreign object impacts.

All of the above distress symptoms can be detected by the magnetostrictive sensor (MsS) technology, pioneered recently at Southwest Research Institute (SwRI) (2); thus, it is ideally suited for health monitoring system applications. Additional merits of this promising new sensor technology include: noncontact and passive sensing; low instrumentation cost; and simplicity, compactness, and durability of the sensor.

Magnetostrictive sensors have many potential advantages over other sensors such as electrical resistance strain gauges, conventional piezoelectric acoustic-emission transducers, and fiber-optic sensors because of their inherent economy, simplicity, and ruggedness. Most important, however, the MsSs can perform essentially ***all*** the functions of combined strain-gauge, acoustic-emission, and fiber-optic sensors. When it comes to application to composite structures, size compatibility is not an issue as it is with fiber-optic sensors, since the diameter or thickness of the magnetostrictive wires or ribbons used as embedded nerves can be matched to the ply thickness of composite materials. And, unlike strain gauges and conventional acoustic emission sensors, the MsS pickup coils can be mechanically fastened or bonded with durable structural adhesives.

Physics of Magnetostrictive Sensors: Ferromagnetic materials such as iron, nickel, and various structural steels exhibit magnetoelastic interaction (3). This mechanism interrelates magnetic and elastic properties of the material and causes many interesting phenomena, including:

- Change in dimensions due to magnetization, called the magnetostrictive or Joule effect (4)
- Change in magnetization intensity due to stress or strain, called the inverse-magnetostrictive or Villari effect (5)

The MsS approach utilizes the Villari effect for monitoring of dynamic stresses and acoustic emissions. Since the MsS relies on the Villari effect, the sensor is directly applicable to components made of ferromagnetic materials. If the component is made of nonferrous metals such as aluminum or titanium, the sensor is still applicable by plating the component with a suitable ferromagnetic material such as nickel. In this case, the plating is necessary only in a local area on which the MsS is placed. The sensor is also applicable to composite structures if small-diameter ferromagnetic wires are embedded in the structure or bonded on the structure surface.

Magnetostrictive Sensor Configuration and Electronics: Figure 1 illustrates the basic configuration of MsS together with accompanying electronics for signal conditioning. The sensor consists of two basic elements, an inductive coil and a permanent magnet. The inductive coil picks up the changes in the magnetization of the ferromagnetic material caused by dynamic stresses or acoustic-emission events such as fatigue-crack propagation or impact by a foreign objects. Although the inductive coil is the simplest configuration, other magnetic-field sensing devices such as the Hall-effect probe can be used to detect the magnetization changes.

The permanent magnet is used to magnetize the ferromagnetic material in the local area where the MsS is placed. The material is kept magnetized for two reasons, to enhance the sensitivity of the MsS and to make the frequency response of the sensor linear (if the material is demagnetized, the frequency of the detected signal will be twice the frequency of the dynamic stress) (6,7). In many applications, the permanent magnet can be substituted by residual magnetization left in the material (accomplished by magnetizing the material using a magnet and then removing the magnet). If possible, the residual magnetization is preferred because the need for installing a permanent magnet is eliminated, which leads to a simpler and more compact sensor design.

Because the sensor is very flexible in its physical shape and size, its configuration can be easily tailored to a variety of applications. For example, on cylindrical components such as rotor shafts, pins, or bolts, a cylindrically shaped MsS encompassing the component may be most suitable. On plate-like components such as rotor hubs or blade spars,

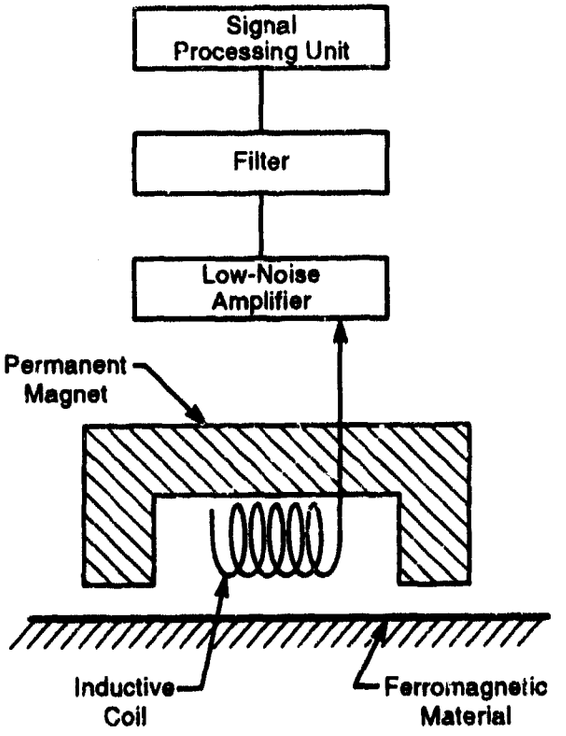


Figure 1. Basic configuration of MsS together with accompanying electronics for signal conditioning

a rectangularly shaped MsS placed on the surface of the component may be most appropriate.

The electronics required to condition the detected signals are standard low-noise amplifiers and filters. Since, in this case, the MsS is used as a passive device, no electronics for actively energizing the sensor are required.

Dynamic Stress Monitoring: Experimental data supporting the feasibility of sensing dynamic stresses using a noncontact magnetostrictive sensor based on the Villari effect are shown in Figure 2 (8). The data, representing a dynamic torque at approximately 35-Hz frequency, were taken from a 1-inch-diameter shaft placed between a four-cylinder automotive engine and a dynamometer using a 2.25-inch inside-diameter encircling coil around the shaft. Residual magnetization, induced in the shaft by placing and then removing a permanent magnet, was used as the static biasing magnetization required for magnetostrictive sensing. The data agreed well with the signals detected using an independent strain-gauge/telemetry torque meter.

Figure 3 shows other supporting data that demonstrate the ability to monitor dynamic stresses in composites by embedding magnetostrictive (in this case, Ni) wires. The data were taken from a graphite/epoxy composite specimen (approximately 4.25 inches long by 0.5 inch wide by 0.04 inch thick with a total of eight graphite plies). The specimen was prepared with six small-diameter (0.004-inch) Ni wires embedded below the first ply and an encircling pickup coil wound near one end of the specimen. Residual magnetization left in the Ni wires was also used in this test as the static biasing magnetization required for magnetostrictive sensing. The composite specimen was mounted in a cantilever configuration with the pickup coil toward the fixed end. The data in Figure 3 show the dynamic stress associated with the damped vibration of the specimen after it was set in motion by simply deflecting and releasing the free end (vibration frequency was approximately 90 Hz). The above data clearly demonstrate that dynamic stresses can be readily detected using the proposed MsS approach.

Acoustic Emission Monitoring: Experimental data supporting the feasibility of sensing acoustic emission (AE) signals using a noncontact MsS are shown in the top trace of Figure 4 (9). These data were taken from a steel rod measuring 0.25 inch in diameter and 20 feet in length. The magnetostrictive pickup coil placed near one end of the rod had a permanent magnet nearby to provide the static-biasing magnetization required for magnetostrictive sensing. AE signals were transmitted into the rod by breaking a 0.5-mm-diameter pencil lead on the far end of the rod—a standard method of simulating an AE event (9). For comparison, signals were detected simultaneously using a conventional piezoelectric AE sensor (with a 300-kHz center frequency) attached to the near-end face of the rod. Data from the conventional AE sensor are also shown in the bottom trace of Figure 4. The first signal in each trace of Figure 4 is the AE signal, and the signals occurring later in time are those traveling back and forth in the rod. Except for the high-frequency components present in the bottom trace, the two signals agree well with each other, indicating that AE events can be detected using the MsS. The high-frequency component was absent in the top trace because the particular MsS used in the test was operable up to approximately 100 kHz frequency (10). The operating frequency of MsSs on small-diameter wires, that is, a few thousandths of an inch in diameter, exceeds 500 kHz (11,12).

Supporting data for sensing AE events (due to either cracking or impact) in composites utilizing embedded magnetostrictive wires are given in Figure 5. The data were acquired from the graphite/epoxy composite specimen mentioned in the previous section. The top trace in the figure is the AE signal produced by a standard 0.5-mm-diameter pencil-lead break (ASTM E-976). The bottom trace is the signal produced by dropping

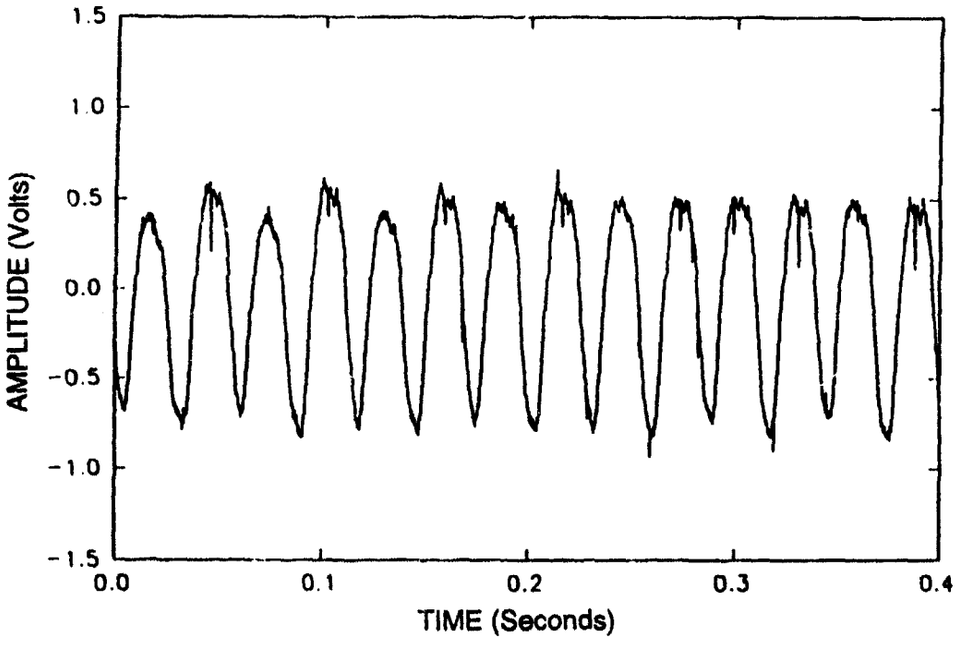


Figure 2. Dynamic torque signal from a 1-inch-diameter shaft using an encircling magnetostrictive sensor. The shaft speed was 1000 rpm, and the steady-state torque was 50 ft-lb.

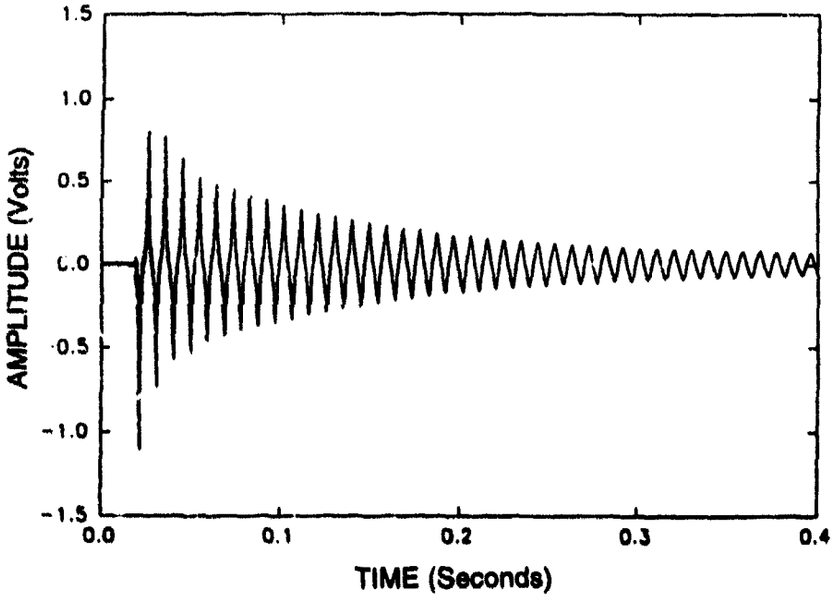


Figure 3. Dynamic stress (or strain) signals acquired from a graphite/epoxy sample using a magnetostrictive sensor

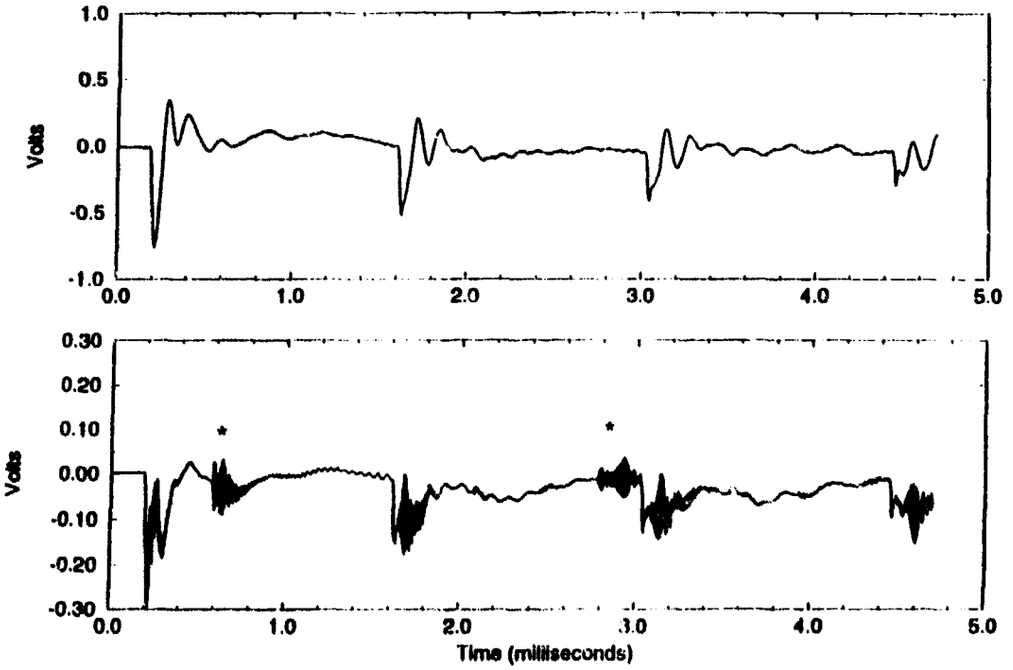


Figure 4. Experimental data supporting the feasibility of sensing AE signals using a noncontact magnetostrictive sensor (top trace) and using a conventional piezoelectric AE sensor (bottom trace)

a 0.375-inch-diameter, 3.5-gram steel ball on the specimen from a height of approximately 4.5 inches. Compared to the dynamic stress signals associated with the damped vibration shown in Figure 3, the pencil-lead-break AE signals contain significantly higher frequency components, while the steel ball impact signals contain intermediate frequency components.

The above data show clearly that the proposed sensor approach can detect AE signals. The results also suggest that, by processing the signals (detected by the same sensors) using bandpass filters with different center frequencies, discrimination among the dynamic stresses associated with service loads, impact signals from ballistic/foreign objects, and AE events from cracking or fracturing can be made.

Comparison with Other Sensors: Strain gauges are a well-established technology for sensing static and low-frequency dynamic surface strain changes (13). For example, in recent years, strain gauges have been used in fixed-wing aircraft for strain monitoring during flight tests. However, their applicability to structural health monitoring is questionable at present because of their lack of long-term reliability and survivability in the aircraft service environment. Their applicability to more complex systems such as helicopter rotors is equally, if not more, questionable.

Fiber-optics embedded in composites can be used for monitoring internal strain changes (both static strains and dynamic strains up to about 500 Hz) (13). Optical fibers appear to be an ideal material to be embedded in composites. However, usefulness of fiber-optic sensors for structural health monitoring systems in other applications is not clear.

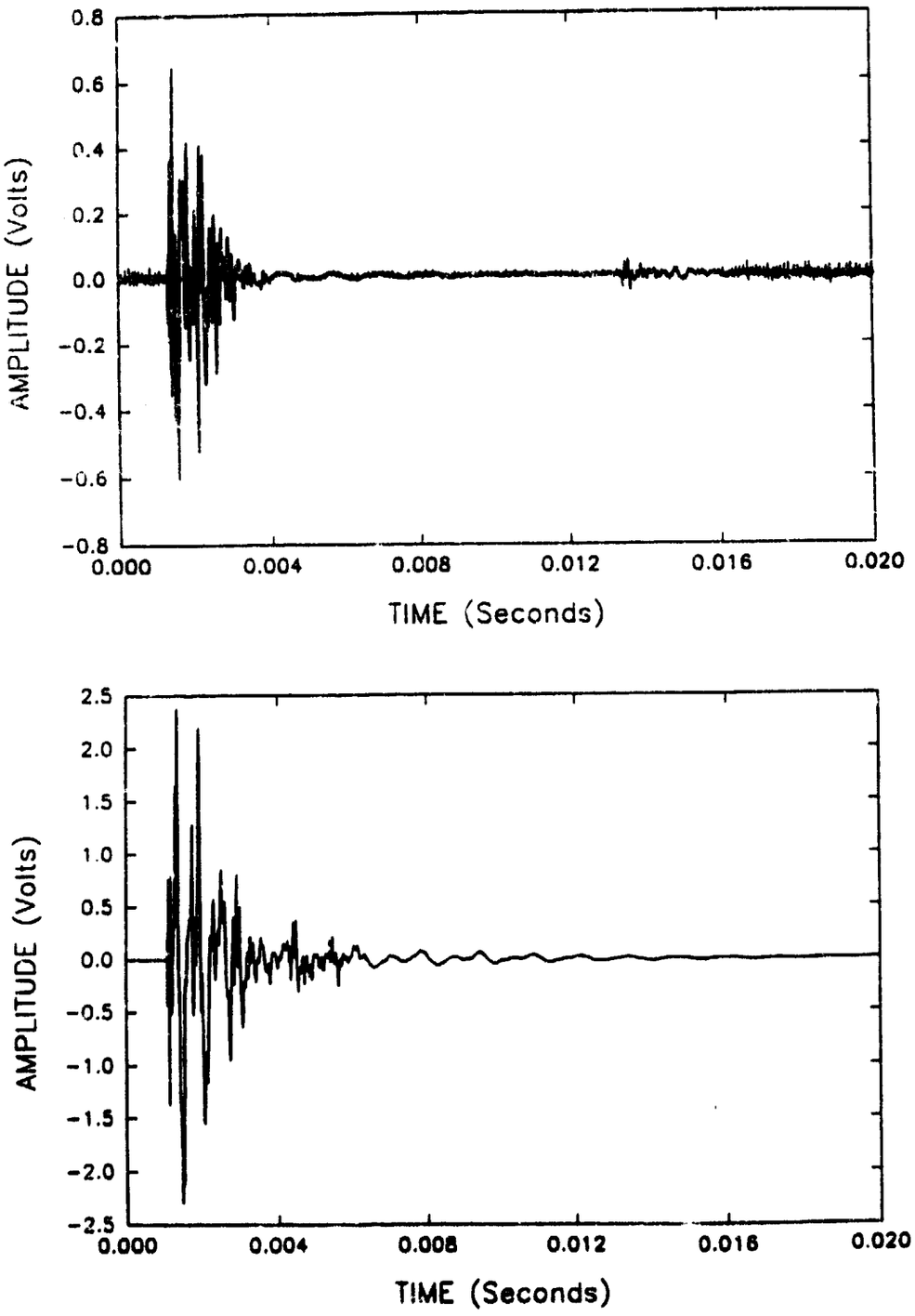


Figure 5. Supporting data for sensing AE events (due to either cracking or impact) in composites utilizing embedded magnetostrictive wires

Piezoelectric sensors mounted on the surface of a structural component can be used to monitor AE events occurring in the structure (13). Piezoelectric AE sensors are widely used for short-term evaluating and testing of metals and composites for crack growth and fractures. However, their use for long-term monitoring has not been successful because of a lack of robustness of the sensors themselves and stability of the coupling (the sensor is acoustically coupled to the structural component using a couplant such as grease and oil). The applicability of these sensors to structural health monitoring is therefore limited.

As previously described, the proposed MsS approach can be used for measuring dynamic stresses and for monitoring AE signals. Thus, the proposed sensor can perform essentially all the functions of the combined strain gauge and piezoelectric AE sensors simply and economically (no high-cost material and electronics are required). The proposed sensor approach also is believed to have a robustness suitable for structural health monitoring applications. Comparison of the above sensor approaches is summarized in Table 1.

Table 1. Comparison of Potential Sensors for Structural Health Monitoring

Sensors	Parameter Detected	Installation	Long-Term Stability	Ruggedness	Applicable Material	Implementation Cost
Strain Gauges	Dynamic and Static Strain	Surface Bonded	Poor	Poor	Metals Composites	Low
Fiber Optics	Dynamic and Static Strain	Embedded	*	*	Composites	Moderate to High
Piezoelectric	AE Signals	Surface Mounted	Poor	Poor	Metals Composites	Moderate
Magnetostrictive	Dynamic and Static Strain AE Signals	Surface Bonded, Surface Mounted, Embedded	*	*	Metals Composites	Low to Moderate

*Needs evaluation

Acknowledgment: This work was supported by the Advisory Committee for Research, Southwest Research Institute.

References:

1. T. J. Gerardi. "Health Monitoring Aircraft." *J. of Intell. Mater. Syst. and Struct.*, Vol. 1, July 1990.

2. Teller, C. M., and H. Kwun. "A Feasibility Study for NDE of Bridge Cables Using Magnetostrictively Generated Ultrasonic Waves." Final Report, Southwest Research Institute (SwRI) Project 17-9691 (SwRI, San Antonio, Texas, December 1992).

3. Bozorth, R. M. *Ferromagnetism* (Van Nostrand, New York, 1951), Chap. 13.

4. Joule, J. P. "On the Effects of Magnetism Upon the Dimensions of Iron and Steel Bars." *Phil. Mag.*, Vol. 30, p. 76 (1847).

5. Villari, E. "Change of Magnetization by Tension and by Electric Current." *Ann. Phys. Chem. LpZ.*, Vol. 126, pp. 87-122 (1865).

6. Komarov, V. A., and P. S. Komonov. "Direct and Inverse Electromagnetic-Acoustic Conversion in Ferromagnetic Rods." *Sov. J. NDT*, Vol. 14, pp. 398-403 (1978).

7. Kwun, H., J. J. Hanley, and C. M. Teller. "The Performance of a Noncontact Acoustic Emission Sensor on a Cylindrical Steel Rod." *J. Acoustic Emission*, Vol. 11, No. 1 (April 1993).

8. Burkhardt, G. L., and H. Kwun. "Development of Noncontacting Torque Sensor for Rotating Shafts." SwRI Internal Research Project 17-9637 (ongoing). Op. cit.

9. American Society for Testing and Materials (ASTM) Standard E 976. "Standard Guide for Determining the Reproducibility of Acoustic Emission Sensor Response." (ASTM, Philadelphia, Pennsylvania, 1984).

10. Kwun, H., and C. M. Teller. "Characteristics of Magnetostrictive Generation and Detection of Mechanical Waves in Steel Rod and Strand." (Submitted to *J. Acoust. Soc. Am.)*

11. Tzannes, N. S. "Joule and Wiedemann Effects—The Simultaneous Generation of Longitudinal and Torsional Stress Pulses in Magnetostrictive Materials." *IEEE Trans. on Sonics and Ultrasonics*, Vol. SU-13, pp. 33-41 (1966).

12. Rothbart, A., and L. Rosenberg. "A Theory of Pulse Transmission Along a Magnetostrictive Delay Line." *IRE Trans. on Ultrasonic Engineering*, Vol. PGUE-6, pp. 32-58 (1957).

13. Mohammadi, J., J. A. Cicero, and K. Kasai. "Smart Structures Application to Aircraft Structural Integrity Programs." *Proc. 1989 USAF Structural Integrity Program Conference*, pp. 283-298 (5-7 December 1989).

FIBER OPTIC SENSORS FOR MACHINERY MONITORING AND CONTROL

Henry K. Whitesel
Christopher P. Nemarich
Carderock Division Naval Surface Warfare Center
Machinery Systems Engineering Branch
Annapolis, MD 21402-5067

Michael J. Ransford
E.C.O. Inc.
1356 Cape St. Claire Rd
Annapolis, MD 21401

Abstract: Optical fiber sensors are being developed for the condition monitoring and fault diagnosis of mechanical and electrical equipment on U.S. Navy ships. In some cases, optical fiber sensors will replace the conventional electrical sensors presently in use. However, much of the optical fiber sensor development work is a response to a need for improved machinery monitoring and control. Compared to conventional sensor technology, optical fiber sensors promise improved electromagnetic interference (EMI) immunity, higher reliability, higher measurement resolution and smaller size and weight. Optical fiber sensors have been or are being developed to measure pressure, temperature, liquid level, strain, acceleration, rotational velocity and shaft position and electrical current and voltage. Some of these sensor developments and their application in monitoring and control systems are discussed. The ability to multiplex optical fiber sensors with a single light source and receiver and the potential use of embedded fiber sensors and integrated optics are also presented.

Keywords: Embedded sensors; fiber optics; interferometric sensors; machinery monitoring; optical fiber sensors; polarimetric sensors; pressure sensors; process sensors; sensor multiplexers; temperature sensors

Introduction: Fiber optics technology is advancing rapidly into new applications that extend beyond the communications and data transfer areas into sensors and computers. This paper addresses some of the recent developments in fiber optic sensor technology and its application to shipboard systems. The information contained herein does not reflect official U. S. Navy application strategies on any particular ship.

The primary benefits provided by fiber optics technology to the ship designer are an easy technical solution to (EMI) problems and the ability to vastly increase the ship operators' awareness of system operating status. This results

directly from the increased information flow available from greater sensor populations and higher data transfer rates in shipboard data systems. The primary focus of this paper is on the sensors needed to acquire this larger volume of information. We are on the threshold of new technology breakthroughs in fiber optic sensors that will allow the development and installation of new and improved shipboard monitoring and control systems.

Introduction of Optical Fiber Sensors into Navy Shipboard Equipment:

The U. S. Navy is developing fiber optic sensors to compete directly with existing conventional sensors. (Sensor is defined herein as a means of measuring a parameter such as pressure, temperature and liquid level, resulting in the transmission and display of a signal proportional to that parameter. By this definition, most transducers are sensors.) A typical fiber optic sensor is shown in Figure 1.

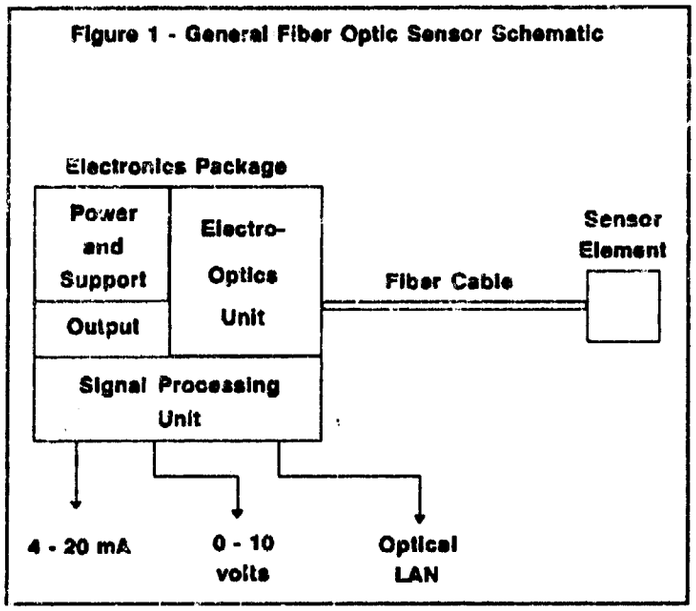


Figure 1 - General Fiber Optic Sensor Schematic

Fiber optic sensor technology offers benefits in the areas of:

- Reduced EMI
- Potential reduced unit measurement cost
- Smaller volume
- Lower weight
- Better accuracy and resolution
- Wider bandwidth
- Distributed and multiplexed sensing
- Embedding capabilities

Each application must be assessed individually, to determine specific benefits.

The application strategy within the U. S. Navy is to develop specifications for fiber optic sensors and to provide them to ship designers for new construction. Backfitting fiber optic sensors will be considered on an individual basis. Private industry will be the primary source of sensor hardware with emphasis on sensors that have commercial application. Off-the-shelf products are being evaluated. In house development programs have been initiated to bridge technology shortfalls.

Fiber Optic Sensor Needs: The size of the Navy sensor market for HM&E systems is small compared to industrial applications. The potential Navy market for fiber optic sensors can be estimated by considering the sensor shipboard population. The total number of sensors on a typical FFG-7 class ship exceed 2200.[1] Continuously indicating types of sensors comprise over 1300 of this total and threshold sensors number just under 900. The number of sensors that could be converted to fiber optic technology was not accurately determined but could be 100% if affordability and performance goals are met. The combination of pressure, temperature, liquid level and electrical voltage and current sensors make up 93% of the total shipboard sensor population.

Fiber Optic Sensor Performance Requirements: Performance requirements for fiber optic sensors to replace electrical equivalents can be summarized from specifications and Navy reports.[2] For machinery monitoring and control, parameters to be measured include pressure, temperature, shaft rotation, position, liquid level, flow rate, strain, acceleration, salinity, torque, load, velocity, relative humidity, dew point, and heat transfer. Pressure must be measured over the range of 0 to 3000 psi with several intermediate ranges allowed. Temperature measurement spans extend from -40°F to 1600°F with several intermediate ranges permitted. Shaft rotation range extends from 0 to 20,000 rpm, over four selected spans; sometimes resolution of a few degrees is required, especially for machinery monitoring applications. Performance requirements for other sensors are presently being reviewed.

Electrical measurements for voltage and current measurements apply to 115 and 450 volts AC at 60 hertz and 400 hertz up to a maximum of about 1500 amperes. DC requirements for submarine systems are nominally 270 volts at up to 3700 amperes. Response times of these sensors must be on the order of a few tens of milliseconds to provide required control of the shipboard electrical systems. In the future, electric drive machinery might extend voltages to 10,000 volts AC at up to 3000 amperes and superconducting machinery systems will require measurements to 500 volts DC at 100,000 amperes maximum.

The performance requirements for damage control monitoring sensors are very similar to those required for machinery control and monitoring systems. A notable exception is that performance in a higher temperature environment is desired to provide monitoring capability during fires. For damage control system applications, temperature should be measured to 1100°C (2012°F) with an accuracy of approximately ±5%. Rate of rise of temperature is also required as this is used to detect the start of a fire. New concepts for damage control monitoring are being presented from both private industry and from several sources within the Navy. We presently have an internal development program to provide measurements of temperature, rate of rise of temperature, smoke, combustion

products, combustible gases, liquid level, flame, hull integrity, fire main pressure and flow rate, hatch and door closure, and air duct flow rate sensors.

In general, the sensors developed for machinery monitoring and damage control must be small, lightweight and affordable. Sensors will be designed with outputs compatible with existing signal transfer formats including direct access to fiber optic local area networks. Operating life should be 40 years with mean time between failures exceeding 10 years. The sensors must also pass requirements for operating in high humidity, salt spray, smoke, and dust.

Fiber optic sensing technology offers a high probability of achieving the above goals, especially in the long term. Some of the test results given later in this paper document progress toward that end.

Fiber Optic Sensing Methods: There are many methods of designing fiber optic sensors that are suitable for shipboard applications. A complete discussion of the methods is beyond the scope of this paper. However, a few of the methods that have been evaluated for shipboard use are illustrated below.

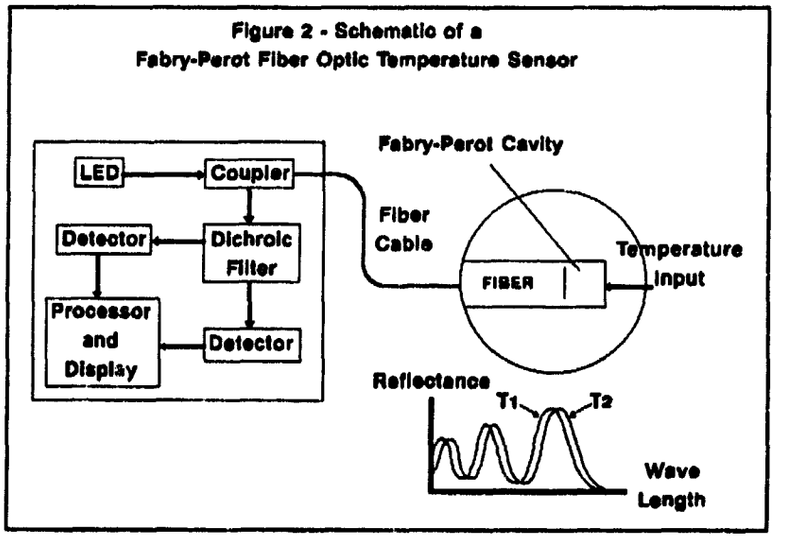


Figure 2 - Schematic of a Fabry-Perot Fiber Optic Temperature Sensor

In general, there are only four optical parameters that can be used to measure machinery control and monitoring parameters. They are optical power, phase, polarization, and wavelength. The usual design strategy is to pick one of these optical parameters that are affected by the parameter to be measured and nothing else. Figures 2 and 3 illustrate how fiber optic technology has been used for measuring temperature and electrical current.

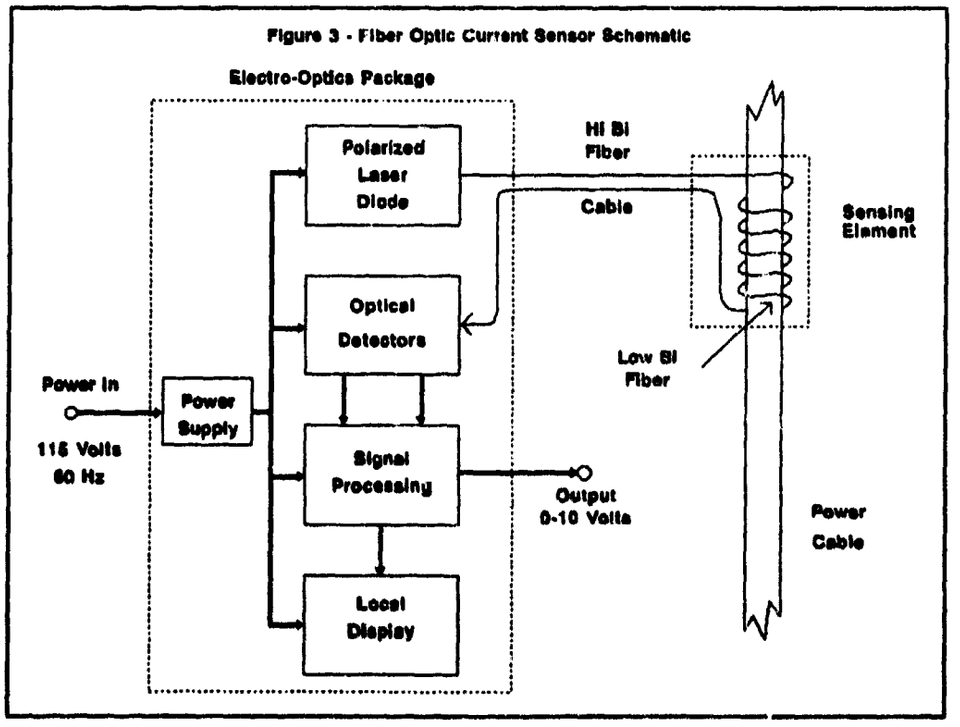
The temperature sensor shown in Figure 2 utilizes interferometric techniques to measure the shift in optical phase across a Fabry-Perot cavity.[3] Two mirrors are designed into the end of a multimode fiber with the material between the mirrors having a known predictable temperature variation of the index of refraction. The reflectance from the cavity resembles an offset sine wave that shifts with wavelength as

a function of temperature. Processing is done by combining the spectrum of the light emitting diode (LED) with the reflectance spectrum of the cavity in a way that splits the resulting spectrum into long and short wavelengths, the ratio of which is proportional to temperature. This sensor has the advantages of very small size and reasonable cost over the range of 0 to 400°C.

There are several potential shipboard applications of the fiber optic temperature sensor. We have used it for fire detection in damage control systems and it can handle most of the machinery control and monitoring applications for HM&E systems.

The electrical current sensor illustrated in Figure 3 uses Faraday rotation of polarized light as the means of detection. Polarized light is generated in a pig-tailed laser diode source and is transmitted through high birefringence fiber to the low birefringence fiber wound around the electrical power cable. The circular magnetic field generated by the electrical current causes a rotation of the state of the polarization of the light in the low birefringence fiber in the sensor spool. Since the magnetic field varies linearly with electrical current, the detection of the polarization in the processing unit is an accurate measurement of the electrical current in the power cable. The advantages of this fiber optic current sensor are complete lack of electromagnetic interference, small size compared with current

Figure 3 - Fiber Optic Current Sensor Schematic



transformers, and a very wide bandwidth with capability of measuring both DC and AC. This sensor was developed jointly with the National Institute of Standards and Technology in Boulder, Colorado.

There are several potential shipboard applications of the fiber optic current sensor. It can be used to collect motor current signatures because of its wide bandwidth and very low frequency response. It can be used in control systems for power generators and distribution systems. Fault current detection is also a natural application for this sensor.

Fiber Optic Sensor Evaluations: In our laboratory at NSWC, Annapolis, we have developed and/or evaluated fiber optic sensors for temperature, electrical current, voltage, liquid level, smoke, and very small fluid flow rates. NSWC, Philadelphia has evaluated fiber optic pressure, temperature, angular velocity and position sensors.

Evaluations of fiber optic temperature and smoke sensors on the ex-USS SHADWELL (a Navy test ship for developing shipboard fire fighting methods) were conducted to determine their ability to detect fires and operate in a shipboard environment. Figure 4 shows results of one of these tests for the fiber optic temperature sensor shown in Figure 2.

The fiber optic temperature sensor was installed in a 1 inch diameter pipe to protect it from structural damage during the fire fighting exercises. The pipe was then threaded through a hole in the deck into the compartment containing the fire. It was mounted in juxtaposition with two thermocouples one of which was in contact with the overhead deck plating and the other was suspended 6 inches below the overhead. Figure 4 shows the temperature indications from both the thermocouple probes and the fiber optic temperature sensors. The test results show favorable comparisons between the thermocouples and fiber optic temperature sensors.

Fiber Optic Sensor Limitations: After conducting these and other evaluations we have reached some conclusions concerning the limitations of applying fiber optic sensor technology on board Navy ships. The performance of fiber sensors is satisfactory in the shipboard environment of high vibration, shock, temperature extremes, and humidity extremes. Fiber sensors are a new technology that have not yet established a credible track record; this will take time and experience. Fiber optic sensors cost more than their conventional counterparts, however, as production volumes increase, their cost is likely to decrease.

The U. S. Navy has recently done a study of fiber optic sensor manufacturability as it relates to availability of sensors and cost. Results show that multiplexing and integrated optics can reduce sensor cost and improve manufacturability.

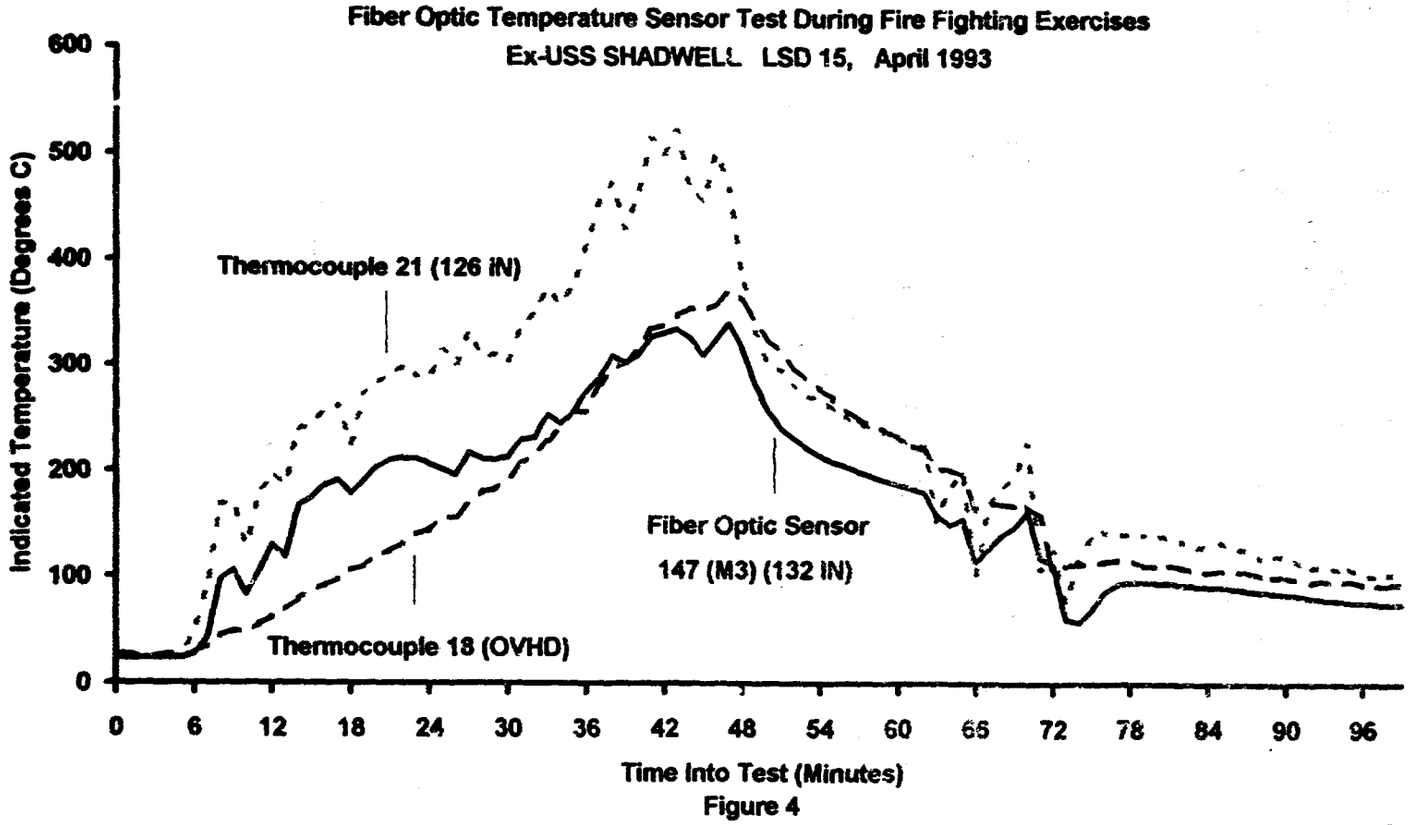

Fiber Optic Temperature Sensor Test During Fire Fighting Exercises
Ex-USS SHADWELL LSD 15, April 1993
Indicated Temperature (Degrees C)
600
500
400
300
200
100
0
Thermocouple 21 (126 IN)
Fiber Optic Sensor
147 (M3) (132 IN)
Thermocouple 18 (OVHD)
0
6
12
18
24
30
36
42
48
54
60
66
72
78
84
90
96
Time Into Test (Minutes)

Figure 4

Multiplexing Fiber Optic Sensors: Five multiplexing schemes were examined for application with Fabry-Perot interferometric sensors[4]. The methods included time division (TDM), frequency division (FDM), wavelength division (WDM), code division (CDM), and coherence (CM) multiplexing. The two methods chosen for further study were TDM and FDM. TDM was chosen for its ease of implementation. A FDM scheme known as FMCW, (frequency modulated continuous wave) was also selected since it could be implemented with shorter fiber delay lines and a CW source. Both methods are proven in communication systems and translate well into optical sensor network designs. Fabry-Perot sensors were selected for multiplexing because their design is relatively mature, they are commercial-off-the-shelf (COTS) products, and some have passed qualifying tests for the shipboard environment.

Time Division Multiplexing: The TDM system configuration uses a sensor tree network with different time delay fiber coils, as shown in Figure 5. A pulse is launched by the light emitting diode (LED) through the 2x2 coupler to the 1x4 splitter sensor tree. Each Fabry-Perot sensor modulates the light pulse corresponding to the parameter sensed e.g. temperature or pressure. Returning pulses are received sequentially at the silicon photodiodes and amplified for processing. Each returned pulse arrives at a time slot associated with a single sensor. Pulses from sensor 1 arrive at time t. Pulses from sensor 2 arrive at time $t+t_d$. Pulses from sensor 3 arrive at $t+2t_d$. Here, t_d is the round trip time of a light pulse through one fiber delay length. By means of accurate timing, the modulated pulses are identified with an individual sensor and the information is processed accordingly.

This multiplexing was successfully tested using a Metricor(TM) pressure sensor. Since the Metricor sensors chosen use a continuous wave (CW) source, it was necessary to provide a pulsed light source. Delay lines were added and an electro-optic amplifier circuit with improved performance was used. Although the tests showed a small amount of hysteresis, they were repeatable, and no problems were incurred taking measurements with 1-2 psi accuracy over a 60 psi range. Presently, we are working to improve the available electro-optic amplifier circuitry in order to achieve the necessary bandwidth and signal to noise ratios required of a fielded system.

Frequency Division Multiplexing: The FMCW system configuration, shown in Figure 6, is similar to the TDM system except that the light source is modulated by a radio frequency (RF) chirp. The sensors are "pulsed" in the frequency domain rather than in the time domain. This modulation technique can resolve individual sensor signals with the use of shorter fiber delay lengths. The modulated sensor signals are recovered in the frequency domain using a phase locked loop demodulator. An inverse Fourier transform of this signal produces a time domain referenced signal. The final result is

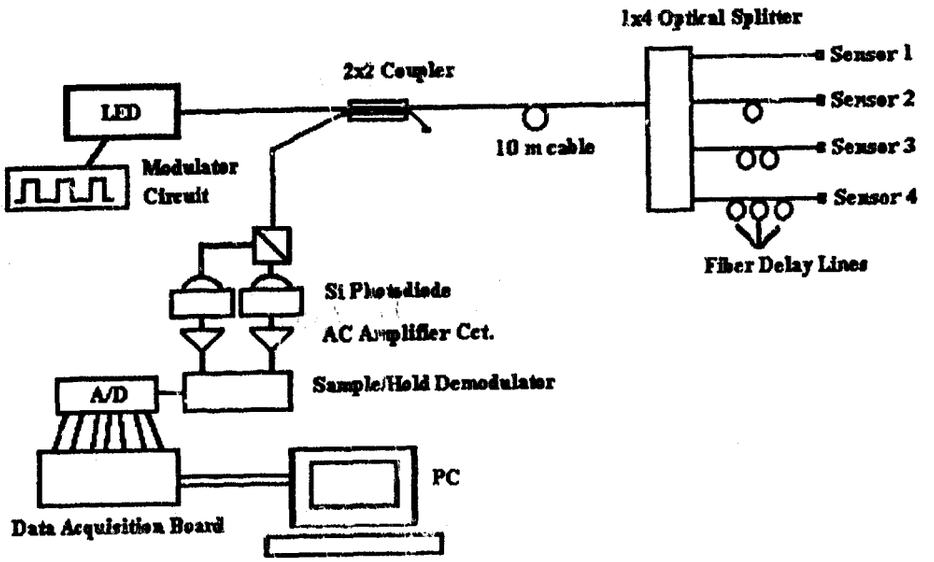


Figure 5 - TDM System Configuration

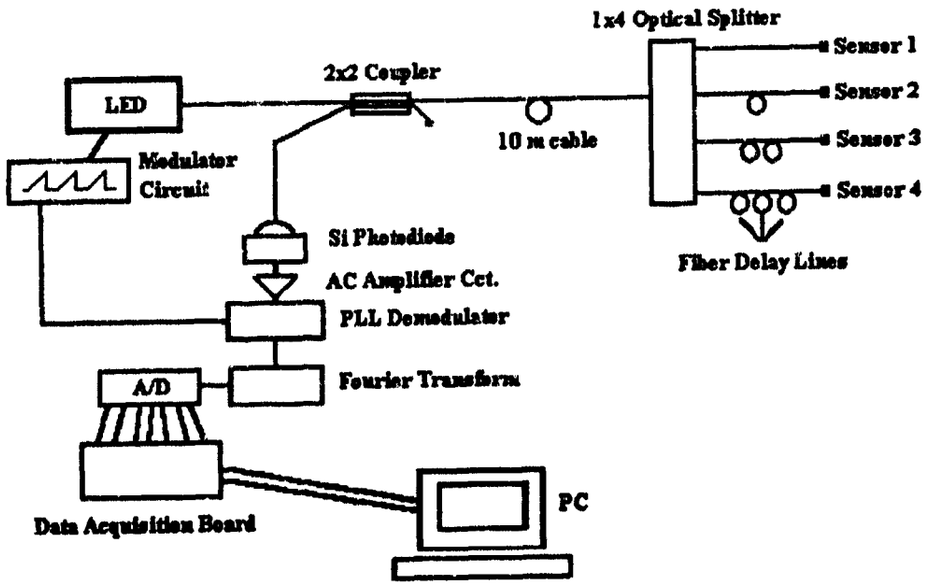


Figure 6 - FMCW System Configuration

a signal which represents the power of the light returned at various distances from the source. For each sensor, the sensed parameter is obtained by taking the ratio of the power reflected from the sensor and the power reflected from a reference point.

The system we tested was implemented with a Hewlett-Packard 8702 lightwave analyzer and a high power LED source. Test results indicate that using the FMCW technique we can accurately resolve to 1 psi with the Metricor pressure sensor. The ability to monitor eight sensors was demonstrated using a 1x8 splitter. This is probably the maximum that can be mulitplexed with this configuration.

A number of issues will need to be resolved before a preferred multiplex architecture can be recommended for shipboard use. Future work includes improving the designs discussed and investigating other techniques.

Integrated Optics: Optical components can be mounted on substrates using similar techniques to those now being used for electronics. The most progress is being made in developing integrated optic components for the telecommunications industry. However, integrated optic circuits are becoming available for fiber optic sensors. One such sensor is the fiber optic gyro that uses an integrated optic circuit to modulate the optical source. As integrated optics are further developed, it is likely that more components will become available that can be used in sensor designs and in associated multiplexing networks.

Embedded Fiber Optic Sensors: Embedded sensors offer the potential to make precise measurements of the operating condition of internal machinery components and provide an indication of wear out or impending failure long before other externally mounted sensors. Our first effort is to demonstrate the ability of an optical fiber pressure sensor to make an accurate measurement inside the cylinder of a high pressure air compressor. This will likely improve the detection and identification of worn valves and rings. The pressure sensor selected is a modified version of a Metricor high pressure Fabry-Perot sensor and is installed through the compressor cylinder head. We measure 1500 PSI at temperatures to 200°C. These tests are presently underway.

Follow-on efforts are focussed on attaching and embedding optical fiber sensors into mechanical structures. To date, optical fibers have been successfully embedded in composite panels. These embedded sensors have been used to measure strain, vibration and temperature inside the host composite. However, not much attention has been paid to embedding optical fibers into metal structures and components. We are currently involved in a joint effort with Virginia Tech to reach these goals.[5] Issues of survivability of optical fibers, the ingress and egress of the fibers and attachment methods such as thermal spray techniques are being addressed.

Summary: Fiber optic sensors are being introduced into the US Navy fleet. These sensors offer a multitude of advantages over conventional sensors in the areas of EMI immunity, cost and size reduction, accuracy, and logistics support. The Navy has defined requirements and implemented programs to specify and develop both machinery monitoring and damage control fiber optic sensors. Finally, research is being conducted in the areas of multiplexing, integrated optics and embedding techniques to further refine the sensors and reduce cost.

References:

1. Whitesel, Henry K., Carl Jacobsen, Frank Leland, Fiber Optic Sensors for Shipboard Applications, Tenth Ship Control Systems Symposium, Ottawa, Ontario, Canada, Vol 4, pp 4-157 to 4-170, 25-29 Oct 1993.

2. Whitesel, Henry K., Sensor Requirements and Availability for Shipboard Auxiliary Machinery, Phase I - Range and Accuracy, David W. Taylor Research and Development Center report DTNSRDC/PAS-79-14, Sept 1979.

3. Hartle, J., and E. Saaski, U. S. Patent No. 4,678,904.

4. Ransford, Michael J., Henry R. Hegner, Design of Optical Multiplexing System for Fiber Optic Sensors, prepared for CDNSWC under contract N651533-88-D-0055, 30 Sept 1993.

5. Sun, D., S. Poland, K. Kali, and R.O. Claus, Optical Fiber Embedment and Attachment Techniques and Survivability Issues, Prepared for CDNSWC by the Fiber and Electro-Optics Research Center at Virginia Tech, 30 Sept 1993.

Acknowledgments: The US DoD sponsored this work and since it represents U. S. government work it is not subject to copyright. Financial support for a portion of the development and evaluation came from James Gagorik of the Office of Naval Research, from Dr. Patricia Tatem of the Naval Research Laboratory, and from Donald Kover of CDNSWC. John Overby, David Kocsik and Eric Hawn of CDNSWC assisted in data acquisition, analysis and presentation of this paper.

FUZZY SET AND LOGIC APPROACHES FOR THE DIAGNOSIS OF A HIGH PRESURE AIR COMPRESSOR (HPAC)

Bruce Friedman
Carderock Division, Naval Surface Warfare Center
3A Leggett Circle (Code 612)
Annapolis, Maryland 21402-5067

Abstract: Fuzzy reliability theory which is a generalization of crisp reliability theory is applied to the diagnosis of a high pressure air compressor (HPAC). Instead of a binary value, one or zero, being associated with each component, the value associated with each component can range from zero through one so that a particular numerical value represents the degree of functioning of a component. The analysis is performed using both series and parallel coherent structure theory. Since the results attained do not show significant improvement relative to an earlier HPAC investigation utilizing fuzzy sets, it is concluded that further research is required into the theory of fuzzy coherent structures.[1]

Key Words: Fuzzy sets; fuzzy logic, diagnosis, air compressor, valves

Introduction: Successful development of techniques to deal with the uncertainty and incompleteness of mechanical diagnostic data is required for the deployment of effective mechanical diagnostic systems. The operation of mechanical diagnostic systems would improve equipment availability at lower costs while reducing dependency on the logistics system for manpower and material.

Changing to a Condition - Based Maintenance (CBM) philosophy is necessary in order to meet the mission requirements of the future in view of decreasing availability of manpower and increasing sophistication of vehicles. The monitoring system allowing unambiguous diagnosis and failure prediction required to maintain user confidence will not be affordable. It is then necessary to develop techniques to reduce uncertainty in the presence of incomplete and probabilistic data.

A high pressure air compressor (HPAC) has been instrumented as a test bed for the CBM program and data acquired. Fuzzy sets and logic have been successfully applied to the analysis of this HPAC data. [1] This analysis entailed the correlation of derived membership

functions with expert opinion in order to obtain estimates of the likelihood of the presence of a fault in a given compressor valve. This expert opinion is formulated in terms of linguistic variables. This type of formulation is an indication that the application of fuzzy sets and logic would be valid and productive.

Objective: The objective of this current study was to ascertain more advanced approaches for utilizing fuzzy set membership functions for the diagnosis of mechanical systems with emphasis on applying these approaches to HPAC operation.

Background: The earlier study was based upon the concept of conditional fuzzy sets as applied to mechanical diagnostics. [1, 2] The rationale behind this approach derives from the recognition that an operating device is composed of numerous interacting components and factors. These interacting components and factors give rise to a description of the device in terms of a set of influences acting on the system which are related to a set of symptoms that depict the degradation of the device. In the real world, it is not unusual for a symptom to act as an influence and vice versa.

The first step in applying the conditional fuzzy set approach (CFS) to the HPAC analysis is to compute the pressures and temperatures of the compressor stages upon the basis of a simplified model.[1, 2] Figure (1) displays a plot of the computed and measured baseline pressure for the first stage as a function of running time. The HPAC baseline represents normal operating conditions. The HPAC is also run with faults either in the suction or discharge valves of the third stage.

The measured value of the pressure and temperature for each stage during a given run is compared with its computed baseline value.[1] Five levels for the values of the ratio of the measured to computed values are designated. According to its level, each pressure and temperature is assigned a degree of membership in the linguistic variable classes: LOW (=very low), low, high, or HIGH(= very high) as shown in Table (1).

Table (2) presents the anticipated effects or symptoms of internal air leakage on discharge pressures and temperatures. [1] This table in conjunction with the determination of the degrees of membership of the pressure and temperature indicators allows the likelihood of a defect in each of the eight valves to be calculated. Each variable in Table (2) has its own membership value. The likelihood value is found by adding the membership

values for a given row yielding the likelihood for the defect on that row.

Figure (2) shows the likelihood of defects for runs with a defective third stage discharge valve where S1 is the numerical value of the likelihood of a fault in the first stage suction valve and D1 is the value of the likelihood of a fault in the first stage discharge valve with S2, D2, S3, D3, S4 and D4 being the likelihood values for faults in the appropriate valves in the second, third and fourth stages, respectively. (1) The diagnostic results for the run show a configuration that correctly indicates the imposed fault for the run.

Approach: To obtain a means for finding likelihood values of enhanced accuracy, it is necessary to commence with the theory of the structural properties and reliability of coherent systems. [3]

It is assumed that the state of a system is completely determined by the states of its components. [3] The function describing the system state is known as the structure function. The order of the system refers to its number of components. A component is irrelevant if the structure function does not change with a change in this component state function. A component is relevant if the structure function does change with a change in this component state function. A system is coherent if its structure function is increasing in each argument and each component is relevant.

In traditional crisp reliability theory a binary value is associated with each component, the value being one if the component is functioning and zero if the component is not functioning. [3] In fuzzy reliability theory, the value associated with each component can range from zero through one so that a particular numerical value represents the degree of performance or quality of functioning of a component. [4, 5, 6]

The test bed HPAC has a series structure. The structure function for a series structure is equal to the product of the individual component states.

For each structure there exist two reliability functions. The traditional reliability function is probabilistic in nature giving the probability that the system is functioning in terms of the probabilities that each of the components is functioning. The fuzzy set reliability function is possibilistic in nature giving the possibility that the system is functioning in terms of

the possibilities that each of the components is functioning. [3, 4, 5, 6]

The series system can be considered to be the least complex of all systems. However, from the probabilistic point of view, even the series system can display subtleties in the course of analysis. Consider a two component series system. If either component fails (malfunctions) then the entire system fails (malfunctions). The components are statistically independent. If one component fails then it is known that the other component did not fail. The failure of one component excludes the failure of the other component implying that knowledge of one component failure event entails a change in the probability of the other component failure event. This state of affairs leads to the seeming paradox that the events represented by the failure of either of the components are not statistically independent. An avoidance of this apparent paradox requires making a careful distinction between the statistically independent competing processes which are transpiring before one of the risks of failure is realized and the mutually exclusive states of the system at a given instant of time. The processes refer to the operation of the two components while the states refer to failed or non-failed modes of existence. [7, 8]

An alternate route to the resolution of the apparent paradox is to consider the operations of the two components not to be independent but to be associated. With associated operations fluctuations in the operation of one component generate fluctuations in the operation of the other component and vice versa. The covariance of the two fluctuations is equal to or greater than zero. [3]

The fuzzy sets approach to reliability and diagnostics utilizing linguistic variables as expressed in Table (2) avoids the seeming paradoxical aspect of the traditional probabilistic approach.[9] The evaluations recorded in Table (2) already incorporate knowledge concerning interactions and associations between the operations of the components of the HPAC. These evaluations record the observations and experiences of designers and engineers.

Let $\mu_{jk\ell}(x_{jk})$ be a fuzzy set membership function where j=1 and j=2 denote the measured observed physical variables of discharge pressure and discharge temperature, respectively.[9] The index k=1, 2, 3, 4 labels the particular stages. The index ℓ =1,... 8 denotes the valves so that ℓ =1 labels the 1st stage suction valve, ℓ =2 labels the 1st stage discharge valve, ℓ =3 labels the

2nd stage suction valve, etc., The symbol $A_{\ell k}$ indicates the fuzzy set defined for a given ℓ value and a given k value. There are 16 fuzzy sets defined by the $A_{\ell k}$ since each of the linguistic variables LOW, low, high, HIGH indicates a fuzzy set which is further decomposed into four fuzzy subsets. There is one subset for each of the four compressor stages. The variable LOW for the 1st stage discharge pressure labels a different fuzzy subset than LOW for the 2nd stage discharge pressure. The decomposition into subsets originates from the situation that the discharge pressures and temperatures differ in their normal values in going from one stage to another. For example, for a discharge pressure of 5000 psi and standard operating and atmospheric conditions, the normal 1st stage, 2nd stage and 3rd stage discharge pressures are 45 psi, 270 psi, and 1110 psi, respectively.[10] A discharge pressure of 150 psi might be linguistically described as HIGH or high if it is displayed by the 1st stage, as low if it is displayed by the 2nd stage or as LOW if it is displayed by the 3rd stage.

Table (3) presents values for $\mu_{A_{\ell k}}(x_{jk})$ for two runs of the test bed HPAC. There are no values for the 4th stage discharge pressure because this output pressure is considered to be a parameter determined by the operating conditions. One run was made with a defective 3rd stage suction valve (3SV), while the other run was made with a defective 3rd stage discharge valve (3DV). Consider the two rows of numbers following the left side label "1st stage suction". The first row is membership values with a defective 3rd stage suction valve present, while the second row is membership values with a defective 3rd stage discharge valve present. The rows of numbers following the other left-side labels are to be similarly interpreted. These membership values in non-normalized form, were used in the earlier investigation.[1] For the third stage suction valve defect the likelihood calculation based upon summing across the row did not yield a reasonable diagnosis since identical maximum estimates are found for the 1st stage suction, 1st stage discharge, 2nd stage discharge and 3rd stage discharge valves. For the third stage discharge valve defect a reasonable diagnosis is obtained since one peak estimate of defect likelihood is found for the 3rd stage discharge valve.

To improve the likelihood estimates and diagnostic capability it is necessary to utilize a knowledge of the properties of complex coherent systems which includes an understanding of the relationship between the physical system components and the events occurring within the system. The structural properties of the physical system

are not necessarily the same as the structural properties of the events generated by that system. This difference is illustrated by the operation of the testbed HPAC.

The series nature of the physical structure of the HPAC leads to both the structure function and the probability of malfunction of the system being described in terms of a product either of component states or probability of component malfunctions, respectively. For the possibilistic interpretation of the membership functions the product description uses the possibility of malfunction of each component. The basic events used for the analysis are deviations of pressure and temperature from their nominal or normal values. These deviations are recorded as actual pressure readings and as linguistic variable ratings. [3-6]

For the initial analysis of HPAC operation, the symptomology is investigated as the combination of two groups of symptomatic events. One group comprises pressure deviations while the other group comprises temperature deviations. The likelihood of a defect or malfunction occurring at a particular air valve is a function of these two groups. For each group of symptoms, the consideration of the symptoms is to be undertaken concurrently or in parallel. From coherent structure theory, [3-6] the likelihood $\phi_{\ell j}$ of a defect at valve $\ell = 1, \ldots, 8$ as deduced from discharge pressure (j=1) or discharge temperature (j=2) is given by

$$\phi_{\ell j} = 1 - \prod_{k=1}^{4} (1 - \mu_{A_{\ell k}}(x_{jk})) \,. \qquad (1)$$

There are two assumptions upon which the formulation of eq. (1) is based in addition to the assumption of parallelism. One assumption is that the relationship between coherent structure theory and possibility or fuzzy set theory is similar to that between coherent structure theory and probability theory. The other assumption is that the symptoms are possibilistically independent in analogy to the concept of statistical independence.

The function $\phi_{\ell j}$ is reminiscent of the work of Jain [11, 12] to obtain an extension principle to provide a general method for extending nonfuzzy mathematical concepts in order to deal with fuzzy quantities. From this viewpoint $\phi_{\ell j}$ would be a fuzzy membership function whose support would be a function of the pressures or temperatures at each of the discharge stages. The fuzzy set corresponding to $\phi_{\ell j}$ is obtained from the four fuzzy sets $A_{\ell k}$ (ℓ fixed, k=1,2,3,4). Dubois and Prade [13]

however, noted that in general a function of the $A_{\ell k}$ is a classical subset of the support provided by the pressures or temperatures when the Cartesian product of the supports of the $A_{\ell k}$ is the set of real numbers and continuous membership functions are considered, so that the result, eq.(1), depends only on the support of the $A_{\ell k}$. For the purposes of this investigation the extent to which $\phi_{\ell j}$ is a manifestation of a fuzzy extension principle is of no concern.

The likelihood ϕ_ℓ of a defect or malfunction as deduced from both pressure and temperature deviations is given by

$$\phi_\ell = \phi_{\ell 1} + \phi_{\ell 2} - \phi_{\ell 1}\phi_{\ell 2} \quad , \tag{2}$$

where again the two groups $\phi_{\ell 1}$ and $\phi_{\ell 2}$ are considered concurrently (parallel).[3]

Expanding the right side of eq.(1) gives

$$\begin{aligned}\phi_{\ell j} = [&\mu_{A\ell 1} + \mu_{A\ell 2} + \mu_{A\ell 3} + \mu_{A\ell 4} - \mu_{A\ell 1}\mu_{A\ell 2} - \mu_{A\ell 1}\mu_{A\ell 3} \\ &- \mu_{A\ell 2}\mu_{A\ell 3} - \mu_{A\ell 2}\mu_{A\ell 4} - \mu_{A\ell 1}\mu_{A\ell 4} - \mu_{A\ell 3}\mu_{A\ell 4} \\ &+ \mu_{A\ell 1}\mu_{A\ell 2}\mu_{A\ell 3} + \mu_{A\ell 1}\mu_{A\ell 2}\mu_{A\ell 4} \\ &+ \mu_{A\ell 1}\mu_{A\ell 3}\mu_{A\ell 4} + \mu_{A\ell 2}\mu_{A\ell 3}\mu_{A\ell 4} \\ &- \mu_{A\ell 1}\mu_{A\ell 2}\mu_{A\ell 3}\mu_{A\ell 4}]_j ,\end{aligned} \tag{3}$$

where the subscript j outside the parenthesis indicates that the support is either a pressure variable or a temperature variable.[3]

Upon examining eq. (3) it is observed that the likelihood function used in the earlier work [1] can be considered to be an approximation to eqs (1)-(3) with the application of normalized membership functions. If the products of two or more membership functions are neglected then as in the earlier work the likelihood function is simply the sum of the membership functions. The product terms could be viewed as being correction factors.

Discussion and Results: For the two sets of data presented in Table (3) the likelihood functions ϕ_ℓ are calculated through terms involving products of two membership function values with the product terms for three and four values being neglected. Table (4) presents the values of $\phi_{\ell 1}^{(r)}$, $\phi_{\ell 2}^{(r)}$, and $\phi_\ell^{(r)}$ for the two sets of data for the two cases of the sum approximation (r=0) to the product function and with the products of two membership functions corrections (r=1). The values for $\phi_\ell^{(0)}$ are the same as those displayed by the bars of Figure (2) taking into account that the bars are based upon non-normalized numbers.

In Table (4a) it is observed that the $\phi_\ell^{(1)}$ values are similar to the $\phi_\ell^{(0)}$ values in that there is no clear indication of a defective third stage suction valve being present. The two lowest $\phi_\ell^{(0)}$ values occur for ℓ =7, 8 while the two highest $\phi_\ell^{(1)}$ values occur for ℓ =7,8.

For the case of the third stage discharge valve being defective, it appears that $\phi_\ell^{(0)}$ yields a better result than $\phi_\ell^{(1)}$. The maximum value of 4.15 for $\phi_\ell^{(0)}$ agrees with the location of the known defective valve, while the two largest values of $\phi_\ell^{(1)}$ occur at ℓ =4,8. (Table (4b))

The apparent lack of improvement in the calculated results in going from $\phi_\ell^{(0)}$ to $\phi_\ell^{(1)}$ can have several reasons. One reason is that the actual state of the HPAC is difficult to determine since it has been used as a test bed for several programs over a number of years and may not have been restored to its original condition. Defects may be present which are not known to the current generation of experimenters. The linguistic variable ratings of Table (2) may have surrendered some of their original validity.

Another reason is that the mathematical route leading to $\phi_\ell^{(1)}$ may not be the most appropriate calculational approach towards improving the values of $\phi_\ell^{(0)}$. A modification that can be made is to use only membership values above a certain level α . An argument for this modification is that the noise level of the experiment tends to obscure the significance of values below a certain level which is taken to be the α level or cut. To examine this modification take α=0.20 . Some of the values in Tables (4a) and (4b) are then changed because α=0.20 means that the membership values of .17 in Table (3) are to be neglected. Since the values of Table (3) are used to calculate the values of Table (4), the rows of ℓ values can have one or more of their boxed values changed. In Table (4a), reading from left to right, the present values for the ℓ=7 row are replaced by

1.33, 1.33, 2.66, 0.75, 0.75, and 0.94, respectively. The other rows in Table (4a) remain unchanged. Similarly, in Table (4b) the ℓ =1 row is replaced by 0.50, 2.17, 2.67, 0.50, 0.42 and 0.71, respectively, the ℓ =2 row is replaced by 0.50, 2.16, 2.66, 0.50, 0.48, and 0.74, respectively and the ℓ =7 row is replaced by 1.99, 1.33, 3.32, 0.75, 0.75, and 0.94 respectively. The other rows in Table (4b) remain unchanged. These altered values in Tables (4a) and (4b), however, do not enhance the discrimination of the defective valves.

It is to be noted that the application of the fuzzy set membership function utilized in this analysis to this point represents only one approach among a number of approaches to the reliability and performance determination of complex machinery using fuzzy sets. (Differing approaches have been formulated by Kaufmann and Gupta [4], Kaleva, [5] and Montero. [6]) In the approach utilized herein the membership function is interpreted as being a possibility distribution in analogy to the more commonly used probability function. This interpretation allows the membership functions to be manipulated in a manner similar to the manipulations of probability functions which similarity has been utilized in this investigation. The possibilistic sum has the same form as the probabilistic sum.

Table (4) shows the assumption that the symptoms are to be treated as if they operated in parallel with each other. Since the results obtained under this assumption do not correspond closely with the known placement of defects, the likelihood functions are now obtained with the replacement of this assumption by the assumption that the symptoms operate in series. This alternate assumption does correspond to the actual physical state of affairs since the compressor stages are connected in series. Under this alternate assumption, the membership values on each line of Table (3) are multiplied together. For the case of a defective third stage suction valve the results in going through the valves from the first through fourth stages are .0069, .0056, .0037, .0056, .0037, .0056, .0012 and .0000, respectively. The corresponding results for a defective third stage discharge valve are .0000, .0000, .0000, .0000, .012, .016, .0032, and .0000, respectively. (See Table 5).

Conclusions: Comparing results between the parallel and series computational models for a defective third stage suction valve, it was found that the results for the series model are no better and are even somewhat worse that those displayed by the values of $\phi_{\ell}^{(0)}$ and $\phi_{\ell}^{(1)}$. For a defective third stage discharge valve, however, the

series model yields a marked improvement relative to $\phi_2^{(p)}$ and $\phi_2^{(v)}$ values.

Further research is required into the theory of fuzzy coherent structures in order to improve the mechanical diagnostics situation utilizing available data.

Acknowledgement: This research was funded by Carderock Division, Naval Surface Warfare Center (CDNSWC), Independent Research Program, sponsored by the Office of Naval Research, ONR10, and administered by the Research Director, CDNSWC 0112, Dr. Bruce Douglas. This work was supervised within the Metals and Welding Department (Code 61) by Dr. O.P. Arora. The numerical values for membership functions were calculated by E.A. Schroeder (CDNSWC, Code 1282).

References

1. B. Friedman, E. A. Schroeder, and G.W. Nickerson, "Reasoning from Uncertain Data for Mechanical Diagnostics," Proceedings of Workshop on Artificial Intelligence for Customer Service and Support, Eighth IEEEE Conference on Artificial Intelligence Applications, March 2-6, 1992, Monterey, California.
2. I. Gazdik, "Fault Diagnosis and Prevention by Fuzzy Sets", IEEE Transactions on Reliability, vol R-34, 1985, pp. 382-388.
3. Richard E. Barlow and Frank Proschan, Statistical Theory of Reliability and Life Testing: Probability Models, (Holt, Rinehart and Winston, Inc., New York, 1975).
4. Arnold Kaufmann and Madan M. Gupta, Fuzzy Mathematical Models in Engineering and Management Science, (North Holland, Elsevier Science Publishers, Amsterdam - New York, 1988)
5. Osmo Kaleva, "Fuzzy Performance of a Coherent System." Journal of Mathematical Analysis and Applications, vol 117, 1986, pp 234-246.
6. F.J. Montero "Fuzzy Coherent Systems," Kybernetes, vol 17, no. 4, 1988, pp. 28-33.
7. R.A. Evans, "Independent or Mutually Exclusive", IEEE Transactions on Reliability, vol R-29, no. 4, October 1980, pg 289.
8. Yuji Nakagawa, "On a Paradox Concerning Independent vs Mutually Exclusive Events", IEEE Transactions on Reliability, vol R-30, no. 5, December 1981, pp. 448-449.
9. George J. Klir and Tina A. Folger, Fuzzy Sets, Uncertainty, and Information, (Prentice Hall, Englewood Cliffs, New Jersey, 1988)
10. High Pressure Class AA Air Compressor (Submarine Service) 13.0 CFH-4500 PSI 440 VAC Motor (Service Manual), (NAVSEA 0349-LP-073-0000), December 1964, change 815, April 1981.

11. Ramesh Jain, "Decision Making in the Presence of Fuzzy Variables", IEEE Transactions on Systems, Man and Cybernetics, vol R-25, October 1976, pp. 698-703.
12. Ramesh Jain, "Tolerance Anaylsis Using Fuzzy Sets" Int. J. Syst Sci. vol 7, no. 12, 1976, pp 1393-1401.
13. Didier Dubois and Henri Prade, Fuzzy Sets and Systems: Theory and Applications. (Academic Press, Inc., Orlando, Florida 1980).

Table (1) DEGREE OF MEMBERSHIP

	level 1	level 2	level 3	level 4	level 5
LOW	1.00	0.67	0.33	0.00	0.00
low	0.50	0.83	0.50	0.17	0.00
high	0.00	0.17	0.50	0.83	0.50
HIGH	0.00	0.00	0.33	0.67	1.00

TABLE (2) EFFECT OF INTERNAL AIR LEAKS ON INTERSTAGE PRESSURES & TEMPERATURES

Leakage at:	DISCHARGE PRESSURE				DISCHARGE TEMPERATURE			
	1ST STAGE	2ND STAGE	3RD STAGE	4TH STAGE	1ST STAGE	2ND STAGE	3RD STAGE	4TH STAGE
Valves								
1st Stage Suction.....	LOW	low	low	----	low	high	high	HIGH
1st Stage Discharge....	LOW	low	low	----	HIGH	high	high	high
2nd Stage Suction.....	HIGH	LOW	low	----	HIGH	low	high	high
2nd Stage Discharge....	high	LOW	low	----	high	HIGH	high	high
3rd Stage Suction.....	high	HIGH	LOW	----	high	HIGH	low	high
3rd Stage Discharge....	high	high	LOW	----	high	high	HIGH	high
4th Stage Suction.....	high	high	HIGH	low	high	high	HIGH	low
4th Stage Discharge....	----	high	HIGH	low	----	high	high	HIGH

Table (3) MEMBERSHIP VALUES FOR RUNS WITH DEFECTIVE VALVES

	DISCHARGE PRESSURE (STAGE)				DISCHARGE TEMPERATURE (STAGE)			
LEAKAGE AT AIR VALVES	1ST	2ND	3RD	4TH	1ST	2ND	3RD	4TH
1ST STAGE (3SV) SUCTION (3DV)	.33 .00	.50 .17	.50 .50	--- ---	.50 .50	.50 .50	.50 .50	.67 .67
1ST STAGE (3SV) DISCHARGE (3DV)	.33 .00	.50 .17	.50 .50	--- ---	.33 .33	.50 .50	.50 .50	.83 .83
2ND STAGE (3SV) SUCTION (3DV)	.33 .67	.33 .00	.50 .50	--- ---	.33 .33	.50 .50	.50 .50	.83 .83
2ND STAGE (3SV) DISCHARGE (3DV)	.50 .83	.33 .00	.50 .50	--- ---	.50 .50	.33 .33	.50 .50	.83 .83
3RD STAGE (3SV) SUCTION (3DV)	.50 .83	.33 .67	.33 .33	--- ---	.50 .50	.33 .33	.50 .50	.83 .83
3RD STAGE (3SV) DISCHARGE (3DV)	.50 .83	.50 .83	.33 .33	--- ---	.50 .50	.50 .50	.33 .33	.83 .83
4TH STAGE (3SV) SUCTION (3DV)	.50 .83	.50 .83	.33 .33	--- ---	.50 .50	.50 .50	.33 .33	.17 .17
4TH STAGE (3SV) DISCHARGE (3DV)	.00 .00	.50 .83	.33 .33	--- ---	.00 .00	.50 .50	.50 .50	.67 .67

Table (4) Likelihood Functions (Parallel Model)

a. 3rd stage suction valve defective						
	$\phi_{\ell 1}^{(0)}$	$\phi_{\ell 2}^{(0)}$	$\phi_{\ell}^{(0)}$	$\phi_{\ell 1}^{(1)}$	$\phi_{\ell 2}^{(1)}$	$\phi_{\ell}^{(1)}$
ℓ=1:	1.33	2.17	3.50	0.75	0.42	0.85
ℓ=2:	1.33	2.16	3.49	0.75	0.48	0.87
ℓ=3:	1.16	2.16	3.32	0.78	0.48	0.88
ℓ=4:	1.33	2.16	3.49	0.75	0.48	0.87
ℓ=5:	1.16	2.16	3.32	0.72	0.48	0.85
ℓ=6:	1.33	2.16	3.49	0.75	0.48	0.87
ℓ=7:	1.33	1.50	2.83	0.75	0.70	0.92
ℓ=8:	0.83	1.67	2.50	0.66	0.75	0.92

b. 3rd stage discharge valve defective						
	$\phi_{\ell 1}^{(0)}$	$\phi_{\ell 2}^{(0)}$	$\phi_{\ell}^{(0)}$	$\phi_{\ell 1}^{(1)}$	$\phi_{\ell 2}^{(1)}$	$\phi_{\ell}^{(1)}$
ℓ=1	.67	2.17	2.84	0.58	0.42	0.76
ℓ=2	.67	2.16	2.83	0.58	0.48	0.78
ℓ=3	1.17	2.16	3.33	0.84	0.48	0.91
ℓ=4	1.33	2.16	3.49	0.92	0.48	0.96
ℓ=5	1.83	2.16	3.99	0.78	0.48	0.88
ℓ=6	1.99	2.16	4.15	0.75	0.48	0.87
ℓ=7	1.99	1.50	3.49	0.75	0.69	0.92
ℓ=8	1.16	1.67	2.83	0.89	0.75	0.97

Table (5). Likelihood Functions (Series Model)

	3RD STAGE SUCTION VALVE DEFECTIVE	3RD STAGE DISCHARGE VALVE DEFECTIVE
ℓ =1	.0069	.0000
ℓ =2	.0056	.0000
ℓ =3	.0037	.0000
ℓ =4	.0056	.0000
ℓ =5	.0037	.012
ℓ =6	.0056	.016
ℓ =7	.0012	.0032
ℓ =8	.0000	.0000

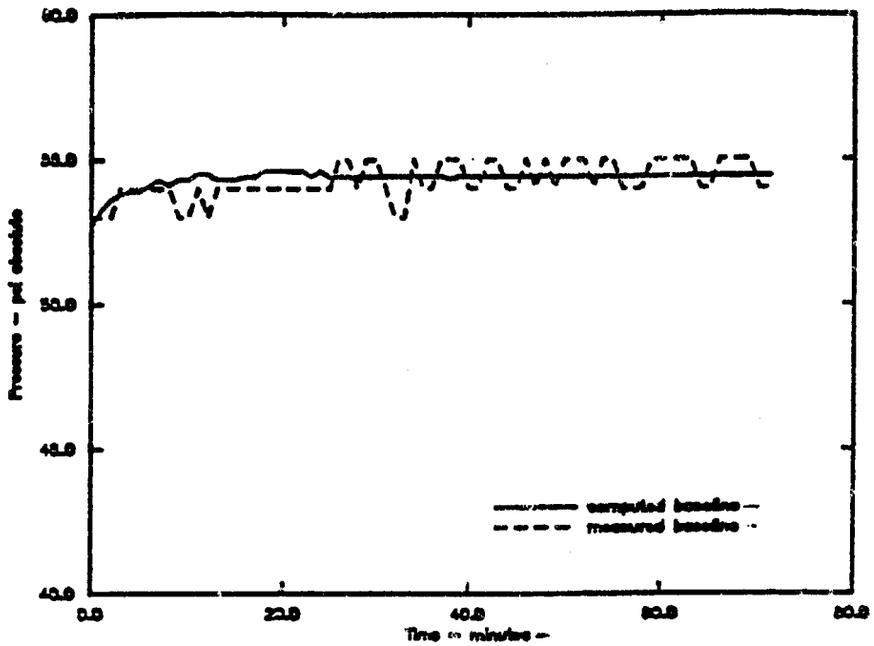
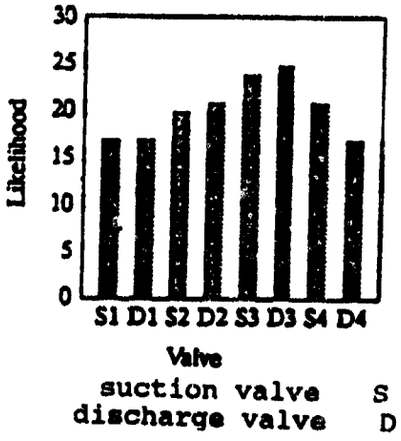


Figure 1. Computed and Measured Pressures - Stage 1



suction valve S
discharge valve D

Figure 2. Likelihood of Defect for Run with Defective Stage 3 Discharge Valve

APPLICATIONS OF DIAGNOSTICS

Cochairmen: Reginald G. Mitchiner
Virginia Polytechnic Institute

Kam W. Ng
Office of Naval Research

AN INTEGRATED APPROACH TO MANUFACTURING EQUIPMENT CONDITION MONITORING AND PRODUCT IN-PROCESS CONTROL

Gary W. Carriveau
Nondestructive Testing Information Analysis Center (NTIAC)
TRI/Austin
415 Crystal Creek Drive
Austin, Texas 78746

Abstract: Among the most important goals of a production facility are the assurance that the production plant manufacturing equipment continuously meets or exceeds acceptable performance levels and that the final product is the highest acceptable quality. Monitoring systems to assess the performance of both independent functions have been developed in the past. These have had a range of levels of sophistication and success. The purpose of this paper is to look beyond these independent functions to an expanded development which capitalizes on the technology infrastructure found in some modern NDE sensor-based process inspection and/or control systems. The objective of this enhanced system is to integrate NDE methodologies found in both "state-of-health" and in-process control monitoring systems with the end result an improvement in process equipment maintenance and repairs. The key innovation in this enhanced system is that it can employ hardware and software components already existing in many in-process inspection or control systems. The primary advantages of this approach, in addition to improved product quality, are to increase equipment productivity by reducing equipment down time, reduce repair part expenditures, and improve equipment maintenance procedures. This paper contains a review of NDE sensors applied to state-of-health monitoring, and in-process inspection and/or control; typical applications of these systems; cost/benefit analysis of a typical system; and a detailed discussion of the integrated system including advantages that may be expected when it is used.

Key Words: Equipment diagnosis; expert systems; NDE sensors; nondestructive evaluation; predictive maintenance; process control; productivity improvement; quality improvement; state-of-health monitoring

Introduction: The objective of using nondestructive sensor-based manufacturing and material process control is to improve the quality and performance of end products. Effective inspection/test/evaluation results in an increase in efficiency in production methods and should also result in the reduction of material wastes and costly rework to correct nonconforming items. Furthermore, product production rates may be accelerated and the time to produce new products and get them into use can be reduced.

Inspection/test/evaluation methods usually fall into three general categories. These are when methods are applied to the end of the process, when they are

concurrent with the process, and when inspection/test/evaluation information is used to control the process.

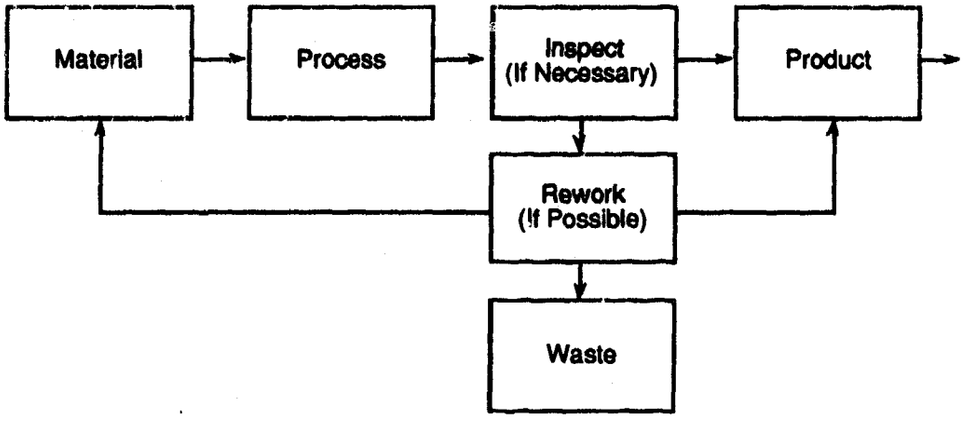
End of process (post process) inspection/test/evaluation, where information is not gathered until the process is completed, has been the most commonly found application over the longest period of time. Inspection/test/evaluation can be accomplished using both nondestructive and destructive methods. A traditional "final test and inspection" suffers in that product deviations are found too late resulting in expensive reworking or scrap, both introducing unnecessary waste. This approach was most often used in the past, however, technology and manufacturing philosophy advances have clearly illustrated the inherent weaknesses and changes are now being made to more advanced systems.

In-process inspection/test/evaluation departs from the conventional systems described above by assessing a number of process parameters during the process, and using this information to assist in the control of the process. Many modern state-of-the-art process systems use some form of in-process inspection/test/evaluation. Through implementation of these, product deviations can be eliminated or corrected before more value is added, thereby reducing rework and waste.

The future is seen in a full application of inspection/test/evaluation information in process control. Here, information from a complete set of parameters is used with a computer-based mathematical model of the process and real-time decisions are made by an expert system to control the process. This results in products that are in full conformance with required specifications, thereby eliminating waste, enhancing availability, and improving quality.

The same hardware/software in-process control infrastructure may also be used as the basis of advanced state-of-health monitoring of the product processing equipment. An integrated system not only achieves reduced waste, it also has the potential for a significant improvement in production efficiency.

Conventional End of Process Inspection: The primary feature of a conventional end of process inspection is that all of the inspection/evaluation/testing is performed at the conclusion of the process, after the processing value has been added to the product. A block diagram of a typical end of process inspection scheme is shown in Figure 1. In this and all subsequent examples, the process and/or inspection may contain many steps. They are shown as only one block for simplicity.

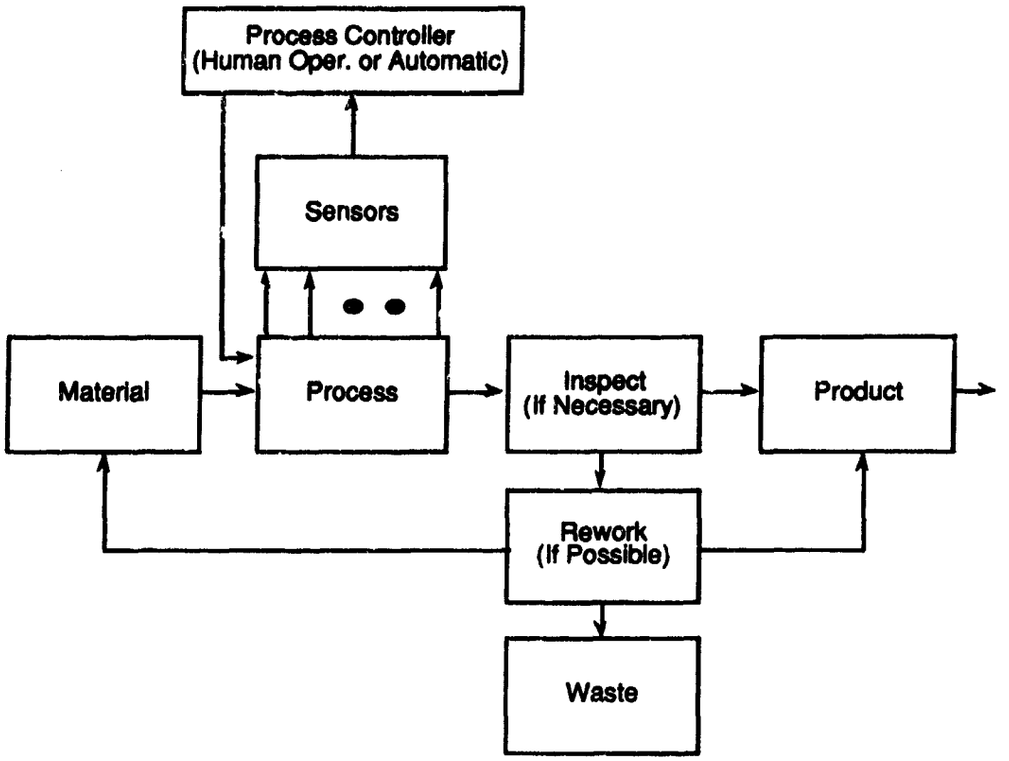


CONVENTIONAL END PROCESS INSPECTION

Figure 1. Block diagram of a typical conventional end process inspection scheme. All inspection occurs at the end of the manufacturing process. The inspection reveals those products that fall outside of acceptance criteria, the remaining products are considered usable. Rejected products may require expensive reworking to be re-used or to meet acceptance criteria or they must be discarded as waste. Rework and waste costs are high.

The inspection should reveal that fraction of the product which falls outside of the acceptance criteria; the remaining fraction is considered usable. Rejected products normally have a large investment in time and effort(value added) as well as initial material costs and they may require additional expensive reworking to meet acceptance criteria. Otherwise they must be discarded as waste. Reworking and waste costs associated with this type of inspection are very high compared to the more sophisticated schemes discussed below.

In-Process Inspection: An improved inspection scheme is outlined in the block diagram found in Figure 2. This illustrates a system approach to an old problem. It must be emphasized that this is not a new technology, in fact, a system of this type normally does not require exotic new equipment.



IN-PROCESS INSPECTION/CONTROL

Figure 2. Block diagram of a typical in-process inspection/control scheme. Evaluation of the manufacturing process occurs during the process through the use of a limited number of sensors monitoring specific parameters. Analyzed data from these sensors (not necessarily in real-time) are used by a human operator (or automatically) to control discrete functions in the process to remain within acceptable values. An understanding of the process (often a priori) is required to set these levels. Final product inspection remains at the end of the manufacturing process as in the end process system shown in Figure 1. The improved control of the manufacturing process ultimately produces a larger fraction of acceptable product, thereby reducing rework and waste costs.

This system employs a limited number of sensors (indicated by dots in the figure). These sensors usually measure the most easily monitored parameters, often not necessarily the most significant. Placement of the sensors is often at locations easily accessed, not necessarily at the best locations. Even considering these shortcomings, the information provided is very beneficial compared to end of process inspection.

Data from the sensors is used in process control, either by a human operator or by automatic (or semi-automatic) controlling devices. The information is not necessarily provided in real-time and, in fact, some time may pass (resulting in unacceptable product) before adjustments are made.

For a meaningful use of the sensor data, a general understanding of the process is required. This may be gained a priori, often with a "best approximation" approach to acquire parameter values. As an example, it may be understood from past operation records that a particular set of pressures and temperatures must be maintained. However, it is not necessary to have a fundamental understanding of the process itself to determine how changing pressure and temperature may change the process results.

As indicated, final inspection is often necessary at the end of the process (or sub-process step). A final inspection produces no added value; it is often used as much for safety considerations and to reduce liability as to improve quality. Implementation of in-process inspection generally reduces waste and rework and this approach is becoming the norm in modern manufacturing.

It is encouraging to note that the number of users of NDE sensors for in-line process inspection and/or control is increasing [1] as the advantages of their use is identified and understood. There is a general evolution to more advanced NDE inspection/control schemes in which more and more inspections are moved "up-stream" in the total manufacturing operation.

Sensors in Manufacturing Process Control: A typical NDE sensor-based process control system is shown in the block diagram of Figure 3. It consists of a larger number of sensors, a computer-based process control model and expert system, and process controllers. It departs from the conventional systems described above by monitoring a large amount of processing environment data, and automatically controlling process parameters in real time, often without the interaction of human operators. In this way, any deviation from pre-set acceptable standards can be corrected as soon as it occurs.

If the process parameters are based on a thorough understanding of the physics and chemistry of the process, and if the process model contains all of the significant process parameters, then no product can go through processing without achieving the desired quality and desired performance characteristics. Through the use of systems of this type with their inherent high quality, waste is greatly reduced.

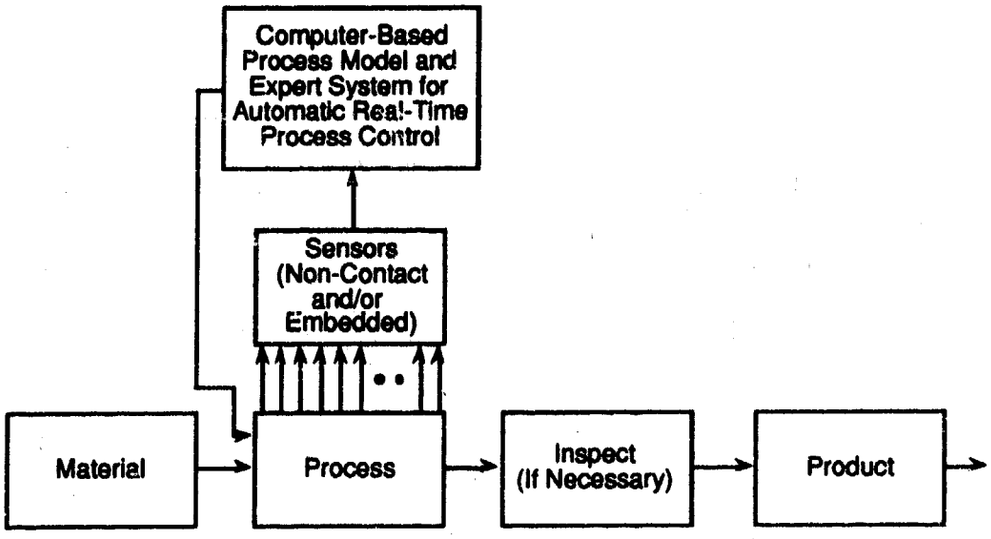


Figure 3. Block diagram of a typical in-process control scheme. Evaluation of the manufacturing process occurs during the process through the use of a large number of multi-parameter sensors. These data, analyzed in real-time, are used by a computer-based process model and expert system. A fully comprehensive understanding of the process, by the process model and decision making ability achieved by the knowledge based expert system, provides real-time process control. Inspection of the product may be required in the early system development stages; however, once the system operation is accepted, inspection may be unnecessary. The final product will be totally acceptable, eliminating rework and waste costs.

As mentioned above, current final inspection procedures do nothing to add to the value of the product, except possibly to limit liability. Their function is primarily to screen out product that have negative value. Some in-process inspection systems may successfully reduce rework and waste, however, costly inspection is often still required. Ultimately, through the use of NDE sensor-based process control, no product that has been produced in conformance with a comprehensive process control model will require inspection. Additional saving may ultimately be found in the reduction and/or elimination of failure analysis to find out why a particular process went wrong and for some kinds of recycling and reworking of process materials.

Requirements for Nondestructive Sensors for Process Control: In seeking to precisely measure those parameters that control the quality of the product, an array of NDE sensors is used. A great deal of effort has been made to develop sensors that provide the best possible information concerning the process parameters. Note that NDE sensors describes a very wide range of devices. Some are very simple, inexpensive, off-the-shelf products, while others may be very costly requiring custom design for specific purposes. Non-contact, non intrusive

NDE sensors which can monitor ongoing processes often must function in very hostile environments.

Sensors to be used in NDE sensor-based control and monitoring systems must be "active", that is they must continuously interrogate the process that they are monitoring. They normally cannot be "passive", waiting for an event to occur to report only when this happens. Data must be available in real-time, rapidly enough to report to the process controller while the process is running. The process model-expert system relies on continuously updated information used for decision making process controls.

Furthermore, the sensors must be nondestructive and non intrusive to the material and processes that they are monitoring. This normally requires that they are non-contact devices, out of physical contact with the material. They normally must use some form of energy that can monitor the material being processed remotely.

In addition, the NDE sensors are often called upon to function in very hostile environments. This may include operation in high or low temperatures and relative humidity, corrosive atmospheres, high or low pressures, and potentially damaging ionizing radiation fields. Because the process control depends on continuous real-time data, the sensors must be very robust and reliable.

Very few currently available NDE sensors meet all of the listed requirements. It is certainly clear that, because of the sophistication often required, these sensors may be expensive. One element of expense will be that most sensors will have to be custom designed for a particular application. For example, a sensor system developed for use in ceramics production which uses neutron radiography may cost between $500,000 and $1,000,000.

The basic thrust of NDE sensors in process control is to use appropriate sensors in sufficient numbers to measure all of the significant parameters and to use this information to fully control the process. The end result will be a product that is of the highest quality and the most cost effective.

State-of-Health/Predictive Maintenance System: Many papers have been written on dedicated state-of-health and predictive maintenance systems [2]. These systems are intended to improve maintenance, increase equipment productivity by reducing equipment down time, reduce equipment repair parts expenses, reduce travel and ancillary expenses associated with centralized support to local maintenance in the repair of processing equipment, and reduce processing labor requirements, particularly labor overtime.

Application of these systems proves especially important in continuous process manufacturing in which unexpected interruptions can become extremely expensive. Examples may be found related to unscheduled downtime involving lost production and in those cases where machinery problems cause damage to product and/or production machinery. In addition, state-of-health and predictive maintenance systems may be employed in an effort to predict and avoid failures that could cause serious injury and/or death to employees.

As processing machinery is made more productive by employing monitoring systems, the level of service can be increased with fewer machines than would otherwise be required. This is a very significant consideration because of the space limitations in many manufacturing centers as well as the high capital costs of new equipment and floor space.

Of course, state-of-health monitoring systems may be very expensive to implement. Their use must create a net cost savings to be considered attractive. A recent cost/benefit analysis for a predictive maintenance system used on electromechanical continuous processing machinery may be used as an example [3]. An abbreviated summary of the results is presented below.

Break even occurs when the total net present value savings from the monitoring system equals the total net present value of the costs of the system. Since the costs of the system are fairly well understood and well bounded, the main issue becomes that of estimating the amount of the potential savings that the system actually captures.

For the system considered in the cited study, using the midpoint capital investment cost estimate of approximately $82.5 million, the system must capture approximately one-third of one percent of the potential savings to break even. With this fractional savings capture, the pay back period (the number of years required for the cash flow in non-discounted dollars resulting form the investment to repay the investment) is approximately 6.5 years. For fractional savings capture approaching one percent the savings to investment ratio approaches four and the pay back period is close to one year. Examples of fractional cost savings approaching ten percent or better are commonly achieved in private industry.

It is highly recommended that a detailed cost/benefit analysis of this type be included as one of the fundamental assessment factors whenever implementation of advanced state-of-health monitoring/predictive maintenance systems is being considered. Successful identification and quantification of resulting savings will provide an important influence when attempting to secure management and/or financial support for such an effort.

Integrated Process Control/State-of-Health/Predictive Maintenance: Figure 4 presents a schematic of the integrated in-process control and state-of-health/predictive monitoring system. Monitoring systems to assess the performance of both independent functions have been developed in the past. These have had a range of levels of sophistication and success. The purpose of the integration is to further improve on production quality and efficiency.

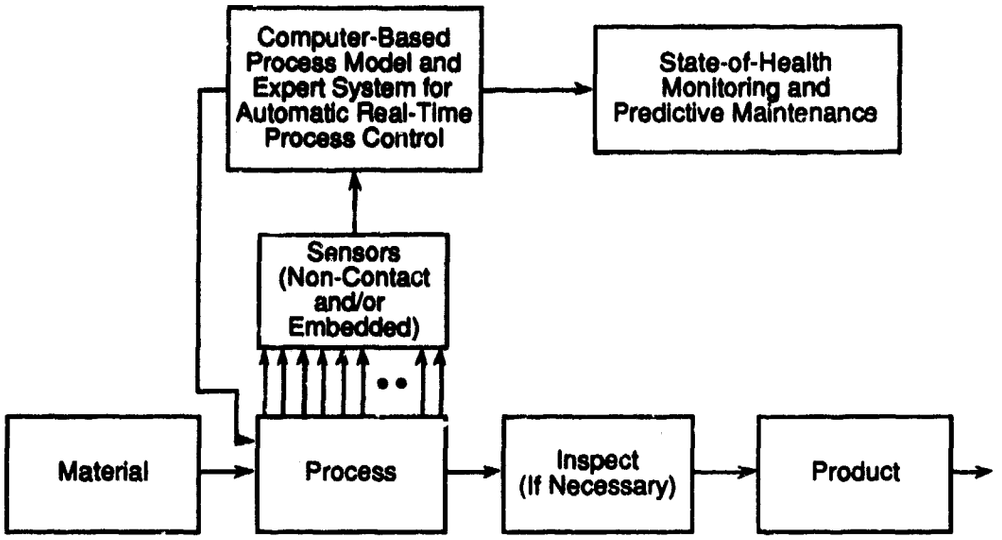


Figure 4. Block diagram of an integrated in-process control/ state-of-health monitoring system. Information from existing sensors (sometimes supplemented by additional sensors) is used by a state-of-health and predictive maintenance system. Results can be used to supply real-time information of process equipment state-of-health. When used with a knowledge-based expert system this combination may also predict potential problems and assist in process equipment repairs. The integrated system employs much of the existing hardware/software infrastructure found in in-process control systems.

This system employs a sensor/analysis system to monitor the state-of-health of the process equipment, often using the same hardware infrastructure (sensors, computers, interfaces, etc.) already available from the in-process control equipment. For some cases, additional sensors for equipment monitoring may be required. The multiple sensor information is used by a knowledge-based expert system to assess how the process equipment is performing. The result is an indication and prediction of potential problems with production machinery. In addition to providing information concerning the state-of-health of the process machinery, a knowledge-based expert system can be used to assist in repairs. This aspect provides an extremely powerful maintenance tool. This is especially important when one considers that hands on experience becomes more and more limited as machinery becomes more and more robust and/or complicated. When a system of this type is developed, it can yield important real-time information which may prevent unscheduled downtime and/or catastrophic failure and furnish assistance to the engineer/technician for required repairs.

Summary:

This advanced concept, which illustrates the direction for the future, is an example of the benefit of system integration. Through the use of this system, all product will conform to acceptable quality specifications. In addition, process equipment will be able to perform in maximum efficiency. Costs will be reduced through minimization of the waste of raw materials and labor(both manufacturing and repair) and in maximization of equipment efficiency.

References:

1. A Series on the U. S. Nondestructive Markets, Maxtech International, Inc., Valley Forge, PA, 1993.

2. For example: see this volume, and/or The Systems Engineering Approach to Mechanical Failure Prevention, The Vibration Institute, Willowbrook, Illinois, and previous volumes in this series.

3. Carriveau, Gary W., et al., "Cost/Benefit Analysis for the POPS Predictive System", Applied Concepts Report AP-190.

SECTOR GATE CASE STUDY: CONDITION ASSESSMENT OF CHICAGO HARBOR LOCK

Kevin L. Rens
Research Asst.
Department of Civil and
Construction Engineering
Iowa State University
Ames, Iowa 50010

Lowell F. Greimann
Professor and Chair
Dept. of Civil and
Construction Engr
Iowa State University
Ames, Iowa 50010

James Stecker
Affiliate Asst. Prof
Dept. of Civil and
Construction Engr.
Iowa State University
Ames, Iowa 50010

Atilano D. Bondoc
Civil Engineer
Construction Operations Div.
Chicago District
U.S. Army Corps of Engr.
Chicago, Illinois 60606

Key Words: Condition Assessment; Dam; Deterioration; Infrastructure; Lock; Maintenance; Management

Abstract: The U.S. Army Corps of Engineers established a Repair, Evaluation, Maintenance, and Rehabilitation (REMR) program to focus more attention on deterioration rates of civil works structures. A team at Iowa State University has focused on the objective to develop an inspection and rating system that uniformly and consistently describes the current condition of sector gate structures. The objectives of this paper are to: (1) Provide an overview of the sector gate inspection and rating system and (2) Present results of a recent inspection of Chicago Harbor Lock which was constructed in 1939 to separate, definitely and effectively, the Chicago River from Lake Michigan which supplies water to the Chicago community.

Introduction: New construction of civil works projects is being overridden by the need to maintain existing projects. Many existing civil works structures are nearing the end of their design life but must continue to serve the public. The U.S. Army Corps of Engineers (US-COE) REMR (Repair, Evaluation, Maintenance, and Rehabilitation) program was designed to focus attention on maintenance and to record and build baseline data to monitor deterioration rates of these structures.

Included within these civil works projects are sector flood gate structures and sector lock gate structures. Sector flood gates function to protect the mainland from hurricanes, high-water situations, and salt water intrusion. Sector lock gates, on the other hand, include the flood gate functions in addition to passing river traffic past dam structures. An Iowa State University research team, applying the REMR project objectives to sector gate structures, has developed an inspection program to detect problems at an early stage. One goal of this project is to develop a rating procedure to describe the overall condition of the structure accurately. Such a rating will flag a potential problem for the engineer. The other goal is to develop maintenance and repair guidelines for these problems. Only the first goal is addressed in the early sections of this paper. A detailed case study is included in the latter sections of this paper.

Overview: During the course of this project, the ISU team has met with Corps personnel and conducted site visits and field investigations at many sector gate facilities. Corps experts conveyed their opinions as to the critical components of sector gate operation and repair, suggested means of quantifying these components, and related them to the overall condition of the sector gates. Then the project team took the experts' comments and formulated them into an inspection procedure and a

tentative set of rating rules. Field tests of the inspection form and rating rules were conducted at four gate sets. At each test site, improvements to the rules and inspection process were suggested by the experts and incorporated into this work.

Field Inspection: The inspection and rating procedure is illustrated schematically in Fig. 1. The entire process is based on a field inspection of the sector gate structure. During this inspection, current physical attributes of the structure are obtained. Data, such as the location of the gate, inspection history, historical water level, and maintenance history, are recorded on the first two pages of the inspection form. Additional pages provide space for entering such field measurements as anchorage movements, elevation changes, gate deflection, cracks, dents, and corrosion. These measurements are used directly to rate the condition of the gate. All data are measured by subjective observations, a tape measure, a level, a ruler, dial gauges, and cameras.

The gate leaves must be jacked at least once for a complete condition index evaluation of a set of sector gate leaves. The leaves are lifted to remove all load from the anchorage system. Three possible jacking scenarios exist during an inspection: (1) gate leaves are jacked during the current inspection, (2) gate leaves are not jacked during current inspection but previous jacking information exists, and (3) gate leaves have not been jacked and will not be jacked during the current inspection. A complete condition index can be calculated for the first two scenarios with the first being more accurate. The final gate leaf condition index cannot be computed for scenario 3.

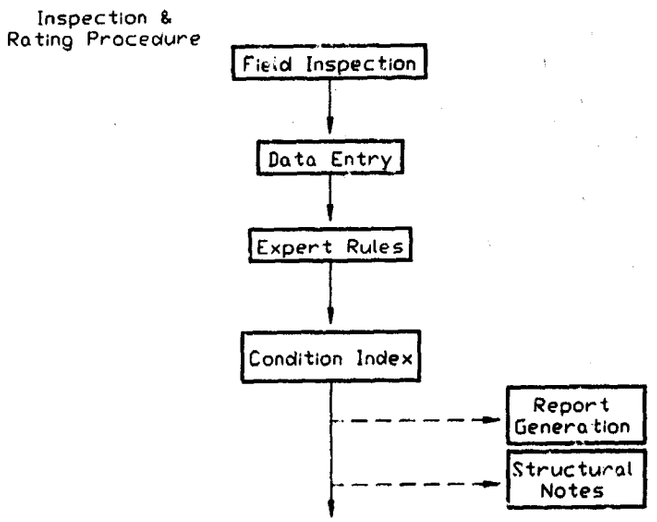


Figure 1. Inspection and Rating Procedure

Condition Index: The condition index (CI) is primarily a planning tool with the index values serving as indicators of the general condition level of the structure. The index is meant to focus management attention on those structures most likely to warrant immediate repair or further evaluation. The CI scale ranges from zero to 100 and is calibrated to group structures into three basic zones: 70 to 100 means no action required; 40 to 69 means maintenance action may be necessary, if economical; and 0 to 39 means maintenance is probably required, after further investigation.

Because the CI involves engineering judgment and depends on the experience of the person making the evaluation, these aspects of the CI were difficult to capture. Experts in this field were interviewed and discussion continued over more than one year until a consensus began to develop. The authors have attempted to blend all the opinions expressed at these meetings into a set of "expert opinion" rules that are embedded in the evaluation that constitutes the CI. The rules have been designed to

interpret straightforward, visual observation data in much the same manner that a seasoned engineer would interpret field observations.

Two general structural criteria for evaluating the CI are available: safety and serviceability. Safety relates to the performance of a structure beyond normal service conditions, for example, under such abnormal conditions as excessive load. Serviceability relates to the performance of a structure under normal service conditions, for example, excessive leakage. The CI for each distress is based on field measurements of the distresses and the opinion of experts. It includes both safety and serviceability aspects.

A series of critical measurements are made on each gate to quantify the CI. Experts were asked to interpret these measurements in light of the serviceability and safety of the gate and to assign the limiting values to the measurements. Specifically, a series of distresses are identified. Each distress is quantified by a measurement X. For example, anchorage movement is a distress quantified by the relative motion between the steel and the concrete at the steel and concrete interface before and after jacking. The individual distress CI is quantified by

$$CI = 100(0.4)^{X/X_{max}} \qquad (1)$$

where X_{max} is some limiting value of X [1-3]. Fig. 2 illustrates the equation and the previously mentioned zones. Experts have selected X_{max} to be the point at which the gate requires immediate repair or at least a more detailed inspection and CI evaluation. In other words, it corresponds to a CI of 40 and is a potentially hazardous situation.

Structural considerations: Many structural considerations are embedded in the CI rules. In addition to functional and operational factors, the experts took structural factors into account when setting limiting values, tolerances, and weight factors. With this in mind, the structural adequacy can be characterized by several of the distress measurements. Some distresses in Table 1 have a more significant impact on safety than others. The structural distress subset is listed in Table 2. The structural distresses are flagged if the structural distress measurement exceeds certain bounds. A structural note along with the corresponding measurement will be included in the summary report for potential structural problems that have been flagged.

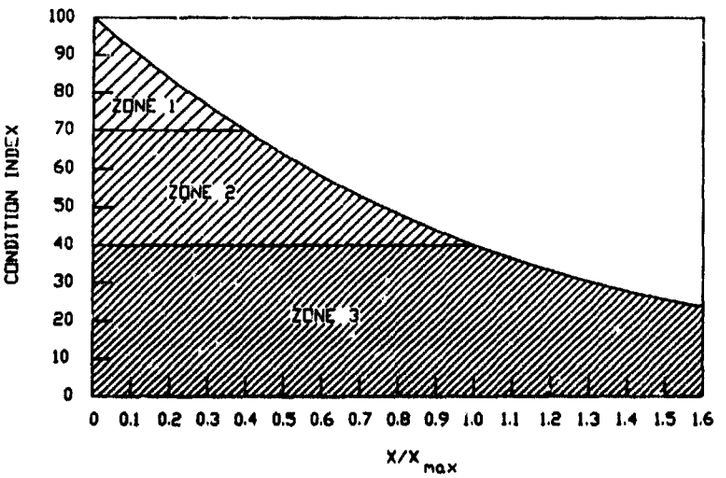


Figure 2. Condition index related to X/X_{max}

Distress descriptions and X_{max}: If a sector gate structure is designed and constructed properly, it has an initial CI of 100. The CI will degrade as various distresses are incurred. A total of ten distresses have been identified for categorization in this project. Each is described briefly in Table 1. Each of these distresses can detract from the safety and serviceability of sector gates.

The CI for each distress depends upon the ratio of a field measurement X to some limit X_{max}. In the following sections, the definition and measurement of X and X_{max} values are described for one distress. For complete documentation, refer to the project report [3]. Potential causes of the distress are also listed. These causes are the problems that must be addressed in the maintenance and repair of the gate.

Table 1. Distresses in Sector Gates

Distress	Description
Top Anchorage Movement	Displacement of embedded anchorage system
Gate deflection	Nose deflection before hinge pin rotates
Levelness	Vertical gate displacement
Cracks	Breaks in structural steel components
Dents	Disfigured structural steel components
Noise, jumping, and vibration	Abnormal noise, jumping or vibration during gate operation
Corrosion	Loss of steel due to interaction with the environment
Hinge wear	Total wear in hinge pin casting (anchor bracket to hinge bracket)
Vertical wear of thrust bushing	Vertical wear of thrust bushing from reference position
Leaks and boils	Water passing through or around the gate

Table 2. Structural Distresses

Distress	Brief Description
Top anchorage movement	Embedded steel movement
Jumping	Abnormal gate jumping
Girder cracks	Breaks in main horizontal girders
Girder dents	Disfiguration of main horizontal girders
Girder corrosion	Loss of girder steel
Gate deflection	Nose deflection before hinge pin rotates

Top Anchorage Movement (definition and causes): Anchorage movement is a parallel and perpendicular displacement of the embedded anchorage system. Movement can occur during opening or closing of the gate leaves and during filling or emptying of the lock chamber. Some anchorage configurations, called flexible anchorage systems, are designed to permit movement, while rigid anchorage systems do not. Anchorage movement can be caused by several factors:

- Corrosion
- Steel elongation or movement
- Movement at shims or nuts
- Concrete cracks
- Additional load

The top anchorage system is the only mechanism that connects the top of the gate leaf to the concrete wall. Hence, the presence of anchorage movement may indicate a significant structural problem, or it could eventually introduce structural problems into other gate components.

Top Anchorage Movement (measurement and Limits): For lock gates, measurement of the anchorage movement will be made between the concrete interface and hinge casting bracket (Fig. 3) while gates are open (O), closed (C), closed with full head (CF), and jacked while half open (J). (Refer to Field Inspection Section, for discussions on gate leaf jacking). For flood gates, the closed full head position is omitted. The maximum motion that occurs at the embedded steel, X, is found by subtracting the smallest of the measurements from the largest in both the parallel and perpendicular directions. If jacking is not performed during the current inspection, the largest movement of a previous jacked inspection will be compared to the current measurements and the maximum selected. If jacking has never been performed on a set of gates leaves, the condition index for anchorage movement and hence the combined gate condition index cannot be computed. The presence of cracked or spalled concrete at the concrete interface and the existence of level 3 (Severe pitting in a dense pattern or thickness reduction in a local area) or greater corrosion on the anchorage configuration is also recorded.

For rigid and frame-type anchorage systems, a displacement of 0.03 in. has been selected as the limiting motion.

$$X_{max} = 0.03 \text{ in.}$$

The experts judged that motion greater than this could indicate a significant structural problem. For flexible anchorage systems, the maximum elastic motion has been selected as

$$X_{max} = 0.0006 \text{ (L) in.}$$

where L is the length of embedded anchorage in inches. Any spalling or cracking of the concrete in this area will reduce the condition index in this area by a factor of 0.85. In addition, a corroded anchorage configuration will reduce the condition index by a factor of 0.85 [2],[3]. The CI for a gate anchorage arm is determined from Eq (1).

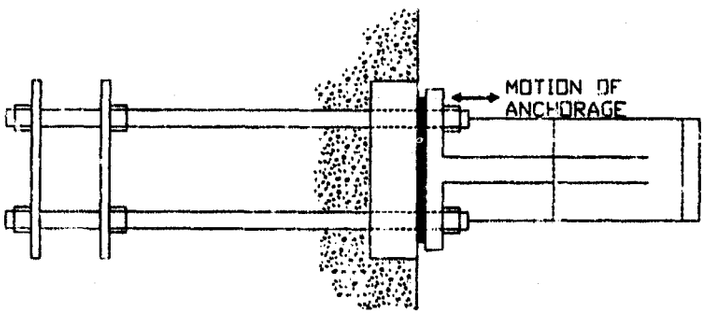


Figure 3. Embedded Anchorage Configuration

Other Distress Codes: Measurement will also be recorded for the nine other distresses shown and briefly described in Table 1. The additional distresses include: hinge wear, gate deflection, levelness, cracks, dents, noise jumping and vibration, corrosion, vertical wear of thrust bushing, and leaks and boils.

Multiple Distresses: When several types of distress occur simultaneously, such as both anchorage movement and cracks, the CIs are combined into a single value. Weighting factors are introduced to reflect the importance of the various distresses. Hence, let w_i be the weighting factor of the CI for distress i (Table 3). The weighting factors assign more value to the more significant distresses. Relative initial weights are listed in Table 3. They reflect, to some degree, the opinion of the Corps experts and also the opinions of the authors. The table illustrates that anchorage movement is the

most important and dents the least important. The normalized values W_i are listed in Table 3 (rounded to add up to 100). The combined CI for all distresses is then given by

$$CI = W_1CI_1 + W_2CI_2 + ... W_{10}CI_{10}$$

As in the previous management systems the relative importance of a distress becomes larger as a distress became more severe. To account for this, the project team introduced a variable adjustment factor to increase the distress weighting factor w by a factor up to eight as its CI approached Zone 3.

Table 3. Condition Index Weighting Factors

DISTRESS	W_i	W_i (%)
Top anchorage movement	10	17
Gate deflection	6	10
Levelness	5	9
Cracks	5	9
Dents	1	2
Noise, jumping, and vibration	5	9
Corrosion	7	12
Hinge wear	8	14
Vertical wear of thrust bushing	7	12
Leaks and boils	4	6

Field Testing and calibration: The performance of the rating rules is evaluated by comparing the calculated CI values with CI values determined by a group of sector gate expert engineers. The inspection and rating module development ends with field testing and final calibration. For example, four preliminary meetings took place before the final seven calibration field tests. Nine US-COE experts were involved in all seven calibration visits. Each expert was asked to rate the individual distresses in each gate leaf, that is , to assign a CI to each of the ten distresses. Additionally, the experts were asked to assess an overall leaf CI. Some adjustments to the limiting values and weighting values were made to better fit the experts' ratings.

The revised version of the overall rating system corresponds well with the experts' ratings, as illustrated in Fig. 4. Variance still exists between the current version as well as between the experts themselves. Therefore, the interpretation of the results must take this into account.

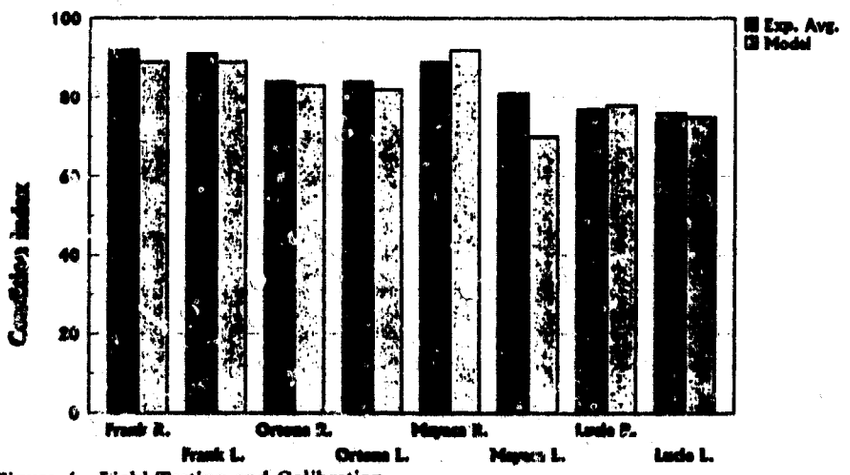


Figure 4. Field Testing and Calibration

Chicago Harbor Lock Inspection: This inspection was conducted by James Stecker, Kevin Rens, and Lowell Greimann under contract with USA-CERL under the direct supervision of Dr. Anthony Kao. The work was coordinated through the U.S. Army Corps of Engineers, Chicago District Office, with Atilano Bondoc and Richard Pickett. Messrs. Bondoc, Pickett, and Ed Johnson, also of the Chicago District office, were present and observed most of the data collection procedures on both the upper and lower sets of gate leaves. The inspection procedures were conducted in accordance with the sector gate technical report [3]. Certain procedures require the gate leaves to be jacked. The equipment for jacking, the divers, and supporting dive crew were provided by the Rock Island District and were under the direction of Mr. Ken Shoemaker and Mr. Mike Beneventi.

The data collected during the inspection procedures are recorded on an Inspection Form and identified as to location and other pertinent general information. The data recorded represents measurements from instrumentation, survey equipment, or subjective observations. A copy of the inspection forms for each gate with raw data for each set of gates as well as the computer evaluation output is included in the case study report [4].

General Observations: The general appearance of both sets of gates was good. They have been well maintained, and the normal cycle operations of both gates were smooth and quiet. At the upper gates, the obvious failure to close indicates there is a major problem, but if the gates were closing properly, a casual observer would think the gates were in good condition. In the course of the inspection and subsequent evaluation of the collected data, it was determined this was not the case. The hands on field inspection helped identify problems that a general inspection plan might miss.

Condition Index Calculation and Analysis: Facing downstream (or west) on the Chicago River, the lower gate is closest to the city whereas the upper gate is closest to Lake Michigan. Left leaves are on the south side of the lock chamber and the right leaves are on the north side.

The overall gate condition indices are calculated to be as follows:

	Right Leaf	Left Leaf
Lower Gate	47	41
Upper Gate	39	40

A full explanation of the meaning of the condition index scales and zones is shown in Table 4. In addition to the overall condition index given for each leaf, the evaluation program also identifies and tabulates a condition index for ten separate distresses for each gate leaf. In the opinion of the authors of this inspection report, it is the distress condition indices that provide the most information for the determination of maintenance and repair plan.

A general observation regarding the overall gate condition index and individual distress condition indices is appropriate. A condition index of 40 is a trigger point as shown in Table 4. If the condition index is below 40, actions are necessary to either restore function or preserve safety. If above 40 but below 69, analysis of alternatives should be done to preserve function and safety. In this specific case, one gate leaf is below 40, two are at 40, and one is approaching 40 on the way down. Evaluation of the condition indices for the distresses on each gate leaf can focus the review of alternatives for maintenance and repair by a qualified management team or engineers.

The following is a narrative of some observations and hand calculated results of recorded data measurements. Refer to the technical report for explanation of terms, evaluation methods for certain components distress documentation, measurements, and limitations [3]. The narrative is outlined by individual distress types for each set of gates. The individual distress CI's are shown in Table 5.

Table 4. Condition Index Scales and Zones

Zone	Condition Index	Condition Description	Recommended Action
1	85 to 100	Excellent: No noticeable defects. Some aging or wear may be visible.	Immediate action is not required
	70 to 84	Very Good: Only minor deterioration or defects are evident.	
2	55 to 69	Good: Some deterioration or defects are evident, but function is not significantly affected.	Economic analysis of repair alternatives is recommended to determine appropriate action.
	40 to 54	Fair: Moderate deterioration. Function is still adequate.	
3	25 to 39	Poor: Serious deterioration in at least some portions of the structure. Function is inadequate.	Detailed eval. is required to determine the need for repair, rehab., or reconstrctn. Safety eval. recommended.
	10 to 24	Very Poor: Extensive deterioration. Barely functional.	
	0 to 9	Failed: No longer functions. General failure or complete failure of a major structural component.	

Table 5. Chicago Harbor Lock Condition Index summary

	Upper		Lower	
Distress	Left	Right	Left	Right
Top Anchorage Mvmt.	68	70	69	71
Gate deflection	87	91	75	75
Levelness	93	90	96	96
Cracks	100	100	100	100
Dents	100	83	83	40
Noise, Jump,Vibration	100	100	100	100
Corrosion	40	32	29	32
Hinge Wear	32	35	20	34
Thrust Bushing Wear	100	100	100	100
Leaks and Boils	0	0	100	100
Overall CI's	40	39	41	47

Lower Gate

Anchorage System: The anchorage system on both leaves showed negligible movement (less than 0.001 in.) during jacking, opening, and closing. Both anchorages had cracked and spalled concrete as well as level 3 anchor corrosion present at the steel embeddment. The left leaf was flagged as a

structural distress because the combination of movement, anchor corrosion, and spalled concrete reduced the CI below 70. This could be a structural problem and should be monitored.

Gate Deflection: A gate deflection of 3/4 in. was recorded for both gate leaves upon gate opening. The total amount of gate deflection on the left leaf was 1 in. This consisted of 1/4 in. initial deflection (with hinge rotation) followed by a pause and then 3/4 of additional deflection (without hinge rotation). Because of the unusual gate leaf behavior, it was decided to record only the 3/4 in. The initial 1/4 in. could be attributed to wear in the operating equipment. The gate deflected less than 3/4 in. upon gate closing for both leaves.

Noise, Jump & Vibration: Overall, both gate leaves operated smoothly and quietly. The left leaf experienced a jumping motion one time at 40 % closed. This was determined to be the result of operating machinery and was ignored by the CI algorithm. It was also pointed out by lock personnel that the left leaf occasionally vibrates during the swinging of the leaf. This inspection did not observe this distress even though several cycles were performed.

Hinge wear: The left leaf hinge wear during jacking was measured to be 0.65 in. and the right leaf hinge wear during jacking was 0.44 in. A vertical movement of 0.18 in. was also measured on the left leaf hinge casting as it was jacked. This became visually evident after the marker for hinge wear on the gate hinge bracket began to rise as the leaf was jacked. The vertical dial gage slipped on the right leaf and no measurement was possible.

The left leaf hinge pin is rotating within the anchor bracket plates, or upper hinge support. The top nut was missing and the pin could be seen to rotate and move laterally in the top plates. The hinge pin and cast steel hinge ball did not appear to rotate in the hinge casting bushing. However, since the hinge casting also moved vertically 0.18 in. when the gate was jacked, one explanation could be that the hinge casting and bushing was sliding up the face of the hinge ball. This would suggest the ball-to-bushing interface is not frozen, but it still may not allow rotation. All the apparent wear appears to be in the hinge support plates. We could not observe a similar movement within the bottom plates but believe wear is also present in the bottom plates.

The right leaf hinge pin appears fixed within the upper hinge support plates. The apparent wearing surface is at the interface between the hinge pin and the cast steel hinge ball. When the gate swings, the hinge ball rotates about the pin. When the gate leaf is jacked, grease is forced out the top of the pin/ball interface and a gap appears on the backside of the pin. The hinge ball and hinge bushing did not rotate with respect to each other but there still appeared to be some vertical movement of the hinge casting similar to the left leaf. As described above, neither gate leaf was rotating as designed at the hinge point. It was not possible to determine if the gates ever did or when they may have stopped rotating in the bushing/ball interface.

Thrust bushing wear: Incremental vertical wear of the thrust bushing is dependent on successive leaf inspections. This inspection fulfills the initial inspection requirements. The bench mark used was SLW1 on the left leaf side (just west of the southwest building). In addition, the divers measured the distance between the pintle ball bushing and the lock floor on the right leaf and indicated roughly 1 1/2 to 2 in. of clear space. The construction drawings indicate 2 in. to be the design clear space. As noted, this distress condition index is dependent on a measurement of accumulated vertical wear to the thrust bushing or pintle ball. An actual measurement of 0.25 inches of accumulated wear on the right leaf would have reduced the distress CI from 100 to 40 and the overall right leaf CI would have reduced from 47 to 43. It is probably safe to assume that there has been at least 0.25 inches of wear and this could be taken into account in the overall repair analysis.

Corrosion: The upstream skin corrosion on both gate leaves was determined to be level 2 and the downstream skin at level 3. The upper girder web on the left leaf was completely corroded through and had been patched by welding plates in on a slight downward angle from upper to lower flange. Including the patch, the corrosion level was chosen to be at a level 4. It was later brought to our

attention by lock personnel that the corrosion was due to salt spread on the steel walkway to melt ice. The corrosion level for the girders on the right leaf was judged to be level 3. The framing plate in the recess corners were completely corroded through on both gate leaves and had not been repaired. The framing was chosen to be at level 5 corrosion. In addition, on several places on both leaves, the steel walkway had significant corrosion and was extremely thin. The very low CI indicates the need for a more detailed inspection for corrosion and a complete structural analysis is recommended on these leaves.

Dents: The left leaf has a small framing dent behind the timber bumpers approximately 25 ft from the nose and 5 feet down from the top girder. The right leaf has a dent that affects both the girder and skin plate approximately 30 feet from the nose near the top girder. The girder dent, although apparently not that serious, caused the dent distress to fall to a CI of 40 on the right leaf. It would be flagged as a structural distress on the summary report recommending further investigation.

Cracks: No cracks were found. Keep in mind that no cleaning was performed and the upstream skin was not checked.

Leaks & boils: Both the left and right recess as well as the nose experienced around a 3 in. gap which allowed leakage. This leakage was attributed to damaged seals and routine maintenance could correct this problem. These leaks were ignored by the CI algorithm and a CI of 100 assigned.

Levelness: The gate leaves experienced only 0.01 feet of elevation change over the full cycle of swinging.

Upper Gate

Anchorage System: The anchorage system on both leaves showed negligible movement (less than 0.002 in.) during jacking, opening, and closing. Both anchorages had cracked and spalled concrete as well as level 3 anchor corrosion present at the steel embeddment. The left leaf was flagged as a structural distress because the combination of movement, anchor corrosion, and spalled concrete reduced the CI below 70. This could be a structural problem and should be monitored.

Gate Deflection: Both gate leaves observed a gate deflection of around 1/4 in. upon gate opening and closing.

Noise, Jump & Vibration: Overall, both gate leaves operated smoothly and quietly through most of the closure cycle. The left leaf experienced noise, jumping and vibration at 95 % closure. This was determined to be the result of normal noises near gate closure and was ignored by the CI algorithm. The right leaf also vibrated at this leaf position. This was also ignored by the algorithm although we suspect it is the result of the roller problem scheduled to be repaired.

Hinge wear: The left leaf hinge wear was measured to be 0.53 in. and the right leaf wear was 0.57 in. A small amount of vertical movement was also observed but not measured. Both the left and right upper hinge behaved like the right leaf of the lower gate and did not rotate correctly.

Thrust bushing wear: Incremental wear of the thrust bushing is dependent on successive leaf inspections. This inspection fulfills the initial requirements. The bench mark used was SLW11 on the left leaf side (just north of the southeast building).

Corrosion: The upstream and downstream skin corrosion was determined to be level 2. The upper girder web on both leaves was determined to be at level 3. The framing plate in the recess corner is completely corroded through on the right leaf and had not been repaired. The framing was chosen to be at level 5 on the right leaf and level 3 on the left leaf. In addition, on several places on both leaves, the steel walkway had significant corrosion and was extremely thin. A more detailed inspection and structural analysis is recommended on these leaves. An analysis of the upper gate must

consider the possibility of high waves and ice from significant storms.

Dents: The right leaf has a small framing dent approximately 50 ft from the nose 6 feet down from the top girder. No other dents were observed.

Cracks: No cracks were found. Keep in mind that no cleaning was performed and the upstream skin was not checked.

Leaks and boils: Both the left and right recess experienced around a 12 in. and 3 in. gap which caused leakage. In addition, the nose has about a 24 in. gap. These leaks were not caused by seals and are the result of the gate not closing properly because of a roller problem. This caused the leak CI to approach 0 and should be near 100 after the roller problem is addressed and seals replaced.

Levelness: The gate leaves experienced only 0.02 feet of elevation change over the full cycle of swinging.

Conclusions: The overall gate condition indices are uniform for all four gate leaves and are consistent with what the authors would expect after a careful review of the inspection data. The CI numbers are around 40 that puts the condition description at the bottom of the fair range or top of the poor range. A CI of 40 suggests the structure is functional but may have problems that impair normal operations. A CI of 40 suggests there may be moderate to severe deterioration of components of the structures.

A more detailed review of the individual distress condition indices specifically call attention to both the excessive hinge wear and corrosion. Condition indices ranging from 20 to 35 for hinge wear correctly reflect the serious deterioration of the hinge system. Condition indices ranging from 29 to 40 for corrosion also correctly reflects the serious deterioration of the steel structure due to corrosion. A complete structural analysis of the girder framing of the lower gates and the added reinforcing plates should be made before changing or ignoring a low condition assessment. As noted before, the condition indices for thrust bushing wear are incorrect because this was not an initial inspection, and this distress is based on accumulated pintle wear. The authors of this report are confident that there is also significant wear at the pintle, but we are unable to quantify it now.

Disclaimer: The current inspection and rating procedure for sector gate structures has had sufficient development and testing to warrant its distribution on a wider basis. However, it should still be considered to be in a developmental stage. Many of the concepts introduced, such as the condition index, structural considerations, X_{max} values, and weighting factors, should be exposed to a broader range of engineers who work in the area. Modifications to the procedure are certainly expected, and suggestions are welcomed.

Study findings, the inspection and rating procedures, and the presented case study are not to be taken as official Department of the Army position, unless so designated by other authorized documents.

Acknowledgements: This study was authorized by Headquarters, U.S. Army Corps of Engineers (HQUSACE), under Civil Works Research Units 32672 and 32673, "Development of Uniform Evaluation for Procedures/Condition Index for Civil Works Structures" and "Development of Maintenance and Repair Guidelines and Management Systems" respectively. These work units are parts of the Operations Management Problem Area of the Repair, Evaluation, Maintenance, and Rehabilitation (REMR) Research Program sponsored by HQUSACE. Dr. Anthony M. Kao is the Problem Area Leader. Mr. James E. Crews (HQUSACE) is the Technical Monitor of the Operations Management Problem Area and is the Chairman of the REMR Research Program Overview Committee. This study was performed by the Engineering Research Institute of the College of Engineering, Iowa State University. The authors acknowledge the input of many Corps personnel in this project.

REFERENCES:

1. Greimann, L and Stecker, J *Maintenance and Repair of Steel Sheet Pile Structures*, 1990, Technical Report REMR-OM-9. U.S. Army Corps of Engineers, Washington, D.C. 20314-1000.

2. Greimann, L, Stecker, J and Rens, K *Management System for Miter Lock Gates*, 1990, Technical Report REMR-OM-08. U.S. Army Corps of Engineers, Washington, D.C. 20314-1000.

3. Greimann, L, Stecker, J and Rens, K *Condition Rating Procedures for Sector Gates*, 1992, Technical Report REMR-OM-112, U.S. Army Corps of Engineers, Washington, D. C. 20314-1000.

4. Stecker, J, Rens, K and Greimann, L *Chicago Harbor Lock Inspection*, 1993, Report prepared for and submitted to U. S. Army Corps of Engineers, Chicago District.

AN ARTIFICIAL NEURAL NETWORK SYSTEM FOR DIAGNOSING GAS TURBINE ENGINE FUEL FAULTS

MAJOR Orlando J. Illi, jr.
Chief, Knowledge Engineering Group (KEG)
U.S. Army Ordnance Center and School,
Aberdeen Proving Ground, MD 21005

Dr. Frank L. Greitzer
Pacific Northwest Laboratory
Richland, WA 99352

Lars J. Kangas
Pacific Northwest Laboratory
Richland, WA 99352

Tracy J. Reeve
Expert Solutions
Stratford, CT 06497

Abstract: The US Army Ordnance Center & School and Pacific Northwest Laboratory are developing a turbine engine diagnostic system for the M1A1 Abrams tank. This system employs Artificial Neural Network (ANN) technology to perform diagnosis and prognosis of the tank's AGT-1500 gas turbine engine. This paper describes the design and prototype development of the ANN component of the diagnostic system, which we refer to as "TEDANN" for Turbine Engine Diagnostic Artificial Neural Networks.

Key Words: Artificial neural networks; expert systems; fault diagnosis; prognostics; turbine engine maintenance.

Introduction: The Army's maintenance practice employs diagnostic procedures that are generally performed manually. Rather than using automated diagnostic and prognostic paradigms, the current practice verifies only whether the operational states are within or out of tolerance. The paradigm does not reflect a real-time operational assessment, nor can it be readily modified to predict failures (perform prognostics).

Technology currently exists to markedly improve both the accuracy and timeliness of the current diagnostic paradigm. These improvements will not require an inordinate expenditure in either development or fielding costs. The purpose of this paper is to discuss improvements to the current maintenance practices and describe a prototype system that employs a real-time automated diagnostic paradigm. The prototype system will be referred to as the Turbine Engine Diagnostic Artificial Neural Networks (TEDANN). TEDANN is being developed to evaluate the feasibility of using Artificial Neural Networks (ANNs) to monitor turbine engine performance and diagnose failures in real-time. In addition, TEDANN will provide a testbed to evaluate the feasibility of developing a real time prognostics capability.

Background: Because the current Army maintenance process is reactive (dependent upon a failure to occur before its initiation), it is incapable of predicting failures. The current diagnostic process depends upon humans to integrate, categorize and analyze currents and voltages. Diagnostics are based on individual experience, heuristics and rules of thumb. After the mechanic arrives at an initial diagnosis, he must analyze his initial hypothesis using a suite of analog and/or digital Test Measurement or Diagnostic Equipment (TMDE). The resultant analysis will either support or refute his hypothesis. The mechanic repeats this iterative process until it produces an accurate diagnosis.

For example, troubleshooting the AGT-1500 gas turbine engine requires the mechanic to take separate Digital Multimeter (DMM) readings at the Electronic Control Unit (ECU). This is accomplished by inserting test probes into an ancillary breakout box (BOB) attached to the ECU's diagnostic connector. Once the mechanic takes the required voltage readings he must manually analyze their significance through a manual computation and compare the results to diagnostic flow charts. This practice is time consuming and error prone due to the dependence upon proper application of mathematical conversion factors and correct placement of test probes into the BOB.

A suggested improvement to current maintenance practices is to automate diagnostics in a real-time system and to use prognostics to detect equipment faults before they occur. This would reduce the occurrences of No-Evidence-Of-Failure (NEOF) outcomes in diagnosing component failures. The NEOF occurs when a component is incorrectly diagnosed as faulty and is evacuated to a maintenance facility for repair. The receiving maintenance facility will, as a matter of course, verify the initial diagnosis. When the defective component is diagnosed as functional, a NEOF is initiated. At the present time certain electronic components exhibit a NEOF rate of approximately 60%. Hence, the requirement for an improved diagnostic and prognostic paradigm.

Objective: The technical objectives of this research are to

- explore improvement opportunities in turbine engine fault diagnosis through application of ANN technology;
- examine the application of ANN technology to an automated diagnostic and prognostic system for turbine engine maintenance.

The long-term operational objectives of this research are to achieve a reduction in NEOF, to reduce maintenance costs, and to increase operational readiness.

Approach: We chose Artificial Neural Network (ANN) technology for our prototype because it is well suited for diagnostics in real-world applications (see the box on Artificial Neural Networks). Each turbine engine is unique in its behavior as a result of its age, manufacturing tolerances, and the environment in which it is operated. All sensors are unique in their response and calibration. A diagnostic system using ANNs can automatically adapt to these individual variables for each turbine engine. Most conventional technologies such as expert systems would monitor each individual turbine engine as a generic engine. These generically-based diagnostic systems will not be as sensitive to individual turbine engines as an ANN-based system.

The TEDANN prototype demonstrates how current practices of manual diagnosis can be replaced by automated diagnostics and prognostics. The prototype analyzes values from on-board sensors in real-time.

Artificial Neural Networks

A brief description of Artificial Neural Networks (ANNs) is given here to help readers who are unfamiliar with this technology to appreciate the computational capabilities of ANNs.

ANNs are algorithmic systems implemented in either software or hardware. The concept of ANNs was inspired by the way the biological brain processes information. ANNs, like people, learn by example. Learning in the biological brain occurs in a network of neurons, which are interconnected by axons. A point of contact (actually most often a narrow gap) between an axon from one neuron to another is called a synapse. Learning is a matter of adjusting the electro-chemical connectivity across these synapses.

An ANN is a network of neurons or Processing Elements (PEs) and weighted connections. The connections correspond to axons and the weights to synapses in the biological brain. A PE performs two functions. It sums the inputs from several incoming connections and then applies a transfer function to the sum. The resulting value is propagated through outgoing connections to other PEs. Typically these PEs are arranged in layers; with the input layer receiving inputs from the real world and each succeeding layer receiving weighted outputs from the preceeding layer as its input. Hence the creation of a feed forward ANN, where each input is fed forward to its succeeding layer. The first and last layers in this ANN configuration are typically referred to as input and output layers. (Input layer PEs are not true PEs in that they do not perform a computation on the input.) Any layers between the input and output layers (usually 0-2 in number) are called hidden layers because they do not have contact with any real world input or output data.

Back propagation is one of several possible learning rules to adjust the connection weights during learning by example. Learning occurs when the network weights are adjusted as a function of the error found in the output of the network. The error is the difference between the expected output and the actual output. The weights are adjusted backwards (back-propagated) through the ANN network until the error is minimized for a set of training data.

A trained ANN, i.e., a network that has learned by example, can be applied to real world problems of considerable complexity. Their most important advantage is in the ability to process data that are too complex for conventional technologies—problems that do not have an algorithmic solution or for which an algorithmic solution is too complex to be found. In general, because of their abstraction from the biological brain, ANNs are well suited for problems that people are good at solving, but for which computers are not. This class of problems includes pattern recognition and forecasting or recognizing trends in data. ANNs have been applied successfully to hundreds of applications.

Figure 1 portrays the major components of the prototype diagnostic system. Initially, the TEDANN prototype will analyze AGT-1500 fuel flow dynamics, that is, fuel flow faults detectable in the signals from the ECU's diagnostic connector. These voltage signals represent the status of the Electro-Mechanical Fuel System (EMFS) in response to ECU commands. The EMFS is a fuel metering device that delivers fuel to the turbine engine under the management of the ECU. The ECU is an analog computer whose fuel flow algorithm is dependent upon throttle position, ambient air and power turbine inlet temperatures, and compressor and power turbine speeds. Each of these variables has a representative voltage signal available at the ECU's J1 diagnostic connector, which is accessed via the Automated Breakout Box (ABOB).

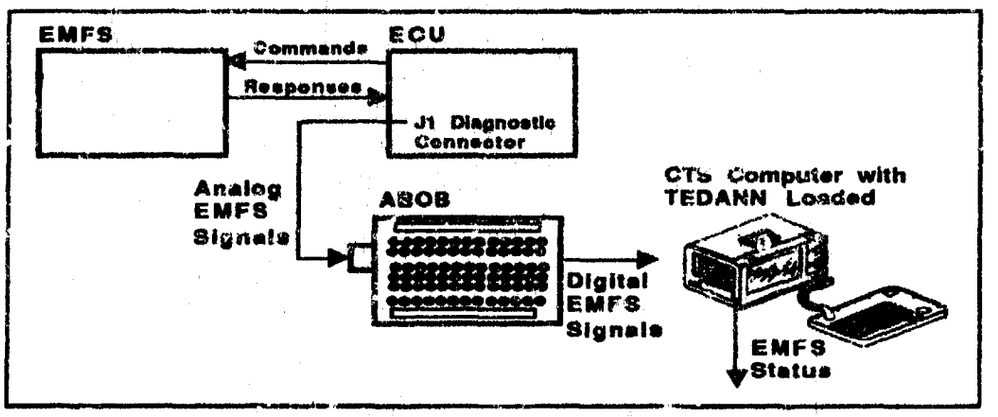


Figure 1. System Design of the TEDANN Maintenance System Concept

The ABOB is a firmware program capable of converting 128 separate analog data signals into digital format. The ECU's J1 diagnostic connector provides 32 analog signals to the ABOB. The ABOB contains a 128 to 1 multiplexer and an analog-to-digital converter, both operated by an 8-bit embedded controller. Army Research Lab (ARL) developed and published the hardware specifications as well as the micro-code for the ABOB Intel EPROM processor and the internal code for the multiplexer driver subroutine. Once the ECU analog readings are converted into a digital format, the data stream will be input directly into TEDANN via the serial RS-232 port of the Contact Test Set (CTS) computer.

The CTS computer is a ruggedized IBM compatible personal computer designed for tactical use on the battlefield. The CTS has a 50MHz 32-bit Intel 80486DX processor. It has a 200MB hard drive and 8MB RAM. The CTS also has serial, parallel and SCSI interface ports. The CTS will also host a frame-based expert system for diagnosing turbine engine faults (referred to as TED; not shown in Figure 1). The TEDANN system will eventually be integrated with the TED expert system, which will receive inputs from TEDANN.

TEDANN was developed using the NeuroWindows ANN simulator software and Visual Basic as a user/computer interface development tool. Both of these software packages run in the MS-Windows environment.

Results: For the prototype development of the TEDANN system, we have chosen to concentrate on three fuel flow faults: Bouncing main valve metering, stuck main valve metering and fuel flow errors. All three problems are difficult to diagnose and lend themselves to real-time analysis using ANNs. The analysis of fuel flow problems is based on an understanding of the operation of the EMFS, which is summarized in the box entitled "Fault Detection in the AGT-1500 Turbine Engine Electromechanical Fuel System."

TEDANN performs diagnostics using values from three fuel flow signals, ambient air and power turbine inlet temperatures, and compressor and power turbine speeds. For our prototype, these signal data were sampled during the first minute of the engine start sequence. Signal values are collected from the ECU's J1 diagnostic connector. The three fuel flow signals are referred to as: WF-request, WF-actual and WF-solenoid. These values determine the specific amount of fuel delivered to the engine according to the fuel schedule in the ECU. WF-request is the ECU's request for fuel flow, WF-solenoid is an ECU signal that positions the main metering valve in the EMFS in response to WF-request and WF-actual. WF-actual is the EMFS' feedback signal indicating the position of the main metering valve.

Fault Detection in the AGT-1500 Turbine Engine Electromechanical Fuel System

In the event a fuel flow fault occurs, the ECU initiates an engine protective mode. A fuel flow fault during the engine start sequence results in engine protective mode 1, which aborts the AGT-1500 start sequence. A fuel flow fault that occurs during normal operation results in protective mode 3, which causes the engine to run at a fixed, severely restricted power level.

AGT-1500 fuel delivery is accomplished by the main metering valve in the EMFS. A Linear Voltage Differential Transformer (LVDT) connected to the bottom of the main metering valve reports the valve's position (WF-actual) to the ECU. WF-request, the ECU's request for fuel flow, controls the metering valve by generating a WF-solenoid current that modulates the solenoid's stem; movement of the solenoid's stem hydraulically positions the metering valve. When WF-request and WF-actual are equal, a nominal null (WF-solenoid) current of 275 milliamps is used to maintain the requested position of the metering valve. When WF-request increases (acceleration) or decreases (deceleration), WF-solenoid current increases or decreases respectively, causing the metering valve to move to the new position. The difference between WF-request and WF-actual during this transition generates an error signal that is used to correct the position of the main metering valve. When WF-actual again equals WF-request, WF-solenoid current returns to its null value of 275 milliamps.

Sensor data for the ANN training and testing were collected from turbine starts by Textron, the tubine engine manufacturer, and at Aberdeen Proving Ground by the U.S. Army Ordnance Center & School. The sensors were sampled at frequences ranging from 3 to 10 per second. Initial data sets were collected from mostly fault-free starts. These data were analyzed to understand how the sensor values behave during fault conditions. Our approach for training the ANNs required the use of data from starts with faults. Because of the difficulties in generating such data with a real turbine engine, we translated some data sets from fault-free starts to faulty starts. Additional data sets from fault-free and faulty starts are being incorporated into TEDANN.

Figure 2 shows the behavior of WF-request, WF-actual and WF-solenoid during a normal start and during the three monitored fuel faults. Normalized data are shown at a sampling rate of three per second. In general, the data tend to be quite messy; this underscores the advantage of the ANN approach. Figure 2a shows the normal condition in which actual fuel flow closely tracks the requested fuel. Figure 2b shows significant vertical motions in WF-actual. This condition of bouncing MMFV is caused by air penetration into the fuel system. Figure 2c (in the last 80 samples) shows a steady level on WF-actual caused by a stuck MMFV. Figure 2d shows irregular WF-actual motions that neither have significant vertical motions nor remain steady on one level. Also, the WF-solenoid varies substantially at several sampling times. We classify this fault type as a fuel flow error. Thus, the sensor behavior during a fuel flow error is characterized by data that cannot be attributed to the bouncing or the stuck MMFV.

TEDANN analyzes the sensor values in the form of "features" computed from the data. During development these features were recognized as discriminators among the three fault conditions. The use of sensor values alone as the input to a simple feedforward ANN do not capture information in the time domain. Thus, to capture changes in individual sensor values over time, we used first derivatives of sensor values and first derivatives of differences between pairs of sensor values as input to the ANN.

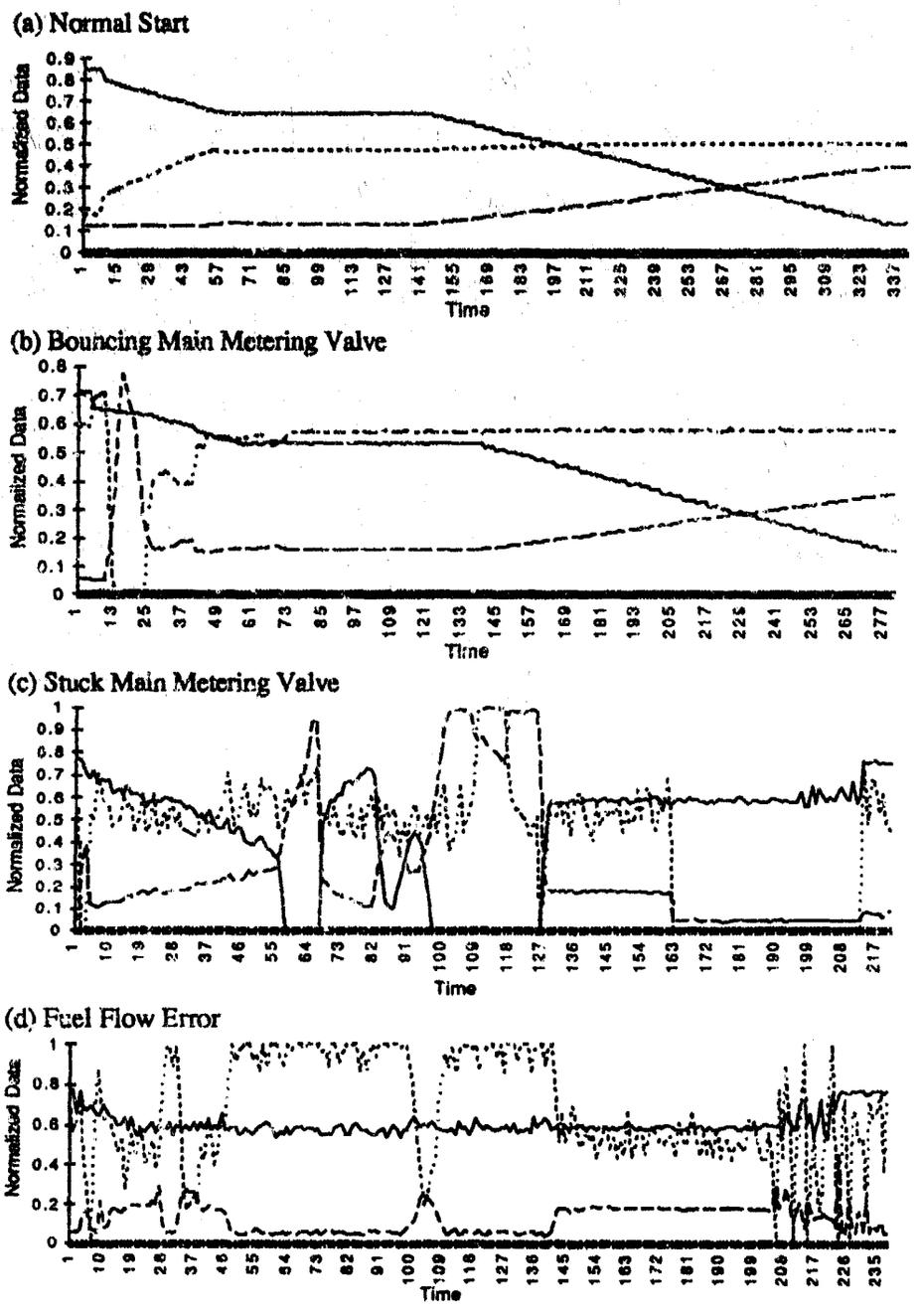


Figure 2. Sensor data plots for three sensor inputs to TEDANN (normalized data) under four engine status conditions: (a) normal start; (b) bouncing MMFV; (c) stuck MMFV; and (d) fuel flow error. Sensor data: WF-Request (solid line); WF-Actual (dashed line); WF-Solenoid (dotted line).

Preliminary studies on recurrent ANNs were performed to determine their applicability to this problem. Recurrent ANNs are specifically suited to capture time dependent data in time series, such as sensor values. These ANNs have not yet been incorporated in TEDANN because they require a large number of data sets for training, which was not available for the first prototype. In the future, when sufficient data are available (e.g., through automatic data collection), recurrent ANNs will be further evaluated for applicability to TEDANN.

Based on the analysis by the ANN system, TEDANN determines which fuel flow voltage readings are out of tolerance with EMFS nominal operational parameters. Having determined the operational condition of the EMFS, TEDANN will display either a fault status message identifying the EMFS faults or a message stating that the EMFS is fully operational. At a future date, when TEDANN and TED are integrated systems, the output of TEDANN will be submitted to TED for further processing.

TEDANN displays diagnostic information for each of the three monitored fuel faults, using a continuous severity scale from zero to one. Values close to zero indicate no fault, and values close to one signify a severe fault. We have arbitrarily assigned the following interpretation of the severity values:

0.00-0.25	no fault (normal)
0.26-0.75	warning (fault)
0.76-1.00	critical (fault)

In the future, we expect that the TED expert system will provide a more realistic interpretation using rule-based post-processing of TEDANN's output.

Preliminary results indicate that TEDANN performs the correct diagnosis. Table 1 shows a confusion matrix that compares TEDANN's output for the three fault diagnoses as a function of actual conditions. The tabulated results are TEDANN s outputs (severity of the faults for each type of fault, shown in the three columns), averaged over several start data sets. The rows represent the actual conditions under which the data were collected. The correctly diagnosed severity levels for the three faults are underlined. For example, in an actual condition of stuck MMFV, TEDANN diagnosed the fault to a severity of 1.0, and correctly failed to find evidence for the other faults. The bottom row shows the output for known no-fault conditions: in this case the diagnosed fault severity levels are appropriately low for the three faults.

Table 1. TEDANN's Diagnostic Performance

	Diagnosed condition:		
Actual conditions:	Bouncing MMFV	Stuck MMFV	Fuel flow error
Bouncing MMFV	0.84	0.22	0.06
Stuck MMFV	0.03	1.00	0.00
Fuel flow error	0.10	0.09	0.47
No fault	0.06	0.19	0.41

The main TEDANN displays are summarized in Figure 3. Two operational contexts are supported in the prototype to demonstrate the proposed concepts. First, a mechanics context is supported in a simple display used to operate the system, summarize results, and access diagrams, schematics, or instructions to aid the maintenance process. Second, a development/expert context is supported by additional windows that display detailed diagnostic information (including the sensor data stream and continuous output of the

ANN); an ANN training window for defining the ANN configuration and setting up or executing the ANN training phase; and a status summary window that provides status lights summarizing various engine components or subsystems. Available from most screens is an electronic maintenance guide, implemented using a hypertext design that can incorporate graphics, schematics, photographic images, and video. For illustrative purposes, the Main Screen is shown in Figure 4.

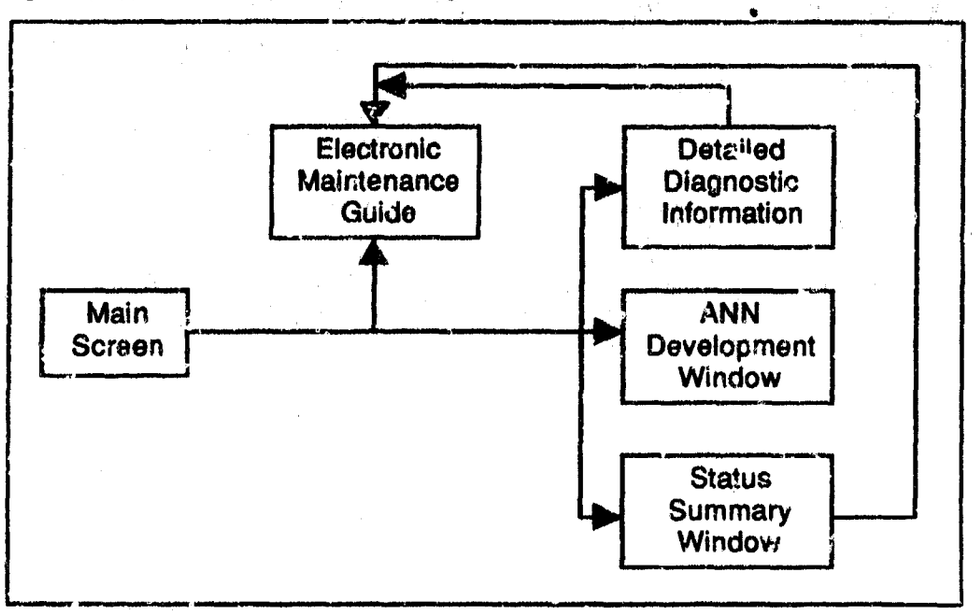


Figure 3. TEDANN prototype window/navigation summary.

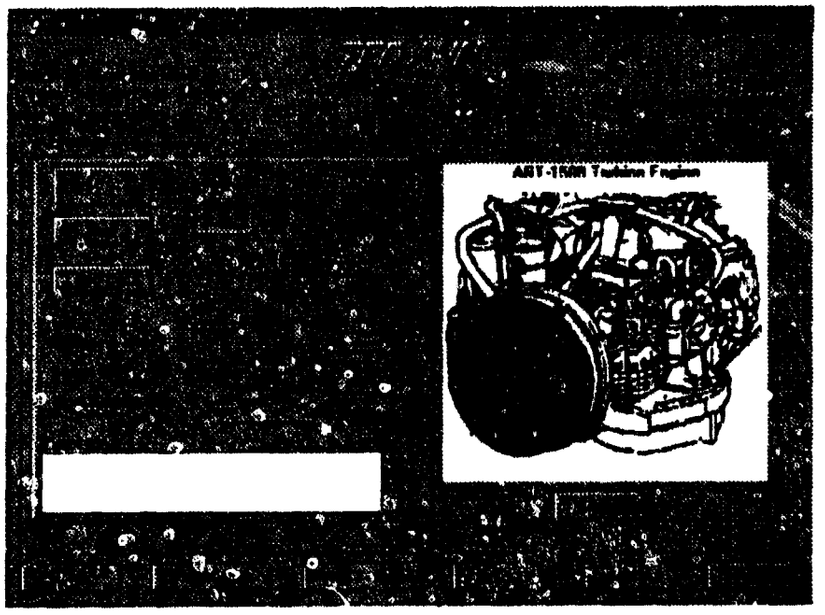


Figure 4. TEDANN Main Screen

Prognostic Capability: ANNs are capable of recognizing trends in data. This fact can be exploited for prognostics. In the short term, it would reduce the amount of man hours spent on diagnosing de facto failures. Further, it would decrease the requirement for ancillary TMDE as the predictive maintenance system is driven by internal sensor information. In addition, a robust predictive maintenance system would decrease the probability of erroneous replacement of operational components due to poor initial diagnosis. This would in turn reduce the number of NEOFs reported at maintenance facilities due to evacuation of operational components that were erroneously removed from systems.

In the long term, predictive maintenance technology would be a powerful and valuable tool to the design engineer as well as the system operator and maintainer. Once a robust and viable predictive maintenance system is developed, it would allow for a system design to be qualitatively analyzed across a range of operational parameters. This would enable design flaws to be corrected before the system is fabricated. In addition, it would enable the operator and maintainer to query the system about its relative operational state prior to the onset of any sustained operation. Any identified flaws could be detected and corrected. This capability would enable combat commanders to evaluate their systems before an engagement and obtain an accurate assessment of available combat power. Probable system failures could be transmitted through available communications media to support maintenance, and the required maintenance assets and repair parts could be dispatched to the unit before the battle. This capability would be a radical departure from the current maintenance support structure that is, by design, reactive in nature.

Conclusions: Preliminary results indicate that our approach to maintenance diagnostics will save time and improve performance. The ABOB's capability to automatically convert analog voltage readings to digital format will save time and markedly increase diagnostic accuracy. With the TEDANN/ABOB interface, a series of manual calculations and decisions may be eliminated to yield further improvements. The application of ANN technology appears to hold great promise for enhancing the effectiveness of Army maintenance practices.

The future of Army Maintenance technology is not dependent upon improvements within our current diagnostic's paradigm. Rather, it lies with the development of a predictive rather than reactive maintenance system. The amount of time and money required to develop this capability is arguably worth the investment, given the potential for reductions in both maintenance man-hours and erroneous replacement of operational system components. In addition, successful integration of predictive maintenance technology would result in the introduction of a new term in the Army lexicon—prognostics rather than diagnostics.

Acknowledgment: This research and development was conducted for the U.S. Army Ordnance Center & School under a related services agreement with the U.S. Department of Energy under Contract DE-AC06-76RLO 1830.

ON-LINE DIAGNOSTICS OF RECIPROCATING MULTI-STAGE AIR COMPRESSORS

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

Scott Delmotte
Dresser-Rand Company
Engine Process
Compressor Division
100 Chemung Street
Painted Post, New York.14870

Abstract: The U. S. Navy is exploring the use of automated monitoring and diagnostic systems for shipboard machinery in the Integrated Diagnostics Demonstration (IDD). The IDD includes a six-month sea trial starting in 1994. As a part of the IDD, The Johns Hopkins University Applied Physics Laboratory (JHU/APL) is developing an Air Compressor Diagnostic System (ACDS) to monitor and diagnose failures in High and Low Pressure Air Compressors (HPACs and LPACs). Dresser-Rand Company, the manufacturer of the target compressors, is providing the diagnostic expertise.

Key Words: Compressor; Confidence; Diagnosis; Expert System; Fuzzy Logic; Sensor Validation; Severity

Introduction: ACDS is one of six subsystems diagnosing various shipboard machines as part of the IDD (Figure 1). The subsystems communicate their diagnostic results to a core system, which integrates the results into a common interface for the user. ACDS must first identify and isolate failing or failed compressor components, and second, provide the user with the information necessary to respond to the failure.

Of the three major subsystems in ACDS, data acquisition, diagnostics, and communication, the diagnostic subsystem is the novel component. The knowledge base uses thermodynamic data from the compressors; 25 pressure and temperature sensors on the HPAC and 18 on the LPAC. Three existing diagnostic systems for air compressors use similar thermodynamic sensors [References (a)-(c)]. The novel features of ACDS are:

1.) the extensive sensor validation,
2.) the use of fuzzy logic principles,

3.) the use of multiple sources of diagnostic confidence, and
4.) the calculation of severity factors.

The user interface and explanation facility are external to ACDS, residing with the core system. JHU/APL, however, has a lead role in designing the standard format for the user interface and explanation facility. Reference (f) fully describes these subjects.

Sensor Validation: User confidence in diagnostic systems degrades rapidly with false alarms. Therefore, ACDS performs sensor validation to distinguish between compressor failures and failures in the ACDS hardware. In addition to the thermodynamic sensors, the diagnostic knowledge base uses data from voltage and power transducers, and AC relays. A total of 32 sensors are used on the HPAC, and 22 sensors on the LPAC. Statistical and analytical-redundancy techniques are used to detect and isolate sensor drift, and failures of sensor, cabling, signal conditioners, multiplexers, power supplies and analog-to-digital converters. ACDS hardware failures are reported to the user for corrective action. In addition, the levels-of-confidence of individual sensor readings are incorporated into the level-of-confidence of the diagnostic conclusions. A comprehensive description of the sensor-validation procedures in ACDS is provided in Reference (g).

A survey of the literature identified three classes of sensor-validation techniques: statistical comparison, analytic redundancy, and neural networks. Statistical comparison techniques are simple to develop and use. These techniques include limit-checking, variance thresholding, data-spike detection, and trending analysis. Data-spike detection is difficult for ACDS since the signal conditioning contains a low-pass filter. In addition, for ACDS, not much is gained from long-term trending analysis that cannot be obtained using limit-checking. In some cases, however, changes in the transient response (short-term trending) of a sensor can identify sensor failure. Therefore, ACDS uses the limit-checking, variance thresholding, and short-term trending statistical techniques.

The complete failure of a sensor, or a piece of data-acquisition hardware, can be detected using the statistical techniques. The ranges of the sensors are chosen so that the minimum and maximum limits of the sensor are beyond what is reasonably expected. For instance, absolute, rather than gage pressure transducers are used since a zero reading indicates a vacuum instead of atmospheric pressure. Therefore, any absolute pressure transducer reading a zero is flagged as failed. Some resolution in the pressure reading is lost, but satisfactory accuracy is obtained with the absolute pressure transducers.

The signal conditioners, multiplexers, power supplies and analog-to-digital (A/D) converters are also checked using simple statistical techniques. The signal conditioners include a 4 Hz low-pass filter, so that large variations in the signal indicate that a failure has occurred after the signal conditioner. Highly-variable sensor readings, indicating a broken wire or a loose connection after the signal conditioner, are identified by thresholding the variance of the sensor readings. The power-supply outputs are connected to the multiplexers and the A/D converters, providing, through the appropriate logic, both a check of the power supplies, and the multiplexers and A/D converters.

The second class of sensor-validation techniques is analytical redundancy. Analytical redundancy uses physical principles or empirical equations to estimate sensor values. The estimated values are used to diagnose failures in both the machine and the sensors. Some diagnostic approaches lend themselves to these approaches more than others. Merrill, et. al., [Reference (h)] describe the use of linear estimators to model machine dynamics. When a reliable linear estimator can be constructed for a machine, the diagnostics and sensor validation are integrated. ACDS is not, however, based on a linear model of an air compressor. The rule base is built on the heuristic expertise of humans. A sensor-validation procedure that is divorced from the air-compressor diagnostics is needed.

An analytical-redundancy technique suited for ACDS is the use of parity equations. These equations relate only two sensors, rather than the entire suite of sensors. An example parity equation is: requiring the air exiting a water-cooled heat exchanger to be hotter than the cooling water. This relationship is mandated by the thermodynamics of the system. ACDS uses over fifty parity equations.

Neural networks offer great potential for both sensor validation and diagnostics. The current lack of a comprehensive training set, however, prevents the use of neural networks. Upgrades to ACDS, however, may include neural networks trained on the data recorded during the IDD.

Knowledge Base: We used the Nexpert Object™ software package to develop the knowledge base. The object data structures provide a convenient method for organizing and storing information in the knowledge base. In addition, the knowledge engine provided by Nexpert Object allows complete control of the execution of the rules. The knowledge base performs a layered diagnosis. The layers of the diagnosis progressively narrow the list of possible faults, reflecting the diagnostic procedures used by the human experts. This approach results in a powerful explanation facility, guiding the user down the diagnostic path.

The components being monitored on the HPAC and LPAC are:

inlet and discharge valves,
piston rings or seals,
air pressure relief valves,
air coolers,
air packing around piston rods,
unloader valve,
general system leak,
backpressure/check valve,
freshwater/seawater heat exchanger,
freshwater and oil pumps,
oil pressure relief valve,
condensate drains and monitors, and
temperature monitors.

ACDS monitors a total of 61 components on the HPAC, and 34 components on the LPAC. Some of the components being monitored are physically composed of several sub-components, any of which might fail. A failure indication for a discharge valve, for instance, could be caused by a faulty valve or a faulty valve gasket. ACDS is unable to distinguish between the sub-components. A finer diagnosis, however, is not required. In general, the same inspection and repair procedure is used for all of the sub-components of a particular component. ACDS presents the user with a list of all sub-components of the failed component as part of the explanation facility.

Although a vibration analysis can better diagnose the rotating parts of the compressor, failure data show that these systems seldom fail. The air end of the reciprocating compressors experiences the most failures. The air end is the collection of parts that contact the compressed air. Since these failures are not readily detected with a vibration analysis, ACDS uses only thermodynamic, and not vibration, data.

The ACDS knowledge base for the HPAC evaluates the health of the components listed above using 143 rules. The rules break into three categories. The first group, the fault identification rules, contains 93 rules that identify a particular failed component. These rules contain the minimum information required to identify a failure. The next category, the confirmation rules, has 27 rules to increase the confidence in the diagnosis. The final group of 23 rules assigns a severity factor to the failure mode. Clearly, not every component requires rules in the latter two categories. These last two groups are discussed in later sections.

Each rule compares measured readings, or values derived from these readings, with a limit. We quickly recognized that a rule base with hard limits would not be adequate. User acceptance is a driving design constraint. The discontinuous behavior of a hard-limited rule base will not engender user confidence. A machine whose health is slowly degrading is not at one instant healthy, and at the next instant, completely failed. A fuzzy representation of the data was developed, resulting in gradual changes in the diagnosis given gradual changes in the state of the machine.

A sensor reading is assigned eleven fuzzy attributes: low6, low5, low4, low3, low2, low1, normal, high1, high2, high3, and high4. Each attribute is a membership function that ranges from zero (non-membership) to one (full membership). The eleven attributes were chosen based on the different limits placed on the sensors for different rules. As in most fuzzy systems, the attributes are not mutually exclusive. A temperature can be both high1 and high2 to some degree. In most fuzzy systems reported in the literature, however, full membership in the attributes are mutually exclusive. That is, a temperature cannot have low1 and low2 values equal to one [left half of Figure 2]. ACDS appears to be unique as a fuzzy-knowledge-base, however, by not enforcing this restriction. Doing so would needlessly complicate the rule base. For example, provided that the attributes are ranked, a temperature that has exceeded the high2 limit has also exceeded the high1 limit. The right half of Figure 2 illustrates example fuzzy membership functions as used by ACDS.

The membership function for the attributes of the sensor readings are individually assigned. The function is a linear increase in membership with increasing (for high attributes) or decreasing (for low attributes) values of the sensor reading. The mid-point and span of the membership function is assigned separately for each limit. The use of fuzzy attributes does not preclude the use of hard limits, where appropriate. The span can be assigned an arbitrarily small value.

Since the rules are written in terms of low, normal and high values, the knowledge base easily transfers to other types of reciprocating air compressor. The fundamentals of compressor diagnostics are constant, only the numerical mid-points and spans change. Since the LPAC compresses air in two stages, and the HPAC in five, the LPAC knowledge base is a subset of the HPAC knowledge base, with minor changes.

Several components, such as relief valves, are diagnosed with one rule. Many others, however, have multiple diagnostic rules, and even multiple levels of rules. Figure 3 shows a sample rule tree leading to the diagnosis of a second-stage inlet-valve leak. The first level of the diagnosis identifies a leak in either the inlet or discharge valve of a particular stage. Two separate rules can lead to the diagnosis of a second-stage valve problem (Valves2 in Figure 3). The second level of diagnosis includes two rules that identify the leak to be in the inlet valve, and not the discharge valve. An additional rule confirms this conclusion by checking for additional possible, but not required, conditions.

Confidence Factors: ACDS calculates a confidence factor for each diagnosis based on four sources of uncertainty. The first, is the reliability of the data being supplied by the sensor. The sensor-validation routines supply a measure of data reliability. The statistical tests for the data acquisition hardware are definitive. If the system fails a test, we can isolate the failed component and assign zero confidence to the data associated with that hardware. The parity tests for the sensors, however, are more numerous and complex. The failure of a particular test does not identify a failed component with absolute certainty. The uncertainty in the results of multiple parity tests for a sensor must be combined into a single statement about the sensor's condition.

A possible cause of the failure of a parity-equation test is a sensor out of calibration. The sensor that is actually out of calibration cannot be determined. Let us return to the heat exchanger example discussed in the Sensor Validation section. If this parity test fails, either the air or cooling water temperature sensor, or both, may be inaccurate. ACDS halves the confidence in a sensor (originally set to 100%) every time it fails a parity equation. This is equivalent to stating that there is a 50% probability that each sensor reading is incorrect. The chance that both sensors are simultaneously incorrect is assumed small and is ignored. Sensors that fail more than one parity equation, have a successively halved confidence.

The second source of uncertainty is in the limits chosen by the diagnostic system developer. Is a discharge temperature high when it reaches 130 or 135 degrees? For machine A the true limit may be 133 degrees, and for machine B it may be 136 degrees. The fuzzy attributes assigned to each sensor reading capture some of this uncertainty.

The center of the high1 limit can be set to 135 degrees, and span a 10 degree interval. A reading of 131 will yield a high1 membership value of 0.1, indicating a small but finite confidence that the high1 limit has been reached. A reading of 139, on the other hand, sets high1 to 0.9, showing high confidence that the limit has been reached.

In summary, fuzzy attributes serve three important functions. The first is to produce gradual and continuous diagnostic results with continuous changes in machine condition. The second is to increase the transportability of the diagnostic system across types of compressors, by removing numerical limits from the rules. The last function is to reduce the tuning required to diagnose different machines of the same model.

The third source of uncertainty exists inherently in the rules. Not every diagnostic reasoning path is 100% accurate, catching all failures and signalling no false alarms. ACDS allows the independent assignment of reliability or confidence measures for each rule.

Lastly, more a source of certainty than of uncertainty, confirmation rules also affect the confidence in a diagnostic conclusion. An example of a confirmation rule is shown in Figure 3. These rules search for information that is expected but not necessary for a diagnosis. Satisfaction of the constraints imposed by these rules increases the confidence in the conclusion.

ACDS combines these four uncertainties into a confidence factor for each diagnostic. Figure 4 shows the rule trees from Figure 3, but with sample confidence calculations indicated. Several constraints are imposed on the calculations of the confidence factor. First, only three degrees of confidence will be communicated to the user: low, moderate, and high confidence. Further resolution in confidences confuse, more than help, the user. Numerical confidences are translated into this coarse grade by splitting the interval from zero to 100% confidence into thirds. Therefore, a numerical confidence from 0% to 33% is displayed as low confidence, 34% to 66% as moderate confidence, and 67% to 100% as high confidence.

A second constraint on the confidence calculations is that the process must be consistent with the user's intuition. Using our own intuition as models, each fault identification rule is assigned a maximum confidence. Fault identification rules that have associated confirmation rules, are given, in general, a maximum confidence of 66%. The confirmation rules are assigned maximum confidences totalling the remaining 34%. Thus, a diagnosis that is not supported by its confirmation rules can have a moderate confidence at the greatest. Those fault identification rules without confirmation rules are given, in general, a maximum confidence of 100%. A few exceptions are made, however, for faults that are difficult to identify. These rules do not have confirmation rules, and have a maximum confidence of only 66%.

Returning to Figure 4, each sensor-reading confidence is multiplied by its corresponding fuzzy-attribute value giving a confidence in the individual antecedent. The minimum of the antecedent confidences for a given rule carries through the calculations. Since all antecedents are "ANDed" in a rule, each antecedent can be considered a link in a chain.

A rule can only be as reliable as its "weakest" link. The minimum antecedent confidence is multiplied by the reliability of the rule, giving a confidence in the conclusion (Rules R1 and R2 in Figure 4). When several rules lead to the same conclusion, the maximum confidence of the rules is preserved (Conclusions C1 and C2 in Figure 4). Individual rules are "ORed" to reach the same conclusion. The rules are therefore parallel reasoning chains leading to the same conclusion. The conclusion is thus as reliable as its most reliable rule. The confirmation rules add an increment to the confidence in the diagnostic conclusion (Conclusion C3 in Figure 4). Each confirmation rule has a confidence calculated in a similar manner as the fault identification rules.

Not all antecedents inspect sensor reading fuzzy attributes. Three other antecedent types are possible. The first is a timing antecedent. Some rules require the compressor to be operating in steady-state. Timing antecedents assure that a fixed amount of time has elapsed since an event before the rule will fire. Although they could be expressed in terms of fuzzy attributes, these antecedents are given 100% confidence. The second type is a value derived from possibly multiple sensor readings. A derived value has the minimum of the confidences in the sensor readings used in its derivation.

The last type of antecedent is the diagnostic conclusion of another rule. For negated diagnostic conclusions (e.g. If there is not a second-stage valve failure and ...) these antecedents have a 100% confidence. Refined diagnostic conclusions are treated differently. When the rules R3 and R4 in Figure 4 refine the conclusion "Valves2 is Faulted", the confidence of the conclusion "Valves2 is faulted" directly reduces the confidence of rules R3 and R4. Thus, the confidence in an second-stage inlet-valve leak is based on the confidence in a leak in either of the second-stage valves.

Severity Factors: Severity factors indicate the degree of the fault. Some components tend to fail abruptly, or at least, their detectable behavior changes abruptly. The diagnostic conclusions associated with these components are always given a high severity rating. Other components, however, degrade more gracefully. The extent of the degradation often can be estimated by the extent that the compressor behavior has deviated from normal. Severity rules are used to measure this deviation. All of the air-end rules, for instance, calculate their severity based on the final-flow-rate reading. The failure of an air-end component leading to an 8% to 17% reduction in final mass flow is assigned a moderate severity. Those failures reducing the final flow by more than 17% are assigned a high severity. Therefore, whereas the confidence factor indicates the reliability of a diagnosis, the severity factor describes the condition of the compressor.

Conclusion: ACDS has been implemented in a Nexpert Object™ knowledge base. The sensor validation, and calculations of fuzzy membership, confidence and severity factors are implemented using the functionality supplied by Nexpert Object. The techniques described in this paper are unique for the diagnosis of reciprocating machinery.

The next step in the development of ACDS is validation and verification. The first-level tests have used historical and simulated sensor data. A fault-insertion test program is planned at a land-based HPAC, testing the diagnostic ability, and usability, of ACDS. Following the land-based tests, ACDS will monitor one HPAC and one LPAC on a U.S.

Navy ship for a six-month demonstration. User satisfaction with ACDS will be evaluated both during and after the shipboard test.

Acknowledgements: This work is supported by the U. S. Navy. Special thanks are due to Wayne Boblitt, Jeff Craighead, and Bob D'Antonio of the Carderock Division of the Naval Surface Warfare Center for sharing their HPAC diagnostic expertise.

References:

(a) Afimiwala, K. A., Woollatt, D., "A Diagnostic System for Reciprocating Compressors," Proc. 1988 Int. Comp. Eng. Conf., August 4-7, 1986, pp. 1063-1072.

(b) Monaco, M., Gardiner, A., "NAVSSES HPAC/HPAD Expert System," Focus on Mechanical Failures: Mechanisms and Detection, Proc. 45th Meeting of the Mech. Failures Prev. Group, April 9-11, 1991, pp. 117-122.

(c) D'Antonio, R. C., Sheedy, P., "A Knowledge Based Approach to a Condition Based Maintenance System for Naval High Pressure Air Compressors (HPAC)," Tech. Proc. of the Fleet Maint. in the 21st Century Joint Symp., October 22-23, 1991, pp. 65-88.

(d) Nemarich, C. P., Boblitt, W. W., Harrell, D. W., "A Condition Based Maintenance Monitoring System for Naval Shipboard Machinery," DTRC-PAS-90/17, June 1990.

(e) Hegner, H. R., Hudak, J. P., Nemarich, C. P., Boblitt, W. W., "System Study of Condition-Based Maintenance System for Shipboard High-Pressure Air Compressors," Focus on Mechanical Failures: Mechanisms and Detection, Proc. 45th Meeting of Mech. Failures Prev. Group, April 9-11, 1991, pp. 123-133.

(f) Coury, B.G., Lin, J.S., Weiskopf, F.B. "System representations, cognitive modeling and the design of user interfaces for an expert diagnostic system." In the Proceedings of the IEEE/SMC '93 Conference (Piscataway, NJ: IEEE Service Center), pp 26-31, 1993.

(g) Lin, J. S., Resch, C. L., Palmer, J. V., "Sensor Validation for On-line Diagnostics," to be published in Proceedings of the 40th ISA International Instrumentation Symposium, May 1-5, 1994.

(h) Merrill, W. C., DeLaat, J. C., and Bruton, W. M., "Advanced Detection, Isolation, and Accommodation of Sensor Failures - Real-Time Evaluation," Journal of Guidance, Control, and Dynamics, Vol. 11, No. 6, Nov.-Dec. 1988

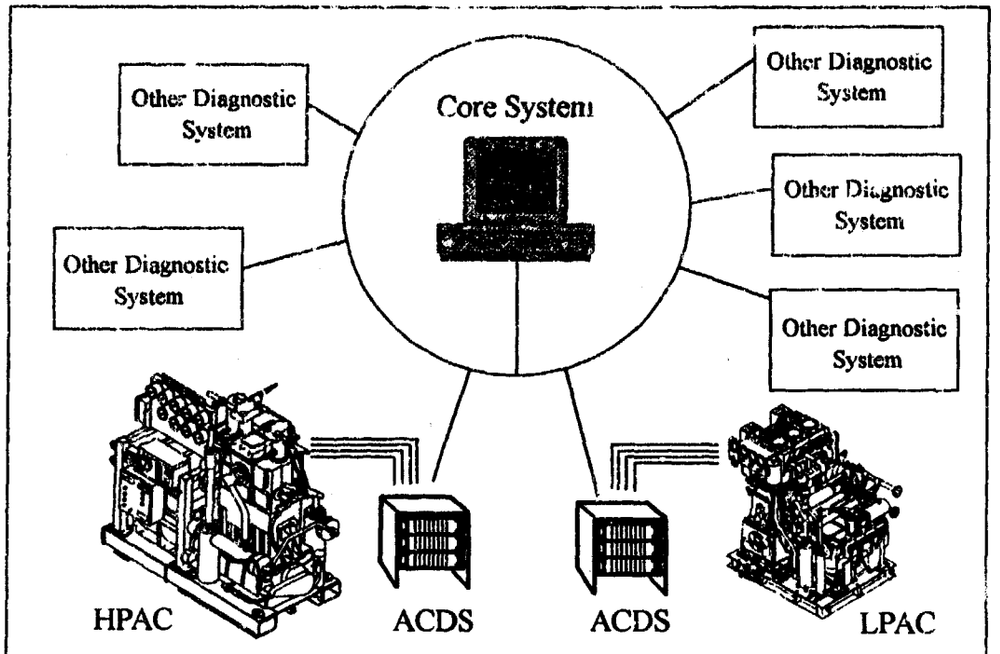


Figure 1. IDD Architecture

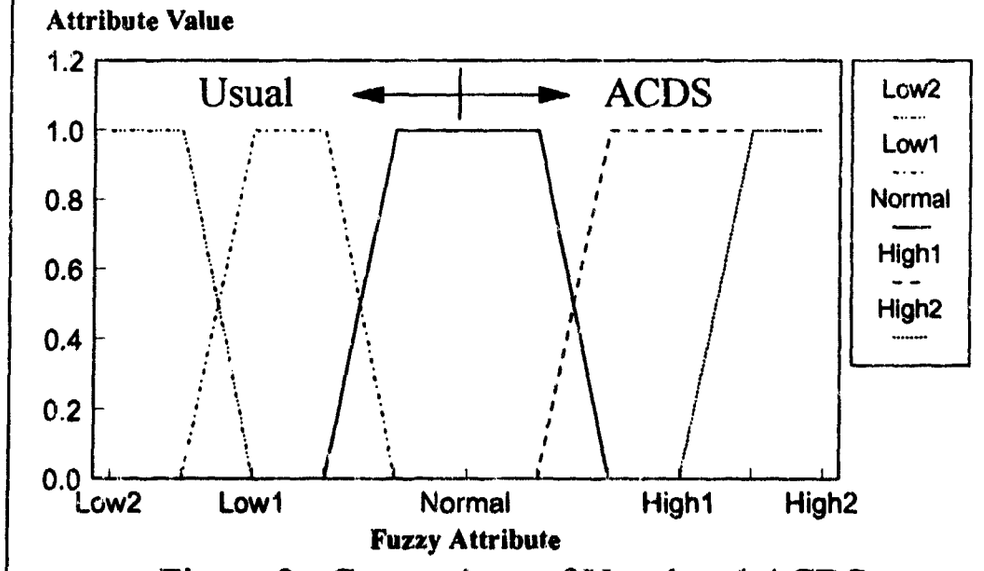


Figure 2. Comparison of Usual and ACDS Fuzzy Attributes

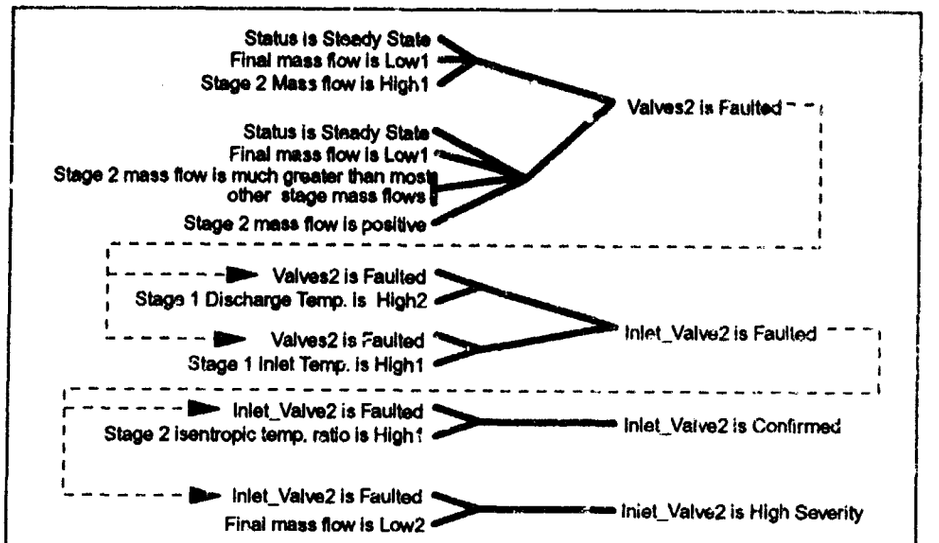


Figure 3. Rules for Stage 2 Inlet Valve Failure

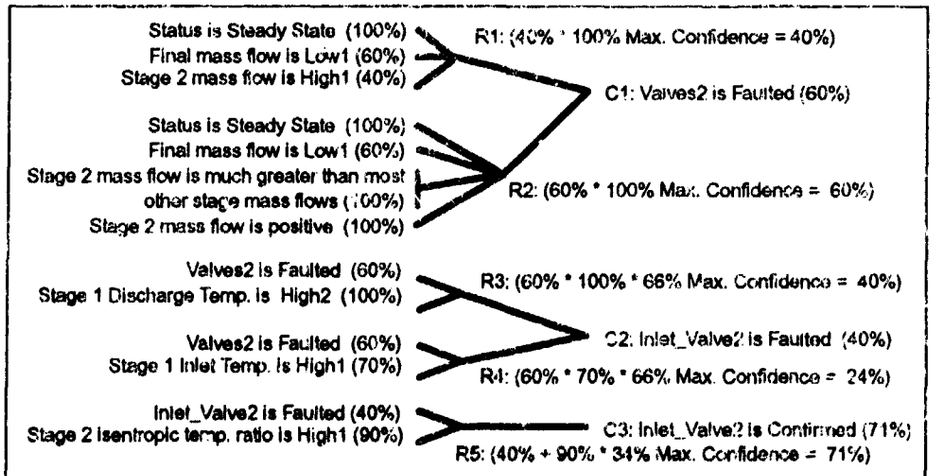


Figure 4. Confidence Calculation for Stage 2 Inlet Valve Failure

NONLINEAR CORRELATION ANALYSIS FOR ROCKET ENGINE TURBOMACHINERY VIBRATION DIAGNOSTICS

Jen Jong
AI Signal Research
904 Bob Wallace Ave, Ste 211
Huntsville, AL 35801

Jess Jones, Preston Jones, Thomas Nesman, Thomas Zoladz
NASA MSFC
Marshall Space Flight Center, AL 35812

Tom Coffin
Wyle Laboratories
7800 Highway 20 W
Huntsville, AL 35807

Abstract: In support of Space Shuttle Main Engine (SSME) development and advanced technology test programs, a number of sophisticated diagnostic algorithms have been developed and applied to specific SSME mechanical symptom investigations.[1-7] This paper reviews some nonlinear applications in mechanical system detection and identification technology. Conventional linear spectral analysis has long been used to identify the signal characteristics associated with machinery faults in vibration signature analysis. However, nonlinearities can play a significant role for signature identification. It has been observed that different rotational mechanisms may interact due to some nonlinear process.[8,9] Application of the bicohedrence and tricoherence to identify nonlinearities of different order in a random signal is described. The hypercoherence function is illustrated to detect the correlation between synchronous frequency characteristics and any *harmonic* component.[3,4] The generalized hypercoherence,[5] also described, permits estimation of the nonlinear correlation between a selected frequency (e.g., shaft speed) and *an arbitrary* frequency component. Investigation of a full wave rectification (FWR) technique is described as applied to the detection of cavitation in turbopump systems. The technique is based on detecting the modulation between periodic rotational components and high frequency noise due to the collapse of cavitation bubbles. This paper briefly summarizes the practical application of the methods through the evaluation of vibration measurements from SSME hot firing tests.

Key Words: Fault detection; machinery diagnostics; nonlinear detection methods; rotational system identification.

Introduction: The Marshall Space Flight Center is responsible for the development and management of advanced launch vehicle propulsion systems, including the Space Shuttle Main Engine (SSME), which is presently operational. The SSMEs provide high performance within stringent constraints on size, weight, and reliability. Based on operational experience, continuous design improvement is in progress to enhance system durability and reliability.

Under the extreme temperature, pressure, and dynamic load environments sustained during operation, engine systems and components, such as pumps, turbines, and

associated hardware, are subject to severe pressure oscillations and damaging mechanical vibrations. Most mechanical failures are preceded by growing tolerances, imbalance, bearing element wear and the like, which may manifest themselves through subtle changes in the waveform observed by dynamic measurements. Diagnostic vibration analysis is based on observing and discriminating between measurable vibration patterns that occur as a result of nominal system operation and those associated with component degradation. In support of SSME development and advanced technology test programs, a number of sophisticated diagnostic algorithms have been developed which are applicable to specific SSME mechanical symptom detection and identification investigations.

In this paper, bi-spectral analysis is illustrated to identify amplitude modulation (quadratic correlation) among spectral components. The characteristics of such quadratic interactions are usually reflected through coherent phase relationship. It has been applied to identify the quadratic phenomenon of a synchronous frequency component modulated by the cage frequency components in a ball bearing system. Tri-spectrum analysis may be used to identify cubic correlation among four different spectral components. A special case of its application is to identify whether an apparent sideband structure is really due to modulation or not. A typical example is a bearing cage frequency component periodically excites a structure mode at its natural frequency.

The hyper-coherence function was developed to detect the correlation between synchronous frequency characteristics and any harmonic component. A major benefit is to determine whether an apparent harmonic in a complex vibration signal is correlated with the fundamental or caused by extraneous noise. This analysis has been extended to arbitrary (noninteger) frequency components, and also as a time-domain algorithm (hyper-coherence filtering) for the extraction of period signals in noisy data. The full wave rectification (FWR) technique can be used to detect the nonlinear correlation associated with cavitation phenomenon in turbomachinery. This technique is based on the modulation process between low frequency periodic rotational components and the high frequency noise due to the collapse of cavitation bubbles. Such a nonlinear modulation phenomenon thus provides a unique signature in the high frequency region conductive to cavitation detection.

In the following sections we briefly summarize the analytical basis for the above algorithms. Practical application of the methods is then demonstrated through the evaluation of vibration measurements from rocket engine hot firing tests exhibiting anomalous behavior.

The Bispectrum and Bicoherence: Given a stationary, zero mean process, the ordinary (linear) spectrum $S_{xx}(f)$, may be defined by

$$S_{xx}(f) = D[X(f)\, X^*(f)]$$

where X( ) denotes the Fourier transform and E[ ] is the ensemble average.

The next higher order is called the bispectrum $B_{xxx}(f_j,f_k)$.

$$B_{xxx}(f_j,f_k) = E[X(f_j)\, X(f_k)\, X^*(f_j+f_k)]$$

Succeeding terms can be written out following the permutation rules for higher order cumulants of random variables.

The bicoherence, a normalized bispectrum, $b(f_j,f_k)$ is defined as

$$b_{xxx}(f_j,f_k) = \frac{|\,B_{xxx}(f_j,f_k)\,|}{\{E[\,|X(f_j)\,X(f_k)\,|^2]\;E[\,|X(f_j+f_k)|^2]\}^{1/2}}$$

By Schwartz' inequality, it can be shown that the bicoherence is bounded by zero and unity. If the wave at f_j+f_k is totally correlated to the waves at f_j *and* f_k, the bicoherence will equal unity. On the other hand, if these three waves or any one of them are statistically independent, the bicoherence will be zero.

Bispectral analysis can be used to identify the existence of amplitude modulation (quadratic correlation) among spectral components. The bispectrum measures the degree of correlation by identifying phase relationship among three spectral components, i.e., frequency sum or difference. It has been applied to identify the quadratic phenomenon of a synchronous frequency component modulated by the cage frequency in a ball bearing and a synchronous component modulated by subsynchronous whirl frequency component. As an example, Figure 1a is the PSD of a vibration measurement taken on the turbopump on engine test 902-436. The peak marked "N" is the sync frequency component N and the peak marked "SS" is the suspected 52-percent subsynchronous whirl frequency component. There is also a component at synchronous frequency N plus SS. The level of SS component is not high enough to be of concern if it is due to independent sources. But it would be critical if it is a subsynchronous whirl component which is synchronous related. Figure 1b shows one slice of bicoherence $b(w_1, 318.75$ Hz) with the second frequency argument fixed at the subsynchronous frequency. Several peaks are detected, and the most significant peak is the one located at (N, SS; N+SS) which indicates that the SS and SS+N components are both synchronous related.

The Tricoherence: The trispectrum can be used to identify cubic correlation among four spectral components. A special case of its application is to determine whether or not an apparent sideband structure is due to modulation or not. Such a sideband structure is another commonly observed nonlinear defect signature. Figure 2a illustrates a preburner pump radial measurement during hot firing. Notice that even though the synchronous and overall RMS levels are nominal, an apparent sideband structure consisting of three peaks marked N-120, N, and N+120 is observed, which could indicate modulation of the synchronous frequency by a lower frequency component. Figure 2b shows a slice through the tricoherence by fixing two of the three independent frequency values. If the sideband and synchronous component are truly correlated, a peak would be expected at f=N+120. Since this does not occur, it indicates the sideband components are due to an independent source and are not synchronously related. In this case, an incorrect conclusion might be made based on the ordinary spectrum.

Hypercoherence Functions: To summarize the nonlinear interaction between harmonically related spectral components in a given stationary, zero mean signal, we define the *Hyperspectrum* of order n by the relation[1]

$$H(n;\,f_1) = E[X^n(f_1)\;X^*(nf_1)], \qquad n = 1,2,3,...$$

where f_1 is an arbitrary reference frequency, and nf_1 is an integer multiple of f_1. Thus, the ascending terms in $H(n;f_1)$ represents a single value from the linear spectrum, bispectrum, trispectrum, etc. at the specific value $f_1 = f_2 = ... = f_n$.

In analogy with the ordinary coherence function, we define the hypercoherence as a normalized hyperspectrum:

$$\Gamma^2(n;\,f_1) = \frac{|\,E[X^n(f_1)\;X^*(nf_1)]\,|^2}{E[\,|\,X^n(f_1)|^2\;E[\,|X(nf_1)\,|^2]} \qquad n=1,2,...$$

The hypercoherence function defines the nonlinear correlation between a reference frequency component in a vibratory signal and its harmonics. A major benefit is determination of whether an apparent harmonic in a complex vibration signal is correlated with the fundamental or caused by extraneous noise. The technique was applied to space shuttle main engine turbopump measurements. The linear spectra of two different tests appear virtually identical, other than the background noise (Figures 3a and 3b). The PSD amplitudes at 3N frequency are very high for both tests. Figures 4a and 4b depict the hypercoherence functions computed for the same two test measurements. Figure 4a indicates that almost all the power at 3N is correlated with the rotational frequency component. On the other hand, the 3N component of the second test is due to a differing physical phenomenon not related to the rotational frequency. This indicates an improved degree of signature discrimination.

The spectral components to be identified by higher order spectra are required to satisfy certain frequency combinations (e.g., the sum of arguments is zero). However, in many situations, we wish to identify the correlation between two arbitrary frequency components that do not satisfy any such requirements. The generalized hypercoherence (GHC) was developed to deal with this situation. The GHC can identify the correlation between two arbitrary spectral components in the sense of frequency synchronization or lock-in.

A vibration signal may be treated as an FM signal with different spectral components at different center (carrier) frequencies. Assume that there is some intelligence being frequency modulated in the signal as the instantaneous frequency about these carriers. To recover the intelligence, we demodulate the FM signal to estimate its instantaneous frequency signal. A narrow-band random process can be modeled as a sine wave with slowly varying amplitude A(t) and phase p(t):

$$x(t) = A(t) \cos [2\pi f_c t + p(t)]$$

The instantaneous frequency $f_i(t)$ is defined by:

$$f_i(t) = \frac{d\, p(t)}{d\, t}$$

A logic diagram illustrating the computational procedure is shown in Figure 5. Figure 6a shows the PSD of a strain gauge measurement taken from a SSME High Pressure Oxidizer Pump (HPOP) hot firing test. The peak marked 5.7N is located at 5.7 times shaft frequency N, near the predicted outer ball pass (OBP) frequency, and a possible defect indication. Figures 6b, c, and d show the instantaneous frequency variation of components N, 5.7N, and an arbitrarily chosen component at 500 Hz, by the GHC method. Strong frequency synchronization is identified between components N and 5.7N and no correlation between components N and noise at 500 Hz. This indicates that the 5.7N is a sync-related component and a potential outer race defect signature.

Full Wave Rectification (FWR) Spectral Analysis: The FWR technique can be used to detect the nonlinear correlation associated with cavitation phenomenon in turbomachinery. This technique is based on a special property associated with cavitation physics. That is:

"When cavitation occurs, the periodic rotational components (such as shaft or blade speed) will amplitude modulate the wide-band high frequency noise generated from the collapse of cavitation bubbles."

The signal generated by cavitation can be modelled as the multiplication of two separate components p(t) and N(t):

$$x(t) = p(t)\, N(t)$$

Here, N(t) represents the wide-band high frequency noise generated from the collapse of cavitation bubbles, while p(t) represents the low frequency periodic pressure fluctuation due to the impeller rotational process.

Such a nonlinear modulation phenomenon thus provides a unique signature in the high frequency region conducive to cavitation detection.[10] However, conventional PSD analysis is unable to identify the signature associated with such a modulation phenomenon. This can be easily deduced from studying its signal model. The operation between N(t) and cos(Wrt) in the time domain is multiplication, but this multiplication becomes a convolution in the frequency domain. Since the PSD of noise is flat and that of a sine wave is a delta function, the convolution of these two functions remains flat without any discrete peak. In this section, the Full Wave Rectification (FWR) spectral analysis method will be discussed for cavitation detection.

For FWR spectral analysis, the raw signal must be high-pass filtered first to remove low-frequency discrete components while keeping the high frequency noise data. Demodulation is then performed in order to recover any existing low frequency periodic component which is modulating the cavitation noise signal. As a result, a new low frequency FWR PSD is generated. Unlike the raw data PSD which included multiple contributions from rotor-dynamics, structure dynamics and hydrodynamics, this new FWR PSD only reflects the high frequency effect due to cavitation-generated modulation. If cavitation occirs, the FWR PSD will show discrete peaks corresponding to the low frequency periodic rotational process modulating the collapsing bubble noise.

A series of Inducer Test Leg (ITL) water flow tests were conducted to study cavitation phenomena in the water flow test facility at Marshall Space Flight Center (MSFC). From video taken during these tests, one can clearly visualize how cavitation develops and changes from one state to another, such as four-blade cavitation to alternate-blade cavitation. The signal from a vibration or high frequency pressure measurement also reflects such a cavitation condition change. However, in complicated operational conditions as experienced during flights or engine hot firing tests, these vibration signals will also contain elements induced by rotor-dynamics, structural dynamics, and hydrodynamics, which make it difficult to identify the sources. Therefore, it is highly desirable to have an effective method for cavitation detection and monitoring using vibration signal analysis.

The FWR spectral analysis for Inducer Test Leg (ITL) water flow test data was performed and cavitation condition predictions compared to actual cavitation conditions visually monitored during the test. In this testing, ordinary PSD analyses accurately reflect cavitation conditions. Therefore, the accuracy and effectiveness of the demodulation signal for cavitation monitoring can best be verified using this water flow test data.

Figure 7 shows a regular PSD isoplot taken from a Kistler pressure measurement from Inducer Test Leg (ITL) Water Flow Test. At the beginning of the test, video information indicates cavitation bubbles attached to each of the four inducer blades (four-blade cavitation). Then at a later time, around S+69s, two alternating bubbles out of the original four disappear. This indicates that the original four-blade cavitation condition has changed into alternate-blade cavitation. Later on at S+92s, the cavitation phenomenon totally disappears. This cavitation condition change is indeed accurately reflected by the

raw PSD isoplot of Figure 7. At the beginning of the test when the inducer is experiencing four-blade cavitation, the 4N PSD component is dominant. At S+69s, when the inducer begins experiencing alternate-blade cavitation, the 4N component diminishes while the 2N component starts to grow and dominate. Finally, when all cavitation phenomenon disappears at S+92s, both the 2N and 4N components diminish. Therefore, in this particular case under isolated laboratory environment, the traditional raw PSD is already accurately monitoring the cavitation condition.

Figures 8a and 8b are the ordinary PSDs of the raw signal and high-pass filtered signal of the Kistler measurement with a maximum frequency of 50 KHz. In this high-pass filtered version, all the low frequency discrete components under 5000 Hz are removed. Demodulation will be performed based on the wide-band high frequency noise floor of Figure 8b. Figure 9a shows the raw data PSD at the beginning of the test, S+0s. The 4N component is dominant since the four bubble regions due to four-blade cavitation are generating a four pulse per revolution pressure fluctuation. Figure 9b shows the corresponding FWR PSD. This FWR PSD recovers several discrete components with the 4N component being the dominant one. This FWR PSD indeed correctly reflects the fact that the inducer is experiencing four-blade cavitation at the beginning of the test.

Figures 10a and 10b are the raw data and the FWR PSDs at S+69s when the four-blade cavitation changes into alternate-blade cavitation. The dominant peak of the FWR signal has now changed from 4N to 2N, which again correctly reflects actual cavitation condition changes the inducer experiences.

Figure 11 shows the corresponding PSD isoplot of the FWR signal. This time-frequency energy distribution agrees quite well with the regular raw data PSD isoplot and correctly monitors the cavitation condition change. However, when dealing with data from real world cases such as flight or hot firing tests rather than an isolated laboratory environment, a raw data PSD will reflect contributions from other effects (e.g. rotor and structural dynamics). These low frequency contributions will make it difficult to identify the existence of cavitation. In this case, the FWR signal becomes an effective tool for cavitation monitoring.

Concluding Remarks: The above discussion provides an overview of some diagnostic efforts in support of the Space Shuttle Main Engine program. The work is strongly driven by engine test observations, but the techniques should find broad application. The FWR technique discussed in this paper suffers from a severe limitation. Since the hidden periodicity is recovered from the envelope signal of a high frequency wide-band noise floor, any discrete components present in this high frequency region will generate false discrete peaks in the FWR signal. Development of an efficient and effective method for wide-band demodulation without such discrete-interference is a matter for future research.

References:

1 Coffin, T., W. L. Swanson, and J. Y. Yong. "Data Analysis Methods and Signal Processing Techniques for Space Shuttle Main Engine Diagnostic Evaluation," Wyle Laboratories Final Report (NAS8-36549) Oct. 1989.

2 Jong, J.Y. "Advanced Signal Analysis Methods for SSME Diagnostic Evaluation," MSFC, S&D Lab Seminar Notes (NAS8-38156). April 1990.

3 Coffin, T., and J. Y. Jong. "A Nonlinear Coherence Function and its Application to Machine Diagnostics," 110th Meeting of the Acoustical Society of America, Nov. 1985.

4 Jong, J.Y., and T. Coffin. "Diagnostic Assessment of Turbomachinery by the Hyper-coherence Method," NASA Cof. Pub. CP2436, May 1986.

5 Jong, J. Y., T. Coffin, J. H. Jones, J. E. McBride, and P. C. Jones. "Correlation Identification Between Spectral Components by Generalized Hypercoherence," 3rd Internat'l Mach. Monitor. and Diag. Conf., Dec. 1991.

6 Jong, J. Y., T. Coffin, and J. H. Jones. "A Non-linear Hyper-coherence Filter for Damage Assessment in Noisy Environments," Model Determination for large Space Systems Workshop (J. C. Chen, Ed.) JPL Doc. D-5576, vol. 3, Mar. 1988, pp. 1129-1142.

7 Coffin, T., and J. Y. Jong. "Some Nonlinear Spectral Methods and Their Aplication to Rocket Engine Diagnostics," First Internat'l Machinery Monitoring and Diagnostics Conf., Sept. 1989.

8 Collacott, R. A. *Mechanical Fault Diagnosis and Condition Monitoring*, Chapman and Hall, London, 1977.

9 Eshleman, R. L. "The role of sum and difference frequencies in rotating machinery fault diagnosis," 2nd Internat'l Conf. on Vibration in Rotating Machinery, Churchill College, 1980.

10 Dwyer, R. S., "Fourth-Order Spectra of Gaussian Amplitude-Modulated Sinusoids," J. Acoust. Soc. Am. 90(2), Pt. 1, August, 1991.

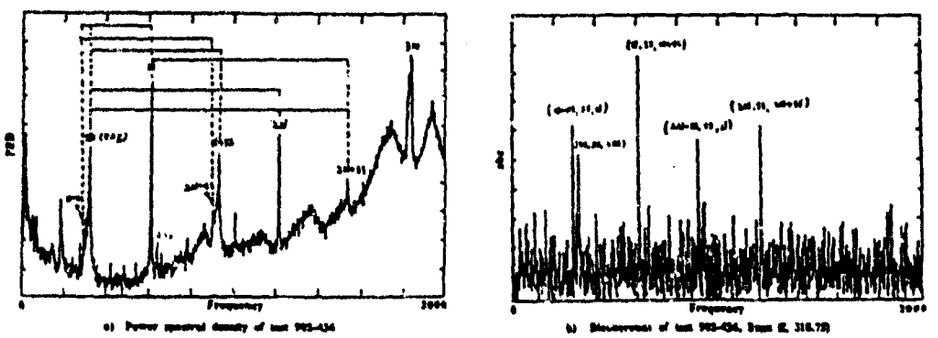


Figure 1. Power Spectrum and Bicoherence from Engine Test 902-436

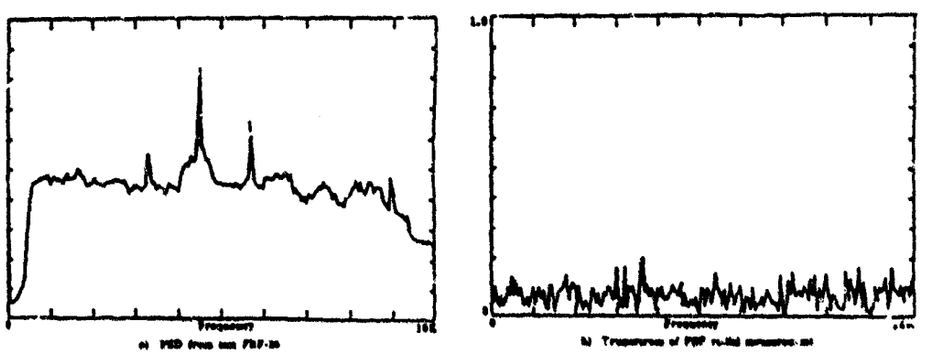


Figure 2. Power Spectrum and Tricoherence from Engine Test FRF-26

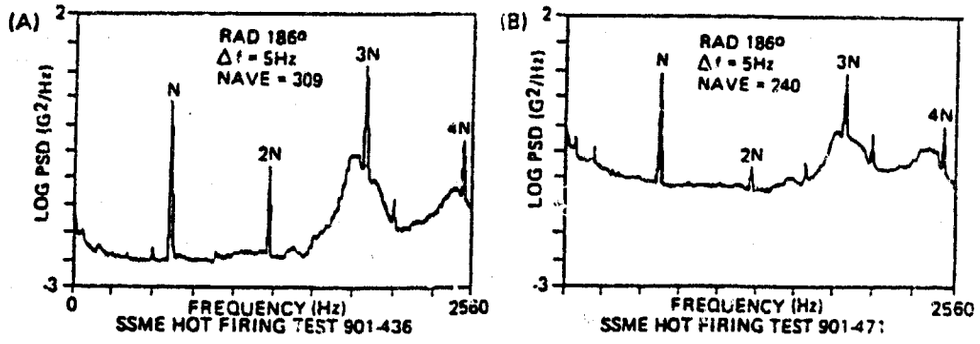


Figure 3. Power Spectra from High Pressure Fuel Pump Measurements

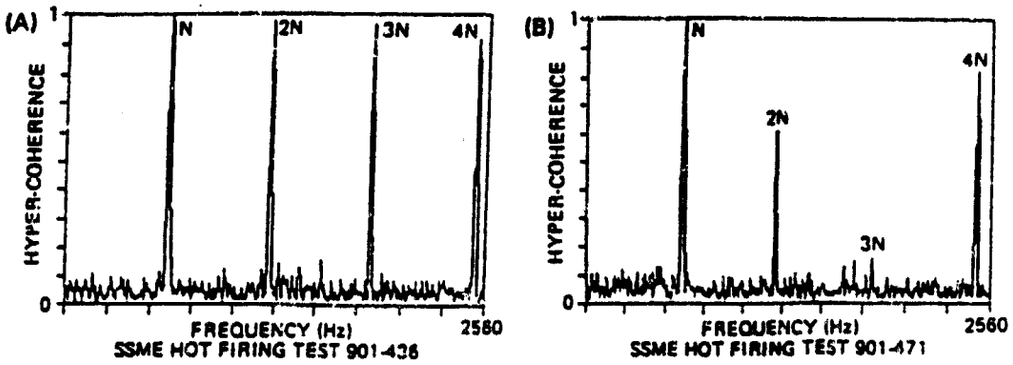


Figure 4. Hypocoherence Spectra

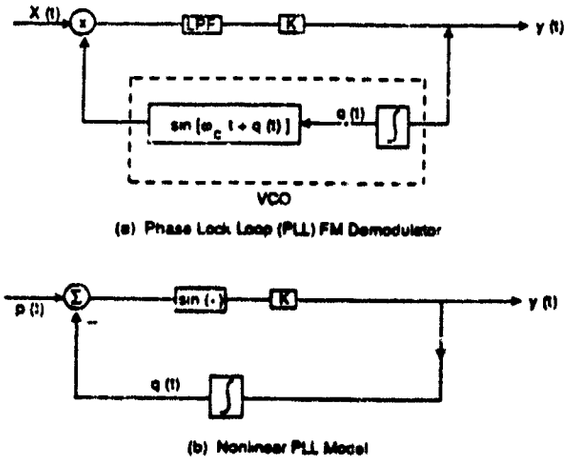
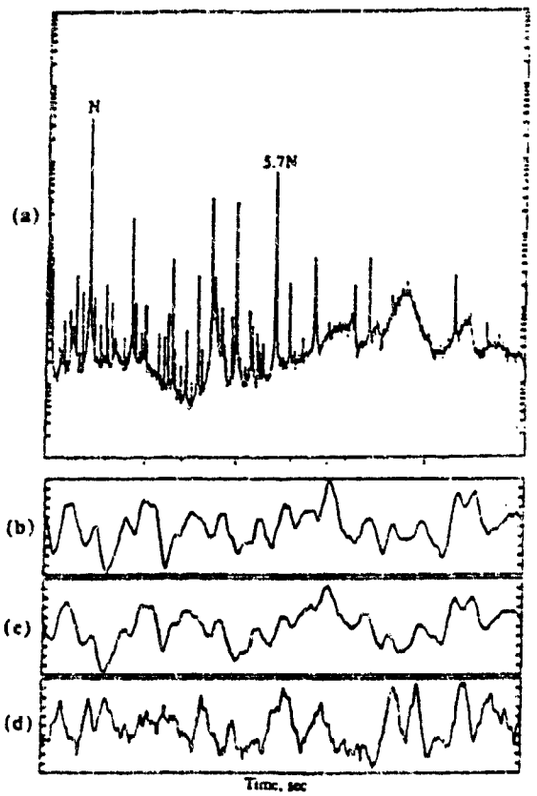


Figure 5. GHC Through FM Demodulator



(a) PSD of SSME HPOP internal strain gauge measurement
(b) Frequency signal of N (sync)
(c) Frequency signal of 5.7N (OBP)
(d) Frequency signal of a component at 500 Hz

Figure 6. Correlation Detection by Generalized Hypercoherence

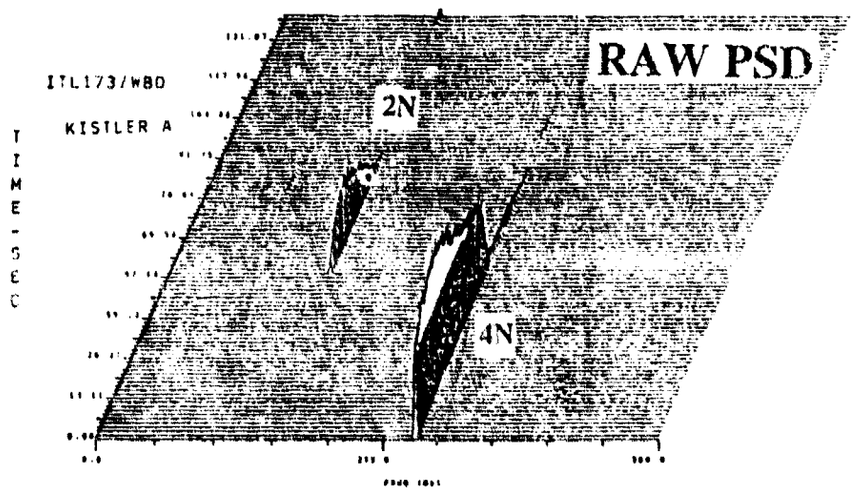


Figure 7. Power Spectra During Flow Test

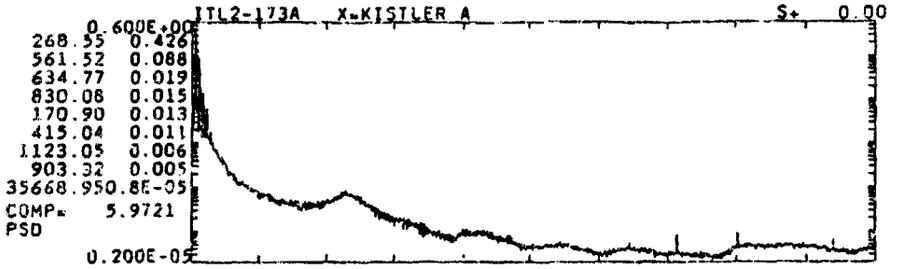


Figure 8a. Pressure Spectrum at Start of Test

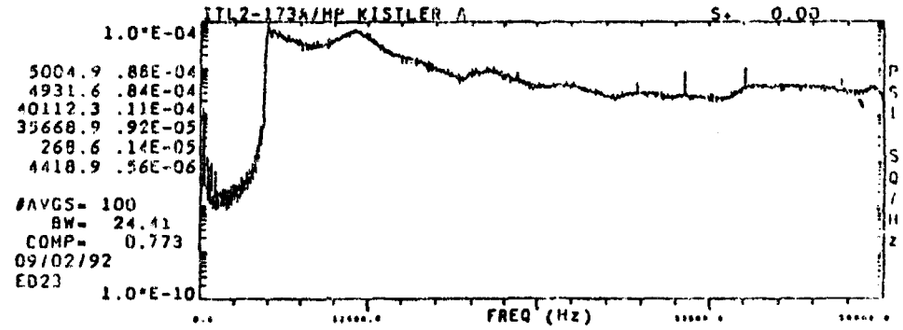


Figure 8b. High-Pass Filtered Spectrum at Start of Test

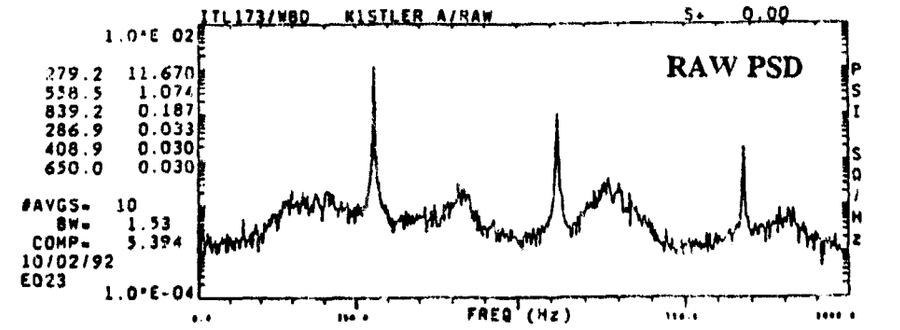


Figure 9a. Raw PSD at Test Start

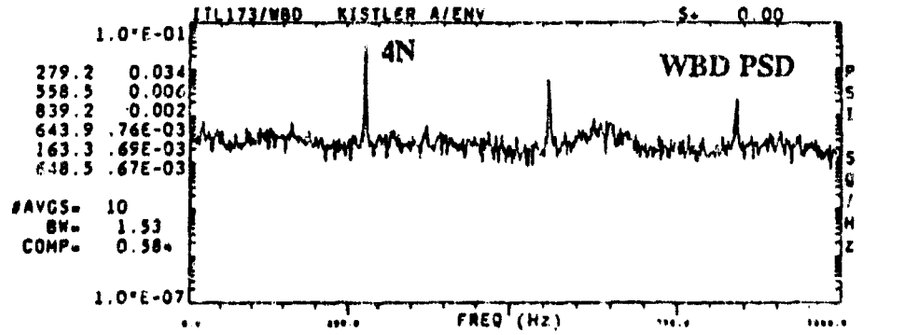


Figure 9b. FWR PSD at Test Start

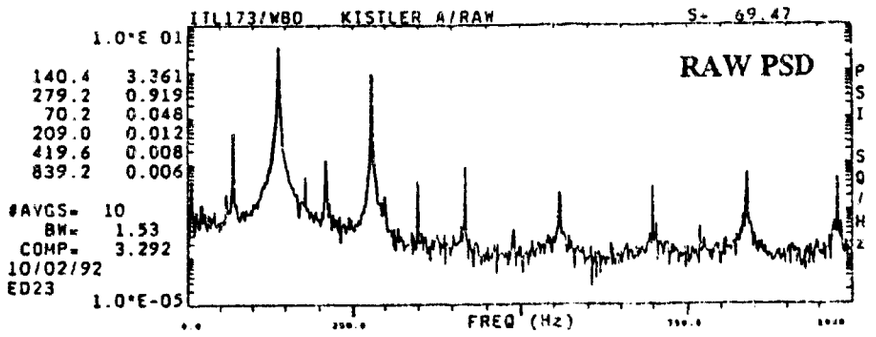


Figure 10a. Raw PSD at 69 Seconds

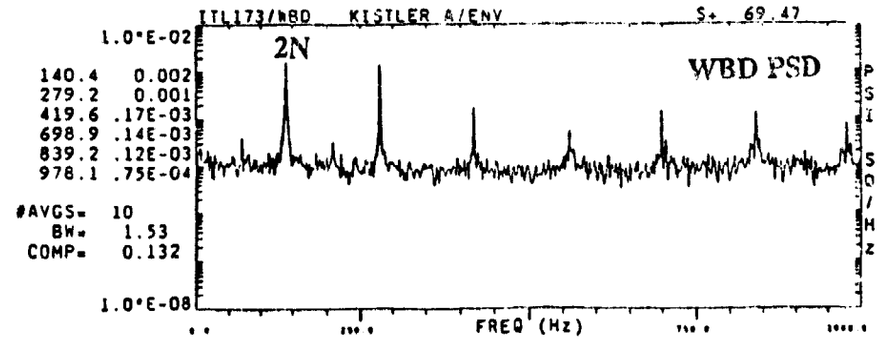


Figure 10b. FWR PSD at 69 Seconds

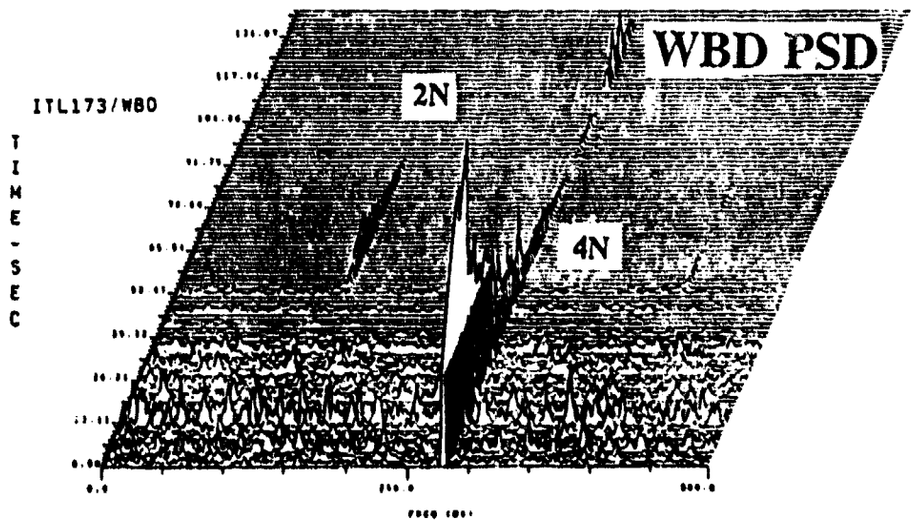


Figure 11. FWR (Demodulated) Power Spectra During Flow Test

NONDESTRUCTIVE TEST AND EVALUATION

Cochairmen: Al Broz
Federal Aviation Administration

Robert Pepper
Textron Specialty Metals

A GENERAL PURPOSE IMAGE ANALYSIS INSPECTION SYSTEM

Patrick J. Sincebaugh
U.S. Army Research Laboratory
Materials Directorate
Watertown, MA 02172-0001

Abstract: This paper describes a General Purpose Image Analysis(GPIA) system that can be utilized to automate the inspection of manufactured assemblies and solve image pattern recognition problems. The objective in developing the GPIA system was to create an adaptive tool that can be utilized to inspect various components and assemblies without making any software modifications. This was accomplished by combining a powerful image analysis library with neural networks technology. Neural networks provide a more robust methodology for inspection than systems implementing other classification techniques such as template matching and statistical methods. A neural network based inspection system differs from other types of systems in that it is trained, rather than programmed. Therefore, knowledge of the component being inspected is more important than being capable of writing software. This allows the system to be utilized by shop floor personnel with only a general understanding of image processing and neural networks, rather than a software engineer. This paper provides an example of how the GPIA system was implemented to inspect artillery fuzes.

Key Words: Automation; backpropagation; image analysis; inspection; neural networks; pattern recognition; testing;

Introduction: Manual inspection techniques are often time consuming, tedious, costly, and prone to errors caused by human fatigue and subjectivity. The manufacturing community has attempted to overcome the inherent problems associated with manual inspection by developing automated inspection systems. However, one of the common characteristics of these automated systems is that they are customized for each application. This results in high costs for developing and modifying each system.

The objective of developing the General Purpose Image Analysis(GPIA) inspection system was to create a robust, user-friendly tool that can be implemented to automate the inspection of various components and assemblies without making software modifications. This was made possible by combining a state of the art image processing library with a neural network based classification scheme. The use of neural networks to classify items allows the GPIA system to be trained rather than programmed. This enables the system to be adapted to inspect different products in a timely, cost efficient manner.

The GPIA system is an 80486 PC based system that is configured with an Alacron Neurosimulator i860 board, a DataTranslation DT2871 color frame grabber board, and a DT2869 video encoder/decoder board(see figure 1). The Neurosimulator i860 board with 8MB of on-board memory is utilized as a coprocessor for the computationally intensive neural network and image algorithm calculations. The use of a coprocessor board frees up the system resources, allowing them to be used for other tasks. This allows real-time image processing applications to be developed. The function of the DataTranslation DT2871 color frame grabber board is to convert analog signals into a digital format that can be manipulated by the computer. For monochrome applications, the less expensive DT2851 board can be substituted for the DT2871. The DT2869 board is used for applications which analyze video images. The function of the DT2869 is to convert RS-170 video signals into a format that can be manipulated by the frame grabber board. The most efficient method for transferring video signals to the DT2871 is via a high speed bus known as the DT-Connect interface.

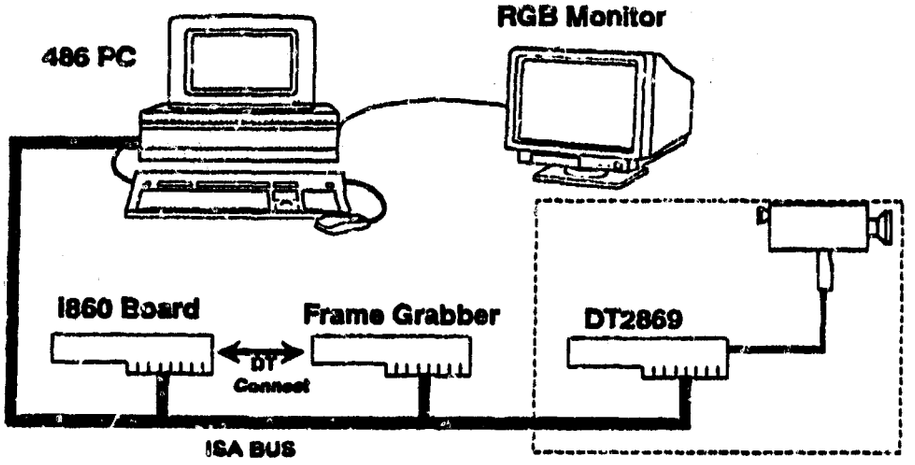


Figure 1. GPIA System Configuration

The GPIA software was developed as a Windows 3.1 application. This allows the user to utilize other useful Windows applications, such as using the Editor to create training files. Windows also provides a user-friendly interface, which allows the user to quickly experiment with different solutions to the inspection problem. The software was designed so that commercial image processing libraries can be easily integrated into the system. Therefore, the software can be upgraded to include the current state of the art algorithms.

Image Processing: The objective of the image processing module is to extract information about the objects that are present within an image. This information is

then used as input into the classification scheme.

The first step in utilizing the GPIA system for inspection is to properly define the problem. It is very difficult to analyze an image that possesses characteristics such as low quality, uncontrolled background scenes, varying image perspectives, and varying object parameters(ie. number of objects). Many of these problems can be avoided by properly constraining the problem. Examples of helpful constraints are the use of artificial lighting and fixing the camera to object distance. Properly constraining the problem will result in less image processing requirements, therefore resulting in a less computationally intensive process.

The next step after defining and constraining the problem is to acquire an image. Images can be acquired by the GPIA system from either a file(Tagged Image File Format or Flat Image File Format), or an external source such as a video camera or Real-time X-Ray machine.

Once the image is stored in the frame grabber buffer, it can be processed by user invoked software algorithms. The first objective in processing the image is to reduce the complexity of the problem. The amount of processing required is dependent upon the type of image being analyzed. The easiest type of image problem to solve is the analysis of a monomorph. A monomorph is an object that can be characterized by one shape, one color, and one level of reflectance. A more difficult problem is the analysis of polymorphs. A polymorph is an object that contains more than one form, color, and/or level of reflectance. The most difficult image analysis problem is the analysis of xenomorphs. A xenomorph is defined as an object that has varying characteristics(ie. shape) that are dependent upon the perspective of the object. The successful processing of complex images using the GPIA system requires that they are reduced to polymorphs, or preferably monomorphs. One technique for reducing a xenomorph to a polymorph is to alter its feature space. This is accomplished by implementing processing tools such as wavelets, or using a non-linear classification scheme such as a Higher Order Neural Networks. A polymorphic image can be reduced to a monomorphic image by invoking the correlation tools that are provided with the GPIA system. For example, a polymorphic image can often be represented by multiple monomorphic templates. The analysis of these templates is often simpler than the analysis of the polymorphic image.

The next step is to further reduce the complexity of the problem by defining what portion of the image is to be analyzed. In most cases, only a portion of the image, called a Region of Interest(ROI) or chip, contains pertinent information. The proper identification of the ROI's will greatly reduce the complexity and computational requirements of the problem. The proper identification of the ROIs will also make training the neural network much easier. The GPIA system allows the user to interactively define chips by using the mouse to draw a box around the ROI.

In many applications, the inspection system can't be successfully trained using the raw pixel values as inputs to the classification scheme. This can be overcome by invoking

image processing algorithms to enhance and extract relevant features from the region of interest. The GPIA system provides numerous algorithms to support geometrical, statistical, point by point, binary, neighborhood, morphological, template matching, correlation, and segmentation processing. These algorithms are automatically implemented by simply adding them to the processing list.

The GPIA systems user-friendly graphical user interface allows for easy experimentation with the various image processing algorithms. The ROIs from each of the training images can be viewed in isolation from the remainder of the image. This allows the user to make an educated guess as to whether the neural network can be trained using the processed chips as input.

Neural Networks: The GPIA system implements neural networks to provide image pattern recognition capabilities. Neural networks were chosen over other methodologies(ie. template matching, statistical methods) due to their inherent robustness for solving different types of pattern classification problems. This robustness is attributed to the fact that a neural network is trained rather than programmed. This allows the GPIA system to be adapted to inspect different items without making software modifications.

The GPIA inspection system allows the user to interactively select which neural network paradigm will be implemented. Current supported models include the backpropagation(BP), enhanced backpropagation(EBP)[1], enhanced enhanced backpropagation(EEBP)[2], and Functional Link Network(FLN)[3] paradigms.

The backpropagation paradigm is a fully connected, feed forward type of network that implements a supervised learning algorithm. The network consists of one input layer, one or two hidden layers, and one output layer. The number of nodes in the input layer is automatically calculated by the image processing module of the GPIA system. This number is equal to the number of pixels that are present in the user defined Region of Interest(ROI). The number of nodes in the hidden and output layers, along with other typical parameters associated with backpropagation networks[4], are selected by the user via the Windows based graphical user interface.

The operation of the backpropagation neural network is described by three phases: the training phase, the testing phase, and the implementation phase. In the training phase, the network is taught to recognize patterns by implementing a supervised learning algorithm. First, a data set that is statistically representative of the problem must be obtained. This data set is then sorted into a training set and a test set. Typically, approximately 85% of the data is used for training the neural network, with the remaining 15% used for testing. The next step is to process the data utilizing the image processing tools described above.

Once the training data is processed, it is presented as input to the neural network. The neural network is trained to classify the image by learning the mapping between the input and output layers. This is done by iteratively showing the neural network

examples of each possible classification scheme along with the desired output. The error between the calculated output and the desired output is then used to adjust the weights so that the error will be reduced. Training is complete when the mean squared error value is less than a user specified limit. Once the network has converged, the test data is used to verify that it has been properly trained.

The GPIA system implements an improvement to the backpropagation paradigm, known as the enhanced enhanced backpropagation(EEBP) paradigm[2]. Utilizing the EEBP paradigm often results in reducing the amount of training time required for a problem. The decreased training time is attributed to the assignment of a learning rate to each individual weight, rather than having the same learning rate assigned to each weight in a given layer. Each learning rate is iteratively updated by applying an exponential decay to the learning rate increase, and adding the momentum term to the result. The learning rate and the momentum term each have a limit in order to assure the proper behavior of the network.

Case Study: The GPIA system is being used to automate the process of inspecting M549 artillery fuzes at the Milan Army Ammunition Plant. Previous inspection techniques included the manual inspection of radiographic images of the M549 fuze. Each fuze is represented by two radiographs, one depicting an axial view(figure 2), the other a transverse view(figure 3). The inspector was responsible for visually inspecting 10 critical regions of interest which were defined for each fuze. This inspection methodology was a tedious, time consuming procedure. Also, many of the inspection errors could be attributed to factors inherent to human inspection, such as fatigue and inspector subjectivity.

An automated inspection system(AFIRS) was developed to overcome the problems associated with the manual inspection process. The robotically controlled system produces digitized radiographs of the axial and transverse view for each fuze. These radiographs are then acquired by a personal computer and analyzed using classical image processing algorithms, such as template matching. The problem with this approach is that it is time consuming and costly to modify the system to inspect different types of fuzes. This modification requires an engineer to develop a different template for each fuze type. Also, the system has not been robust enough to account for slight differences between fuzes produced by different manufacturers. This has resulted in some fuzes being misclassified as either faulted or nominal.

The GPIA system is being integrated with the AFIRS system in order to provide a more robust classification scheme. The use of a neural network classification method enables the system to be easily trained to inspect different types of fuzes without making time consuming modifications. Neural networks also generalize more than template matching techniques, and are therefore less susceptible to slight variations between fuzes produced by different manufacturers.

The first step in implementing the GPIA system was to obtain enough data that was representative of the fuze inspection problem. This data was obtained from the

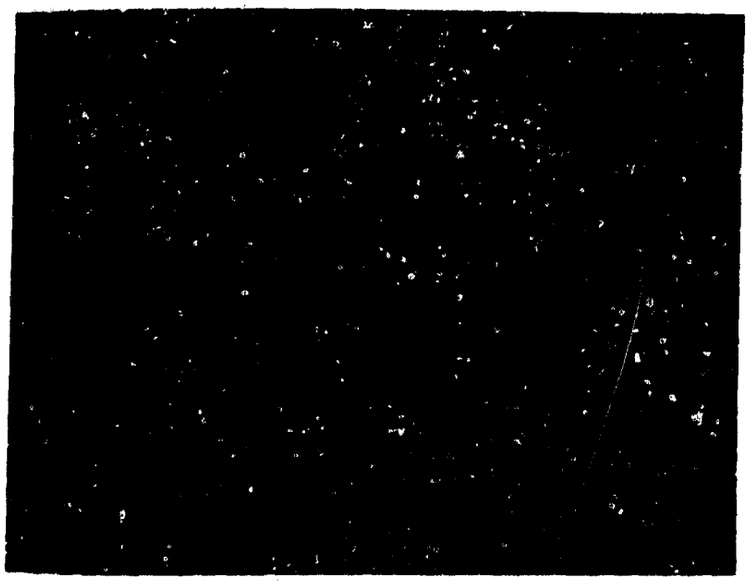


Figure 2. Axial View of M549 Fuze X-Ray

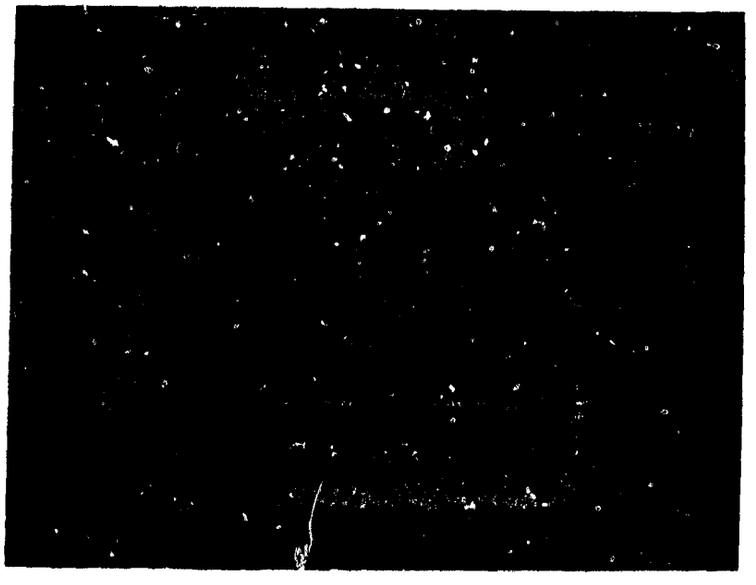


Figure 3. Transverse View of M549 Fuze X-Ray

AFIRS system. The fuze inspection problem is very well constrained by the AFIRS system. Factors such as fuze position, fuze to x-ray source distance, and x-ray quality are kept constant. This greatly reduces the complexity of inspecting the fuzes with the GPIA system.

The critical regions of interest that were used as a basis for inspecting the fuzes manually and with the AFIRS system are also used by the GPIA system. However, the size of the ROIs have been significantly reduced in order to lower the number of pixels used as inputs to the neural network. Figure 4 shows the ROI that was defined in order to inspect whether the actuator was properly assembled. An ROI that would enclose the complete actuator consists of 272x78(21,216) pixels. This corresponds to 21,216 input nodes for the neural network. The ROI that was used to solve this problem contained only 6x40(240) pixels. Experience has shown that this approach not only results in a less computationally intense process, but consequently leads to more trainable neural networks.

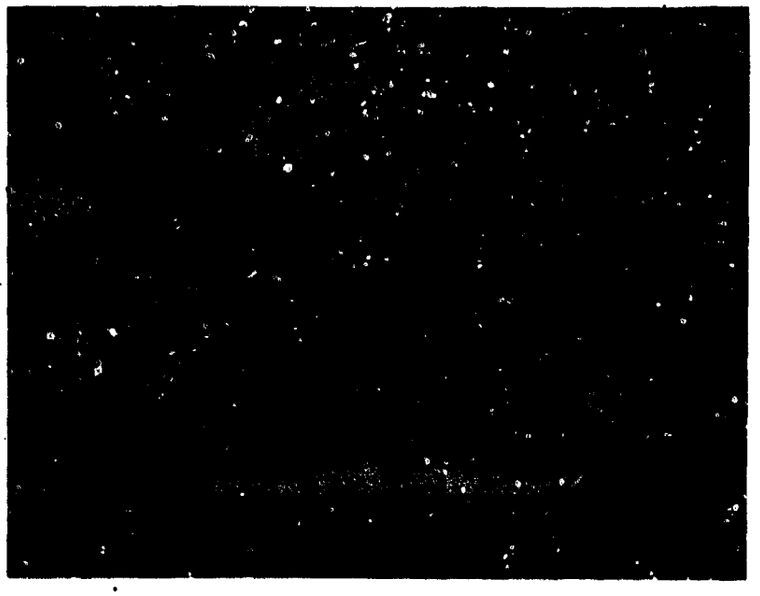


Figure 4. Defined ROI for Inspecting M549 Fuze Actuator

Once the regions of interest were identified, they were processed in order to extract relevant information. The user friendly interface of the GPIA system allowed for the rapid prototyping of solutions. Therefore the user was able to experiment with the effects of applying different image processing algorithms. The algorithms used to

solve this problem were histogram equalization, vertical edge detection, horizontal edge detection, high pass filtering, and median filtering. Figure 5 shows the difference between the unprocessed actuator ROIs and the effects of processing the ROIs with the histogram equalization algorithm. The objective of this processing was to enhance the image in order to determine if more than one actuator had been accidently installed. This can be determined by examining the thickness of the black pixels in the middle of each chip. The effect of the histogram equalization algorithm was to enhance the contrast between the light and dark intensities of the pixels. Training the neural network with the processed data was much easier than training with raw data.

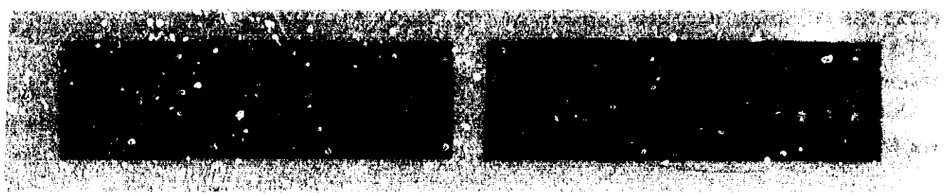


Figure 5. Unprocessed ROIs vs. Processed ROIs for M549 Fuze Actuator

Once the optimum processing algorithms were discovered for each ROI of the fuze, they were implemented on the training data set. An enhanced enhanced backpropagation neural network was then trained for each ROI using the processed data. The EEBP networks were comprised of one input layer, one hidden layer, and one output layer. The number of nodes for the input layer was dependent upon the number of pixels within the defined ROI. For example, the network that was trained to inspect the actuator was comprised of 240 input nodes, 8 nodes in the hidden layer, and 2 output nodes. The output nodes classified the ROI as faulted or nominal. The network was trained until the Mean Squared Error between the desired output and the actual output was less than 0.10.

After the networks were trained, they were tested using the data from the test set. A report of the network output resulting from the test set is automatically generated by the GPIA system. The report lists the image that was used as input to the network, along with the corresponding network output. The user can view the image by selecting it with the mouse. This allows the user to verify the results obtained from the neural network. If the results are satisfactory, then the network is ready for implementation.

Future Applications: The GPIA inspection system is a user-friendly, interactive tool that can be implemented to automate the inspection of various assemblies and components. There has been considerable interest in applying the GPIA system to automate the inspection of microelectronics. Manual electronic inspection is becoming a more critical, costly, and difficult process as electronic technology progresses. A recent survey of electronics manufacturers has shown that 82% of the companies reported that inspection costs are becoming a larger percentage of the manufacturing cost(currently at 24%). Seventy percent of the companies reported

problems associated with manually inspecting electronics according to standards such as Mil-Std-833. Implementing the GPIA system for electronic inspection would alleviate many of the problems associated with manual inspection. Examples of electronic applications for which the GPIA system can be trained to inspect include identifying faulty components(ie. blistered), verifying correct component placement, and inspecting the condition of the conductive paths.

The number of applications for the GPIA system is virtually limitless. The main consideration for determining whether the GPIA system should be used to automate the inspection process is the availability of training data. In order to assure the successful application of the GPIA system, it is imperative that a data set that is statistically representative of the problem to be solved be used to train the neural network.

Conclusions: This paper described how the General Purpose Image Analysis(GPIA)inspection system can be utilized to automate the inspection of manufactured assemblies and solve image pattern recognition problems. Details were provided pertaining to how the system is trained rather than programmed to inspect different items. This allows the system to be used by shop-floor personnel who are familiar with the product being inspected, rather than software engineers. The concepts of combining image analysis techniques with a neural network classification scheme were demonstrated by giving an example of how the GPIA system was implemented to inspect M549 artillery fuzes. Examples of possible future applications of the GPIA system were also provided.

Acknowledgements: The author would like to acknowledge the Aeronautics Department at the Johns Hopkins University Applied Physics Laboratory for their efforts in the development and implementation of the General Purpose Image Analysis system for inspecting M549 fuzes.

References:

1. Jacobs, Robert, A., "Increased Rates of Convergence Through Learning Rate Adaptation", Neural Networks Vol. 1
1988

2. Minai, Ali, A. and Williams, Ronald, D., "Acceleration of Back-Propagation through Learning Rate and Momentum Adaptation", IJCNN-90-WASH DC Proceedings, International Neural Network Society and Institute of Electrical and Electronics Engineers Inc., Lawrence Erlbaum Associates, 1990

3. Pao, Yoh-Han, Adaptive Pattern Recognition and Neural Networks, Addison-Wesley, 1989

4. Nelson, McCord, Marilyn and Illingworth, W.T.,
A Practical Guide to Neural Nets, Addison-Wesley, 1991

A REAL-TIME VISION SYSTEM TO MONITOR/ANALYZE THE CHANGES IN COMPOSITE SPECIMENS DURING MECHANICAL TESTING.

United States Army Research Laboratory
Materials Directorate
405 Arsenal St.,
Watertown, MA 02172-0001
Attn: AMSRL-MA-PB
James D. Kleinmeyer

Quest Integrated Inc.
21414 - 68th Avenue South
Kent, Washington 98032
Gary B. White

Abstract: Research has demonstrated the feasibility of using digital image processing techniques to determine dimensional changes and analyze qualitative and quantitative fracture processes in fiber-reinforced composite material test specimens that have experienced environmentally and mechanically induced stresses. The application of image analysis techniques during mechanical testing has been shown to facilitate test automation and to provide information essential to interpretation and quantification of mechanical test data. Since fracture is a dynamic process, real-time image acquisition and special image processing techniques are needed to determine specimen size and shape changes, identify where fracture is initiated, and describe how fracture develops in specimens during mechanical testing. Images must be classified and interpreted in terms of appropriate fracture models to facilitate the analysis of mechanical test data and provide a quantitative assessment of damage.

Key Words: Digital image processing; Vision system; Damage assessment: Composite Materials; Fracture; Mechanical testing; Environmental durability

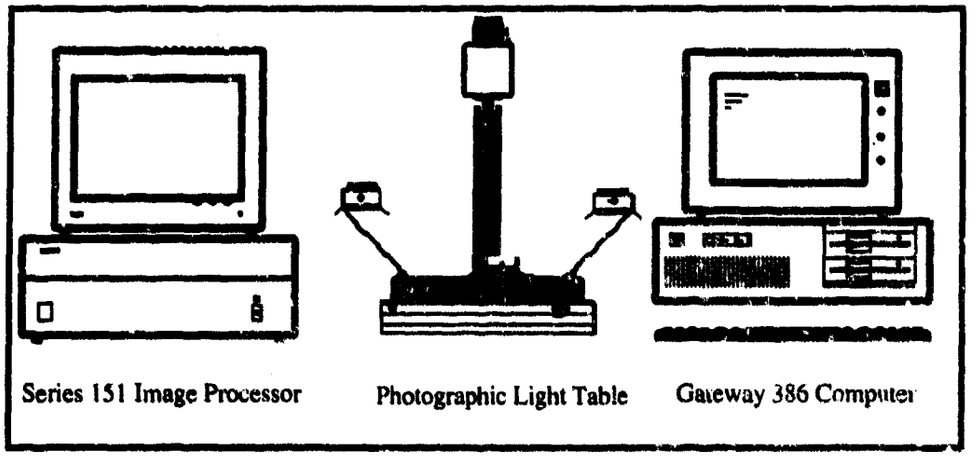
Introduction: This report describes the development and application of digital image analysis (DIA) techniques for analyzing the fracture behavior and environmental deterioration of test specimens prepared from fiber-reinforced polymer matrix composite materials. Digital image analysis techniques were employed to investigate the effects of moisture and temperature on the environmental aging of composite materials. A first generation digital image processing system developed at the Army Research Laboratory has been implemented to capture images which document the time sequence of damage accumulation and structural changes that occur during the environmental exposure and mechanical testing of polymer matrix composite materials. A more advanced system is currently being developed in collaboration with Quest Integrated, Inc. under an SBIR program sponsored by the Army Research Laboratory. The Quest system will have expanded digital image capturing and processing capabilities, as well as provisions for high speed, in-situ image acquisition and analysis of specimens undergoing deformation and fracture during mechanical testing. These techniques provide a nondestructive

method for evaluating environmental deterioration and offer an innovative and versatile capability for more accurately analyzing the failure behavior of complex materials and structures. Digital image analysis complements other techniques for characterizing the physical and mechanical properties needed to guide the application and design of advanced composite materials.

Experimental Approach:

Army Research Laboratory Testing And Image Processing System: An imaging system consisting of a CCD RS-170 video camera and a Series 151 Image Processor (Imaging Technology, Woburn, MA) under control of a Gateway 33 MHz 386 computer was used to analyze polymer matrix composite test specimens in two ways (see figure 1). The first was to record and characterize the effects of accelerated aging on the test specimens while the second was to identify and quantify the extent of damage as a result of mechanical testing. Special routines were developed using the image processor's C-interpreter for generating pixel histograms to help evaluate the digitized images.

Figure 1. Imaging System Configuration



The laminates used in this study were prepared from unidirectional S2-glass fiber, epoxy resin (SP250) prepreg materials (3M Corporation). The laminates consisted of 6 plies of prepreg in 4 different layup configurations. Defining lamina orientation as the angle (degrees) between the prepreg fiber direction and a reference axis (the X-axis, or flexural test specimen long axis, which is perpendicular to the load exerted during the 3-point bend test), the laminates are designated -

0°	fiber orientation	$[0_6]_T$
60°	" "	$[60_6]_T$
+- 45°	" "	$[+45_2-45_2+45_2]_T$
0/90°	" "	$[90_2\,0_2\,90_2]_T$

Flexural test specimens (nominally, 100 mm long x 19.3 mm wide x 1.45 mm thick) were carefully machined from the laminate materials and conditioned by drying at 60° under vacuum for 5 days. Test specimens were immersed in distilled water at 80° C for 144 hours to accelerate environmental aging. The percentage of moisture uptake was

determined by measuring the change in specimen weight before and after immersion. Upon cooling to room temperature, wet specimens were blotted with filter paper and immediately weighed. After immersion and weighing, the specimens were dried under vacuum at 60° C for at least 5 days or until all traces of moisture were removed. The conditioned specimens were then stored in a desiccator. Using the image processor, histogram analysis was carried out to quantify the effects of the accelerated aging process, by comparing the pixel count distributions of the aged specimens with that of the control specimens.

An automated Instron Universal Testing Instrument (Instron Corp., Canton, MA). was used for performing three-point loading, flexural tests. [1] The mechanical testing work cell includes the Instron (Model 4206) with hydraulic grippers and load cells selected for flexural testing of polymer matrix composite materials, a Zymate robot arm and controller (Zymark Corp., Hopkinton, MA), specimen racks, a bar code reader and a Gateway 386 computer. The robot is programmed to automatically transfer test specimens from the specimen racks to the test apparatus and then to remove specimens from the apparatus after the test is completed. All data acquisition, handling and reporting is fully automated. Stress was applied to fracture the specimens; however the strain was limited to prevent specimens from breaking apart. Recognizing that the location of the fracture damage region is well defined by the flexural test procedure, image acquisition and analysis was generally limited to a 19 mm region surrounding the area of probable damage. A histogram equalization function was used to enhance the contrast of the damaged region in each test specimen. Realizing that the image consists of 256 shades of gray, it was expected that most shades would fall away from the very dark (black) regions and the very light (white) regions. The damage regions, in fact, were only a slightly lighter shade of gray than the undamaged regions. The histogram equalization function equalizes the brightness of the entire image based upon the area of interest. It then uses LUTs (Look Up Tables) to force a normal distribution onto the area specified. This has the effect of greatly enhancing the contrast of the area of interest. A stretch frame function was used to finalize the area to be calculated. Similar to the histogram equalization function, the stretch frame function forces a distribution of gray levels onto a range of values. Specifically, the range of gray values from 150 to 175 was chosen for this application. All values below 150 were turned to zero (black), and any value above 175 was equated to 255 (white). The stretch frame made the contrast sharper, by turning the fractured region white and the unfractured region black. Thus the only part of the image remaining was the fracture region which appeared on the monitor as sharp white in a field of black. Finding the extent of damage in the fractured region was relatively simple. The remaining white pixels were counted by the image processor and the fractional or percentage of damage could then be calculated.

Quest Integrated In Situ Testing and Image Processing System: In collaboration with ARL, Quest is developing an in situ imaging system to acquire, interpret and archive images of composite materials during flexural testing. The system will include a digital camera subsystem capable of capturing images over a 10-cm by 2.5-cm test section with 40 micron resolution. This capability will be provided by a conventional digital camera configured as either a camera array or a single camera on a translation stage. The resulting stream of image data will be captured by a frame grabber equipped with a 128MB RAM buffer. Up to 80 images will be captured and analyzed during a typical 250 second test. The maximum acquisition rate in burst sampling mode will be at least 30 frames/sec.

Image data will be analyzed using a multi-processor plug-in board on a desktop computer. Historically, it has been impossible to obtain the power required for high performance image interpretation on a desktop machine. However, the emergence of high speed buses, multitasking operating systems and multi-processor DSP and RISC

based coprocessors has made it possible to consider the desktop computer as an alternative. These platforms represent the best price/performance value which is a prime consideration for commercial applications.
The emphasis of Quest's development program is on rapid image interpretation. Specifically, image interpretation which leads to an understanding of the failure dynamics in composite materials. The goals of the interpretation process are to:

- Determine specimen size and shape change.
- Locate the region(s) where mechanical failure is initiated.
- Define fracture type.
- Identify the sequence of damage accumulation and structural change.

Case Studies:
Army Research Laboratory Aging/Life Prediction Study: For this study insuring that image processing conditions and test procedures remain constant was essential for obtaining reproducible results. The digital image analysis of wet specimens was very sensitive to the time interval between sampling and imaging and did not show the large shift in gray level values observed for aged specimens that were conditioned (dried). Excellent reproducibility was obtained when specimens were conditioned and maintained in the dry state. Digital image analysis of conditioned specimens aged under identical conditions and stored in a desiccator generated similar pixel histograms. Fiber orientation, relative to the camera position, however, did not significantly affect the average gray level values of aged and unaged specimens. Specimens could be rotated 90° for image analysis with little effect on pixel count distribution under the prescribed test conditions. Image analysis was performed after the aged specimens were conditioned (dried at 60°C). Histogram analysis shows that the pixel count distribution of aged specimens was shifted to higher gray level values when compared to that of the control specimens. The calculated mean weighted average gray level for control and aged (144 h) specimens are listed in Table 1. The average gray level values for the control specimens are 160.5-161.9 compared with 188.7-191.0 for the aged specimens.

Table 1. Damage Area, Mean Gray Level and Percentage Moisture Uptake

Specimen	Mean Gray Control	Mean Gray Aged (144 Hrs.)	Moisture Uptake (weight-%)	% Damage (Unaged)
0°	160.5	191.0	2.82	90
60°	161.9	188.7	2.90	13
+45°	161.1	189.4	2.98	38
0/90°	161.9	188.7	2.70	20

As mentioned in the experimental section, the ASTM D790 method was performed. A support span of 25.4 mm (1 inch) was chosen because of the specimen thickness. The crosshead speed and temperature were 0.100 in/min and 25°C, respectively. Flexural modulus, flexural strength and the maximum load at yield were determined. The extent of fracture (% Damage) as indicated by digital image analysis was essentially identical for unaged and aged specimens with the same fiber orientation. It was also shown that the specimens have a distinctive fracture pattern based on fiber layup and test orientation. The fracture direction parallels the lamina fiber orientation direction and the fracture patterns are highly reproducible. The damage mode in the unaged specimens is a combination of delamination and fiber separation within plies. Pixel analysis of white/gray versus black regions shows that the percentage white/gray pixels (apparent percent damage) is greatest for the 0° specimen and least for the 60° specimen (Table 1).

Flexural test results for conditioned (dry) laminate specimens before and after accelerated aging are summarized in Table 2. In general, immersion of test specimens in water at 80°C for 144 hours caused a reduction of 10-30% in their flexural properties. As expected, fiber orientation has a significant effect on mechanical properties. However, the relative changes in flexural properties with accelerated aging were not significantly affected by fiber orientation. This study showed that accelerated aging of the SP250 epoxy/glass fiber test specimens in water at 80° C for 144 hours is comparable to environmental exposure results for the same material under tropical rain forest conditions for five to ten years.

Table 2. ASTM 0790 Flexural Test - SP250 Laminates

Specimen		Modulus (MPa)	Strength (MPa)	Max. Load (kN)
0°	control	32,800	1,440	1.69
	aged (144h)	29,700	1,046	1.36
60°	control	14,700	376	0.475
	aged (144h)	10,200	345	0.461
45°	control	7,320	299	0.339
	aged (144h)	5,290	240	0.303
0/90°	control	1,520	280	0.294
	aged (144h)	1,170	211	0.253

Table 1 shows that aged test specimens were all found to have similar percentages of moisture uptake. Thus, fiber orientation does not seem to affect moisture absorption. This is consistent with the fact that water absorption is a matrix dominated property and infers that, except for fiber orientation, resin-fiber interface characteristics of the specimens are essentially identical.

Quest Integrated Carbon-Fiber Composite Study: In preliminary experimental studies, Quest demonstrated that in situ images could be analyzed to locate and identify the formation of damage in carbon-fiber composite material. During these experiments, a ± 45° AS4/3501-6 carbon/epoxy test specimen was subjected to a flexural load which resulted in the formulation of micro cracks along the specimen's surface and edge. An image of this specimen is shown in Figure 2.

In this image, the matrix cracks appear as long, thin white lines. This image also includes a variety of other surface features including ink markings and scratches which are typical of unprepared test specimen surfaces. The objective of the image analysis operation was to separate the useful image data (matrix cracks) from the image noise (scratches and ink markings) and obtain the distribution of matrix cracks.

To accomplish this objective a morphology based algorithm was developed to process the image. This algorithm is based on the application of a "structure element". The structuring element is used to decompose the image into those features that match the expected size and shape of the matrix cracks. In many composite material testing applications it is possible to predict the general shape and orientation of localized damage phenomenon such as matrix cracking based on the ply configuration and test type. Consequently, morphology based image analysis provides a powerful tool to isolated damage modes. Figure 3 provides an example of the results obtained using the algorithm on the image from Figure 2.

Interpretation of Matrix Crack Damage in a Carbon Fiber Flexural Test Specimen.
Figure 2. Original Image ± 45 ° AS4/3501-6 Carbon/Epoxy Specimen

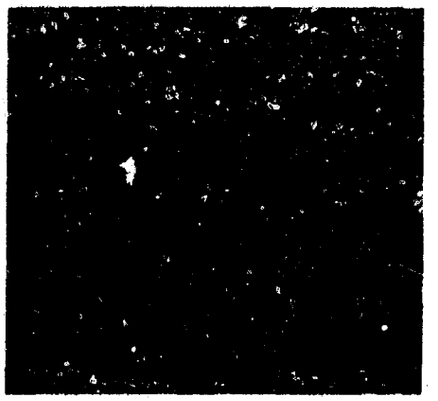
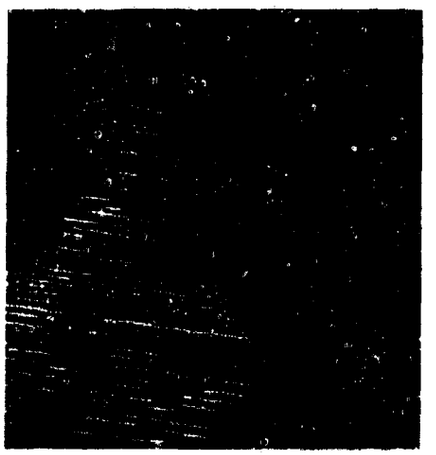


Figure 3. Microcrack Shapes Extracted From Figure 2 Using Morphology Based Image Processing Algorithm



The extension of this technique to streams of in situ image data would make it possible to obtain the following relationships:

- Crack length vs. location.
- Distribution of fractures per unit length.
- Distribution of crack separation distance.
- Distribution of fracture size and shape.
- Rate of fracture generation per applied load.
- Rate of fracture size and shape change per applied load and/or per location.
- Fiber rotation per applied load.

Conclusions and Recommendations: Digital image processing technology was employed to capture and analyze images of aged and unaged composite test specimens before and after fracture. The digitized images were evaluated using pixel histograms generated through routines developed on a C-interpreter. Pixel histograms of environmentally aged test specimens were broader and shifted to higher gray level values compared to histograms of the unaged specimens. Under the test procedures developed, DIA results were reproducible and fiber orientation did not significantly affect the average gray level values of either aged or unaged specimens. Laminate fiber orientation and aging conditions were found to have a significant effect on the observed fracture patterns and surface characteristics of the test specimens after fracture. Digital image processing is a promising technique for investigating the fracture and aging behavior of composite materials. Digital image processing can provide reliable and quantitative information relating to visual changes induced in composite test specimens due to accelerated environmental aging. Additional studies are required to improve analytical procedures for describing and quantifying fracture characteristics of flexural test specimens and to understand how and why environmental aging affects changes in pixel histograms. A detailed systematic study to determine possible correlation's between DIA and results obtained from mechanical, physical and chemical characterization of composite test specimens as a function of environmental aging conditions is needed.

Possible application of this technology to other types of materials should be investigated. Quest is currently seeking partnerships with groups involved with materials development including composites, plastics, thin films, rubbers, coatings, and adhesives. Quest is also seeking to expand the application of real-time image interpretation techniques to other in situ imaging modalities including radiography, 3D profiling and the various microscopy techniques

References:

1. ASTM D790 in Annual Book of ASTM Standards, 08.01, Philadelphia, PA: American Society for Testing and Materials, 1987.

Acknowledgment:

The Quest Integrated study was performed under a Phase I Small Business Innovation Research (SBIR) program for the U.S. Army Research Laboratory under Contract No. DAAL04-92-C-0012.

USE OF AURALLY PROCESSED NDE DATA TO IMPROVE THE PROBABILITY OF DETECTION

Glenn M. Light, Amos E. Holt,
Kent D. Polk, and William T. Clayton
Southwest Research Institute
P.O. Drawer 28510
San Antonio, Texas 78228-0510

Abstract: Southwest Research Institute has developed technology for converting electronic signals generated by conventional nondestructive evaluation (NDE) equipment (i.e., ultrasonic, eddy current, and acoustic emission) into audible information so that the inspector can make use of both the conventional NDE signals (which are often confusing) and the audible information to make a flaw/nonflaw decision. The objectives of the program were to: (1) identify useful sound properties and to determine how they could be applied to NDE signal representation; (2) identify and evaluate aural data presentation methods which could be applied to typical NDE electronic signal information; (3) initially develop aural techniques for specific inspection scenarios for ultrasonic, eddy current, and acoustic emission methods; and (4) present the technical results of the research to the NDE community to encourage further development and use of aural capabilities of signal interpretation. The aural technology has been applied to composite impact damage, composite delamination detection, and corrosion detection. In addition, the aural technology was used to detect and discriminate intergranular stress corrosion cracking. Examples of application of aural NDE technology are described. In addition, an audio-visual presentation was made during the paper presentation to demonstrate the technology to the audience.

Key Words: Aural; composite damage; corrosion detection; inspection; NDE; stress corrosion cracking; ultrasonic(s).

Introduction: Nondestructive evaluation (NDE) and inspection technologies as used today rely primarily on some form of visual presentation of the inspection information for human interpretation. The signals and images for some classes of inspection are easily interpreted (e.g., thickness measurement). However, a number of classes of inspection can provide signals that are tremendously difficult to interpret and often result in what is called a "false call" or a "missed defect call." A false call can result in expensive removal of a good part, and a missed defect call could result in a catastrophic failure. Both of these results are not acceptable. Therefore, some way is needed to aid the inspector in correctly perceiving the presence or absence of flaws with a greater reliability than is now available with only the interpretation of the visual signals presented on the NDE instrument or computer-generated image.

NDE technologies often are used to characterize materials for quality. Usually they rely on visual interpretation of electronically displayed signals generated from conventional NDE instrumentation [e.g., ultrasonic testing (UT) A-scan trace on an oscilloscope or a Lissajous figure on an eddy current testing (ET) instrument]. In some inspections, such as ultrasonic wall-thickness measurement, the signals can be easily interpreted. In many other inspections, however, the signals observed on the oscilloscope have low signal-to-noise and are difficult to visually interpret because of inherent noise.

One method for improving the accuracy of data interpretation would be to utilize multiple human senses simultaneously to characterize the material (e.g., use both eyes and ears for data evaluation). In the past, cursory attempts have been made at using this concept. For example, the electronic signals received by an NDE instrument have been used to generate alarms if their amplitude exceeded a certain threshold. This was done to alert the inspector with an aural as well as a visual signal. The technology, however, was only an "off/on" process and did not make use of any information inherently contained in the NDE signal except amplitude (UT/ET) or phase (ET).

If it is assumed that the NDE signal contains complete information about the condition of the material under inspection and that the information could be converted into audio (aural) signals (while keeping the content of the NDE signal intact), then both the ear and eye can be used to enhance the accuracy of data interpretation.

Auralization and Aural Conversion of NDE Signals: Since 1985, Southwest Research Institute (SwRI) has been conducting internally funded research to develop the capability to convert NDE signals from conventional NDE instrumentation into aural signals. The intent has been to improve the quality and accuracy of NDE inspections and material characterization. This paper discusses work done to convert conventional NDE inspection signals into aural signals for improved defect detection.

Auditory studies show that principles used in listening perception are analogous to visual perception principles. These, in turn, are parallel to classic Gestalt perception concepts used in psychology (1-4). The Gestalt concepts (1) interpret the perception processes by the relationships among all human senses and (2) allow each of the human senses its own unique way of interacting with the surrounding environment. Applying the Gestalt concepts to the aural perception process has yielded some important conclusions including:

(1) All attributes of sound such as pitch, timbre, loudness, and frequency are not independent; that is, each aspect of sound production influences another.

(2) Listener training and a priori knowledge of aural sources play an important role in successful identification of aural information.

(3) The human auditory system is very sensitive to changes in sensory input and can detect and discriminate these changes from background noise. Evidence suggests that the auditory mechanisms that allow discrimination and detection of these changes are fundamental to auditory operations and even excel with certain types of noise stimuli.

These and other concepts were used to determine how to best use sound to represent NDE signal information from different inspection methods. A more thorough discussion of the Gestalt concepts and how they have been applied to auralization of NDE data is given by K. D. Polk, et al. (5).

Conversion of NDE Signals: Developing algorithms to convert high-frequency electronic signals from NDE instruments into meaningful aural signals is the most difficult challenge. The task requires, first, an appropriate data acquisition and analysis system (DAAS). For example, the prototype DAAS designed for this work uses a conventional UT instrument appropriately interfaced to the computer decision system, as illustrated in Figure 1.

Several different ultrasonic instruments have been used including the MetroTek modular system that allows multiple receivers and pulsers to be operated simultaneously in a

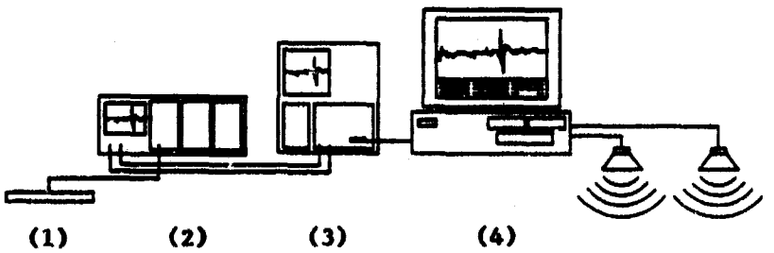


Figure 1. Aural NDE system configuration consisting of (1) sample/ transducer, (2) conventional ultrasonic equipment, (3) signal digitizer, and (4) Amiga 2500 computer

controlled fashion. The heart of the system is the signal digitizer. One of the most important issues that had to be addressed for development of the aural ultrasonic process was the definition of a procedure for time dilating the ultrasonic data. The signal digitizer system can digitize ultrasonic waveforms at rates from 10 Hz to 200 MHz.

The selected computer is a realtime, icon-driven, modular DAAS decision tool (software/hardware package). Its modular construction provides graphical selection and control of individual processing modules including gating, filtering, scaling, data source, and audio playback, depending on the specific design of the NDE process under study. Current programming modules offer selection of up to 16 channels of waveform data with the capability of changing the number of channels and some channel content dynamically. The inspector also can use one or two audio output channels for either mono or stereo (binaural) presentation of the aural information.

Auralization of Ultrasonic Signals: One basic form of aural conversion involves time dilation. This process increases the time interval between the digitized samples of UT information and a new waveform with frequencies in the audible range. For example, if a waveform is digitized at a 10-MHz sampling rate, then it is sampled every 0.1 microsecond. By taking each of these sample points and increasing the time between them from 0.1 microsecond to 0.1 millisecond, the waveform appears to have a frequency 1000 times lower than the actual waveform before time dilation. Using the user-interface display, the operator can hear this time-dilated signal information through the two audio channels in mono or stereo. When presented in two separate channels, the operator can take advantage of binaural hearing capabilities.

To develop the aural UT system, a trained UT operator is asked to position a transducer on a test sample and locate a region with a signal of interest. The waveform (A-scan, amplitude-time) is acquired and digitized. The actual digitized waveform is displayed on the computer screen (which is set up to look like an oscilloscope). The time-dilated signal derived from the ultrasonic signal is sent through an audio receiver and can be played in mono or stereo. To form an initial set of data to begin algorithm development, signals from other areas, both good and defective, are then acquired and digitized.

The modular elements of the system software are key to the developing technology because they permit the operator to produce a variety of waveform processing techniques. Using realtime interactive computer feedback, the operator can experiment with different features to enhance aural perception of the aural signal information. The acquired UT signal information is played back audibly in several different analysis forms. The operator then can begin to discern differences aurally in the data obtained from good and defective regions.

The procedure used to develop the aural signals is very important to the final aural output. The signal of interest is usually gated and digitized at a rate that faithfully reproduces the waveform. The aural output is a combination of the time-dilated signal and the window (as shown in Figure 2).

Most of the work done during this project used a 50-MHz digitization rate. Then the data were time dilated at a 2-kHz rate. The data were usually windowed so that the ultrasonic signal of interest filled approximately 40 percent of the window. The aural signal heard by the operator was a composite of the time-dilated ultrasonic signal and the window width. For example, if the ultrasonic signal had a frequency of approximately 2.3 MHz and was time dilated at a rate of 2 kHz, then the time-dilated frequency was calculated in the following way:

$$2.3 \text{ MHz} \times (1/2000) = 1150 \text{ Hz}.$$

The aural signals obtained from various types of NDE ultrasonic signals were evaluated by several different observers.

From empirically obtained information, it was known that the aural signal generated by the DAAS was a composite of the actual time-dilated ultrasonic signal and the window width. For example, for a certain window width, different time-dilated signals generated different sounds. Also, if different window widths were used for the same time-dilated ultrasonic signal, different sounds were generated.

The approach taken to better understand the generation of the aural signals was to perform a frequency analysis of the aural signal. An HP 3582A spectrum analyzer was used to obtain a frequency plot of the data. The frequency spectrum consisted of discrete peaks which themselves formed an envelope having a peak frequency. An example of this envelope is illustrated in Figure 3, as obtained from an IGSCC.

The data-collection process consisted of digitizing the ultrasonic data at 50 MHz (equivalent to 20 nanoseconds per point) using a window that was 114 points long (this

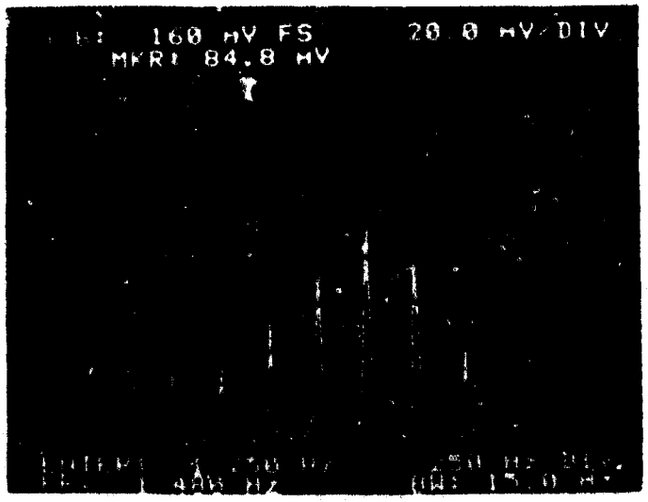


Figure 2. Plot showing the frequency spectrum of aural data. The discrete peaks (harmonics) are a result of the window, while the peak frequency of the overall envelope is due to the actual UT data.

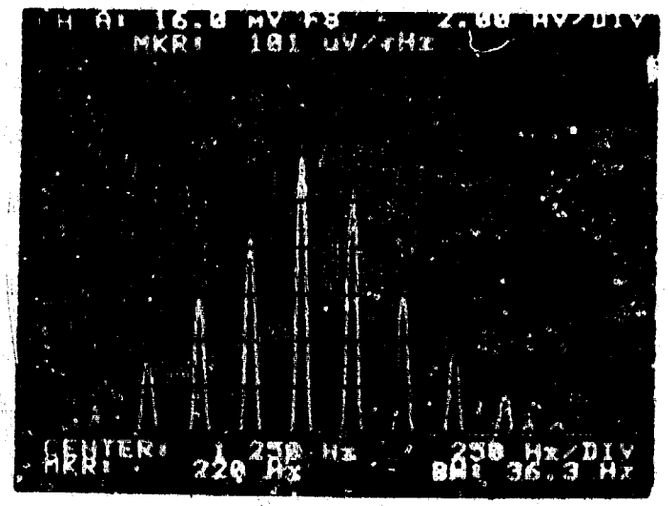


Figure 3. Frequency spectrum obtained from the aural signal generated from an IGSCC ultrasonic signal digitized at 50 MHz and windowed with a 2.3-microsecond window

window length was arbitrarily chosen by approximating what was believed to be a good window length). If the peak frequency for the time-dilated (auralized) ultrasonic data were calculated from the peak ultrasonic frequency, then the peak aural frequency should be 1150 Hz. If the window frequency is calculated in a similar way, its frequency is approximately 219 Hz. This can be shown in the following way. If the window is 114 sample points wide, then the time-dilated window time is given by

$$114 \text{ points} \times 20 \text{ nanoseconds} \times 2000 \text{ Hz} = 4.5 \text{ ms};$$

and the audible frequency caused by the window is given by the inverse of the time or

$$1/(4.5 \text{ ms}) = 219 \text{ Hz}.$$

An example of the composite frequency is shown in Figure 4, which is a frequency spectrum obtained from an ultrasonic signal from a flat-bottom hole in a carbon steel plate windowed with a 114-point window and a time-dilation rate of 2000 Hz.

Discussion of Results: The following discussion concerns the use of aural NDE technology to detect corrosion on metallic plates, debonds and impact damage in composite panels, and intergranular stress corrosion cracking (IGSCC) in stainless steel piping. The preliminary results obtained for each experiment are discussed in this section.

Hidden Corrosion: Detection of hidden corrosion in aircraft structures is difficult using conventional techniques. Figure 5 shows an ultrasonic A-scan of two regions from an aluminum wingskin mockup, with and without corrosion. The UT signals are very similar. By using multiple reflections from the internal surface of the material and converting that signal into aural signals by time dilation, the operator can easily hear the difference between good and corroded material. The UT waveform used for the time dilation is shown in Figure 6. When both the A-scan and aural UT data are used simultaneously, the presence of the hidden corrosion is easily detected.

Characterization of Composite Material: In the aerospace industry, aircraft aluminum structure is often replaced by composite material because of the latter's stiffness

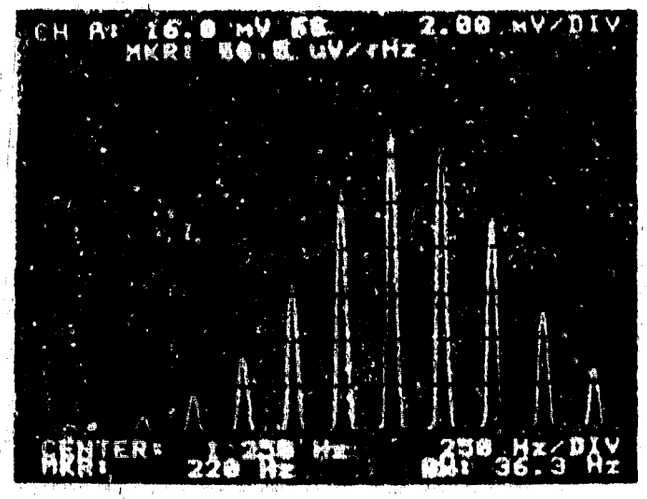
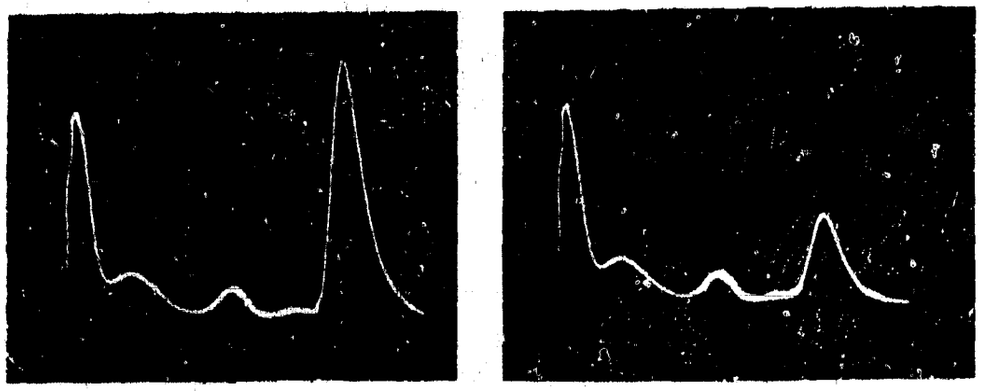


Figure 4. Frequency spectrum obtained from an ultrasonic signal from a flat-bottom hole in a carbon steel plate windowed with a 114-point window and a time-dilation rate of 2000 Hz

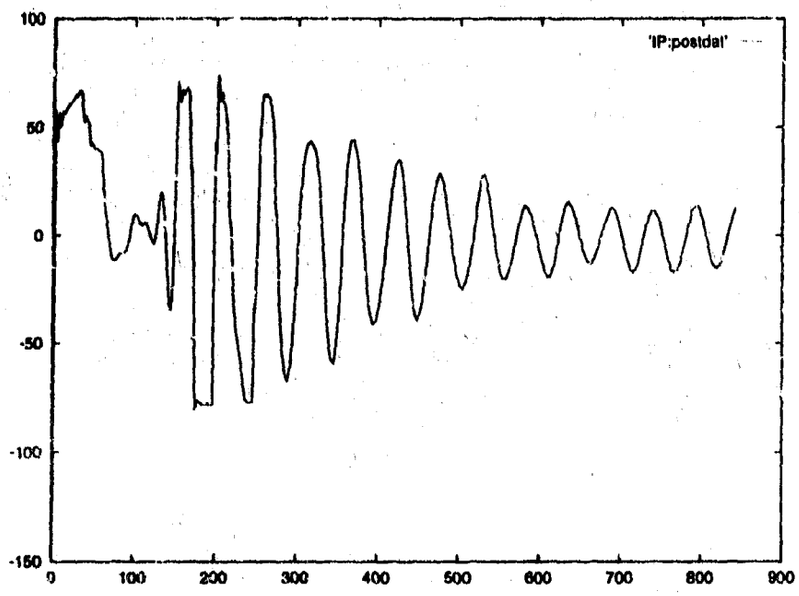


(1) (2)

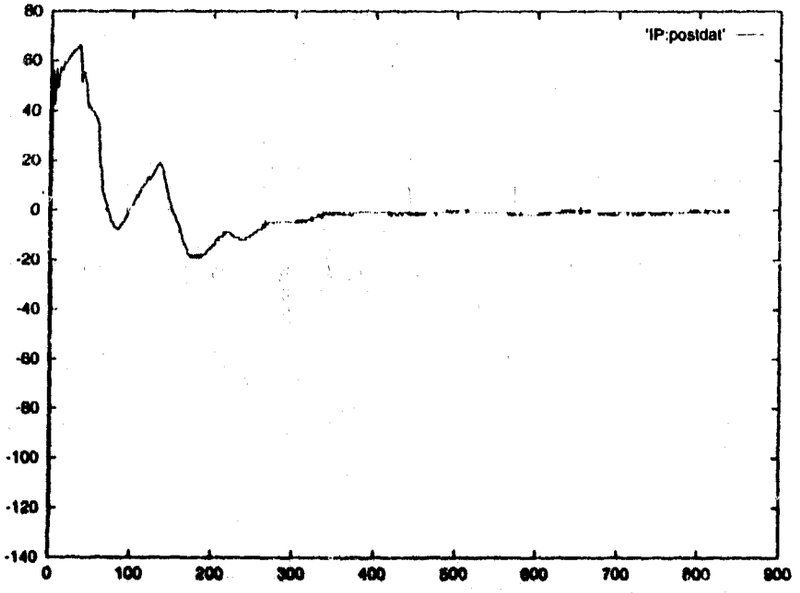
Figure 5. Ultrasonic A-scans obtained from a wingskin mockup which has (1) no corrosion and (2) corrosion

and light weight. The composite material, adhesively attached to the aluminum parts, must be well bonded for aircraft safety. The quality of the bond, therefore, has to be verified throughout service life. Composite material also can lose much of its strength if its layers delaminate or the matrix cracks. These two problems often are caused by an impact of the aircraft with objects such as runway debris, hail, or birds.

To inspect composites, data from conventional UT using pulse-echo contact techniques can be difficult to interpret. When the UT signal was time dilated to audible signals, the difference between good and defective material was easily heard. Again, by using both the UT and aural data, the quality of the inspection was increased.



(1)



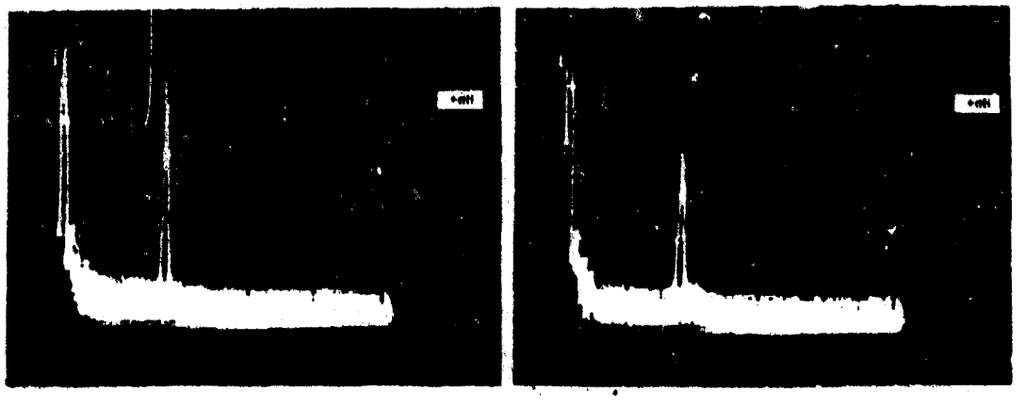
(2)

Figure 6. Ultrasonic data that were time dilated into aural data and used to aurally determine (1) no corrosion and (2) corrosion

Detection and Discrimination of IGSCC: Detection and discrimination of intergranular stress corrosion cracks (IGSCC) in stainless steel recirculation piping in boiling water reactors are of major importance to the nuclear power-plant industry. IGSCC is particularly difficult to discriminate correctly from root geometry because IGSCC develops in the heat-affected zone of the weld at the root region. Examples of root geometry and IGSCC signals obtained using a 1.5-MHz, 45-degree, shear-wave transducer are shown in Figure 7. The Electric Power Research Institute NDE Center has been working with ultrasonic inspectors that have had many years' experience in training and testing their abilities to detect and discriminate IGSCC successfully. The first-time failure rate on this test by these highly trained inspectors is as high as 60 percent (6). This attests to how difficult evaluating stainless steel material for IGSCC is.

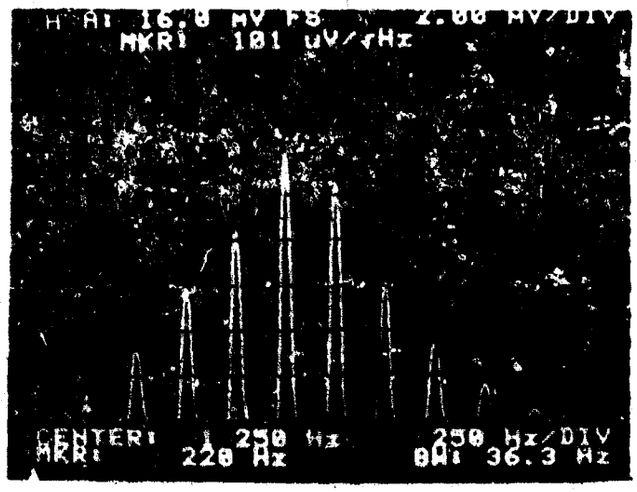
One goal of the SwRI aural NDE development is to improve detection and discrimination of IGSCC. Only preliminary work has been conducted so far, but results seem promising. By properly setting up the auralization system, both mono and stereo data can be obtained. Two channels of the four-channel aural system have been used to present the data to the inspector, with one channel sent to each ear, respectively. To test this concept for IGSCC, a special transducer fixture was developed that contained three transducers. One transducer provided for 1.5-MHz, 45-degree, shear-wave pulse-echo data to be obtained normal to the weld. The other two provided for 1.5-MHz, 45-degree, shear-wave received-only data to be obtained at angles between approximately 20 and 45 degrees of normal to the weld. These data were time dilated and separated into right and left channels on the audio output presented to the inspector. Using this approach, the inspector could discern the IGSCC. Figure 8 shows the frequency content of a geometrical reflector compared to an IGSCC.

Conclusions and Future Work: SwRI has developed an inspection method to provide simultaneously to the eyes and ears of an inspector material-characterization information obtained through UT techniques. This method converts the actual ultrasonic signal into an aural signal so that it can be presented to the ear of the inspector in conjunction with the visual signal. The method, called Aural NDE, has been successfully demonstrated for ultrasonics, eddy current, and acoustic emission (7). This paper illustrated uses of the aural NDE method for ultrasonic detection and discrimination of hidden

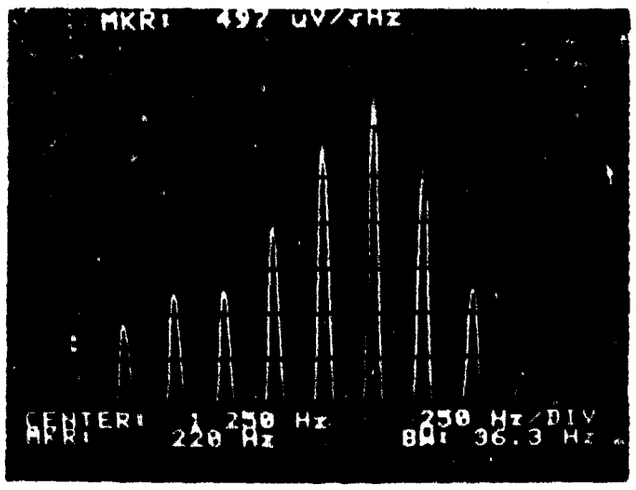


(1) (2)

Figure 7. Ultrasonic A-scan data from (1) root geometry and (2) IGSCC obtained using a 1.5-MHz, 45-degree, shear-wave transducer



(1)



(2)

Figure 8. Frequency content of (1) geometry and (2) actual IGSCC

corrosion in aerospace material, delaminations and impact damage in composites, and IGSCC in stainless steel piping. Aural NDE should play a major role in significantly improving the quality of NDE inspections performed in a wide variety of industries.

SwRI plans to transfer this technology to the power and aerospace industries for ultrasonics, eddy current, and acoustic emission applications. Successful transfer will involve development of appropriate data acquisition procedures and aural conversion techniques specific to each class of inspection.

References:

1. Dooling, R. J., and S. H. Hulse, eds., "The Comparative Psychology of Audition: Perceiving Complex Sounds," L. Erlbaum Associates, Hillsdale, New Jersey (1989).

2. Handel, S., "Listening: An Introduction to the Perception of Auditory Events," MIT Press, Cambridge, Massachusetts (1989).

3. Warren, R. M., "Auditory Perception: A New Synthesis," Pergamon Press, New York (1982).

4. Moore, B. C. J., "Introduction to the Psychology of Hearing," University Park Press, Baltimore, Maryland (1977).

5. Cervantes, R. A., K. D. Polk, and A. E. Holt, "Aural Perception Techniques for NDE Signal Information," Southwest Research Institute (SwRI) Final Report No. 17-9625, SwRI, San Antonio, Texas (1992).

6. Stephens, H., Training activities at the NDE Center, in "Nondestructive Evaluation Research Progress in 1992: Proceedings of the Thirteenth Annual EPRI NDE Information Meeting," NP-7560-S, Electric Power Research Institute, Palo Alto, California (1992)

7. Polk, K. D., G. M. Light, R. H. Peterson, and J. L. Jackson, "Investigation of Audible Recognition Benefits for Ultrasonic Echo Detection and Characterization," SwRI Final Report Nos. 17-1400-609 and 17-3131-606 (extension), SwRI, San Antonio, Texas (1990).

FRACTURE TOUGHNESS MEASUREMENTS OF TWO SPECIMEN GEOMETRIES CONSIDERING STABILITY

Kyu Cho[1], Isa Bar-On[1], and Francis I. Baratta[1]
[1]Army Research Laboratory, Materials Directorate, Watertown, MA 02172
[1]Worcester Polytechnic Institute, Worcester, MA 01609

Abstract: It has been shown that a lack of stable crack extension can influence the critical stress intensity factor (SIF) measured for brittle materials. Previously performed stability analyses of two specimen geometries were used to design experiments, which would make it possible to observe the transition from unstable to stable fracture as a function of the specimen compliance. This transition was observed to be in agreement with the predictions. The lack of stability gave higher critical SIF values for the material with the higher fracture toughness, while this difference was lost in the experimental scatter for the other materials.

Key Words: Fracture toughness; crack stability; bend bar geometry; compliance; alumina; silicon nitride

Introduction: Fracture toughness of metallic materials is typically determined at the start of quasi-static crack extension [1]. The quasi-static condition is achieved during fracture by the low rate of increase in stress intensity, 0.55 to 2.75 MPa$\sqrt{m}$/sec [2]. In addition, the quasi-static extension is aided or even facilitated by local crack tip plasticity in these materials.

For ceramic materials, however, quasi-static crack extension is much more difficult to obtain. These inherently brittle materials do not benefit from the crack tip plastic zone. Also, fracture toughness tests in ceramics are typically performed at higher loading rates, in order to avoid the effect of potential environmental interactions with the grain boundary phase. At low loading rates, this interaction can lead to artificially low fracture toughness measurements [3,4].

The lack of quasi-static or stable crack extension has been recognized to affect fracture testing even for relatively ductile materials [5]. Unstable fracture as encountered, for example, in notched specimen tests will frequently lead to artificially high fracture toughness values. Similarly, unstable fracture tends to occur in very stiff specimens such as ceramics, when the test set-up is not sufficiently stiff relative to the specimen. For those materials, crack stability is often extremely difficult to obtain even for specimens containing naturally sharp cracks. Underwood et al. [6] and Baratta et al. [7] have shown that this lack of stability can lead to inflated critical stress intensity factor (SIF) measurements for liquid-phase sintered tungsten and polymethylmethacrylate (PMMA) materials, respectively.

Baratta and Dunlay [7] analyzed the crack stability of the rectangular bend bar (RTBB) in three point and four point loading. Their analysis and test results for quasi-brittle specimens showed that the three point loading geometry had greater stability potential than the four point loading geometry. Underwood et al. [6] analyzed the crack stability of a round bend bar (RBB). Their analysis and test results for tungsten specimens showed that the round bend bar specimen to be significantly more stable than the RTBB loading configuration. This would make the RBB specimen geometry an attractive candidate for fracture toughness testing of ceramic materials, as it might promote stability in this case.

In ceramic materials, however, it is difficult to create sharp pre-cracks in a reproducible and controllable manner. Recently a pre-cracking method has been developed [8] and systematically studied [9] by which cracks of varying lengths can be introduced in RTBB specimens. For this method, a Vickers micro-indentation is placed on one of the specimen's longitudinal surfaces. A through thickness straight crack is then created by loading the specimen in compression between a double anvil fixture. This bridge indentation method [10] can be applied to the RBB only if the cross section of the bar is modified to create two parallel flat surfaces as shown in Fig. 1. Cho et al. [11] modified the RBB specimen geometry to allow pre-cracking while obtaining potentially greater crack stability. This new specimen geometry has been called the modified round bend bar (MRBB). Crack stability of the MRBB has been analyzed and compared to that of the RBB and RTBB [12].

As previously mentioned, the recent works [6,7,13] have shown that unstable crack extension can result in an apparent increase in critical SIFs compared to those measured during stable crack growth of quasi-brittle polymer and brittle metallic materials. For both materials, the transition from stable to unstable fracture behavior was predicted based on stability analyses. The effect of crack stability on measured fracture toughness of ceramics, however, is unclear. Therefore, the objective of this work is the determination of the fracture toughness of ceramic materials while focussing on crack stability. The analytical stability prediction is compared to the fracture behavior observed for RTBB and MRBB specimens of alumina and silicon nitride. The experimental results were in excellent agreement with the analytical prediction. Critical SIF measurements suggested that it would be necessary to select a material with a relatively high fracture toughness and a specimen with larger dimensions to observe a noticeable difference in critical SIFs due to crack instability. Also, it was predicted that stability in the MRBB can be obtained for shorter crack lengths than for the RTBB.

Stability Summary: The stability equation for bend bars has been derived previously [6,7,12] and is based on the requirement that at fracture:

$$dG/dA \leq dG_{CR}/dA \qquad (1)$$

where G is the elastic strain release rate, A is the crack face area, and G_{CR} is the critical elastic strain release rate. For materials with a flat crack growth resistance curve, i.e., $dG/dA = 0$, then Eq. (1) becomes:

$$dG/dA \leq 0 \qquad (2)$$

Bluhm [14] has shown that stability for beams can be obtained for displacement control (i.e., fixed grip) conditions only. The stability equation for this condition is:

$$dG/dA = 1/2\{P^2(d^2\lambda_T/dA^2)+2P(dP/dA)(d\lambda_T/dA)\} \leq 0$$
$$= G\{d^2\lambda_T/dA^2-2/\lambda_T(d\lambda_T/dA)^2\}d\lambda_T/dA \leq 0 \qquad (3)$$

where λ_T is the total compliance of the system, consisting of the specimen compliance, λ_s, and that of the machine including ancillary fixture, λ_M, and P is the applied load.

The nondimensional load-line compliance for the cracked RTBB and MRBB are taken from the literature [12,15]. For the RTBB, the nondimensional plain strain compliance, $\bar{\lambda}_{RTBB} = (\delta EB/P)_{RTBB}$, is [15]:

$$\bar{\lambda}_{RTBB} = 2(S/2W)^2\{S/2W+[2.85/(S/2W)-0.42/(S/2W)^2]/4+9(1+\nu^2)\int \alpha f^2(\alpha)d\alpha \qquad (4)$$

where δ, E, P, B, S, W, ν are the load-line deflection, the elastic modulus, the applied load, the specimen thickness, the span length, the width of the specimen, Poisson's ratio, and the dimensionless crack length, respectively, and $f(\alpha)$ [13,16,17] is the dimensionless stress intensity factor expression for the RTBB. For the MRBB, the nondimensional plain strain compliance, $\bar{\lambda}_{MRBB} = (\delta ED/P)_{MRBB}$ is [12]:

$$\bar{\lambda}_{MRBB} = (S/W')^3\{3.4862\times10^{-1}+8.0862\times10^{-1}(1+\nu)/(S/W')^2\}$$
$$+2.2992(S/W')^2/W'^2\int [\alpha'f^2(\alpha')g(\alpha')/\{(1-\alpha')^3(\Omega+1-\alpha')\}d\alpha' \qquad (5)$$

where D, W', and α' are the diameter, the width, and the dimensionless crack length of the MRBB, respectively. $f(\alpha')$ [11] is the dimensionless stress intensity factor for the MRBB, $g(\alpha')$

= dA/dα' [12], and Ω [11] are geometry constants for the MRBB.

The stability solution for the RTBB and MRBB in the three point bending can be derived using the nondimensional machine compliance, $\overline{\lambda}_M = \delta_M EB/P$ for the RTBB and $\overline{\lambda}_M = \delta_M ED/P$ for the MRBB, respectively. The results of the stability calculation relevant for the experiments are summarized in Fig. 2 in terms of threshold crack length. The threshold crack length is the minimum nondimensional crack length for which stable crack growth cab be predicted. Nondimensional machine compliance including test fixture and span to width ratios used in the experimental tests were $\overline{\lambda}_M \approx 60$ (see experimental procedure for detailed description), and S/W = 5 for the RTBB and S/W' = 6 and 6.25 for the MRBB, respectively.

Experimental procedure: The RTBB and MRBB geometries were used to measure fracture toughness of an alumina (AD99[†]) and silicon nitride (NC132[‡] and HS130[‡]). Table I summarizes the mechanical and physical properties of the materials. The alumina was a Coors grade AD99 and was manufactured in the early to mid 1970's. AD99 is a nominally 99% Al_2O_3 with SiO_2 as a sintering additive. The average grain size is 12 μm (range of 2 to 50 μm) and the density is 3.83 g/cm³ as reported by the manufacturer. A comparison of the theoretical density for 99% Al_2O_3 with the 3.83 g/cm³ reported for AD99 suggests that a considerable amount of porosity can be expected. AD99 was primarily used as refractory thermocouples and electrical insulators. It had been extruded and then fabricated into a rod of 6.35 mm diameter. The rods were cut into 50 mm long MRBB specimens with a D/W' ratio of 1.1346 as shown in Fig. 1.

Table I. Mechanical and Physical Properties of AD99, NC132, and HS130

	Grain Size μm (range)	density g/cm³ (range)	Elastic Modulus GPa	Poisson's Ratio	Modulus of Rupture MPa
AD99[†]	12 (2 - 50)[†]	3.83[a]	350[†]	0.24[a]	262[a]
NC132[‡]	Maximum 3[‡]	3.25[‡]	320[‡]	0.27[‡]	825[‡]
HS130[‡]	2.5 (1 - 10)[b,c,d]	13.17 - 3.21[b]	300[b,c]	0.26 - 0.27[b]	516 - 681[b,c,e]

[†]Coors Ceramics, Golden, CO. [‡]Norton Co., Worcester, MA. [†]Coors product literature. [‡]Norton product literature. [a]Ref. 18. [b]Ref. 19. [c]Ref. 20. [d]Ref. 21. [e]Ref. 22.

One of the silicon nitride was a Norton grade HS130 (later developed into NC132) and was manufactured in the early to mid 1970's. HS130 is a nominally 98% pure Si_3N_4 utilizing MgO as a sintering agent. The major crystalline phase is β-Si_3N_4 and traces of α-Si_3N_4 and Si_2N_2O were identified by x-ray diffraction [19]. Three types of grains had been reported in the literature [19-21]: equiaxed grains ranging from 1 to 4 μm in size; equiaxed grains of the order of 8 to 10 μm; and elongated grains of 2 x 10 μm. A density of 3.17 to 3.21 g/cm³ has been reported [19] while that of NC132 is reported by the manufacturer as 3.25 g/cm³. This suggests that HS130 contains some porosity. The only measurement of this, however, consists of occasional layered shading of the x-ray radiographs [19,22]. The comparatively wide range in density values is due to tungsten contaminations in the form of either WC or WSi_2. The mechanical properties are anisotropic due to the hot-pressing process. The modulus of rupture ranges from 516 to 681 MPa, the elastic modulus is 3.0 x 10⁵ MPa, and Poisson's ratio is 0.26 - 0.27 [19,20,22]. This is primarily a high strength high temperature silicon nitride. The material was machined into5.52 mm diameter rods from 6 x 6 x 1½ inch billets. The rods were cut into 50 mm long MRBB

[†]Coors Ceramics Co., Golden, CO.
[‡]Norton Co., Worcester, MA

specimens with a D/W' ratio of 1.1346.

The silicon nitride used for the RTBB geometry fracture toughness tests was a Norton grade NC132. NC132 is a 100% theoretically dense hot pressed silicon nitride also using MgO as a sintering agent. The α-Si_3N_4/β-Si_3N_4 phase composition ratio is 20/80 [23]. The material has a maximum grain size of 3 μm. A density of 3.25 g/cm^3, modulus of rupture of 825 $\pm$ 137 MPa, Vickers hardness of 16 GPa, elastic modulus of 320 GPa, and Poisson's ratio of 0.27 have been reported by the manufacturer. The material was cut into 6 x 8 x 45 mm RTBB specimens from 6 x 6 x 1 inch billets with the hot pressing direction perpendicular both to the long direction of the specimen and to the crack plane.

The stability calculations were used as a guideline in designing a test system which would be stiff enough so that stable and unstable crack growth could be realized. The testing system consists of an Instron 250 kN servo-hydraulic load frame with a 25 kN load cell. Frame stiffness of 585 kN/mm, 250 kN capacity load cell stiffness of 2560 kN/mm, and 25 kN capacity load cell stiffness of 1020 kN/mm were all specified by the manufacturer. The resulting stiffness of the frame the two load cells can be calculated as 322 kN/mm. Initially, a conventional fully articulating three point bend test fixtures, similar to one that specified in MIL-STD 1942A [24], were used. With this set-up, however, stable crack growth was unattainable even for very long pre-cracks. The stability solution, which provided guidance to the experiments, indicated that the fixture had to be stiffened for stability to be obtained in this system. Thus, the fixture was replaced by a stiffer semi-articulating one. The compliance of the machine, load cells, and test fixture was determined experimentally using an uncracked silicon nitride bend bar. The measured compliance of this test set-up was 3.07×10^{-8} m/N, corresponding to a stiffness of 32.57 kN/mm. This measured compliance corresponds to a nondimensional machine compliance of $\overline{\lambda}_M = 58.94$.

The specimens were indented with a Vickers diamond indenter using loads ranging from 69 to 490 N. Indentation was performed on a screw driven Instron 5 kN load capacity testing machine with a 0.5 kN load cell at a displacement rate of 0.1 mm/min. The indenter was immediately released after the set loads had been reached. These specimens were pre-cracked to dimensionless crack lengths of 0.2 - 0.8 a/W using the bridge indentation technique [10]. Pre-cracking was performed on a servo-hydraulic Instron 250 kN capacity testing machine using a 25 kN load cell at a loading rate of 1 kN/sec. The cracks were marked with dye penetrant ink which was dried in a furnace after pre-cracking. Crack length was measured after the test on the fracture surface at three equidistant points along the crack front [2].

The fracture toughness tests were performed in three point bending with three different span to width ratios: the span to width ratio for the RTBB was 5; for the MRBB, 6.25 for the alumina and 6 for the silicon nitride specimens. The tests were performed on a servo-hydraulic Instron 250 kN capacity testing machine using a 25 kN load cell. The scale was set to 1/64 of full range. The fracture toughness tests were performed at a displacement rate of 0.1 mm/min for all materials. Two additional displacement rates of 0.25 and 0.5 mm/min were used for the alumina, since alumina tends to be sensitive to subcritical crack growth in air (i.e., static fatigue) [3]. Load displacement records were taken for all tests.

Results: The load displacement records showed three distinctly different traces. For initial critical crack lengths much less than the threshold value, the load displacement record was linear to the point of fracture as shown in Fig. 3a. Fracture occurred instantaneously across the entire cross section as indicated by the load drop to zero. For initial critical crack lengths close to the critical crack length, some stable crack extension occurred as documented by the load displacement curves, which were non-linear near the maximum load (see Fig. 3b). Specimens with very long cracks exhibited more pronounced non-linear load displacement records (see Fig. 3c)

For the silicon nitride (HS130) MRBB specimens, an average fracture toughness value of 2.84 ± 0.13 MPa√m was measured. This value agrees with reported literature values [23,25]. Fig. 4 shows the critical SIFs subdivided into stable and unstable results based on the appearance of the load displacement record. It can be seen that the analytically predicted threshold crack length, $(a'/W')_o$ = 0.59, agrees with the experimentally observed transition from unstable to stable behavior.

For the alumina (AD99) MRBB specimens, an average fracture toughness value of 2.27 ± 0.10 MPa√m was measured. Because of the lack of fracture toughness data for this alumina, three 3 x 4 x 40 mm rectangular beam specimens were fabricated from the supplied round bar and critical SIFs were measured by the SEPB method [8]. The test was performed on a screw driven Instron testing system with 5 kN load capacity using a 0.5 kN load cell. A conventional fully articulating three point bend fixture of low stiffness with a span length of 16 mm was used. The measured average fracture toughness was 2.44 ± 0.11 MPa√m which agrees very well with the fracture toughness measured on the MRBB specimen. Fig. 5 shows the critical SIF values subdivided into unstable and stable results. Again the analytically predicted threshold crack length, $(a'/W')_o$ = 0.60, divides the results into stable and unstable tests, with the exception of two data points which are stable for a somewhat shorter crack length than predicted.

Alumina is typically susceptible to subcritical crack growth in air under static loading [3]. This effect would make the fracture toughness results sensitive to the displacement rate at which the test was performed, but would also promote stable crack extension for a different reason. Fig. 6 summarizes the results of critical SIFs for three different displacement rates. While there appears to be no discernible effect on the mean fracture toughness value for the different loading rates it is note worthy that the two specimens which broke stably below the threshold crack length were tested at the lowest displacement rate. This would allow for some environmentally assisted stable crack growth to occur prior to fracture for an a/W' which is less than the predicted threshold crack length, $(a'/W')_o$.

For the silicon nitride (NC132) RTBB specimens, critical SIF values vary from a high of 4.66 MPa√m to a low of 4.00 MPa√m. The average measured fracture toughness was 4.36 ± 0.21 MPa√m. This value agrees well with reported literature values measured by several methods [26-29]. Fig. 7 shows the critical SIFs subdivided into stable and unstable results. The analytically predicted threshold crack length, $(a/W)_o$ = 0.66, agrees well with the transition observed in the experiments. One specimen with a crack length of a/W = 0.75 was tested in a more compliant conventional fully articulating fixture. The measured critical SIF (4.65 MPa√m) was in the same range as those obtained from the other unstable specimens. Using the more compliant fixture increased the overall machine compliance to a value much above 60. For higher machine compliance values, the stability analysis predicts a longer threshold crack length or completely instability. Thus, this compliant fixture was expected to result in unstable fracture which agreed with the observed load displacement record and the higher critical SIF.

Discussion: Baratta et al. [7] and Underwood et al. [6] have found that for PMMA and tungsten the critical SIF values obtained in stable tests are lower than those obtained in unstable tests. Similar behavior has been observed for the RTBB silicon nitride (NC132) specimens in this study. No such conclusions can be drawn based on the results obtained from the MRBB alumina (AD99) and silicon nitride (HS130) specimens. The difference between unstable and stable tests for the RTBB silicon nitride is about 10% of the average measured critical SIF. Tables II to IV summarize the average measured critical SIFs of the MRBB silicon nitride (HS130) and the alumina (AD99) specimens, and the RTBB silicon nitride (NC132) specimens, respectively, grouped into unstable, semistable, and stable results based on the appearance of the load displacement record.

Table II. Critical SIFs of Silicon Nitride (HS130) MRBB Grouped by Degrees of Stability

	Unstable	Semistable	Stable
Average Critical SIF (MPa√m)	2.82 ± 0.13	2.82 ± 0.02	2.82 ± 0.16
No. of Specimens	4	2	2
a'/W'	0.27 - 0.51	0.55 - 0.64	0.67 -0.79

Table III. Critical SIFs of Alumina (AD99) MRBB Grouped by Degrees of Stability

	Unstable	Semistable*	Stable
Average Critical SIF (MPa√m)	2.27 ± 0.10	2.31 ± 0.05	2.15 ± 0.01
No. of Specimens	9	4	2
a'/W'	0.22 - 0.58	0.52 - 0.64	0.67 -0.76

*Include specimens susceptible to subcritical crack growth

Table IV. Critical SIFs of Silicon Nitride (NC132) RTBB Grouped by Degrees of Stability

	Unstable	Semistable	Stable*
Average Critical SIF (MPa√m)	4.54 ± 0.12	4.23 ± 0.13	4.19 ± 0.08
No. of Specimens	7	5	5
a/W	0.26 - 0.56	0.53 - 0.65	0.64 -0.78

*Does not include the specimen tested in the compliant fixture

For ceramic materials, it is typical to obtain a standard deviation of ± 0.15 to 0.3 MPa√m when fracture toughness is determined by the single edge pre-cracked beam (SEPB) method [4] regardless of the compliance of the machine and the specimen [30]. For the materials used in this MRBB geometry study, 10% would be 0.2 - 0.25 MPa√m, which is well within the experimental scatter of the results. To observe a noticeable difference in critical SIFs due to the instability (i.e., higher estimate of critical SIFs because of the unstable test), it would be necessary to select a material with a higher fracture toughness and a specimen with larger dimensions, so that systematic differences would not be obscured by the lack of resolution of the experiments. This lack of resolution stems largely from the large load cell with rather low resolution that is necessary in order to obtain the necessary stiffness. A specimen with higher fracture toughness and larger dimensions will give higher load readings thus reducing the scatter.

A comparison of Figs. 5 to 7 show that stability in the MRBB is obtained at a short crack length than for the RTBB in agreement with the prediction of Fig. 2. The stress intensity expression for bend bars increase rapidly as a function of dimensionless crack length, α, for $\alpha > 0.55$. Thus, for specimens with pre-cracks long enough to be in the stable region small errors in the crack length measurements can cause large inaccuracies in critical SIF measurements. Previous studies [12,13] indicate that using a span to width ratio between 7 and 8 would lead to a shorter threshold crack length. This, in turn, would make the results less sensitive to errors in crack length measurements thus giving more accurate fracture toughness results.

Conclusions: Fracture toughness tests were performed for bend bars of rectangular and modified round cross sections. The specimen pre-cracking and test conditions were selected in such a way that a transition from unstable to stable crack extension could be observed. This transition agreed well with previously published analytical predictions and showed the MRBB to be more stable for shorter pre-crack lengths than the RTBB. The stable tests gave lower critical SIFs for the tougher material. A similar difference could not be discerned in the other two materials possibly due to experimental scatter.

Acknowledgments: The authors wish to thank Mr. Brian Pothier (U.S. Army Research Laboratory) and Dr. David Jablonski (Instron) for their assistance with the mechanical testing. The support of this work by the U.S. Army Research Laboratory is greatfully acknowledged. Part of this work was performed while one of us was a National Research Council Fellow at NASA LeRC, Cleveland.

References

[1] Srawley, J.E. and Brown, W.F., Jr.. "Fracture Toughness Testing Methods," Fracture Toughness Testing and Its Applications, ASTM STP 381, ASTM, Philladelphia, PA, 133-198 (1966).

[2] ASTM E-399, "Standard Test Method for Plane-Strain Fracture Toughness of Metallic Materials," Annual Book of ASTM Standards, Vol. 03.01., ASTM, Philadelphia, PA 487-511 (1989).

[3] Fett, T. and Munz, D., "Differences between Static and Cyclic Fatigue Effects in Alumina," J. Mater. Sci. Let., **12** [5] 352-354 (1993).

[4] Nose, T. and Fujii, T, "Strain Rate Dependence of Fracture Toughness for Ceramic Materials," in S62 Annual Meeting of Japan Ceramic Society, Nagoya, May 12, 1987.

[5] Clausing, D.P., "Crack Stability in Linear Elastic Fracture Mechanics," Int. J. of Fract. Mech., **5**, 211-227 (1969).

[6] Underwood, J.H., Baratta, F.I., and Zalinka, J.J., "Fracture Toughness Tests and Displacement and Crack Stability Analyses of Round Bar Bend Specimens of Liquid-Phase Sintered Tungsten," Experimental Mechanics, **31**, 353-359 (1991).

[7] Baratta, F.I. and Dunlay W.A., "Crack Stability in Simply Supported Four-Point and Three-Point Loaded Beams of Brittle Materials," Mechanics of Materials, **10**, 149-159 (1990).

[8] Nose, T. and Fujii, T., "Evaluation of Fracture Toughness for Ceramic Materials by a Single-Edge-Precracked-Beam Method," J. Am. Ceram. Soc., **71**, 328-333 (1988).

[9] Bar-On, I., Beals, J.T., Leatherman, G.L., and Murray, C.M., "Fracture Toughness of Precracked Bend Bars," J. Am. Ceram. Soc., **73** [8] 2519-2522 (1990).

[10] Warren, R. and Johanneson, B., "Creation of Stable Crack in Hard Metals Using 'Bridge' Indentation," Powder Metall., **27** [1] 211-227 (1969).

[11] Cho, K., Hantz, B.F., IV, and Bar-On, I., "Stress Intensity Factor Calculation for a Modified Round Bend Bar by 3-D Finite Element Analysis," Int. J. Fract., **62**, 163-170 (1993).

[12] Cho, K and Bar-On, I, "Crack Stability Analysis and Fracture Toughness of Ceramic Bend Bars with a Modified Circular Cross Section", to be published in Experimental Mechanics.

[13] Bar-On, I, Baratta, F.I., and Cho, K., "Crack Stability and Its Effect on Fracture Toughness of Hot Pressed Silicon Nitride Beam Specimen," to be published in J. Am. Ceram. Soc.

[14] Bluhm, J.I., "Stability Consideration in the Generalized Three Dimensional 'Work of Fracture' Specimen," Fracture 1977, **3**, ICF4, Waterloo, Canada, 409-417 (1977).

[15] Baratta, F.I., "Load-Point Compliance of a Three-Point Loaded Cracked-Notched Beam,"

J. Test. and Eval., 16, 59-71 (1988).

[16] Srawley, J.E., "Wide Range Stress Intensity Factor Expression for ASTM E-399 Standard Fracture Toughness Specimens," Int. J. Fract., 12, 475-476 (1976).

[17] Brown, W.F., Jr. and Srawley, J.E., "Plane Strain Crack Toughness Testing of High Strength Metallic Materials," ASTM STP 410, ASTM, Philadelphia, PA, (1966).

[18] Quinn, G.D., Corbin, N.D., and McCauley, J.W., "Thermomechanical Properties of Aluminum Oxynitride Spinel," Am. Ceram. Soc. Bull., 63 (5) 723-729 (1994).

[19] Miller, D.G., Anderson, C.A., Singhal, S.C., Lange, F.F., Diaz, E.S., and Kossowsky, R., "Brittle Materials Design, High Temperature Gas Turbine Material Technology," AMMRC CTR 76-32 Volume IV, Final Report, Dec. (1976).

[20] Bratton, R.J. and Miller, D.G., in Ceramics for High Performance Applications - II, ed J.J. Burke, E.N. Lenoe, and R.N. Katz, Brook Hill Publishing Co., Chestnut Hill, MA (1978).

[21] Lange F.F., and Iskoe J.L., in Ceramics for High-Performance Applications, ed J.J. Burke, A.E. Gorum, and R.N. Katz, Brook Hill Publishing Co., Chestnut Hill, MA (1974).

[22] Kossowsky R., in Ceramics for High-Performance Applications, ed J.J. Burke, A.E. Gorum, and R.N. Katz, Brook Hill Publishing Co., Chestnut Hill, MA (1974).

[23] Ritter, J., Nair S., Gennari P., and Dunlay W., "High-Strength Reaction-Bonded Silicon Nitride," Adv. Ceram. Mat., 3 (4) 415-417 (1988).

[24] "Flexural Strength of High Performance Ceramics at Ambient Temperature," Military Standard 1942A (MIL-STD-1942A), Army Research Laboratory, Materials Directorate, Watertown, MA, 1983.

[25] Baratta, F.I., Driscoll, G.W., and Katz, R.N., in Ceramics for High-Performance Applications, ed J.J. Burke, A.E. Gorum, and R.N. Katz, Brook Hill Publishing Co., Chestnut Hill, MA (1974).

[26] Evans, A.G. and Charles, E.A., "Fracture Toughness Determination by Indentation," J. Am. Ceram. Soc., 59 [7-8] 371-372 (1976).

[27] Salem, J.A. and Shannon, J.L., "Fracture Toughness of Si_3N_4 measured with Short Bar Chevron Notched Specimens," J. Mater. Sci., 22, 321-324 (1987).

[28] Anstis, G.R., Chantikul, P., Lawn, B.R., and Marshall, D.B., "A Critical Evaluation of Indentation Techniques for Measuring Fracture Toughness: I, Direct Crack Measurements," J. Am. Ceram. Soc., 64 [9] 533-538 (1981).

[29] Chantikul, P., Anstis, G.R., Lawn, B.R., and Marshall, D.B., "A Critical Evaluation of Indentation Techniques for Measuring Fracture Toughness: II, Strength Method," J. Am. Ceram. Soc., 64 [9] 539-543 (1981).

[30] Quinn, G.D., Salem, J., Bar-On, I., Cho, K., Foley, M., and Fang, H., "Fracture Toughness of Advanced Ceramics at Room Temperature," J. Res. Natl. Inst. Stand. Technol., 97, 579-607 (1992).

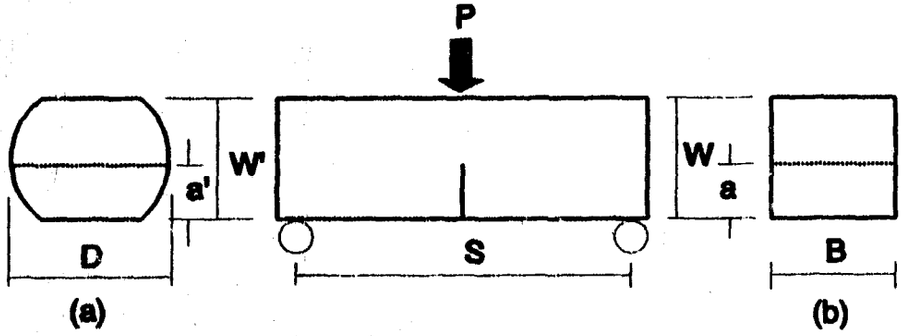


Fig. 1. Geometries of (a) the modified round bend bar (MRBB) and (b) the rectangular bend bar (RTBB).

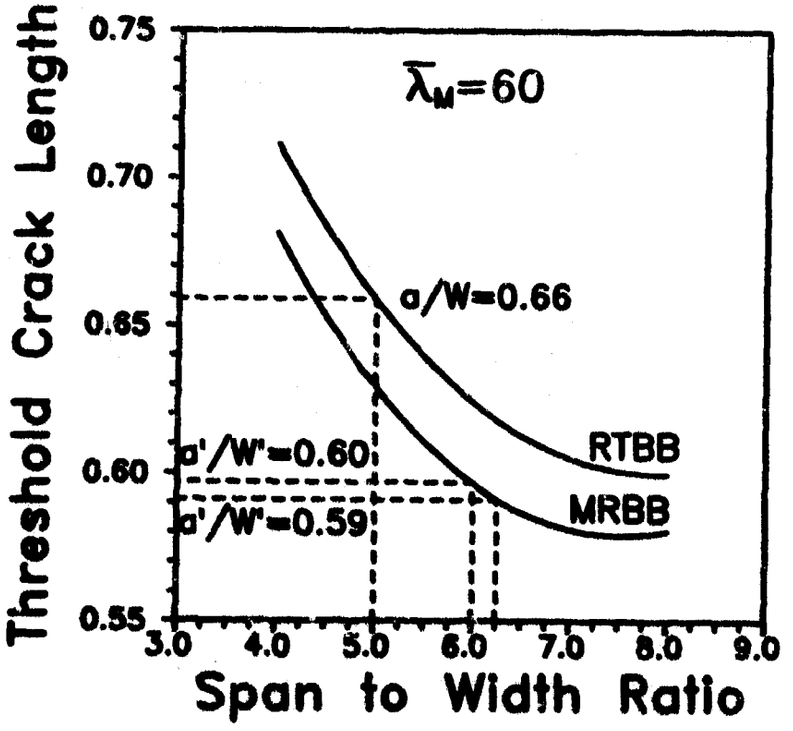


Fig. 2. The threshold crack length for the rectangular and modified round bend bars with a nondimensional machine compliance, $\bar{\lambda}$, of 80.

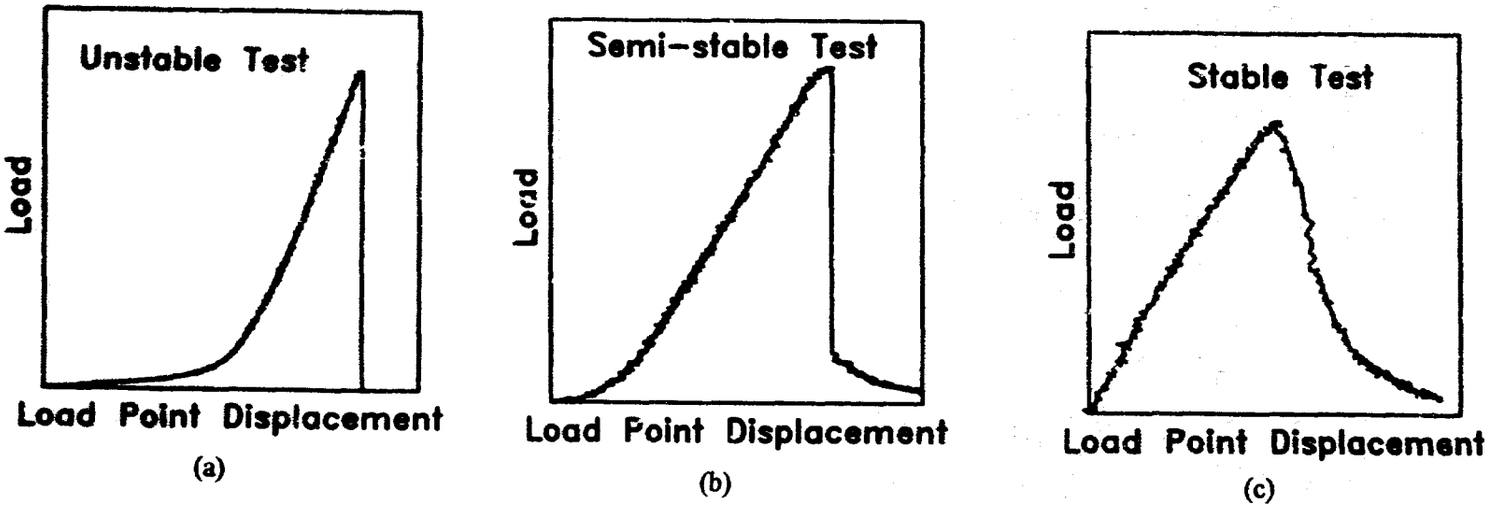


Fig. 3. Typical load displacement records for (a) an unstable test, (b) semi-stable test, and (c) stable test.

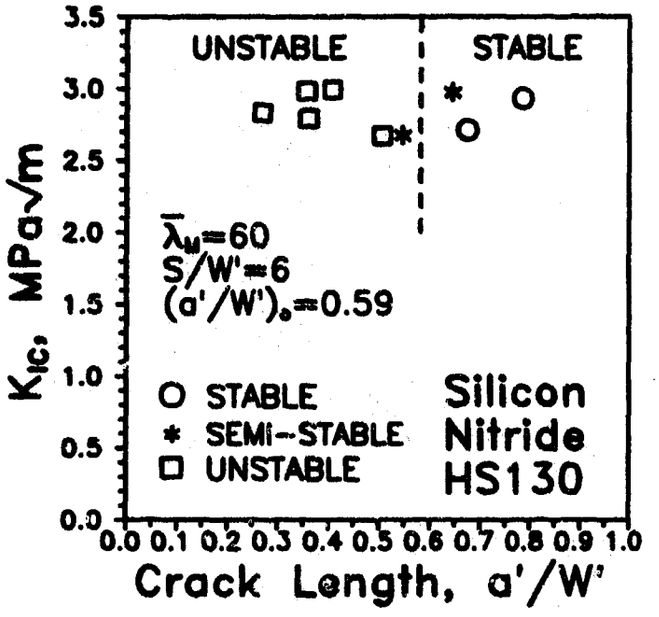


Fig. 4. Measured critical stress intensity factors of the silicon nitride (HS130) modified round bend bars for varying pre-crack lengths. The predicted threshold crack length is $(a'/W')_o = 0.60$.

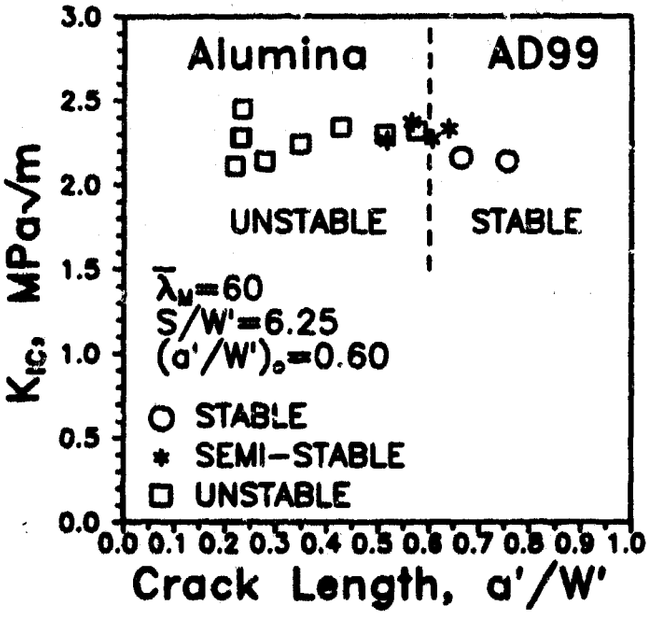


Fig. 5. Measured critical stress intensity factors of the alumina (AD99) modified round bend bars for varying pre-crack lengths. The predicted threshold crack length is $(a'/W')_o = 0.59$.

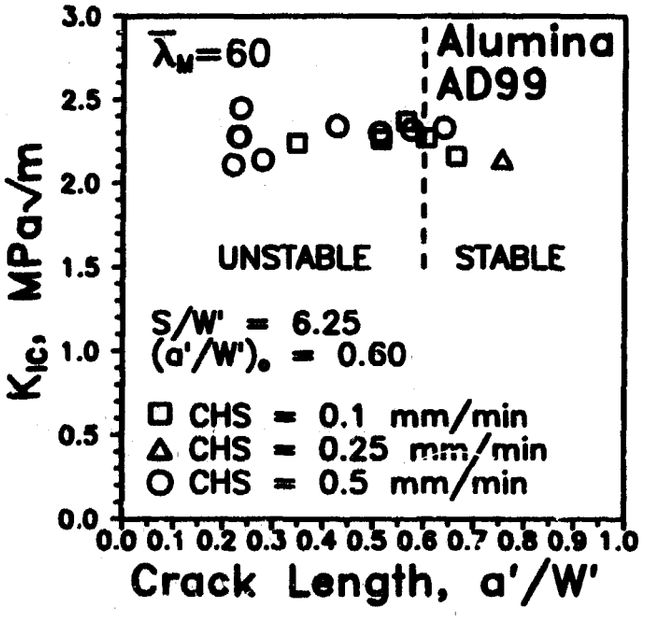


Fig. 6. Measured critical stress intensity factors of the alumina (AD99) modified round bend bars grouped by displacement rates.

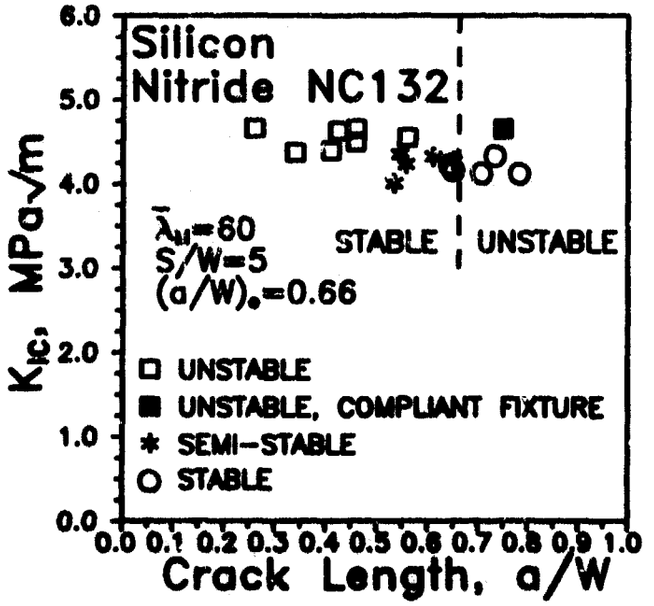


Fig. 7. Measured critical stress intensity factors of the silicon nitride (NC132) rectangular bend bars for varying pre-crack lengths. The predicted threshold crack length is (a/W) = 0.66.

THE IMPORTANCE OF DIESEL OIL FILTRATION DEMONSTRATED BY SURFACE LAYER ACTIVATION WEAR MONITORING

John Truhan and Charles Covington
Fleetguard Corporation
Cookeville, TN 38501

Claudia Colerico and Charles Blatchley
Spire Corporation
One Patriots Park
Bedford, MA 01730-2396

Abstract: Surface Layer Activation (SLA) is a technique which allows convenient wear monitoring of a selected part from outside a running engine. A particle accelerator activates the part by bombarding it with high energy ions. Gamma rays emitted from this volume have characteristic energies that can be detected and analyzed to indicate surface material remaining in real-time while the engine operates. The extreme precision of SLA in measuring wear is ideal for evaluating ordinary mechanical interactions, for high temperature environments, and for lubrication testing. Fleetguard is applying SLA markers, rather than tracers, in an *in situ* tribological evaluation of filter effectiveness. The marker method measures the decrease of radioactivity remaining in the activated part as particles of wear debris are carried away. A detector, mounted external to the engine, directly monitors the intensity of a selected gamma ray energy. Many of the penetrating gamma rays pass from the activated part through the engine to reach the detector and be counted. Currently, a top piston ring, cylinder liner and sliding tappet cam follower are being monitored simultaneously using different isotopes in a medium duty diesel engine to evaluate the relative effects of oil chemistry and cleanliness on wear. SLA has proven to be a cost and time effective tool to evaluate the performance of oils and filters in operating engines.

Key Words: Diesel engines; filtration; lubrication; surface layer activation; wear testing

Introduction: Operators of diesel engines, for both on-highway, off-highway , and stationary applications, are attempting to reduce operating costs by increasing routine service intervals, in some cases, far beyond current recommendations. For light duty cycles, this can often be accomplished with current lubricating oils and filtration with little danger to the engine. For heavier duty cycles, however, severe degradation of engine life could result due to increased wear. To reliably specify oil and filtration quality for operating in these heavier service conditions, extensive testing is needed to relate engine wear to oil chemistry and cleanliness. Unfortunately, traditional field testing with engine tear down and inspection is too costly in both time and labor involved.

As an alternative to field testing, Surface Layer Activation, or SLA, can be an effective on-line monitoring tool, which avoids most mechanical inspection labor and greatly reduces engine time needed for reliable measurements. The ability to monitor minute increments of surface loss can be particularly powerful when coupled with information about oil condition and duty cycle. In SLA, particle accelerator bombardment of a selected engine part tags a shallow near-surface zone with a gamma ray emitting radionuclide. Very small amounts are used, usually in the microcurie range, and often below the level requiring a license. A detector at one of several possible locations outside the engine then monitors the gamma ray signal from either the tagged part (marker approach) or from debris collected in a filter or oil line (tracer approach). Several pedestrian introductions to SLA have been published,[1-3] and additional technical details are included in a recent ASM Handbook.[4]

The operating principle is that activity in a monitored location is proportional to a counting rate determined by counting gamma rays of the selected energy and dividing by the length of the counting interval in seconds. As long as the detector position is fixed, this proportion remains unchanged, although its exact value is usually not needed for monitoring. Instead, counting rates for measurements corresponding to worn conditions are compared to the starting or "no wear" counting rate.

The gamma ray energy is selected by fitting a curve to a spectrum of counts accumulated per energy channel or bin. After an automated fitting routine separates the counts under a gamma ray peak from the background contribution, division by the time interval produces a counting rate, expressed in counts per second, which is corrected for the radionuclide's natural decay before dividing by the (half-life corrected) starting counting rate. Since decay events are random, counts under a gamma ray peak, and subsequent counting rates, are controlled by Poisson counting statistics. The main implication is that at least 10,000 counts must be accumulated in each measurement to assure a 1% statistical uncertainty, which is roughly comparable to instrumental uncertainties. For many applications, particularly those for which a linear regression may later be applied, a larger range of variability in individual measurements can be tolerated, provided many more readings are taken.

Wear deduced from these counting rates critically depends on a calibration curve plotting wear depth as a function of fraction of activity remaining. Calibrations are determined in laboratory wear simulations using identical activation conditions as those applied to the monitored parts. The maximum depth for which these calibration curves produce reliable readings contains about 90% of the activity. This point on the curve, with 10% activity remaining, is arbitrarily termed the "irradiation depth." The greatest accuracy in individual readings occurs at about half of this so-called irradiation depth. Incident particles, of course, continue well beyond the point where activity is produced.

Fleetguard has employed the SLA marker technique for several years to relate wear in diesel engines to lubricant filtration levels and lube oil chemistry as part of an ongoing program to understand the relative effects of each on the various engine parts subject to wear. The studies completed under this program were intended to provide technical guidance to both engine designers and end users. Additional results were reported earlier,[5-7] confirming that each tribological system in the engine has a different sensitivity

to cleanliness and chemistry. For example, lead overlay bearings are very sensitive to acid buildup in the oil but can embed particulate contamination to some extent, whereas the chromium plated piston ring was found to be relatively insensitive to oil chemistry but sensitive to particulate contamination levels.[6]

Experimental Procedures: A top piston ring, the corresponding cylinder liner, and sliding tappet cam follower were irradiated to produce Mn-54, Co-56, and Co-57 respectively. A summary of the bombardment conditions are given in Table I. A complete discussion of gamma spectrometry procedures used to monitor these isotopes simultaneously has been published elsewhere.[7] The main complication is deconvolution or subtraction of peaks that coincide in a NaI spectrum. These separations must occur before half-life corrections and comparison to starting counting rates. To provide adequate counting statistics for these subtractions, each wear datum required a counting interval of one hour.

Table I *Summary of irradiation conditions for the engine parts tested.*

COMPONENT	ISOTOPE	HALF LIFE (days)	IRRADIATION SPECIES	ACTIVITY LEVEL (microCi)	IRRADIATION DEPTH (microns)
Ring	Mn-54	303	6.5 MeV Deuterons	3	35
Liner	Co-56	77	6.52 MeV Protons	10	15
Tappet	Co-57	270	6.52 MeV Deuterons	3	25

94141

The oils were aged either in the test cell or in the field, however, no artificial contaminants were added, so that contaminant levels are realistic. The oils are run in the test engine, which is a medium duty diesel engine with a sump capacity of about 25 liters. In order to determine the aggressiveness of an oil to produce wear, the test engine is run at full rated speed and load for at least 24 hours. The oil is run as-received with no additional lube filtration on the test engine. A wear vs time curve is generated for each part, the slope of the curve being the wear rate for that set of conditions. In some cases, a cleanup filter can then be installed on the test engine without draining the oil and the engine rerun so that the effect of filtration level can be studied separately from oil chemistry. Full chemical analysis of the oil includes a determination of wear metal concentrations, additive level, total acid and total base numbers, viscosity and soot content. Since used oil is opaque, traditional particle counting techniques are not applicable, so contamination levels are determined by measuring the differential pressure buildup across screens of different mesh sizes using a Coulter LCM2. It provides particle concentrations for sizes greater than 25, 15, and 5 micrometers.

Results and Discussion: In earlier testing with a top piston ring only,[6] ring wear rates were measured as a function of oil chemistry and particulate concentration. Figure 1 shows ring wear rate as a function of concentration of particles 15 micrometers and greater. As would be expected, the cleaner the oil, the lower the wear rates, with about a factor of two reduction in the wear rates for an order of magnitude reduction in contamination level. The wear rate was almost reduced to new oil levels by filtering a used oil with a high efficiency system, so that even with a depleted additive package, ring wear could effectively be controlled through filtration.

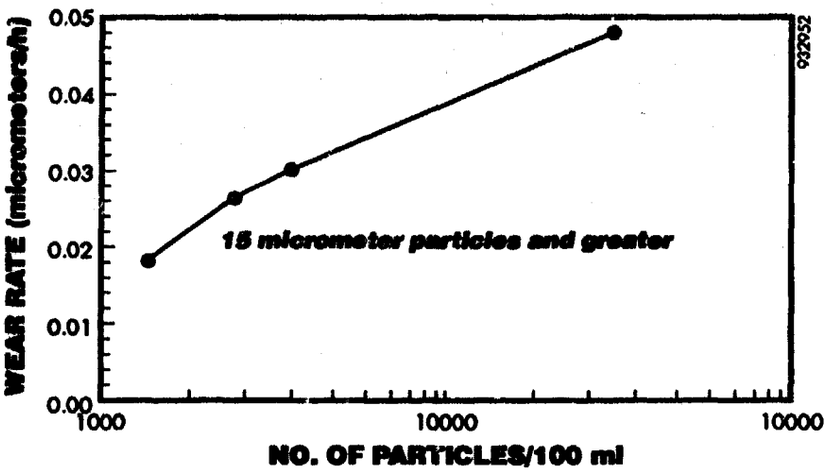


Figure 1 *The relationship between ring wear rate and oil cleanliness in a heavy duty diesel engine running at full load and speed.*

Since the marker technique provides wear rates in terms of a dimensional change, the wear rate can be transformed into part life by assuming an arbitrary amount of wear that can be tolerated in that particular part. In the case of a top piston ring, ring life is roughly equivalent to engine life since diesel engines are usually rebuilt when oil consumption becomes excessive. Using the data from Figure 1, two wear rates comparing the range of commercially available filtration efficiencies were transformed to part lives and shown as a function of filter efficiency for 15 micrometer particles in Figure 2. The extremes in ring life represent the range of expected engine life for most on-highway applications. The type of data represented in Figure 2 allows engine designers to specify the proper filtration levels depending on application and also allows end users to determine the cost effectiveness of applying more expensive advanced filtration products to their engines.

The accuracy of the SLA marker technique is illustrated in Figure 3. A profile trace of the face of the top piston ring used in the study described above was taken before and after use. The ring had about 1200 hours of operation in a variety of load and speed cycles before it was removed from the engine. There is very good agreement between the amount of wear predicted by SLA and the actual wear as measured by a comparison of the profile traces. This demonstrates that, with proper care, the marker technique can provide not only relative wear comparisons, as does the tracer technique, but absolute data as well.

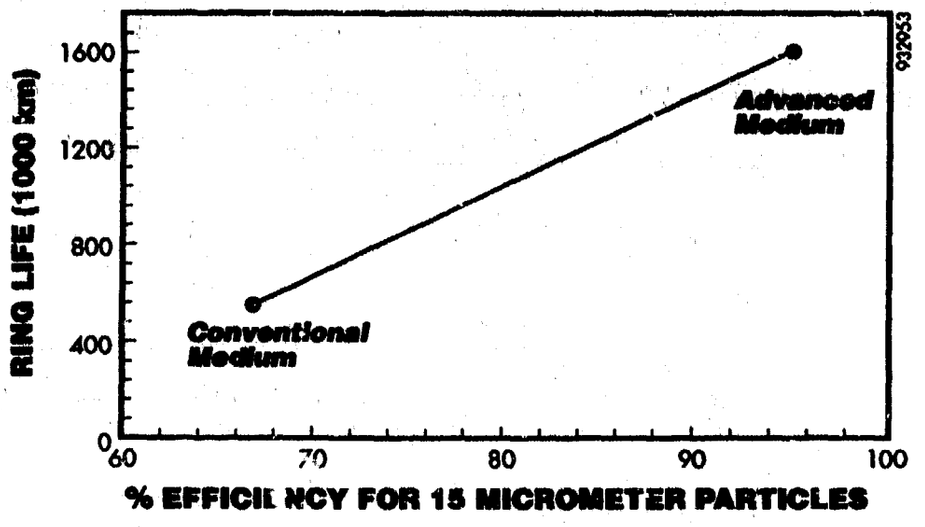


Figure 2 *The relationship between ring life and filter efficiency for 15 micrometer particles.*

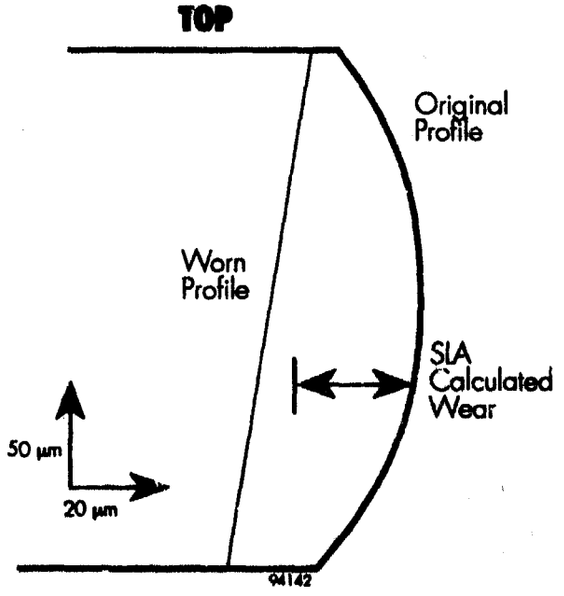


Figure 3 *A comparison of the before and after profile trace of the ring face with the amount of wear calculated by SL4.*

A typical wear vs time curve for the tappet, ring and liner is shown in Figure 4. Individual data points from each hour of counting are shown along with the linear regression lines to determine wear rates for that given set of conditions. The individual data points show little scatter from measurement to measurement and are an indication of the reproducibility of SLA even in running engines. The oil represented in Figure 4 was run on-highway for about 32K kilometers in light to medium duty applications. The wear rates measured under these conditions were uniformly low and the oil showed only moderate chemical degradation.

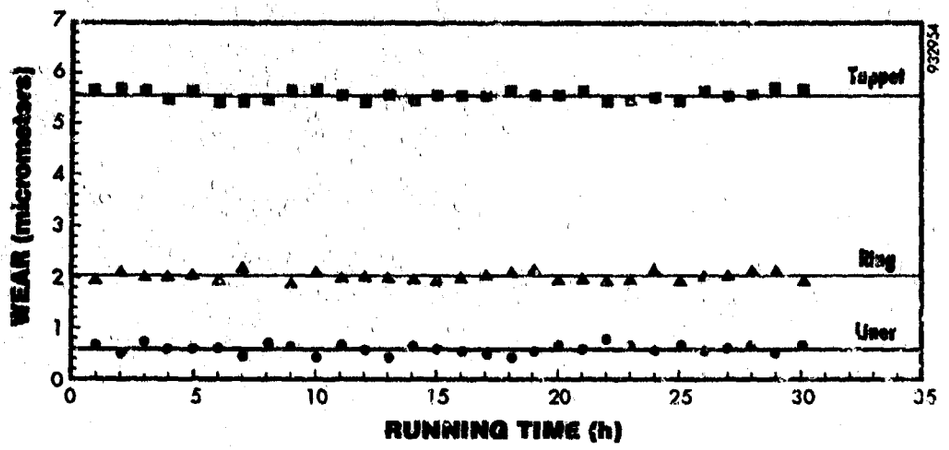


Figure 4 *A typical wear vs time run for an oil aged in the field for 32K kilometers.*

Figure 5 shows the increase in wear resulting from extending the service interval from 32K kilometers to 96K kilometers. Although the liner seemed to be insensitive to this extension, the ring and tappet showed sharp increases in the rate of wear. The liner in the test engine is made from a hardened cast iron and the ring is only in periodic contact with any given location, so it should not show the sensitivity to operating conditions as would the ring and tappet. The ring and tappet are in constant contact with their opposing surfaces and the contact stresses are much higher for the tappet, so it should be the part most sensitive to operational variations.

As stated above, the field return oils are run in the test engine without additional filtration to determine their aggressiveness in the as-received condition. Figure 6 shows a comparison of the as-received oil wear rates to wear rates obtained while running through a high efficiency cleanup filter on the test engine for another 24 hours. The oil shown here was run in the field for 96K kilometers, so it is considered to be fully degraded by any typical measure. Despite its degraded chemistry, a cleanup of this oil resulted in drastically reduced wear rates for the ring and tappet. The liner, once again, showed a lack of sensitivity to operating conditions. Figure 7 shows the same comparison of wear rates plotted against the concentration of greater than 15 micrometer particles demonstrating the cleanup effect more directly.

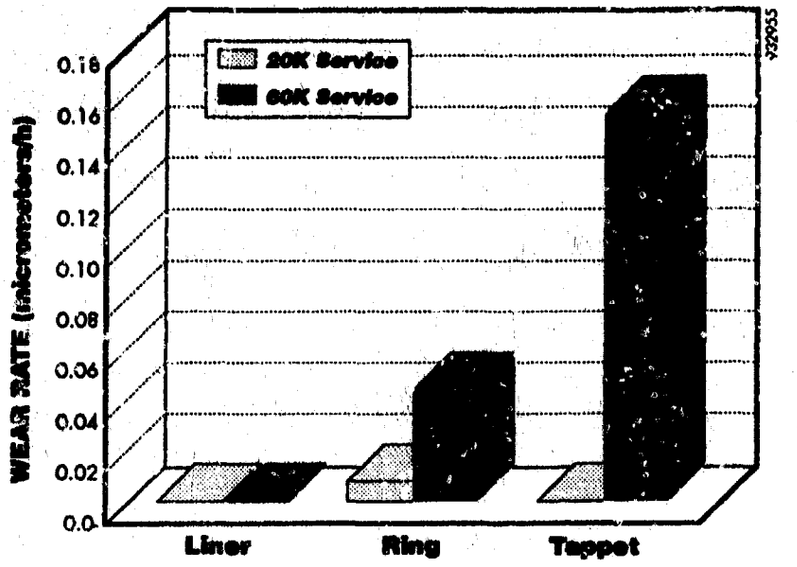


Figure 5 *A comparison of wear rates as a function of service interval.*

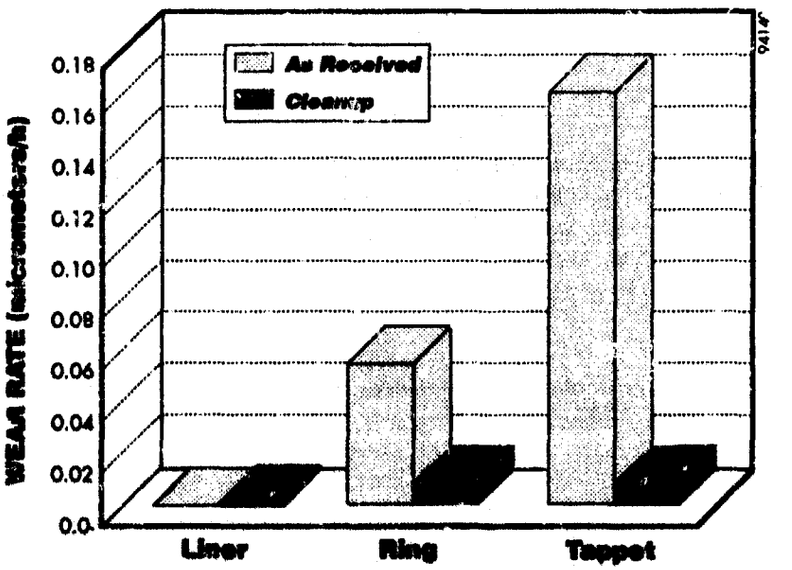


Figure 6 *The reduction of wear rates obtained by filtering a 96K kilometer oil during the SLA test with a high efficiency filter.*

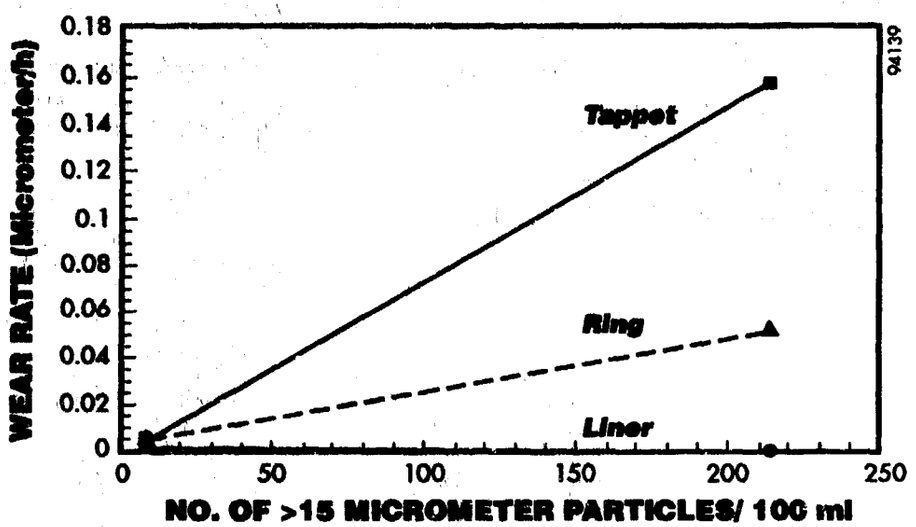


Figure 7 *The relationship between wear rate and concentration of greater than 15 micrometer particles for the 96K kilometer oil before and after cleaning.*

Summary and Conclusions:

1. SLA markers successfully measured wear rates in an operating diesel engine on a top piston ring, cylinder liner, and sliding-tappet cam-follower to evaluate effects of filtration and lubricant chemistry.

2. Surface loss (loss of measured activity due to wear) from all three components was monitored simultaneously during engine operation using distinct isotopes. Increments of wear derived from these online readings were combined in later analysis to determine wear rates associated with each operating condition. This test configuration achieved reduced costs and time compared to conventional field testing.

3. Oils aged in the field were returned to study their aggressiveness in producing wear in the test engine. Oil chemistry and cleanliness levels were measured to determine the relative importance of each in producing wear in the different engine systems.

4. Despite having depleted additive packages, wear was reduced in heavily used oils by high efficiency filtration.

References:

1. Blatchley, C. and Sioshansi, P., "Monitoring Wear with Gamma Rays," *Machine Design*, October 25, 1990, pp. 99-102.

2. Blatchley, C., "Monitoring Wear in Engine Components by SLA: Extension to Nanometer Accuracy," SAE Fuels and Lubricants Conf. (Toronto), November, 1987.

3. Fritz, S.G. and Cataldi, G.R., "*In Situ* Piston Ring Wear Measurements in a Medium Speed Diesel Engine," *Lubr. Eng. 4*, (6) 1989, p. 365.

4. Blatchley, C., "Radionuclide Methods," in *ASM Handbook, Volume 18, (Friction, Lubrication, and Wear Technology)*, Henry, S. Ed. (Blau, P.J., Volume Chairman), ASM International, 1992, pp. 319-329.

5. Truhan, J., Covington, C., Colerico, C., and Blatchley, C., "Surface Layer Activation for Wear Studies," NUCL-065, Am. Chem. Soc. Nat. Meeting, San Diego, March, 1994.

6. Truhan, J. and Covington, C., "Effect of Filtration on Top Ring Face Wear in Heavy Duty Diesel Engines," Coordinating European Council, Paper number CEC/93/EL18, May, 1993.

7. Truhan, J. and Covington C., "The Use of Direct Method Surface Layer Activation to Measure Wear in Diesel Engines," Society of Automotive Engineers, SAE 931627, Oct., 1993.

APPENDIX

MFPG PUBLICATIONS

Both printed and microfiche copies of the following MFPG publications whose catalog numbers start with either **AD** or **COM** may be obtained from the

National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161

Glossary of Terms	AD 721 354
Proceedings of Meeting Nos. 1-9 (set of five) Meeting Nos. 1-5 Papers and Discussion on Failure Analysis and Control	AD 721 359
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" January 21-22, 1970	AD 721 912
Meeting No 11 "Failure Mechanisms: Fatigue" April 7-8, 1970	AD 724 475
Meeting No 12 "Identification and Prevention of Mechanical Failures in Internal Combustion Engines" July 8-9, 1970	AD 721 913
Meeting No 13 "Standards as a Design Tool in Surface Specification for Mechanical Components and Structures" October 19-20, 1970	AD 724 637
Meeting No 14 "Advances in Decision-Making Processes in Detection, Diagnosis and Prognosis" January 25-26, 1971	AD 721 355
Meeting No 15 "Failure Mechanisms: Corrosion" April 14-15, 1971	AD 725 200
Meeting No 16 "Mechanical Failure Prevention Through Lubricating Oil Analysis" November 2-4, 1971	AD 738 855

Meeting No 17 "Effects of Environment Upon Mechanical Failures, Mechanisms and Detection" AD 750 411
April 25-27, 1972

Meeting No 18 "Detection, Diagnosis and Prognosis" AD 772 082
November 8-10, 1972

Meeting No 19 "The Role of Cavitation in Mechanical Failures" (NBS SP 394)
October 31-November 2, 1973 COM-74-50523

Printed copies of the following publications may be obtained from the

U.S. Government Printing Office
Superintendent of Documents
Washington, DC 20402

Microfiche copies of these publications may be obtained from the **NTIS**.

Proceedings

Meeting No 20 "Mechanical Failure - Definition of the Problem" (NBS SP 423)
May 8-10, 1974 SN003-003-01451-6

Meeting No 21 "Success by Design: Progress Through Failure Analysis" (NBS SP 433)
November 7-8, 1974 SN003-003-01639-0

Meeting No 22 "Detection, Diagnosis and Prognosis" (NBS SP 436) SN003-003-01556-3
April 23-25, 1975

Meeting No 23 "The Role of Coatings in the Prevention of Mechanical Failure" (NBS SP 452)
October 29-31, 1975 SN003-003-01664-1

Meeting No 24 "Prevention of Failures in Coal Conversion Systems" (NBS SP 468)
April 21-24, 1976 SN003-003-01760-4

Meeting No 25 "Engineering Design" (NBS SP 487) SN003-003-01829-5
November 3-5, 1976

Meeting No 26 "Detection, Diagnosis and Prognosis" (NBS SP 494) SN003-003-01844-9
May 17-19, 1977

Meeting No 27 "Product Durability and Life" (NBS SP 514) SN003-003-01935-6
November 1-3, 1977

Meeting No 28 "Detection, Diagnosis and Prognosis" (NBS SP 547) SN003-003-02083-4
November 28-30, 1978

Meeting No 29 "Advanced Composites" (NBS SP 563) SN003-003-02120-2
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 SN003-003-02272-1

Meeting No 31 "Failure Prevention in Ground Transportation Systems" (NBS SP 621)
April 22-24, 1980 SN003-003-02428-7

Meeting No 32 "Detection, Diagnosis and Prognosis: Contribution to the Energy Challenge" (NBS SP 622) SN003-003-02361-2
October 7-9, 1980

Meeting No 33 "Innovation for Maintenance Technology Improvements" (NBS SP 640)
April 21-23, 1981 SN003-003-02425-2

Meeting No 34 "Damage Prevention in the Transportation Environment" (NBS SP 652)
October 21-23, 1981 SN003-003-02488-1

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 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

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: UNCLASSIFIED/UNLIMITED
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): MFPG 48	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): Advanced Materials and Process Technology for Mechanical Failure Prevention

12. PERSONAL AUTHOR(S): Compiled by Henry C. Pusey and Sallie C. Pusey

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Proceedings	FROM 94/1/1 TO 94/4/21	94/4/21	446

16. SUPPLEMENTARY NOTATION

17. COSATI CODES: FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Rotating equipment, machinery, condition monitoring, failure analysis, failure detection, diagnostics, sensors, neural networks, signature analysis and nondestructive evaluation.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This document is the proceedings of the 48th Meeting of the Mechanical Failures Prevention Group (MFPG) which was held in Wakefield, Massachusetts on April 19-21, 1994. The proceedings contains featured papers from the Opening Session and a Plenary Session. There are session papers on Machinery Monitoring and Diagnostics, Failure Analysis, Evaluation of Materials Properties, Sensors and Neural Networks, Applications of Diagnostics and Nondestructive Test and Evaluation.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL: Peter Schmidt	22b. TELEPHONE (Include Area Code): (703) 696-4410	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: 1988—547-046