DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

NEW FRONTIERS IN INTEGRATED DIAGNOSTICS AND PROGNOSTICS

Proceedings of

THE 55th Meeting

of the

SOCIETY FOR MACHINERY FAILURE PREVENTION TECHNOLOGY

VIRGINIA BEACH, VIRGINIA
April 2 – 5, 2001

Compiled by

HENRY C. and SALLIE C. PUSEY
and
W. R. HOBBS

A Publication of the

(Society for Machinery Failure Prevention Technology (MFPT)
(A Division of the Vibration Institute)

20020920 025

Copyright © 2001 by
Society for Machinery Failure Prevention Technology (MFPT)
(A Division of the Vibration Institute)
1877 Rosser Lane
Winchester, VA 22601
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.

MFPT 55

PREFACE vii

MFPT SOCIETY BOARD OF DIRECTORS viii

FEATURED PAPER

Vibratory Locomation Revisited 3
H. A. Gaberson, P. L. Stone, J. B. Curry and R. S. Chapler

DIAGNOSTICS

Gear Fatigue Crack Diagnosis by Vibration Analysis Using Embedded Modeling 17
C. J. Li and J. Lee

Comparative Evaluation of Structural Surface Intensity to Statistical Features for Gearbox Failures 27
J. C. Banks, R. L. Campbell and C. Byington

A Hybrid Stochastic-Neuro-Fuzzy Model-Based System for In-Flight Gas Turbine Engine Diagnostics 37
D. M. Ghiocel and J. Altmann

Diagnostic Feature Comparisons for Experimental and Theoretical Gearbox Failures 47
R. L. Campbell, J. C. Banks, C. Begg and C. Byington

Real-Time Condition Based Maintenance for High Value Systems 57
W. W. Matzelevich

Review of Vibration-Based Helicopters Health and Usage Monitoring Methods 69
V. Giurgiutiu, A. Cuc and P. Goodman

FAILURE MODES AND ANALYSIS I

Aftermarket Parts: Are They All They Are "Cracked" Up to Be? 81
V. K. Champagne

Failure Modes and Predictive Diagnostics Considerations for Diesel Engines 93
J. Banks, J. Hines, M. Lebold, R. Campbell, C. Begg and C. Byington

The Role of Manufacturing Defects in Munition Component Failures 103
M. Pepi

Evaluating the Impact of Environmentally Friendly Cleaners on System Readiness 117
W. Ziegler and A. Walker, Jr.

Estimation of Reliability Growth Determination in Cracked Specimens Under Fatigue Failure 127
M. Riahi and M. Aslanimanesh

PROGNOSTICS

A Prognostic Modeling Approach for Predicting Recurring Maintenance for Shipboard Propulsion Systems 139
G. Kacprzynski, M. Gumina, M. J. Roemer, D. E. Caguiat, T. R. Galie & J. J. McGroarty

Prognostic Enhancements to Naval Condition-Based Maintenance Systems 151
M. J. Roemer, G. J. Kacprzynski, A. Palladino, T. Galie, C. Byington and M. Lebold

Prediction Methods and Data Fusion for Prognostics of Primary and Secondary Batteries 165
J. D. Kozlowski, M. J. Watson, C. S. Byington, A. K. Garga and T. A. Hay

FAILURE MODES AND ANALYSIS II

Effects of Shot Peening Processing on the Fatigue Behavior of Three Aluminum Alloys And TI-AL-4V 177
J. Campbell

Ejection-Seat-Quick-Release-Fitting – Quantitative Fractography and Estimation of the Local Toughness Using the Topography of the Fracture Surface 185
K. Wolf

A Comparison of Fatigue Design Methods 193
R. J. Scavuzzo

SENSORS AND AUTOMATED REASONING

Robust Laser Interferometer Findings Relative to Condition Monitoring and Diagnostics/ Prognostics Engineering Management 205
M. F. Karchnak, A. J. Hess and T. Goodenow

Application of Torsional Vibration Measurement to Shaft Crack Monitoring in Power Plants 217
K. Maynard, M. Trethewey and C. Groover

Methods to Estimate Machine Remaining Useful Life Using Artificial Neural Networks 227
M. A. Essawy

Automated Recognition of Advanced Vibration Features for Machinery Fault Classification 237
K. McClintic, R. Campbell, G. Babich, A. Garga, J. Banks, M. Thurston and C. Byington

Development of Diagnostic and Prognostic Technologies for Aerospace Health Management Applications 247
M. J. Roemer, G. Kacprzynski, E. Nwadiogbu and G. Bloor

TOTAL OWNERSHIP COST

Enhanced FMECA: Integrating Health Management Design and Traditional Failure Analysis 261
G. Kacprzynski, M. J. Roemer, C. Byington and R. Campbell

Issues in the Design and Optimization of Health Management Systems 275
M. Yukish, C. Byington and R. Campbell

Cost Benefit Analysis Models for Evaluation of VMEP/HUMS Project 285
V. Giurgiutiu, G. Craciun and A. Rekers

FEATURE EXTRACTION

Extraction of Bearing Fault Transients from a Strong Continuous Signal via DWPA Multiple Band-Pass Filtering 297
J. Altmann and J. Mathew

Minimizing Load Effects on NA4 Gear Vibration Diagnostic Parameter 307
P. J. Dempsey and J. J. Zakrajsek

The Use of Histograms for Detection of Electrical Insulation Breakdown 319
J. Wang and S. McInerny

Detection and Severity Assessment of Faults in Gear Boxes from Stress Wave Capture and Analysis 329
J. C. Robinson

Assessment of Data and Knowledge Fusion Strategies for Diagnostics and Prognostics 341
G. J. Kacprzynski, M. J. Roemer and R. F. Orsagh

Machinery Diagnostic Feature Extraction and Fusion Techniques Using Diverse Sources 351
C. Lee and J. Pooley, III

STANDARDS DEVELOPMENT

Standards Developments for Condition-Based Maintenance Systems 363
M. Thurston and M. Lebold

Development of Performance and Effectiveness Metrics for Mechanical Diagnostic Technologies 375
R. F. Orsagh, M. J. Roemer, C. J. Savage and K. McClintic

SMART SENSORS

The Clients' View of CBM in 2001 389
L. Watt

Conductive Polymer Sensor Arrays – A New Frontier Technology for CBM 393
J. N. Schoess

Intelligent Sensor Nodes Enable a New Generation of Machinery Diagnostics and Prognostics 405
F. M. Discenzo, D. Chung, A Twarowski and K. A. Loparo

SensWeb: A Wireless Self-Organized Cooperative Sensor Network Topology 415
J. N. Schoess and S. Menon

INDUSTRIAL CASE HISTORIES

Industrial Case Histories 423
K. R. Guy

Troubleshooting Vibration Problems – A Compilation of Case Histories 467
N. L. Baxter

AUTHOR INDEX 485

PREFACE

The 55th Meeting of the Society for Machinery Failure Prevention Technology (MFPT) was held April 2-5, 2001 in Virginia Beach, Virginia. MFPT sponsors include the Office of Naval Research, U.S. Army Research Laboratory, National Aeronautics and Space Administration and the Naval Surface Warfare Center. The cooperation of several professional organizations with mutual interests is acknowledged. **Mr. Carl S. Byington** from the Applied Research Laboratory at Pennsylvania State University was Technical Program Chairman and Chair of the Opening Session.

The Keynote Speaker was **Mr. Patrick Stevens**, Program Manager with the U. S. Army Aviation and Missile Command at Redstone Arsenal, Alabama. He presented an Overview of the Army Diagnostics Improvement Program. Other distinguished invited speakers in the Opening Session were **Squadron Leader Richard Friend (RAF)**, assigned to the Air Force Research Laboratory at WPAFB in Ohio, speaking on *Turbine Engine Research in the United States Air Force*; **Mr. Thomas McCloskey** from the Electric Power Research Institute on *Failure Prevention and Reliability Improvements of Turbomachinery in the Power Generation Industry*; and **Dr. Howard A. Gaberson** whose paper, *Vibratory Locomotion Revisited*, is published in these proceedings.

On the first afternoon the ever popular **New Products and Services** session was held. This session allows ten minutes for each exhibitor to describe his service or product. This is an additional venue for the exhibitors to interact with the other attendees. Since the exhibits are an integral part of the conference, such a session is very useful. **Mr. Kenneth P. Maynard** of Penn State/ARL presented an excellent half-day Tutorial on *Feature Extraction for Smart Sensors.* Numerous industrial case histories were presented on the last day of the meeting. I am grateful to **Nelson L. Baxter** and **Kevin R. Guy** for accepting this task. Note that their case histories are published in these proceedings.

I offer my sincere appreciation to my wife Sallie for her continuing work above and beyond the call of duty. Thanks to Carl Byington and the Program Committee for a job well done and to all speakers and session chairs for their contribution to another successful MFPT meeting.

Henry C. Pusey
Executive Director
MFPT Society

MFPT SOCIETY BOARD OF DIRECTORS
(The Society for Machinery Failure Prevention Technology is a Division of the Vibration Institute.)
OFFICERS

Chair
Michael J. Roemer, Ph.D.
Impact Technologies, LLC
125 Tech Park Drive
Rochester, NY 14623

Vice Chair
Victor K. Champagne
U.S. Army Research Laboratory
Weapons & Materials Research Directorate
Bldg. 4600
APG, MD 21005-5069

Secretary
MarjorieAnn E. Natishan, Ph.D.
University of Maryland
Mechanical Engineering Dept..
College Park, MD 20742

Treasurer
Rudolph J. Scavuzzo, Ph.D.
University of Akron
College of Polymer Science, Bldg. 329
Akron, OH 44325-3909

STANDING COMMITTEE CHAIRS
Finance
Rudolph J. Scavuzzo, Ph.D.
University of Akron
College of Polymer Science, Bldg. 329
Akron, OH 44325-3909

Membership
Marc Pepi
U.S. Army Research Laboratory
Weapons & Materials Research Directorate
Bldg. 4600
APG, MD 21005-5069

Programs
G. William Nickerson
Oceanor Sensor Technologies, Inc.
1346 South Atherton Street
State College, PA 16801

Research
Kam W. Ng, Ph.D.
Office of Naval Research
Code 333
800 N. Quincy Street
Arlington, VA 22217

TECHNICAL COMMITTEE CHAIRS
Diagnostics & Prognostics
Howard A. Gaberson, Ph.D.
NAVFAC Engrg Service Center
Code 231, 1100 23rd Avenue
Port Hueneme, CA 93043-4370

Failure Analysis
Victor K. Champagne
U.S. Army Research Laboratory
Weapons & Materials Research Directorate
Bldg. 4600
APG, MD 21005-5069

Distributed System Architecture
Jeffrey N. Schoess
Honeywell Technology Center
MN65-2600 3660 Technology Drive
Minneapolis, MN 55418

Sensors Technology
Henry R. Hegner
Magyar & Associates Inc.
812 Lakemount Drive
Moneta, VA 24121

Education & Training
John P. H. Steele, PH.D
Colorado School of Mines
Division of Engineering
Golden, CO 80401-1887

Members-at-Large
Paul L. Howard
Paul L. Howard Enterprises
PO Box 362
Newmarket, NH 03857

John C. Pooley, III
AMTEC Corporation
Suite 314
500 Wynn Drive
Huntsville, AL 35816

Carlos M. Talbott, DSc.
Talbott & Associates
641 Nor Oak Court
West Chicago, IL 60185

FEATURED PAPER

Invited Speaker

VIBRATORY LOCOMOTION REVISITED

Howard A. Gaberson, Ph.D., Oxnard, California
Philip L. Stone, Santa Barbara, California
John B. Curry, Oxnard, California
Robert S. Chapler, Oxnard, California

Abstract: Vibratory Locomotion is an old unused method of moving over terrain we invented at the Naval Civil Engineering Laboratory over 25 years ago. The patent [1] has expired. We had great hopes for it, but they never materialized. It is being presented here to remind readers it exists in hope that someone will find an application for it. The paper describes some applications, presents a simplified design method for the devices, and discusses the effectiveness of several vibratory locomotion prototypes we built and tested.

Key Words: Reciprocators; oscillator; skid; vibratory locomotion

INTRODUCTION:

The method uses a reciprocating weight to cause an object to incrementally slide or shuffle over the ground surface. It becomes perfectly reasonable to mount the weight inside a box, and have the box shuffle over the ground surface, and equally reasonable to put the reciprocating weight inside a boat and have the boat shuffle across the beach. An oscillating mass can be fixed to skids in place of the tracks on a bulldozer and make a different kind of a tractor. The peak drawbar pull of such a tractor is twice the product of its weight and local coefficient of friction. Our work demonstrated all of this and provided a detailed theoretical analysis that proved it all had to be true.

First of all, vibratory locomotion is a method for accomplishing land locomotion by causing a mass to reciprocate, back and forth, in a straight line that is inclined horizontally. To visualize the concept, imagine a skid that contains machinery that can reciprocate a heavy weight back and forth. The path of the weight's motion, viewed from aboard the skid, is a straight line inclined at 45 degrees for example, the path is such that the weight moves up and forward, down and backward as shown in Figure 1. When shaken at appropriate amplitude, the weight provides reaction forces on the skid that lifts and slides it along the ground. Specifically, when the weight is at the top of its stroke it lifts the skid and slides it forward; at the bottom of the stroke it is pushing downward and backward on the skid, but since the downward force increases friction, no back sliding takes place. The net result is a forward shuffling motion of the skid. Control of the skid can be accomplished by using two reciprocators, one on each side, and controlling the forward thrust of each. If one reciprocator is thrusting forward and the other aft, the skid can pivot about its center. One application considered was to propel a large solid concrete barge over a road for mine clearing. Figure 2 is a conceptual drawing of this idea.

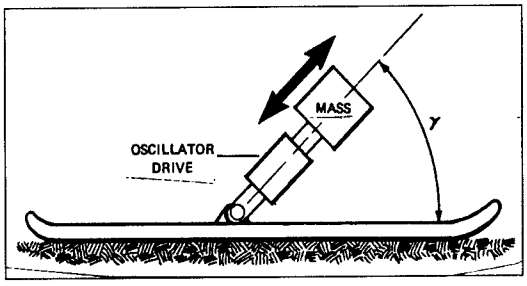


Figure 1. The vibratory locomotion concept.

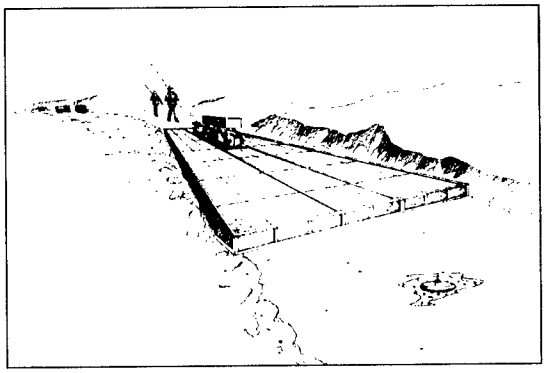


Figure 2. A massive concrete barge for mine clearing propelled by vibratory locomotion.

DEVICES BUILT AND TESTED:

After we completed the theoretical analysis, which we'll discuss later, we designed and built several prototype models to test the locomotion and drawbar pull capabilities. The first was a rocker crank oscillator with a 100-pound weight shown in Figure 3. It worked well and was the test skid used to provide the data for the published theoretical study [2]. This skid was tested in the arctic at Point Barrow and performed quite well. We tested several bottom configurations and the smooth bottom worked best. A compressed air bin shaker vibrator shown in Figure 4 also powered this small skid. The idea of a heavy piston vibrating inside a cylinder made the concept very compact and safe, especially compared to counter rotating eccentrics which are convenient but dangerous.

Figure 3. The crank rocker oscillator on the small skid with various cleat bottom arrangements for snow testing.

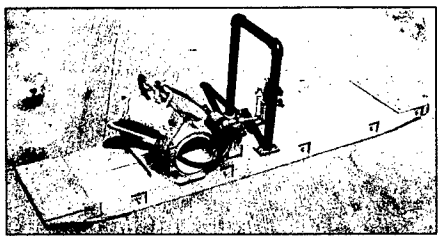


Figure 4. A small skid with a compressed air vibrator to accomplish vibratory locomotion.

We also built a large skid with a spring-supported platform (Figure 5). It was first powered by a resonant spring oscillator. A hydraulic motor rotated an eccentric weight to excite the resonant vibration, which smoothly propelled the large skid. Unlike the rocker crank oscillator it was easy to change the shake angle of the oscillator.

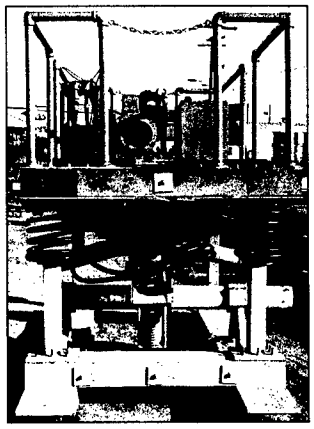


Figure 5. The large skid with a spring supported platform and a resonant spring oscillator. The engine powered a hydraulic pump to energize the hydraulic motor exciting the vibrating mass.

We also powered the big skid with our most versatile oscillator, a concentric shaft, counter-rotating eccentric oscillator with a phase or shake angle changer. A car differential was used to change the phase or shake angle (Figure 6). The oscillator on the skid is shown in Figure 7. The skid could climb modest hills and could tow a half-ton Navy pickup truck with its wheels locked. Once while touring the Navy Lab, a group of about 15 children was invited to climb aboard the skid for a ride, which they thoroughly enjoyed. The ungainly big skid made the 11 o'clock news nationwide one night. It

always attracted attention as we drove it around the compound. Even though it had only one oscillator it could be slowly steered or turned by shifting your weight to the desired turning direction.

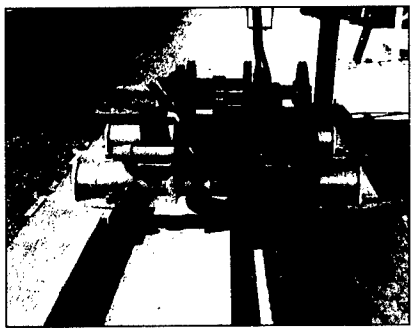


Figure 6. The concentric shaft counter rotating oscillator. The center weight turns in the opposite of the two outer weights. The drive shaft coming out of the differential is for changing the shake angle.

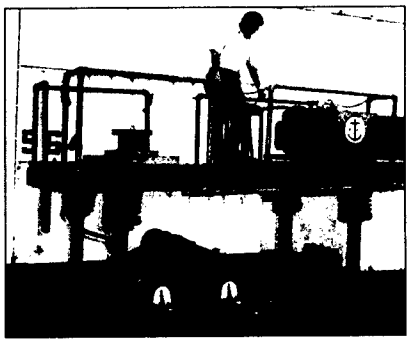


Figure 7. The large skid with the concentric counter rotating oscillator

We called one of the uses we proposed for the technology, the Beach and Launch Unit. The concept was to provide Marine landing craft with capabilities of assured satisfactory beaching, subsequent relaunch, limited land locomotion, and broach recovery. We were able to demonstrate the first three. Anti-broach capability was to be provided by two independently controllable reciprocators, mounted outboard on the boat. We proposed to develop a large free piston engine (Figure 8) to power a landing craft up the beach and back it down back into the surf. A drawing of the free piston engines installed in the boat is shown in Figure 9. To develop this amphibious use, we mounted our concentric counter-rotating oscillator in a small Marine Corps Logistic Support Boat. The boat was 20 ft long and 7-1/2 feet wide; it weighed 1,350 lbs. The bottom is double-v shaped and the deck was flat; a substantial foam filled cavity laid between the deck and the hull bottom. We dug a small pond and lined it with plastic for testing; a beach was at one end. Figure 10 shows the boat coming out of the pond. Our third author became proficient driving that boat in and out of the pond at will. To demonstrate the beauty of the free piston engine concept, we mounted our air vibrator in a smaller boat and our third author is shown driving that boat out of the water in Figure 11. Testing in the actual surf didn't work as well, and the air cushion vehicle came along and solved the problem better than we could. We ran out of funds before we could master the technique.

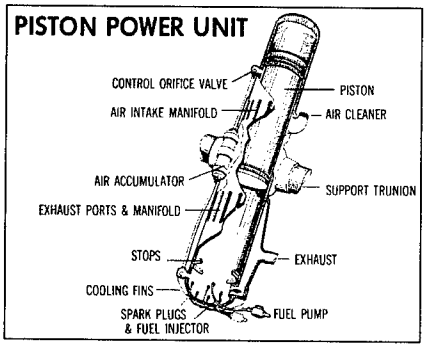


Figure 8. The free piston engine concept for use with the Beach and Launch Unit and the bulldozer thrust doubler.

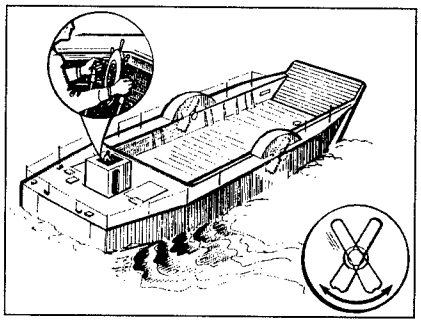


Figure 9. The vibratory locomotion free piston engines installed in a landing craft.

Figure 10. The vibratory locomotion boat emerging from the pond up onto the beach.

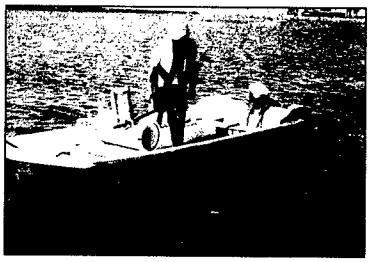


Figure 11. The compressed air vibrator installed in a small boat climbing up the beach. The long rod held by the operator is used to change the shake angle.

The final problem we attacked with our solution was doubling the drawbar of a bulldozer. A tractor can only pull with a force equal to the product of its weight and the coefficient of friction. It is easy to show that the peak pulling force of a vibratory locomotion vehicle is twice this value. Figure 12 shows what we believe to be the largest concentric counter rotating weights ever built. They could shuffle that 12,500-pound tractor through the dirt and definitely pull with a peak force twice its weight. Figure 13 shows a close up of the small weights, which were actually used to document the thrust doubling. Again, we had hoped to be able to develop the free piston oscillator for use on the Doubler in place of the dangerous counter rotating weights.

Figure 14 shows the artist's concept drawing we used to try to convince our sponsors to procede.

Figure 12. A drawbar pull test of the bulldozer with the largest set of weights.

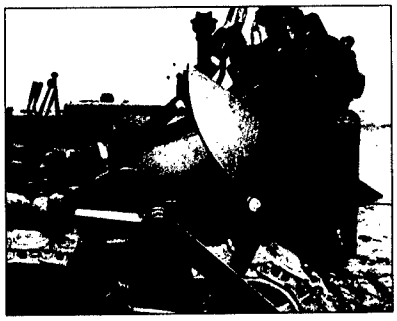


Figure 13. A close up of the small weights on the bulldozer.

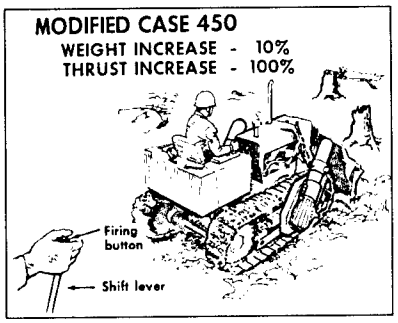


Figure 14. Concept drawing of free piston engines on a bulldozer to accomplish thrust doubling.

THE THEORY SUMMARY:

A theoretical explanation of the solution of the piece-wise linear differential equations involved in vibratory locomotion is given in Reference [2]. The solutions involved stability and had to be computed for a wide variety of non dimensional operating conditions. The results of those computations are given in a solution value map that yields the nondimensionalized step size or the net forward advance per cycle of mass oscillation for all anticipated operating conditions. Conceptually, everything but the vibrating mass is considered attached to a skid of mass m_1. The skid can slide over a terrain inclined to the horizontal amount, β, with a coefficient of friction, μ. A mass m_2 is vibrated sinusoidally with amplitude, a, and frequency ω in radians per unit time, in a straight line inclined to the skid at angle, α, as shown in Figure 15. The motion of m_2 with respect to m_1 is taken to be:

$$z = a\sin\omega t \qquad (1)$$

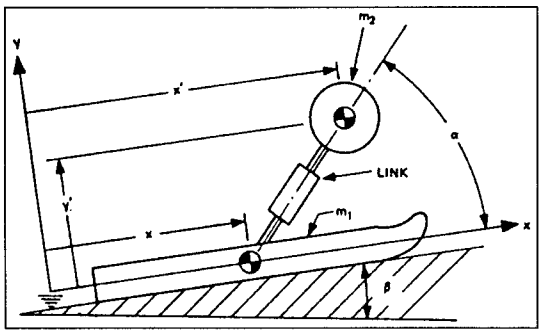


Figure 15. The theoretical model, the coordinate system and the angles.

The intensity or vibration amplitude, A, is given by:

$$A = \frac{a\omega^2 \sin\alpha}{g\cos\beta} \tag{2}$$

where, g, is the acceleration of gravity. The relative mass, M, is given by:

$$M = \frac{m_2}{m_1 + m_2} = \frac{w_2}{w_1 + w_2} \tag{3}$$

Shuffling mode vibratory locomotion takes place when the following two conditions are met:

$$MA \leq 1.0 \tag{4a}$$

$$\phi MA > 1.0 \tag{4b}$$

where:

$$\phi = \frac{\mu + \cot\alpha}{\mu + \tan\beta} \tag{5}$$

When m_2 is vibrated such that MA >1., small flights occur once per cycle so long as:

$$MA > \sqrt{\pi^2 + 1} \approx 3.297 \tag{6}$$

Beyond this limit the motion cannot be once per rev periodic.

The value of MA, if greater than one, yields three times of flight and impact [2]. It distinguishes a compactor from a vibratory locomotion vehicle. Compactors require a flight to develop an impact and thus must operate such that the flight occupies a substantial portion of the cycle; minimum values of MA are about 1.5 for compactors, with most operating close to MA= 3.0. In contrast, vibratory locomotion vehicles are not built to suffer such impacts; in fact they operate 90% of the time at values of MA less than unity. The highest drawbar pull is obtained with MA < 1. Thus devices with MA > 1 are compactors and devices with MA < 1.1 are the subject of vibratory locomotion.

To compute to the cyclic advance for any set of operating conditions from the solutions map in Reference [2] you need:

$$\psi = (\mu + \tan\beta)\sin\alpha \tag{7}$$

Then for values of MA and ϕ, the design chart gives the value of $S/(\psi M)$. From the computed values of ψ and M, calculate s, where:

$$s = aS \tag{8}$$

and "s" is the actual net displacement for each cycle of oscillation. The average velocity will be the product of the step size, s, and the frequency in cycles rather than radians per unit time, then:

$$v = sf, \tag{9}$$

SIMPLIFIED DESIGN:

If a device works on level ground, it will easily go down hill, and will climb uphill to a certain extent, so at first, we only consider level operation. You will have to fabricate a more complicated oscillator that can conveniently vary its shake angle to be able to climb uphill better. The hardware we built took such a beating when we "flew" it, that we seldom ran it that hard. Therefore, we only designed it to a maximum condition of MA = 1, which means that at the peak of the stroke, the weights are just lifting the full vehicle weight. Since this occurs for just an instant, no flight occurs. We make one further simplification for design; the theory gives a minimum shake angle for which no back slide can occur. This is probably the most efficient shake angle, for no power is wasted in backward sliding and yet the shake direction is leaning forward as much as possible to tend to the largest step possible without any back slide. Given these conditions for design (level terrain or $\beta = 0$, minimum shake angle for no back slide, MA = 1), the simplified procedes as follows. With MA = 1, and $\beta = 0$, Equation (27a) from Reference [1] gives the limiting condition for no back slide to be $\phi = 3$. Using this value and $\beta = 0$, in Equation (5), the shake angle must be:

$$\tan\alpha = \frac{1}{2\mu} \tag{10}$$

Proceeding to the design chart in Reference [2], the design parameter $S/\Psi M$, is obtained for the values $\phi = 3$, and $MA = 1$, to be:

$$\frac{S}{\Psi M} = 7.11 \tag{11}$$

S and Ψ are (for $\beta = 0$)

$$S = \frac{s}{a}, \text{ and} \tag{12}$$

$$\Psi = \mu \sin\alpha \tag{13}$$

where "s" is the length of a single step.

Using Equations (12), and (13) in (11) yields:

$$s = 7.11 M a \mu \sin\alpha \tag{14}$$

For MA equal unity, and $\beta = 0$, Equation (2) yields:

$$M = \frac{g}{a\omega^2 \sin\alpha} \tag{15}$$

Frequency in cycles per unit time is related to ω by $2\pi f = \omega$; using this and Equation (15) in (14) we finally obtain:

$$s = 7.11 \frac{\mu g}{(2\pi f)^2} \tag{16}$$

Taking $g = 32.17$ ft/sec^2 this becomes:

$$s = 69.53 \frac{\mu}{f^2} \tag{17}$$

where "s" is in inches and "f" is in Hz (cycles per sec.). For a surprisingly common number of situations, 0.5 is a good value to take for the coefficient of friction (e.g., timbers on sand, loose earth, wood on pavement), and for this case (10) becomes:

$$s = \frac{34.77}{f^2} \tag{18}$$

where again "s" is in inches and f is in Hertz. The velocity of the skid is the step size times the number of steps per unit time or the frequency, thus Equation (17) yields

$$v = 69.53\frac{\mu}{f}, \tag{19}$$

and Equation (18) becomes:

$$v = 34.77/f \tag{20}$$

where, once more, "v" is in inches per second and "f" in Hz. The above is a striking result; the step, size only depends on frequency. The lower the frequency the larger the step size and the velocity. Unfortunately very low frequencies cost a great deal and high frequencies are cheap.

So, the way you design is by selecting a step size, velocity and frequency from Equations (17) or (18). Using the design friction coefficient, the shake angle is selected from Equation (3), and then a trade-off between heavy weights, m_2 and a short shake amplitude, a, ensues. Sometimes the resulting oscillator size is too big for the skid to be moved, so you relax your requirement for so great a velocity, increase the frequency, and try again.

Only a few power calculations have been made, but these indicate 60 to 75% of the theoretical power to drag the skid at the design velocity. Therefore, we suggest that, since one does not want to attempt hardware with insufficient power that a good design criteria is to have the full theoretical power available to drive the weights. The theoretical power is the product of the force and the velocity; the force is the product of the coefficient of friction and the total weight, W_t, and the velocity is given by Equation (12), thus:

$$P = 69.53\frac{\mu^2 W_t}{f} \tag{21}$$

where "P" is in in-lbs/.sec, W_t in lbs, and "f" in Hz. Converting the above to the units, Hp, gives:

$$P = 0.0153\frac{\mu^2 W_t}{f} \tag{22}$$

When μ has the common value 0.5, the above becomes:

$$P = 0.002634\frac{W_t}{f} \tag{23}$$

where W_t is in lbs and "f" is in Hz, and "P" is in horsepower.

The above summarizes the simplified design. It is interesting to use the above to design an oscillator for the LCM-8. The LCM-8 is a 74 ft long landing craft that weights 250,000 lbs fully loaded; it is powered by two 325 Hp diesel engines. Assume we want it to crawl across the beach with 2.5 inch steps. Taking the friction coefficient to be 0.5, Equation (17) yields the frequency to be 3.75 Hz. With $\mu = 0.5$, Equation (10) gives the shake angle to be 45°: Equation (20) yields the velocity to be 9.27 inches per second. Equation (23), indicates that the power required to drive the weights will be about 176 Hp. The hull of the boat is about 9 ft deep with quite straight sides. Assume we might fit two sets of counter-rotating weights, one on each side, about amidships, each set having two weights as we did with the doubler. The distance, A, is the distance from the axis of the weight out to its center of gravity; this is assumed to be 40 inches in Equation (2) which yields an "A" of 40.7. The product MA must equal unity, so M = 0.0246. Substituting this value into Equation (3), with m_1 equal to 250,000 lbs one computes m_2 equal to 6,297 lbs. Since there are to be four weights, each must weigh 1,574 lbs. Such a weight with a 4-ft outside radius can easily be cut from 4-inch steel plate.

There would still be many design problems to solve. The weights would have to be synchronized because if one lifts while the other pushes down, nothing happens. The two weights synchronized would not permit any directional control. The only present conception is by varying the shake angle of one of the two sets; one could be pulling forward while the other was pulling backward. As can be imagined, such large weights have a good deal of energy stored in them when up to speed; it would take a great deal of power to bring them up to speed fast and thereby offer quick response. It is our feeling that counter-rotating weights could be developed into an acceptable system, but we are unsure. That the boat would come out of the water and walk on the sand, there is no question; the only question would be concerning the clumsiness and responsiveness of such a system. It would have to be built and tried for an accurate answer.

CONCLUSION:

To our knowledge, no one, with the exception of us, has ever built any of these. Two articles were published attempting to extend our analyses [3], [4]. The first author of this paper has copies of Reference [2] and can provide a limited number of them. The Technical Information Center at the Naval Facilities Engineering Service Center, 1100 23rd Avenue, Port Hueneme, California, 93043-4370, can provide copies of References [5] and [6]. If you consider applying the technology and have questions, the first author can be reached at hagaberson@att.net

ACKNOWLEDGEMENT:

Those marvelous drawings of our concepts for applying vibratory locomotion were drawn by Dan Nunez of Oxnard, California, an artist now retired like the rest of the authors from the Naval Civil Engineering Laboratory.

REFERENCES:

1. Gaberson, H.A., "Vibratory Locomotion Means", U.S. Patent 3,916,704; issued 4 November 1975.

2. "Vibratory Locomotion," H. A. Gaberson and P. L. Stone 1974; Journal of Engineering for Industry, Transaction ASME v 96 Ser B n 2 May 1974 p 644-652. Also published as NCEL Technical Note N-1292, "Vibratory Locomotion," July 1973

3. Brower, W. B., "Analysis of the Vibra-Lo," J. Appl. Mech. ASME Series E. 40, 1138 (1973)

4. Sharma, R. S., "Analysis of the Vibra-Lo," Mechanism and Machine Theory, 1978, Vol 13, pp 109-212. Pergamon Press; Great Britain

5. "Vibratory Locomotion for Landing Craft," CEL Interim Report 63-76-4, July 1975.

6. "Doubling the Drawbar of Marine Corps Bulldozers," CEL Technical-Note N-1444, July 1976.

DIAGNOSTICS

Chair: Mr. Mark L. Hollins
Naval Air Warfare Center/AD

GEAR FATIGUE CRACK DIAGNOSIS BY VIBRATION ANALYSIS USING EMBEDDED MODELING

C. James Li and Hyungdae Lee

Dept. of Mechanical Engineering, Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
110 8th St., Troy, NY 12180
E-mail: lic3@rpi.edu

Abstract: This paper describes an embedded modeling methodology for identifying gear meshing stiffness from measured gear angular displacement or transmission error. An embedded model integrating a physical based model of the gearbox and a parametric representation, in the form of truncated Fourier series, of meshing stiffness is established. A solution method is then used to find the meshing stiffness that minimizes the discrepancy between model output and the measured output. Both simulation and experimental studies were conducted to evaluate if identified tooth meshing stiffness can reveal a tooth crack more effectively.

Key words: Crack; Embedded model; Gear diagnosis; Meshing stiffness; Torsional vibration; Transmission error.

INTRODUCTION

Figure 1 shows a dynamic model of a single-stage gearbox where θ represents angular displacement, I represents inertia, T represents torque, and k_m represents meshing stiffness between the two gears. Even if the gears are perfect, the meshing stiffness varies periodically as the number of teeth in contact and the contact point change. This non-constant meshing stiffness becomes a parametric excitation for gear vibration during gear rotation. Localized gear faults such as cracked or fractured teeth, that affect only a few teeth, introduce variations in meshing stiffness which, in turn, increase amplitude and phase modulations in gear vibrations.

To detect and assess tooth crack, existing gear fault diagnostic algorithms accentuate or extract magnitude and/or phase modulations from gear vibrations. Dalpiaz et al. [1] gave references for a number of early and modern gear diagnostic algorithms and Choy et al. [2] gave references on Wigner-Ville Distribution (WVD) and some statistical based methods including FM4, NA4 and NB4.

There are limitations for vibration based gear diagnostic methods. Vibrations are secondary effects in the sense that they are dynamic responses of a gearbox excited by meshing stiffness and other excitations. The effect of irregular meshing stiffness

associated with a cracked tooth is filtered by the gearbox dynamics and contaminated by other vibrations. Even when a cracked tooth is revealed by detectable amplitude/phase modulations, there is no straightforward relationship between the modulation amplitude and the crack size which is necessary for severity assessment and residual life prognosis.

In order to alleviate these limitations, this study has developed a method to identify meshing stiffness from measured gear vibration.

PROPOSED METHOD

The goal of this study is to establish the utility of embedded modeling method [3] to identify the meshing stiffness which is more directly related to the fault than vibration. A summary of the embedded modeling method is provided below.

Embedded modeling method: Let's say that the dynamics of a real system is described as

$$\begin{aligned} \dot{x} &= F(x, f) \\ s &= h(x) \end{aligned} \tag{1}$$

where x is the system state vector, f is the forcing function, s, also a vector, stands for system physical variables that are measured by transducers, h stands for the function relating the state and the measurable, s. It is also assumed that x, and s are n, and m dimensional, respectively. Let's further assume that the model of the system takes the following form

$$\begin{aligned} \dot{\hat{x}} &= \hat{F}(\hat{x}, f; \omega) \\ \hat{s} &= h(\hat{x}) \end{aligned} \tag{2}$$

where the circumflex stands for approximation, ω (p-dimensional) stands for the vector of model parameters that, at least, is partially unknown. Note that h is assumed to be known here to simplify the discussion.

One can show that model output $\hat{s}$ will approach the real system output, s if $\hat{F}$ approaches F provided the initial error is small. Therefore, it is natural to find the $\hat{F}$ that can minimize the difference between system and model outputs, i.e.,

$$\begin{aligned} E &= \frac{1}{2}\int_{t_0}^{t_f} e^T e dt \ (\textit{continuous}) \quad \textit{or} \\ E &= \frac{1}{2}\sum_{i=1}^{M} e_i^T e_i \ (\textit{discrete}) \end{aligned} \tag{3}$$

where $e = s - \hat{s}$, and t_0 and t_f define the interval. Equation (3) is the so-called objective function.

Using nonlinear programming, Equation (3) can be minimized by an iterative procedure starting from an initial guess

$$\omega^{k+1} = \omega^k - \alpha R^k g^k \tag{4}$$

where g^k is the gradient of E at ω^k, R^k is a positive-definite square matrix and α defines the step size. The product of g^k and R^k gives the search direction. Using different R matrix yields different gradient-based updating schemes such as steepest descend and Newton's method with different efficiency, robustness and computation cost. This study used Levenberg-Marquardt method for its robustness.

To obtain the gradient one takes derivative of E with respect to ω which yields

$$g = \sum_{i=1}^{M} e_i^T \frac{\partial e_i}{\partial \omega} = -\sum_{i=1}^{M} \left(\frac{\partial \hat{s}_i}{\partial \omega} \right)^T (s_i - \hat{s}_i) \tag{5}$$

Note from above equation that gradient g is calculated from e and an $m \times p$ Jacobian matrix that contains the partial derivative of $\hat{s}$ with respect to ω. The Jacobian is denoted as J hereinafter. Taking derivative of $\hat{s}$ with respect to ω yields J as

$$J = \frac{\partial \hat{s}}{\partial \omega} = \left(\frac{dh}{d\hat{x}} \right)^T \frac{\partial \hat{x}}{\partial \omega} \tag{6}$$

However, the partial derivative of $\hat{x}$ with respect to ω is not readily available. Let's take partial derivative of the model or Eq. (2) with respect to ω to obtain the so-called sensitivity equation

$$\begin{aligned} & d\left(\frac{\partial \hat{x}}{\partial \omega} \right) / dt = \frac{\partial \hat{F}}{\partial \hat{x}} \frac{\partial \hat{x}}{\partial \omega} + \frac{\partial \hat{F}}{\partial \omega} or \\ & \dot{\xi} = \hat{F}_x' \xi + \hat{F}_\omega' \end{aligned} \tag{7}$$

where $\xi = \partial \hat{x} / \partial \omega$; $\hat{F}_x'$ and $\hat{F}_\omega'$ are the partial derivatives of $\hat{F}$ with respect to x and ω, respectively. Solution of Eq. (7) is the needed $\frac{\partial \hat{x}}{\partial \omega}$ which has an initial value as a zero matrix because the initial condition, $\hat{x}(t_0) = \hat{x}_0$, does not depend on the parameters of the model.

Formulating an embedded model for a gearbox : The model of a simple spur-gear transmission can be represented by a collection of masses, springs, and dampers as in

Figure 2 [4].

The non-linear equation of motion can be written as:

$$
\begin{aligned}
&J_M\ddot{\theta}_M + C_{s1}(\dot{\theta}_M - \dot{\theta}_1) + K_{s1}(\theta_M - \theta_1) = T_M \\
&J_1\ddot{\theta}_1 + C_{s1}(\dot{\theta}_1 - \dot{\theta}_M) + K_{s1}(\theta_1 - \theta_M) + C_g(t)(R_{b1}(R_{b1}\dot{\theta}_1 - R_{b2}\dot{\theta}_2)) + K_g(t)(R_{b1}(R_{b1}\theta_1 - R_{b2}\theta_2)) = T_{f1}(t) \\
&J_2\ddot{\theta}_2 + C_{s2}(\dot{\theta}_2 - \dot{\theta}_L) + K_{s2}(\theta_2 - \theta_L) + C_g(t)(R_{b2}(R_{b2}\dot{\theta}_2 - R_{b1}\dot{\theta}_1)) + K_g(t)(R_{b2}(R_{b2}\theta_2 - R_{b1}\theta_1)) = T_{f2}(t) \\
&J_L\ddot{\theta}_L + C_{s2}(\dot{\theta}_L - \dot{\theta}_2) + K_{s2}(\theta_L - \theta_2) = -T_L
\end{aligned}
\tag{8}
$$

where θ_M, θ_1, θ_2, and θ_L represent the rotations of the motor, the pinion and the gear, and the load; J_M, J_1, J_2, and J_L represent the mass moments of inertia; C_{s1}, C_{s2}, and C_g are damping coefficients of the shafts and the gear mesh; K_{s1}, K_{s2}, and K_g are stiffness of the shafts and the gear mesh; T_M, T_L, $T_{f1}(t)$, and $T_{f2}(t)$ are motor and load torques and frictional torques on the gears; R_{b1} and R_{b2} are base circle radii of the gears.

The meshing stiffness, i.e., K_g is assumed to be periodic. For a good gear that has regularly-spaced identical teeth, the meshing stiffness is largely repeated from one tooth to the next. In this case, a truncated Fourier series with a fundamental frequency of tooth meshing is used to represent K_g. (Note that Eq. (9) is formulated as a function of time. While it is more accurate for it to be a function of angular position, it is more convenient to have it as a function of time along which the gearbox dynamics evolves.)

$$
K_g(t) = \frac{a_0}{2} + \sum_{n=1}^{N}\left(a_n \cos(2\pi n f_m t) + b_n \sin(2\pi n f_m t)\right) \tag{9}
$$

where N is the number of harmonics included.

On the other hand, a faulty tooth gives a meshing pattern that is repeated only once a revolution. Therefore, a Fourier series with a fundamental frequency as the rotation frequency is more suitable.

$$
K_g(t) = (1 + \sum_{k=1}^{K}[c_k \cos(2\pi k f_s t) + d_k \sin(2\pi k f_s t)])K_g(t) \tag{10}
$$

where K is the number of harmonics considered. Equation (9) can be considered as a special case of Eq. (10) which can be used for both good and faulty gears. However, to cover the same number of meshing harmonics, Eq. (10) uses more terms than Eq. (9). This means more unknowns and therefore higher computational cost.

The damping, C_g, is assumed to be a function of the meshing stiffness

$$
C_g(t) = 2\zeta\sqrt{K_g(t)R_{b1}^{\ 2}R_{b2}^{\ 2}J_1J_2/(R_{b1}^{\ 2}J_1 + R_{b2}^{\ 2}J_2)} \tag{11}
$$

where ζ is the damping ratio [4].

To identify gear meshing stiffness from gear vibration, one takes the gearbox model, i.e., equations (8) – (11), and solves for Fourier coefficients a_i and b_i or c_i and d_i in the way ω_i is solved as described in the last section.

SIMULATION STUDY

A gearbox dynamic simulator, DANST (Dynamic Analysis of Spur Gear Transmissions) was extended to simulate a good gearbox and another with a cracked tooth. Given inputs such as geometry of the tooth and crack, a Finite Element Method (FEM) program is first used to compute individual tooth stiffness. Then, DANST computes meshing stiffness by superimposing the stiffness of teeth according to gear meshing kinematics. Using numerical methods, the gearbox governing equation (8) is then solved for a complete rotation of the gear repeatedly until the initial condition that results in an identical terminal condition is found.

Broken lines in Figure 3 show the meshing stiffness and its corresponding transmission error of a good gearbox calculated by the DANST. Figure 4 shows the same for a gearbox with a crack in tooth 13.

To identify the meshing stiffness from the transmission error of the simulated gearbox, the gearbox model Eq. (8) is formulated into the state space form (2).

$$\dot{x} = F(x,u;\omega) = \begin{bmatrix} x_2 \\ (T_M - C_{s1}(x_2 - x_4) - K_{s1}(x_1 - x_3))/J_M \\ x_4 \\ a_{41} \\ x_6 \\ a_{61} \\ x_7 \\ (-T_L - C_{s2}(x_8 - x_6) - K_{s2}(x_7 - x_5))/J_L \end{bmatrix}$$

$$a_{41} = [T_{f1}(t) - C_{s1}(x_4 - x_2) - K_{s1}(x_3 - x_1) - C_g(t)(R_{b1}(R_{b1}x_4 - R_{b2}x_6)) - K_g(t)(R_{b1}(R_{b1}x_3 - R_{b2}x_5))]/J_1$$

$$a_{61} = [T_{f2}(t) - C_{s2}(x_6 - x_8) - K_{s2}(x_5 - x_7) - C_g(t)(R_{b2}(R_{b2}x_6 - R_{b1}x_4)) - K_g(t)(R_{b2}(R_{b2}x_5 - R_{b1}x_3))]/J_2$$

$$\hat{s} = h(x) = R_{b1}x_3 - R_{b2}x_5 \qquad (12)$$

where $x = [\dot{\theta}_M, \ddot{\theta}_M, \dot{\theta}_1, \ddot{\theta}_1, \dot{\theta}_2, \ddot{\theta}_2, \dot{\theta}_L, \ddot{\theta}_L]$, u is the external excitation, and K_g takes the form of (9) or (10) depending on if a tooth crack is suspected (or one can use equation (10) all the time at the expense of higher computational cost in the case of a good gear.)

The sensitivity equations for (12) are then derived so that gradients can be calculated to carry out a search for optimal value of Fourier coefficients. Solid lines in Figures 3 and 4 show the meshing stiffness found by the embedded modeling method for good and cracked gears, respectively. The 5X errors are shown in the bottom of each figure. It is obvious that the meshing stiffness is accurately identified in both cases.

EXPERIMENTAL STUDY

Figure 5 shows the gear test rig. It is consisted of a 40 HP variable speed motor and a 75 HP generator between which a testing gearbox is installed. Transducers are available to measure vibration, input torque, and transmission error with a resolution of $7*10^{-5}$ rad. Additionally, crack gauges and bore scopes are installed to track the evolution of gear faults such as tooth crack and pitting. The 10 HP single-stage spur gearbox used in this study contains a pinion of 23 teeth, and a gear of 54 teeth. The nominal pinion speed is 450rpm and the maximum input torque is 920in-lb. In a test, a small notch was made with wired electrical discharge machining at the root of tooth 17 to create a stress concentration which eventually led to a propagating crack. The transmission errors measured when the tooth is healthy and when it has a crack of 0.18 inches are shown in Figure 6 in solid lines. Notch filtering is then applied to remove the once-per-revolution frequency due to eccentricity.

Figures 7b and 7d show the meshing stiffness identified by the proposed method and their corresponding transmission error in Figures 7a and 7c along with the measured transmission error. When the gear is good, the identified meshing stiffness (7b) is roughly repeated from tooth to tooth as expected. Because of the cracked tooth, the transmission error in Figure 7c has a clear transient around 1,100 along the horizontal axis. This transient becomes even more pronounced in the meshing stiffness. Such increased sensitivity will enable one to detect the crack at an early stage.

CONCLUSIONS

This paper describes the development of an embedded model for a gearbox, which, in turn, enabled the identification of gear meshing stiffness from its vibration. When applied to a simulated gearbox, the method resulted in accurate identification of meshing stiffness which gives an accurate account of the state of gear. When applied to real data taken from a gear test, the meshing stiffness appears to be more sensitive to the crack than the vibration. In addition to increased sensitivity, the ability to identify meshing stiffness makes it possible to determine the physical size of a crack which is needed for severity assessment and residual life prediction.

REFERENCES

1. Dalpiaz, G., Rivola, A. and Rubine, R., 2000, "Effectiveness and Sensitivity of Vibration Processing Techniques for Local Fault Detection in Gears," Mechanical Systems and Signal Processing, 14(3), pp. 387-412.
2. Choy, F. K., Braun, M. J., Polyshchuk, V., Zakrajsek, J. J., Townsend, D. P., and Handschuh, R. F., 1994, "Analytical and Experimental Vibration Analysis of a Faulty Gear System," NASA Technical Memorandum 1.15:106689.
3. Yimin Fan and C. James Li, "Nonlinear System Identification Using Lumped Parameter Models with Embedded Feedforward Neural Networks," Symp. on Sensor-Based Control for Manufacturing, ASME International Mechanical Engineering Congress and Exposition, Anaheim, CA, Nov. 15-20, 1998, Manufacturing Science and Engineering, MED Vol. 8, Editor J. Lee, ASME, New York, 1998, p 579-587.
4. H-H. Lin, R. L. Huston and J. J. Coy, 1988, "On Dynamic Loads in Parallel Shaft Transmissions: Part I-Modeling and Analysis," Journal of Mechanisms, Transmissions, and Automation in Design, Vol. 110, pp. 221-225.

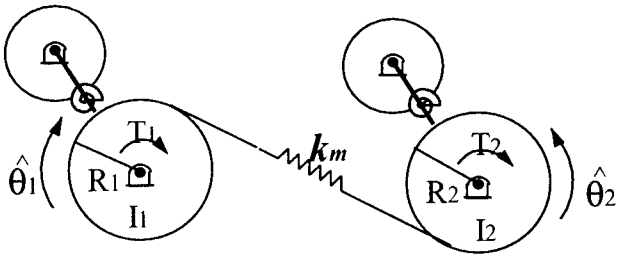
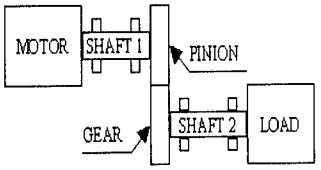
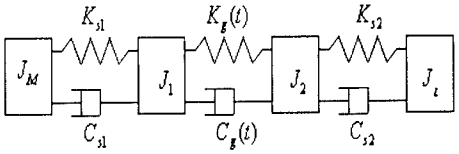


Figure 1. Gear dynamic model

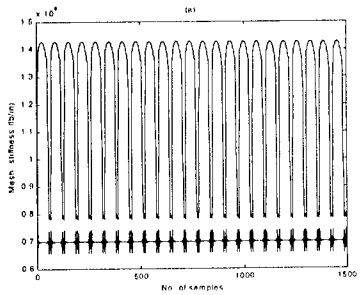


(a)

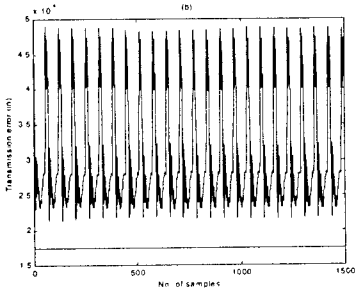


(b)

Figure 2. (a) A spur gear system, (b) A model of the gear system

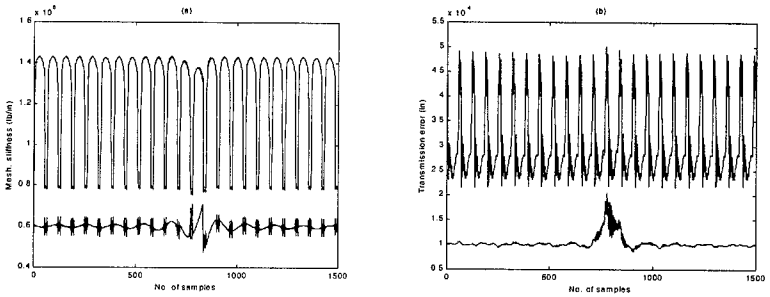


(a) The meshing stiffness $K_g(t)$



(b) The transmission error.

Figure 3. Meshing stiffness and simulated transmission error of a good gear



(a) The meshing stiffness $K_g(t)$ (b) The transmission error.

Figure 4. Meshing stiffness and simulated transmission error of a cracked gear

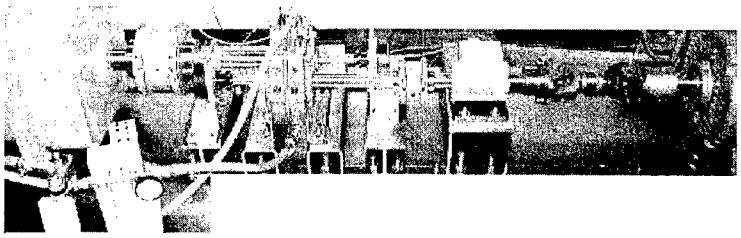
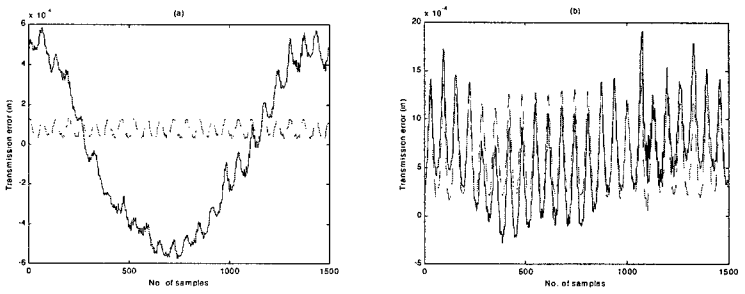
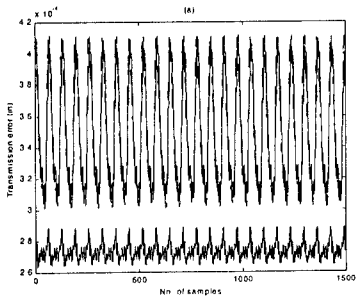


Figure 5. The gear test setup

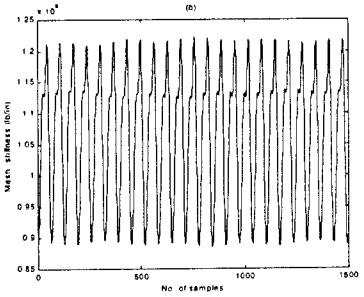


(a) normal gear (b) cracked gear

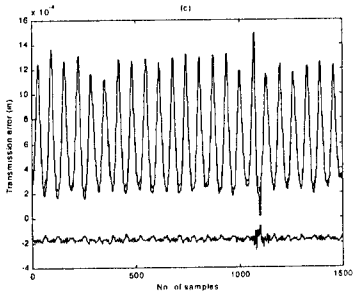
Figure 6. Transmission error (-measured, --filtered)



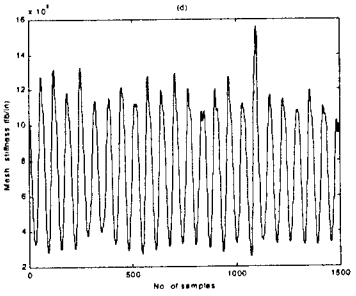
(a) Transmission error (gear is good)



(b) Identified meshing stiffness



(c) Transmission error (tooth has a .18" crack),



(d) Identified meshing stiffness

Figure 7. Transmission errors and meshing stiffness of the test gear (-experimental, -- identified, 0.5X error)

COMPARATIVE EVALUATION OF STRUCTURAL SURFACE INTENSITY TO STATISTICAL FEATURES FOR GEARBOX FAILURES

Jeff Banks, Rob Campbell and Carl Byington

The Pennsylvania State University
Applied Research Laboratory
Condition Based Maintenance Department

Abstract: The key to an effective Condition-Based Maintenance (CBM) program lies in the ability to extract machinery health information through diagnostic and prognostic indicators. In an effort to develop such indicators, the CBM department at the Penn State Applied Research Laboratory (ARL) has evaluated the use of structural surface intensity (SSI) for diagnostics and prognostics. In order to characterize structural surface intensity's effectiveness as a machinery diagnostic indictor, transitional fault data for three failure modes of an industrial grade gearbox was generated and SSI parameters are extracted and compared to the more widely used statistical-based features. The comparisons were focused on early detection capability and the relative change of the indicators subsequent to fault initiation. Results of such comparisons are provided for the three test runs. The comparisons show that in certain cases, SSI, as a diagnostic indicator, may provide an earlier detection capability and result in higher decision confidence than those obtained using the "traditional" statistical-based features.

Keywords: Condition-Based Maintenance (CBM); diagnostic and prognostic indicators; feature extraction; power flow; Structural Surface Intensity (SSI).

Introduction: Machinery and system maintenance is one of the key areas that contribute to industrial production effectiveness, transportation reliability and military readiness. The primary function of such maintenance is to maximize availability of operational assets through systematic evaluation and repair practices. The philosophies that affect maintenance methods have improved iteratively over the years based upon a better understanding of the failure mechanisms of mechanical components and systems and technological improvements. The evolution of machinery maintenance has led to the idea of a Condition-Based Maintenance (CBM) philosophy. In practice, this type of maintenance requires methods to assess the current and future states of 'health' of a mechanical system. The key to CBM lies in the development of robust diagnostic and prognostic indicators that facilitate determining the functional readiness of a system. Much effort has been focused on developing such indicators over the past several years. This paper will discuss a novel method for machinery health diagnostics using an indicator based on structural surface intensity (SSI) parameters. This method was developed at Penn State Applied Research Laboratory (ARL) in an ongoing effort to

improve diagnostic accuracy and prognostic capabilities necessary for machinery CBM. The performance of the SSI indicators is compared to the more common statistical-based diagnostic indicators using transitional data from the Mechanical Diagnostics Test Bed (MDTB) at ARL. Three fault types will be investigated for the MDTB industrial-grade single-reduction helical gearbox: gear tooth breakage, bearing failure, and shaft fracture.

Statistical Feature Extraction: In principle, information concerning the relative condition of monitored machinery can be extracted from a vibration signature, and inferences can be made about the health by comparing the vibration signal with previous signals to identify any anomalous conditions that may be occurring. In practice, however, such direct comparisons are not effective mainly due to the large variations between subsequent signals. Instead, several more useful techniques have been developed over the years that involve feature extraction from the vibration signature [1]. Generally these features are more stable and well behaved than the raw signature data itself. In addition, the features constitute a reduced data set since one feature value may represent an entire snapshot of data, thus facilitating additional analysis such as pattern recognition for diagnostics and feature tracking for prognostics. Moreover, the use of feature values instead of raw vibration data will become extremely important as wireless applications, with greater bandwidth restrictions, become more widely used.

The feature extraction method may require several steps, depending on the type of feature being calculated. Some features are calculated using the "conditioned" raw signal, while others may use a time-synchronous averaged signal that has been filtered to remove the "common" spectral components. ARL developed a CBM Features Toolbox that allows these features to be calculated systematically. Additional information regarding various feature extraction methods and the many types of diagnostic features available, see References [1,2].

Structural Surface Intensity: Structural intensity (SI) or power flow measurement techniques have traditionally been used to measure vibrational energy fields, determine strong transmission paths and locate vibration sources in simple beam and plate-like structures. SI as a vector quantity (magnitude and direction) may offer insight into the state of health of a mechanical system, based upon the changes of the flow of energy through that system. The idea of using power flow as a diagnostic indicator has seen limited application for several reasons. One of the primary reasons is that many factors restrict the application of traditional structural intensity measurement methods to practical structures with complex geometries. This is partially due to curvature and thickness variations of these structures, which invalidate the flat plate or beam assumption. These structural simplifications allow the straightforward implementation of the finite difference approximations that are necessary to estimate power flow.

G. Pavic originally developed the intensity measurement approach used in this research[3]. Pavic's method, which is not limited to structures, estimates the active intensity vector using the surface vibration and strain of the structure. Although SSI does not indicate the total power level within a structure (total power is an integration of the intensity through the entire cross section), it does provide insight into the energy flowing

through its surface as a vector quantity. Using SSI as a diagnostic indicator focuses on the relative changes in the intensity magnitude and direction as the structural properties of the gearbox components change during the fault development process, such as a gear tooth crack. Using surface intensity as opposed to total power is adequate for a diagnostic indicator, since a change in SSI will be indicative of a change in the system's structural characteristics, such as a developing fault.

In order to estimate in-plane surface intensities, five parameters must be measured including strain in the X and Y directions, shear strain, and velocity in the X and Y directions. The SSI method requires an array of transducers, including two accelerometers in the X and Y directions (planar to the surface) and three strain gauges in a rosette pattern of 0-45-90 degrees to the Y-axis (which is inline with the drive axis). The acceleration and strain data is manipulated to give estimates for the intensity magnitude and intensity direction angle using several algorithms. Preliminary research concerning the application of the SSI method is described in detail in a previous research paper [4].

MDTB Experimental Apparatus: In order to develop and evaluate diagnostic and prognostic indicators, seeded and transitional machinery fault data must be generated. In an effort to provide this data, the Mechanical Diagnostic Test Bed [5] was built by the Penn State University Applied Research Laboratory to experimentally simulate the accelerated fault evolution of a single reduction gearbox. The test bed, shown in Figure 1, consists of a 30 horsepower AC variable speed drive motor and a 75 horsepower AC load motor to load the gearbox at variable torque levels. The test gearbox is instrumented with input and output torque cells to monitor the loading conditions throughout the test cycle. The MDTB has been instrumented with a variety of sensors including accelerometers, strain gages, thermocouples, acoustic emission sensors, and oil debris sensors to acquire data for post-test processing.

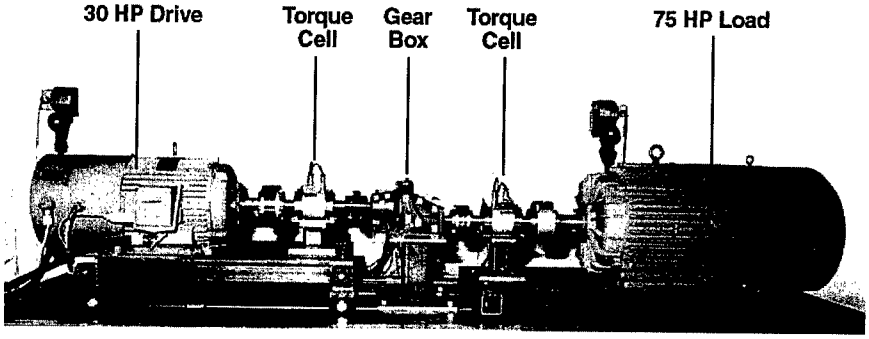


Figure 1: Mechanical Diagnostics Test Bed facility located at the Pennsylvania State University Applied Research Lab

The MDTB can create a variety of duty cycle profiles desired for testing within the physical limits of the motors. For the subject research, the MDTB was run at 1750 RPM (input side) and at 3 times the maximum rated load of the test gearbox.

Test and Analysis Results: A comparative analysis between SSI results and selected statistical features was conducted using the transitional fault gearbox data generated on the MDTB. This study was conducted to evaluate the use of SSI parameters as diagnostic and prognostic indicators for geartooth, bearing and shaft failure modes for the single reduction gearbox. The magnitude and direction angles of the SSI were estimated using an array of sensors attached to the gearbox housing. The features were extracted from one of the same accelerometers used to estimate SSI. The indicator parameters are post-processed and trended for the entire test run and significant changes in level are used as indications of an onset of a fault condition.

Gear Tooth Failure: MDTB test run 20 culminated with one broken gear tooth at 11 hours into run. Several features where extracted from the accelerometer data, but only two of these features are used for comparisons reported in this paper as shown in Figure 2. Due to constraints of the data acquisition system, the time period between successive data points for test run 20 varies from a few minutes to an hour as shown in Table 1.

Data Point	Date	Time	Speed (RPM)	Torque (in-lbs.)
1	12/9	15:58	1750	1665
2	12/9	16:00	1750	1665
3	12/9	18:00	1750	1665
4	12/9	20:00	1750	1665
5	12/9	22:00	1750	1665
6	12/10	5:15	1750	1665
7	12/10	6:15	1750	1665
8	12/10	7:15	1750	1665
9	12/10	8:10	1750	1665
10	12/10	8:15	1750	1665
11	12/10	8:20	1750	1665

Table 1: Experimental Test Set Run Conditions for Run 20

The first feature (BPMRUNVAR) shows a 35% change in level between data points 2 and 3. This feature is extracted by using a running variance of the band-passed gearmesh frequency including sidebands. The second feature (MRWINTK) is extracted by using the interstitial noise of the raw signal. This involves band passing the noise floor data in the region between the higher orders of the gearmesh frequencies. Then kurtosis is applied to this interstitial signal. See reference [1] for more details on this processing. This feature shows a 60% change in level between data points 2 and 3.

Structural surface intensity magnitude and direction angle for MDTB Run 20 were also measured and is shown in Figure 3. The intensity magnitude shows a significant change of 47% between data points 3 and 4 and the direction angle shows a 33 degree change in level between data points 5 and 6.

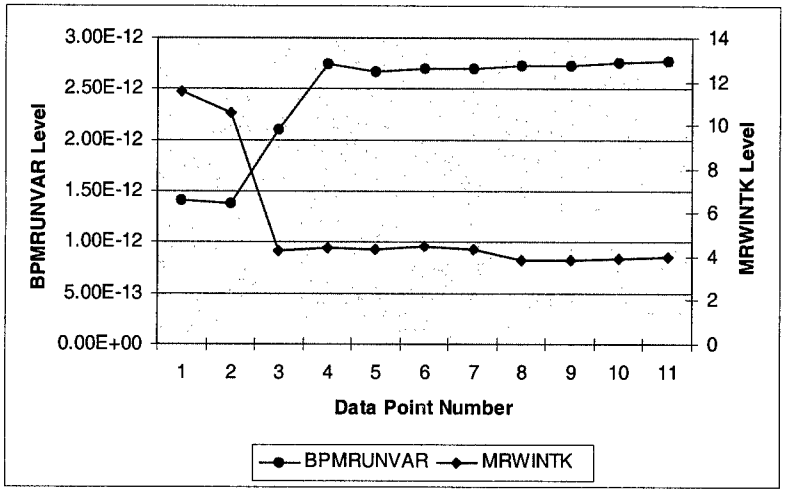


Figure 2: Feature Extraction for Geartooth Fault Mode

The test run for the geartooth failure mode is relatively short and it difficult to establish a baseline level, which is important when looking for a relative change in the parameters. Based on the available data though, the features appear to react earlier than the SSI, which would make them better diagnostic indicators for this failure mode test.

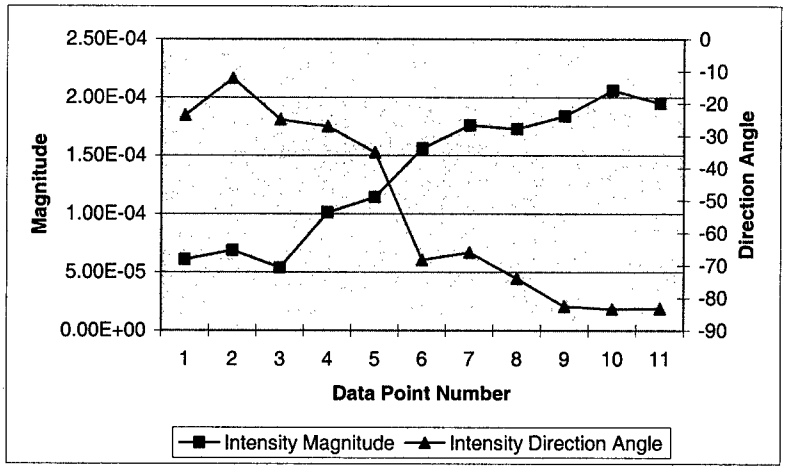


Figure 3: Intensity Magnitude and Direction Angle for Geartooth Fault Mode

The intensity magnitude does show an incremental increase in level that could possibly provide a prognostic indication. Generating more data sets for this failure mode would help define SSI's characteristic reaction to geartooth fault conditions.

Bearing Failure: A rolling element bearing failure occurred on the MDTB test run number 21 after 150 hours of run time. Data from before and during the onset of bearing failure was extracted (99 files) from the entire set of test data to use for the calculation of the statistical features and the SSI parameters. The data points for this test were taken in approximately fifteen-minute increments.

Figure 4 shows two features extracted from the accelerometer data: MRWRMS and MRWEVRMS. MRWRMS is the mean RMS level of the "raw" vibration signal, and MRWEVRMS is the mean RMS level of the enveloped signal. The enveloped signal involves isolating the high-frequency resonance response of the mechanical system to periodic impacts such as those generated by bearing faults [1]. As illustrated in Figure 4, MRWRMS changes by roughly 8% and MRWEVRMS shows an 11% change in level between data point numbers 109 and 110. These abrupt changes can be construed as a change in the gearbox health. The feature levels then trend upward subsequent to data point 133 due to further degradation of the bearing condition.

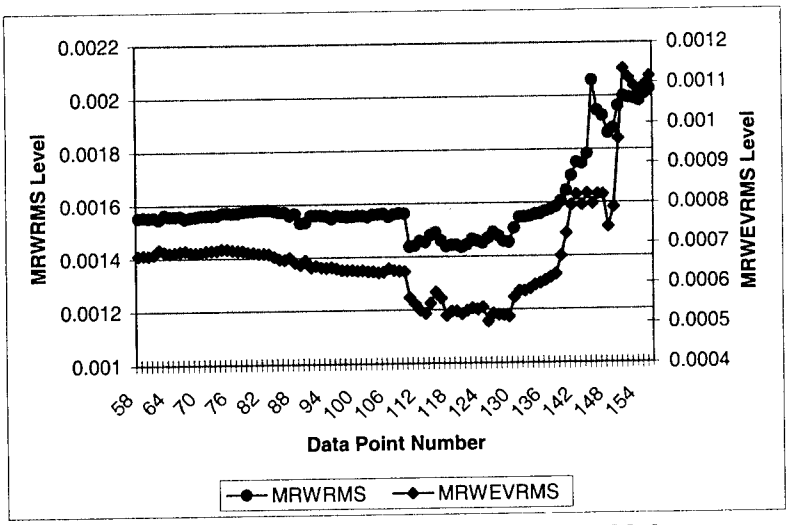


Figure 4: Feature Extraction for Bearing Fault Mode

Structural surface intensity magnitude and direction angle for MDTB Run 21 was also extracted and is shown in Figure 5.

The intensity magnitude and direction angle both show a significant change in level of 7% and 4 degrees respectively between data points 109 and 110.

The reaction of the features and of the SSI parameters to the initial onset of the bearing fault is shows a significant coincident change, which is typically a good diagnostic confirmation. The features also produce a significant increase in level as the bearing condition deteriorates, which can provide a good tracked parameter and possible prognostic indicator for the bearing damage.

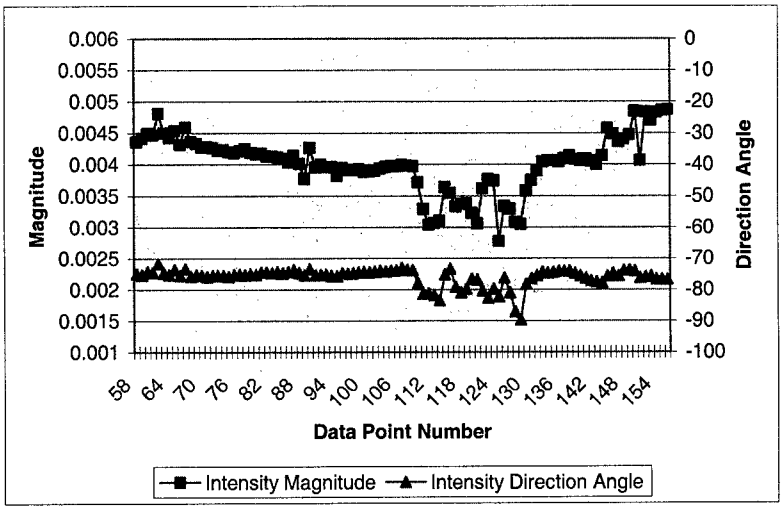


Figure 5: Intensity Magnitude and Direction Angle for Bearing Fault Mode

Shaft Failure: A shaft failure occurred after 150 hours of run time during MDTB test run 22. Similar to the bearing test, only data near the onset of shaft failure was extracted from the test data set to use for the calculation of the statistical feature parameters as indicated in Figure 6 by the data point numbers. Again, the time increment between data points is roughly 15 minutes.

Figure 6 shows two features that were extracted from the accelerometer data: MRWKURT and MRWCRST. MRWKURT is the kurtosis of the mean raw signal and MRWCRST is the crest factor of the mean raw signal [1]. As illustrated in Figure 6, MRWKURT shows a 5% and 16% change in level between data points 63 and 64, and 64 and 65, respectively. Kurtosis is the fourth moment of the distribution and it represents the relative peakedness of a distribution compared to a normal distribution [1]. Similarly, MRWCRST shows a change in level of 8% and 13% between data points 63 and 64, and 64 and 65, respectively. Both features continue to increase until data point 67 where the kurtosis level flattens while the crest factor becomes erratic/noisy.

Structural surface intensity magnitude and direction angle for MDTB Run 22 was extracted for the same data files and is shown in Figure 7.

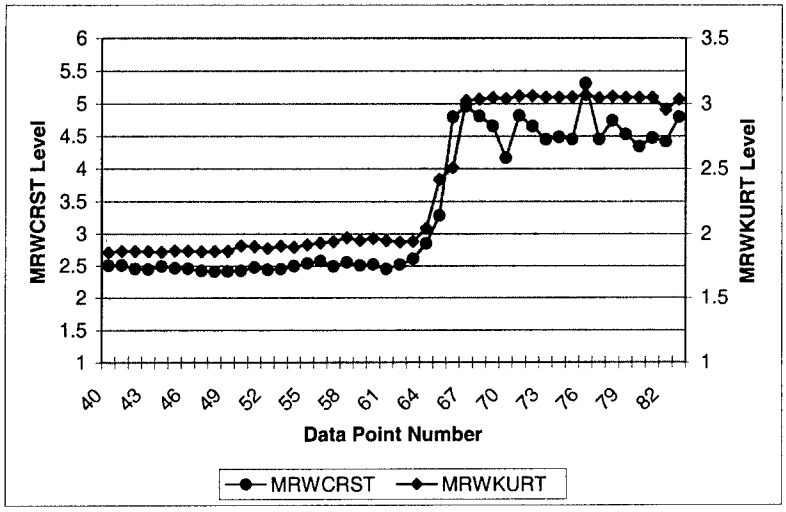


Figure 6: Feature Extraction for Shaft Failure Mode

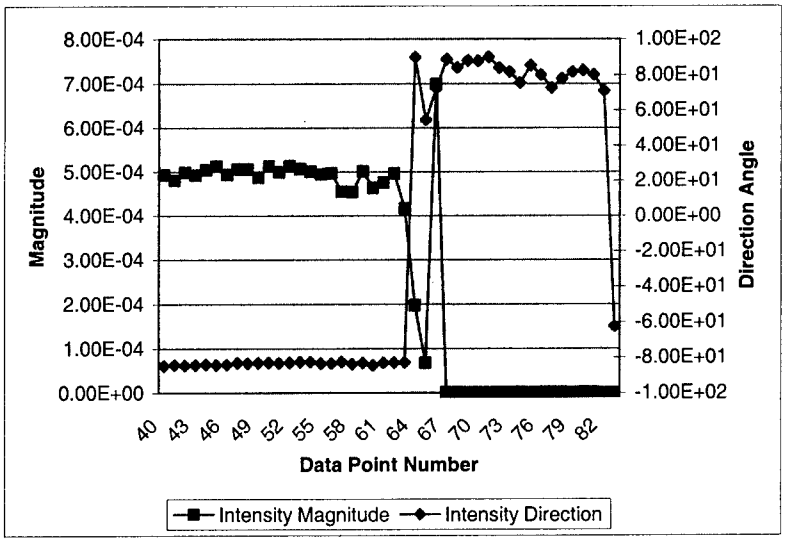


Figure 7: Intensity Magnitude and Direction Angle for Shaft Failure Mode

The intensity magnitude shows a 16% change in level between data points 62 and 63 and a 53% change in level between data points 63 and 64. The direction angle shows a 173-degree change in level between data points 63 and 64.

Comparison of these results shows that the SSI parameters appear to be more sensitive to a gearbox shaft failure condition than the statistical-based features shown in Figure 6. The intensity magnitude shows a dramatic change roughly 15 minutes prior to any indication from the statistical-based features. Though the intensity magnitude level changes at the same data point as the kurtosis and crest factor, its relative change is much more significant.

Summary: The development of diagnostic and prognostic indicators that are sensitive to mechanical faults is paramount when creating an effective CBM system. In an effort to evaluate the performance of structural surface intensity parameters as diagnostic indicators, they were compared to commonly applied statistical-based diagnostic features. The results of the comparisons vary slightly for each failure mode analyzed. For the geartooth failure case the selected features perform better than the SSI, but the SSI vector parameters may have some attractive attributes for tracking and prognosis. During the bearing failure test, all of the parameters indicate a change in condition at the same time with relatively small deviations in level. The shaft failure test showed the most promising results for the SSI parameters with a large change in the intensity magnitude roughly 15 minutes prior to the statistical features.

A more in-depth evaluation of SSI is necessary to validate its use for equipment diagnostics. The hope is that future research will provide more data sets for each failure case to better understand how SSI changes as a fault develops within a mechanical system and progresses toward failure.

Acknowledgment: This work was supported by the Office of Naval Research through the *Multidisciplinary University Research Initiative for Integrated Predictive Diagnostics* (Grant Number N00014-95-1-0461). The content of the information does not necessarily reflect the position or policy of the Government, and no official endorsement should be inferred.

References:

1. McClintic, K., et al, *Residual and Difference Feature Analysis with Transitional Gearbox Data*, 54th Meeting of the MFPT, Virginia, May 2000.

2. Lebold, M., *et al, Review of Vibration Analysis Methods for Gearbox Diagnostics and Prognostics*, 54th Meeting of the MFPT, Virginia, May 2000.

3. Pavic, G., *Structural Surface Intensity: An Alternative Approach in Vibration Analysis and Diagnosis*, Journal of Sound and Vibration, 115(3), pp. 405-422, 1986

4. Banks J., Hambric S., *Structural Surface Intensity as a Diagnostic Indicator*, 54th Meeting of the MFPT, Virginia, May 2000.

5. Kozlowski, J. D. and C.S Byington, *Mechanical Diagnostics Test Bed for Condition Based Maintenance,* ASNE Intelligent Ships Symposium II, November 25. 1996.

A Hybrid Stochastic-Neuro-Fuzzy Model-Based System for In-Flight Gas Turbine Engine Diagnostics

Dan M. Ghiocel and Joshua Altmann

STI Technologies
1800 Brighton-Henrietta TL Rd.
Rochester, New York, USA
Ph: (716)424-2010

Abstract: One key aspect when developing a real-time in-flight risk-based health management system for jet engines is the development of accurate and robust fault classifiers. Regardless of the complex uncertainty propagation in the data fusion process, the selection of fault classifiers is the critical aspect of a health management system. The paper illustrates the application of a hybrid *Stochastic-Fuzzy-Inference Model-Based System (StoFIS)* to fault diagnostics and prognostics for both the engine performance. The random fluctuations of jet engine performance parameters during flight missions are modeled using multivariate stochastic models. The fault diagnostic and prognostic risks are computed using a stochastic model-based deviation (using a gas-path analysis model) approach. At any time the engine operation for the future is approached as a conditional reliability problem where the conditional data are represented by the past operational history monitored on-line by the engine health management (EHM) system. To capture the complex functional relationships between different engine performance parameters during flight fast an adaptive network-based fuzzy inference system is employed. This increases significantly the robustness of the EHM system during highly transient in-flight conditions. Both the monitored and fault data uncertainties are considered in a multidimensional parameter space, with two probabilistic-based safety margins employed for fault detection and diagnostic, as follows: (i) Anomaly Detection Margin (ADM) and (ii) Fault Detection Margin (FDM).

Key Words: ANFIS, Engine Health Monitoring, Gas Path Analysis, and Stochastic Analysis

Adaptive Network-based Fuzzy Inference System (ANFIS) GPA Modeling: A schematic model of the investigated turbofan engine, including the performance parameters used for fault diagnostic is shwoimplemented in the ANFIS GPA/Stochastic engine health monitoring system are shown in figure 1. Figures 2 and 3 show pressure variations as a function of the high-pressure shaft speed. It is obvious from these figures that although for slowly varying ground tests the pressure closely follows a non-linear relationship with shaft speed, for in-flight testing the pressure deviates from this non-linear path due to highly transient conditions and changes in the inlet conditions. This means using deviations from a fitted polynomial curve for diagnostics, as commonly used in earlier engine health monitoring software based on ground-test data, is not suited to in-flight conditions. This is the driving force to move towards deviations from a GPA engine model as the basis for health monitoring and sensor validation in state-of-the-art engine health monitoring systems.

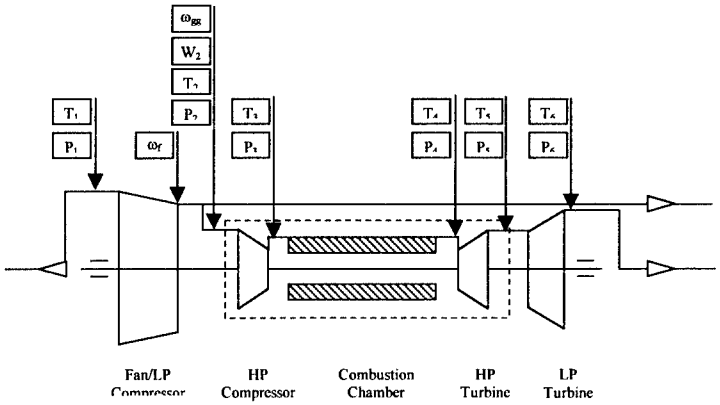


Fig. 1: Schematic of Turbofan Engine

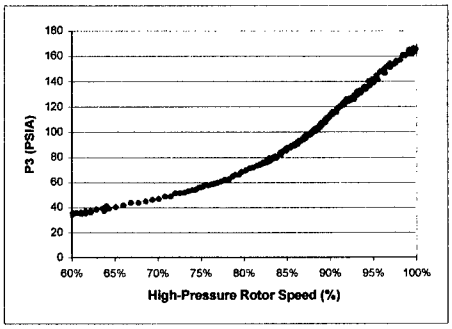


Fig. 2. Plot of P3 versus ω_{gg} for a ground test.

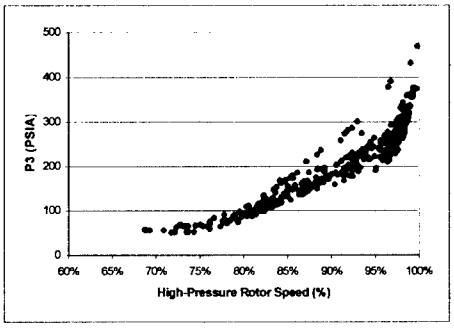


Fig. 3. Plot of P3 versus ω_{gg} for a typical mission in-flight measurement.

As analytical GPA models only cater for quasi-stationary engine operation, an alternative scheme capable of including the highly transient in-flight conditions has been developed. This lead to the formation of a diagnostic system based on parameter statistical deviations from an adaptive network-based fuzzy inference system GPA model. This is essentially a GPA model developed from a hybrid neuro-fuzzy approach based on training from typical in-flight data for a given engine. Both an overall model of the turbofan engine and models of the individual compartments have been constructed. The functional basis for the GPA models are described below.

- Transient Overall ANFIS GPA model
  The following relationship is assumed in this model. It takes transients into account by the inclusion of the low-pressure shaft speed and flow-rates. The air mass-flow is denoted by $\dot{m}_{gg}$, and the fan and gas-generator rotor speeds are denoted by ω_f, ω_{gg} respectively.

$P_n, T_n = fn(P_1, T_1, \dot{m}_{gg}, \omega_f, \omega_{gg})$ (1)

- Transient Compartmentalized ANFIS GPA model
  The following relationship is assumed in this model. It takes transients into account by the inclusion of the low-pressure shaft speed and flow-rates. Increased accuracy is obtained at the compartment level by basing output pressures and temperatures on the inlet temperatures and pressures of the compartment in question.

  $P_n, T_n = fn(P_{n-1}, T_{n-1}, \dot{m}_{gg}, \omega_f, \omega_{gg})$ (2)

  The exception to this is the modeling of T4, which would include the fuel mass flow in the model. This is an attempt to model the transients due to fuel flow variation in the combustion chamber. This would also enable the detection of abnormal fuel flow ($\dot{m}_f$) levels, and any impact to be assessed.

  $T_4 = fn(P_{n-1}, T_{n-1}, \dot{m}_{gg}, \omega_f, \omega_{gg}, \dot{m}_f)$ (3)

  P_n and T_n represent the static pressure and temperature parameters indicated in figure 1.

Figures 4 and 5 illustrate the importance of being able to have a system able to discriminate with a high degree of accuracy any deviation from normal operating conditions. The effects of high-pressure turbine efficiency drop on the deviations are shown on the parameters P_5, T_6 and P_6. Figure 4, which is based temperature and pressure deviations from a fifth-order polynomial fitting of parameters against the high pressure rotor speed, is inadequate for fault detection due to the large degree of uncertainty in measurements. In comparison, the ANFIS GPA model shrinks that uncertainty to a level that small changes in operating efficiency or capacity from normal status can be detected. Even with this improved fault resolution, it is important to use this approach in conjunction with stochastic modelling, as a relatively high degree of uncertainty will exist if individual measurements are used as the basis for assesment of the engine condition.

Figure 6 illustrates the output from the equivalent compartmentalized model for the data used in the overall model shown in figure 4. There are two major advantages that are provided by including a compartmentalized ANFIS model to complement the overall ANFIS model:

a) Increased resolution of parameters is available, as uncertainty introduced by other compartments is illuminated.
b) Ability to discriminate between the presence of single and multiple compartment faults in an engine. An overall GPA model can lead to difficulty in diagnosing the cause and mapping the progression of faults if two or more compartments are operating out of specification.

The main drawback of the compartmentalized model is that one loses the multidimensional parameter space, i.e. interaction between engine parameters, that is available in the overall model. At most two parameters, the pressure and temperature, are available in the transient ANFIS GPA model. Both methods should be incorporated to make use of the synergies that can be attained from these models. The examples illustrated are based on in-flight performance based measurements from engines without faults, which can be used to provide the baseline for generating the ANFIS GPA models. The advantage of implementing the model in the form of an adaptive network-based fuzzy inference system is that while the system can be initially formed using fuzzy sub-clustering and then trained using a hybrid neural network (least squares and backward propagation), it can be implemented as a standard fuzzy inference system.

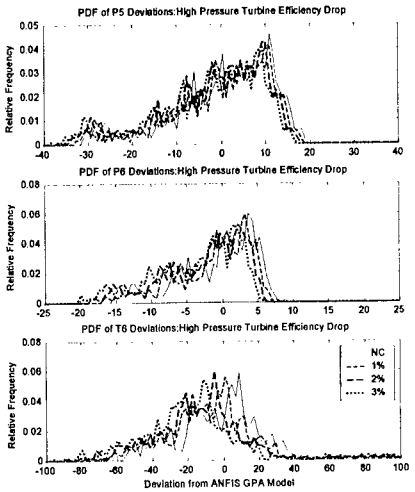


Fig. 4. PDF of deviations from fifth-order polynomial fitting $Fn(\omega_{gg})$ for in-flight missions.

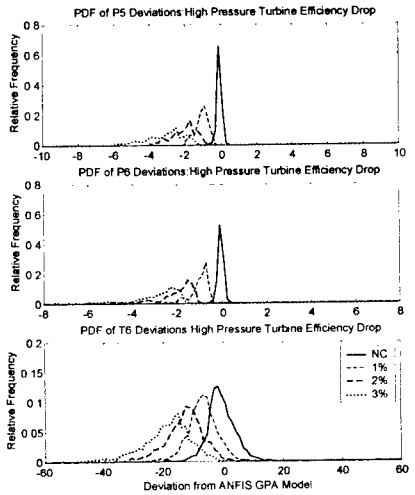


Fig. 5. PDF of deviations from overall ANFIS GPA model $Fn(P_1,T_1,\omega_{gg}, \omega_f, \dot{m}_{gg})$ for in-flight missions.

This reduces the overheads associated with deploying the system into an onboard monitoring environment, increases the operational speed of the system, and enables reasoning for the parameter deviations to be assessed.

A generic analytical GPA model was used to simulate faults in the engine, with the corresponding parameter deviations mapped onto the ANFIS model. The statistical parameter deviation database was be modeled using advanced stochastic modeling techniques to provide tools for fault detection, diagnosis and prognosis. The ability of this approach to model the deterioration of a fault statistically over time is of major benefit, as this facilitates a method for assessing the deterioration of a fault, thus providing prognostic tools that were previously extremely limited in their capabilities.

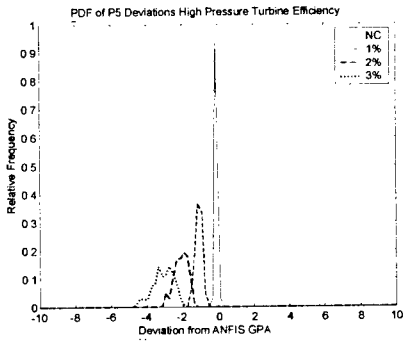


Fig. 6. PDF of deviations from compartmentalized ANFIS GPA model $Fn(P_1,T_1,\omega_{gg}, \omega_f, \dot{m}_{gg})$ for in-flight missions.

Risk-Based Fault Diagnostic and Prognostic Using GPA Model: Typically, performance parameter data measured on-line include pressures, temperatures and fuel flows in different compartments of the jet turbine engine. A pictorial representation of the proposed probabilistic fault diagnostic/prognostic procedure is given in Figure 7.

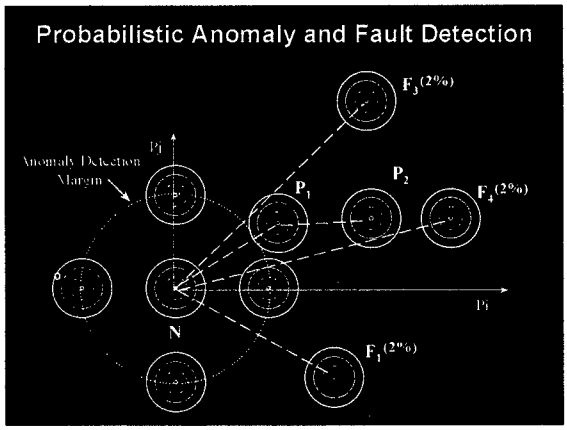
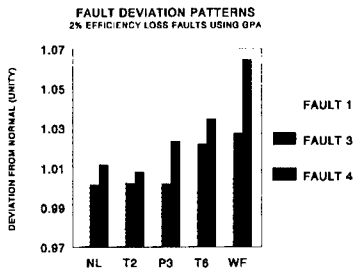


Fig. 7. Pictorial Representation of Probabilistic Fault Diagnosis Procedure

Initially, the probability distributions of measured performance parameters are defined based on statistical data acquired during normal operating conditions assuming that the turbine has no usage (represented by the origin of reference system is the multidimensional parameter space). The probability distributions of faults are defined for a given severity level in the engine efficiency loss. They are determined experimentally and/or numerically by "seeded" faults using test results and/or gas-path-analysis simulation results. During lifetime the turbine performance departs from the parameter space origin (zero bias in the measurement space) due to usage and the measured parameter probability distributions begin to shift as performance degradation occurs. After hundreds of flights when a specific anomaly detection level is reached an anomaly warning becomes active. Then, after other several hundreds of flights the measured parameter joint distribution moves to the point P1 as shown in the Figure 7. At this point, the fault looks like it might be classified with near equal probability as either a Fault 3 or a Fault 4 scenario. However, after a continued operation the measured data moves toward point P2, which will be classified with a highest degree of confidence as Fault 4. These evolutionary scenarios of turbine performance degradation shown in Figure 6, can be handled mathematically accurately by using the proposed probabilistic fault diagnostics/prognostic procedures.

Figure 8 illustrates the mean parameter deviations in five parameter spaces, such as power level, NL, pressure P3, temperatures T2 and T6 and fuel mass flow, WF, for three fault types expressed by a 2% efficiency loss. Figure 9 shows how actual engine performance Faults 1 (fan outer efficiency degradation), Fault 3 (high-pressure

compressor efficiency degradation, and Fault 4 (high-pressure turbine efficiency degradation) manifest in the parameter space.



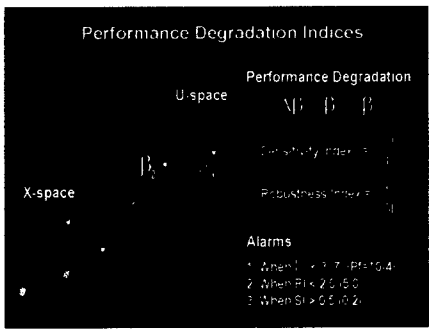


Fig. 8 Mean Deviations of Faults 1, 3 and 4 Fig. 9 Performance Degradation in U-space

The probabilistic fault diagnosis system uses as a safety margin for reliability probabilistic calculations the distance between the measured condition and the fault condition in a five-dimensional parameter space. This distance is becoming gradually smaller with respect of some faults as the performance degrades.

The turbine performance is a specific time-variant reliability problem due to usage and aging effects affecting all components. To probabilistically investigate and characterize the turbine performance evolution two types of reliability indices are introduced herein. Specifically, a cumulative and an evolutionary reliability sensitivity index for safety (performance) degradation are introduced. The cumulative reliability sensitivity index (CRSI) is defined by the "global" non-dimensional variation of the FORM reliability index, β (the relation between "failure probability", here read fault diagnostic probability, and reliability index is discussed on the next page) from initial state, at 0, to the final state, at t (over the interval [0, t]):

$$C_{0,t} = -\frac{\beta_t - \beta_0}{\beta_0} = -\frac{\Delta\beta_{ot}}{\beta_0} \tag{7}$$

The evolutionary reliability sensitivity index (ERSI) is defined by the "local" non-dimensional variation of the reliability index, from an intermediary state, at time ti, to another intermediary state, at time ti+1 (over the interval [ti, ti+1]

$$E_{ti,ti+1} = -\frac{\beta_{ti+1} - \beta_{ti}}{\beta_{ti}} = -\frac{\Delta\beta_{ti,ti+1}}{\beta_{ti}} \tag{8}$$

These two reliability sensitivity indices indicate in percents the change of the safety (performance) measure due to the time changes in the basic variables. A zero value indicates no safety (performance) degradation, while a positive value indicates a safety (performance) degradation and a negative value indicates safety improvement. Robustness indices (RI) can be defined as inverse of sensitivity indices (SI). For the engine performance degradation problem the "red" alarms can be set at a reliability index

below 3.7 (fault probability of 0.0001), for a CRSI of 0.5, or equivalently for a CRRI of 2.0 and for a ERSI of 0.2, or equivalently for a ERRI of 5.0 as shown in Figure 9.

Herein, specific quantitative measures of safety, such as safety margin, failure probability and reliability index, are used for diagnosing the faults. The conceptual framework for structural reliability and probability-based design is provided by the classical reliability theory. A general reliability model relates the "capacity" (fault conditions) and "demand" (measured conditions) variables in a limit state function (or failure equation or mode), also often called g-function of the form:

$$g(x_1, x_2, \ldots, x_n) = 0 \tag{9}$$

where $x_i, i = 1, n$ are the design variables (or life drivers). Failure occurs when $g < 0$ for any ultimate or operability (performance) limit state (or failure mode) of interest. Herein, for the engine performance reliability calculations, the failure equation is defined by the distance in the multidimensional parameter space X between a fault location, defined by vector R, and a measured usage state location, defined by vector S, i.e. g(X) = R(X) - S(X).

Then safety is assured by assigning a small probability p_f to the event that the limit state will be reached

$$p_f = \int \ldots \int f_X(x_1, x_2, \ldots, x_n) dx_1, dx_2 \ldots dx_n \tag{10}$$

in which f_X is the joint probability density function for $x_1, x_2, \ldots$, and the integration is performed over the region where $g < 0$.The failure event occurs when R - S < 0, where R is the "capacity" or "resistance' and S is the "demand" . The probability of failure is computed by

$$p_f = p(R < S) = \int_0^\infty F_R(x) f_S(x) dx \tag{11}$$

in which F_R is the cumulative probability distribution function (c.d.f.) of R and f_S is the probability density function of S. If R and S both have normal distributions the probability of failure can be computed (R-S model)

$$p_f = \Phi\left(-\frac{\overline{R - S}}{\sqrt{\sigma_R^2 + \sigma_S^2}}\right) = \Phi(-\beta) \tag{12}$$

where $\overline{R}$, σ_R are the mean and standard deviation (σ_R^2 = variance) for R and similarly for S. The function $\Phi[\ ]$ is the standard normal cumulative distribution. The notation β is the reliability index. If R and S both have lognormal distributions the failure probability can be computed (R/S model)

$$p_f \approx \Phi\left(-\frac{\ln(\overline{R}/\overline{S})}{\sqrt{V_R^2 + V_S^2}}\right) = \Phi(-\beta) \tag{13}$$

where V_R, V_S < about 0.30, in which V_R, V_S = coefficient of variation (c.o.v.) in R and S. If R and S are not normal or lognormal variables then the probability of failure can be determined using a computer algorithm.

The above equations provide a basis for quantitatively measuring structural reliability such as the measure being given by the probability of failure, p_f, or by the reliability index, β. The use of reliability index is convenient as long as the limit function is not highly nonlinear. When the limit state function is highly nonlinear the above expressions for evaluating probability of failure become crude approximations. This was the basic reason for developing the first-order reliability method, denoted FORM. While any continuous mathematical form of the limit state equation is possible, in FORM, it must be linearized at some point for purposes of performing the reliability analysis. Linearization of the failure criterion

$$Z \approx g(X_1^*, X_2^*, \ldots X_n^*) + \sum_i (X_i - X_i^*)\left(\frac{\partial g}{\partial X_i}\right)\Big|_{X_i^*} \qquad (14)$$

where $(X_1^*, X_2^*, \ldots X_n^*)$ is the linearizing point. In the classical FORM procedure, this linearizing point is obtained by computing the minimum distance between the mean point and the limit state function in a standard normal space (this distance is equal to the reliability index).

Figure 10 shows the computed turbine (reliability) performance indices, β, for the initial no-usage condition and usage conditions, P1 and P2, with respect to the Faults 1, 3 and 4. Figure 11 shows the failure probabilities computed with respect to the investigated seven fault conditions assuming initial normal condition and degraded conditions P1 and P2. It should be noted that from the start the fault diagnosis (failure) probabilities are very different for the seven faults. In the process of performance degradation these fault diagnosis probabilities may also change severely as illustrated in Figure 11. Initially, Fault 1 was the most likely fault being the closest to normal operating conditions, but after the performance degraded to condition P1 and then P2, the most likely faults have become Fault 3 and Fault 4. In the performance degradation process the turbine state departs from initial normal condition and from Faults 1, 2, 5, 6 and 7 for which the fault diagnosis probabilities become low due to direction of usage trajectories in parameter space. The evolution of the turbine performance toward Fault 4, in an opposite direction of Fault 1, is obvious from both Figures 10 and 11.

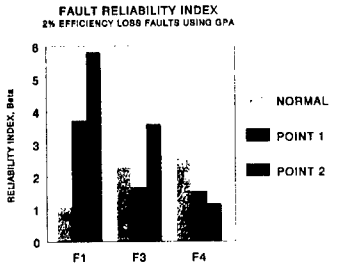


Fig. 10. Fault Reliability Index

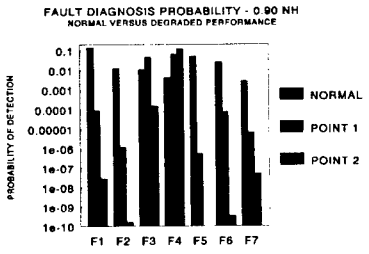


Fig. 11. Fault Probabilities

The cumulative and evolutionary reliability sensitivity indices, CRSI and ERSI, which describe the performance degradation rate in time in terms of safety indices were computed. These indices can be interpreted as Cumulative and Evolutionary Performance Degradation (RPD) index, respectively.

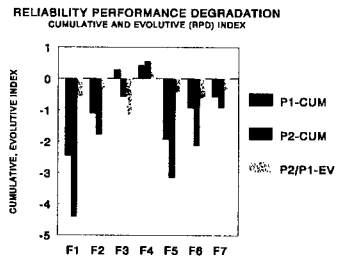


Fig. 12. Cumulative/Evolutionary Indices

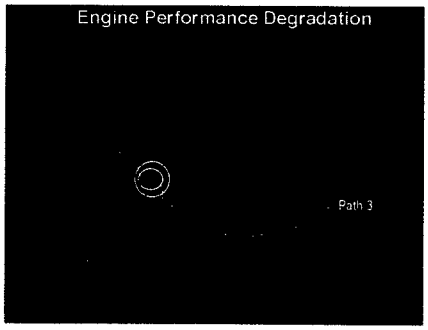


Fig. 13. Engine Performance Fault Map

Figure 12 illustrates the computed values of Cumulative RPD index for location P1 and P2, and Evolutionary RPD index for interval P1-P2. The negative values of CRPD and ERPD indicate a departure from a specific fault, while the positive values indicate a movement toward a specific fault. The CRPD and ERPD show that performance degradation was initially in the direction of Fault 3 and Fault 4 for Normal-P1 variation, and then for P1-P2 variation only in the direction of Fault 4, departing form the other six faults. The positive nature of ERPD index for Fault 4 is a decisive way to examine the movement of the measurement set in the multi-dimensional performance parameter space. The computed reliability index and the reliability sensitivity indices are used for fault diagnostic and prognostic, respectively. At instant time, depending on the fault safety margins to different faults and on the performance degradation rates in terms of safety, the remaining life of the turbine can be predicted for a given severity of efficiency loss between different times and probability levels. A minimum risk level is accepted before taking a maintenance action (in Figure 9 suggests a reliability index of 3.7).

A more realistic risk-based engine performance diagnostic-prognostic system is shown in Figure 13. The engine usage paths and faults are represented by stochastic complex trajectories and maps in multidimensional parameter space. However, this risk-based model that is currently in implementation is not discussed in this paper.

A key aspect for getting realistic results for in-flight operating conditions is to separate the *true statistical variabilities* (random part) from the *functional variabilities* introduced by engine transient behavior. For real, in-flight transient conditions the functional dependence between engine performance parameters becomes complex. If these, transient complex functional dependencies between parameters are ignored then the statistical variability is overestimated. To incorporate the complex functional dependencies for transient in-flight conditions the ANFIS has been employed.

It should be noted that important information for fault prediction is incorporated in the parameter statistical deviations after ANFIS was applied. These statistical deviations are not noises. The multiple-parameter statistical deviations can be described using a multidimensional stochastic process model. Figure 14 shows auto- and cross-correlation functions for different pair of parameters for normal conditions and fault conditions. For fault conditions the correlation length of parameter deviations is larger than for normal operating conditions.

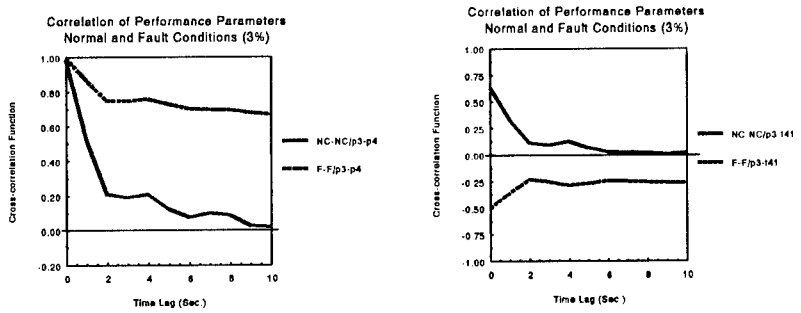


Figure 15: Correlation of Performance Parameters

The zero-time lag cross-correlation estimations between different performance parameters are incorporated in probabilistic fault diagnostic model.

Concluding Remarks:

The conceptual framework for a Probabilistic *Prognostic Health Management* (PHM) system has been outlined. The proposed *Stochastic-Fuzzy-Inference Model-Based System (StoFIS)* has been demonstrated to be capable of performing accurately under highly transient in-flight conditions, as encountered during military operations, through the exploitation of an ANFIS based GPA model. Probabilistic tools then enable deviations from the model to be used for risk-based fault diagnostics and prognostics.

References:

[1] Altmann J., Mathew J. "Automated DWPA Feature Extraction of Fault Information from Low Speed Rolling-Element Bearings", Proceedings of Asia-Pacific Vibration Conference, Singapore, 1999

[2] Ghiocel, D.M., "Advanced Stochastic Classifiers for Jet Engine Performance and Vibration", 55th MFPT Conference, Virginia Beach, NC, May 4-6, 2000

[3] Ghiocel, D. M., Roemer, M J. "A New Probabilistic Risk-Based Fault Diagnosis Procedure for Gas Turbine Engine Components" 40th AIAA/ASME/ASCE/AHS/ASC Non-Deterministic Approaches Forum, St. Louis, MO, April 12-15, 1999

DIAGNOSTIC FEATURE COMPARISONS FOR EXPERIMENTAL AND THEORETICAL GEARBOX FAILURES

Robert Campbell, Jeffrey Banks, Colin Begg, Carl Byington
The Pennsylvania State University
Applied Research Laboratory
P.O. Box 30 (North Atherton Street)
State College, Pennsylvania 16804-0030

Abstract: As part of a combined experimental-theoretical analysis investigation effort related to equipment diagnostic and prognostic feature development, an experimental gearbox testbed was previously developed and transitional failure data was obtained for several runs. A finite element model representing the rotating components on the testbed was developed to perform simulations for both healthy and selected gearbox fault conditions. The fault simulations are focused on gear tooth fracture since this was witnessed to be the primary gearbox failure mode during the test runs. Comparisons between the response of several common diagnostic features using both the gearbox testbed experimental data and the simulated data are provided. The knowledge obtained from evaluations of the simulated data sets and the feature comparison studies can be used to develop features with improved physical understanding of underlying mechanisms and optimize preprocessing methods for the existing diagnostic indicators. The results of the comparisons are presented and recommendations for future enhancements to the model are provided.

Key Words: Condition-Based Maintenance; dynamic systems; model-based diagnostics; simulation; statistical feature; transitional failure data.

Introduction: Condition-Based Maintenance (CBM) has been driven in part by the demand to increase system readiness/availability and reduce operations and maintenance costs. CBM accomplishes this through timely identification of equipment failures and elimination of unnecessary maintenance. Numerous authors have highlighted the cost and safety benefits of using CBM. [1, 2] This approach to maintenance relies on monitoring the condition of a system in order to detect and isolate anomalous conditions in a timely manner. An ultimate goal is to develop a health prognosis or prediction of Remaining Useful Life (RUL) with an associated functional impact assessment considering contextual information so that appropriate maintenance can be optimally scheduled. Technology maturation in the areas of measurement sensors, signal processing, digital processing hardware, dynamic system simulation, multi-sensor data fusion, and approximate reasoning have enabled the recent advancements in CBM.

Machinery fault detection generally involves comparing historical and nominal values to identify any statistically significant changes. During the diagnosis process, specific fault recognition

parameters (figures of merit) are calculated and often compared to threshold limits. [3] Additional processing may be used to enhance the diagnostic robustness using data fusion and reasoning modules to automate fault classification and damage assessment. Equipment health prognostics builds upon the diagnostic assessment with a tracked parameter that is related to damage and a future damage state prediction. These diagnostic and prognostic analyses can be based on extensive statistical experimental data with an associated empirical model of the particular system, an estimate made using predictions from a detailed system model, or a hybrid approach using a combination of both methods. The current work investigates comparisons between figures of merit calculated using results of a dynamics model and empirical results from transitional failure tests.

The transitional failure tests were conducted at the Penn State Applied Research Laboratory (ARL) using the Mechanical Diagnostics Test Bed (MDTB). The MDTB was built as an experimental research station for the study of fault evolution in mechanical gearbox power transmission components. [1, 2] It consists of a motor, gearbox, and generator mounted on a steel platform. The gearbox is instrumented with accelerometers, thermocouples, acoustic emission sensors, and oil debris sensors. A dynamics model of the MDTB was developed and is used to perform simulations of the system for both healthy and faulty gearbox conditions. [4] The simulations are focused on gear tooth failures since these are observed to be the most common type of fault encountered with the MDTB test runs.

Model-Based Methods and Considerations: The development of a model-based diagnostic/prognostic capability for CBM requires a proven methodology to create and validate physical models that capture the system response under normal and faulted conditions. For a majority of systems, operational demands induce a slow evolution in material property and/or component configuration changes. The potential thus exists to track the fault during the failure progression and provide an advanced warning of impending failure with a RUL estimate.

Model-based diagnostics, one of many CBM techniques, can be an optimum method for damage detection and condition assessment because empirically validated mathematical models at many state conditions are still deemed the most appropriate knowledge bases. [5] One approach for using model-based diagnostics is shown in Figure 1. The figure illustrates a conceptual method to identify the type and amount of degradation using a validated system model. The actual system output response (event and performance variables) is the result of nominal system response plus fault effects and uncertainty. The model-based analysis and identification of faults can be viewed as an optimization problem that produces the minimum residual between the predicted and actual response.

A consideration that differentiates the modeling of the MDTB from more common rotordynamic systems is the fact that the rotor system contains a pair of meshing gears. One of the most powerful and popular tools for modeling a rotordynamic system has been the finite element method (FEM). [6] Gearbox dynamics problems differentiate themselves from other structural dynamic systems by the branching of transmitted power through a gear mesh that leads to parametric excitation.

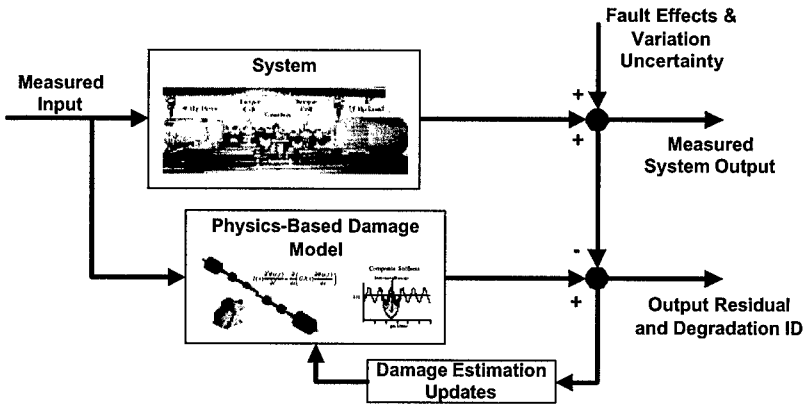


Figure 1. Model-Based Diagnostics Process

Some common practices have been established in dynamic modeling of geared rotor power transmission systems with full-face width hub gearing. [7, 8] For instance, the base rotor hub is treated as a rigid disk with gear tooth contact, body, and root deflections lumped together to represent a dependent function of both pinion and gear rigid rotational motion. The dynamic response between gear pairs can be treated as a transmission error [9] or by defining the dynamic forces using effective gear tooth deflection forces and apparent variable stiffness. [10] The latter more accurately characterizes a system in terms of effective parameters for dynamic system analysis.

MDTB Dynamics Model: The topology of the MDTB mechanical structure is shown in Figure 2. The rotor system finite element model of the MDTB is made up of five subsystems: 1) drive motor, 2) torque transducer at gearbox input, 3) single reduction helical gearbox, 4) torque transducer at gearbox output, and 5) load motor. The subsystems are linked with 1 chain and 3 gear couplings, which are modeled using lumped mass polar moments of inertia and elastic gear tooth mesh compliance.

The system rotor model is comprised of 36 structural finite elements and 38 nodal points. The structural finite elements include: rotational axisymmetric, axial translational, and 2-dimensional bending type elements for circular shafts. [11] A translational spring (representing gear mesh tooth stiffness) is incorporated into a rigid hub/elastic tooth gearbox pinion and gear coupling matrix. [12] The nodal points include: 16 single degree-of-freedom axisymmetric rotational nodes at rotary torsional element connections of the driveline outside of the gearbox, and 22 six degree-of-freedom nodes along the gearbox shafts. Nodes are placed at discrete steps in shafts, at the axial center of shaft couplings, and at the center of gearbox shaft bearing seats. Only torsionally driven axisymmetric rotations about the system driveline shaft are considered. Shaft axial and bending type displacements of the rotor train are eliminated at the input and output gear couplings due to the effective kinematic joint associated with the gear coupling.

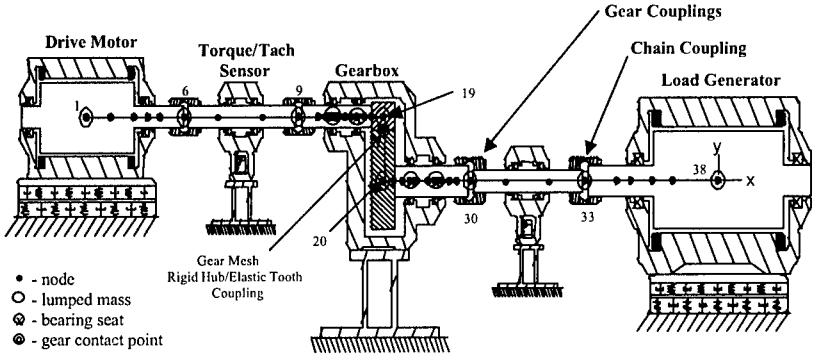


Figure 2. MDTB Topology and Rotordynamic Model with Node Points

The nominal lumped parameter (FEM) system model parameters (inertia-[M], damping-[C], gyroscopic-[G], and stiffness-[K]) can be modified to incorporate system faults for response simulations. However, the fault simulations in this study were limited to gear tooth faults, and thus only perturbations due to a time varying stiffness were present, see Equation (1).

$$[M]\ddot{\bar{s}} + ([C]+[G])\dot{\bar{s}} + ([K]-[\Delta K(t)])\bar{s} = \omega^2 \overline{Re}^{i\omega t} + \overline{S}_g \qquad (1)$$

Few structural dynamic models of dynamic, *in situ*, gear tooth fracture appear in the literature. However, variable stiffness tooth profiles have been modified for use in dynamic simulation of a root fracture in a gear tooth. [13] The damaged tooth's stiffness profile is lessened by some degree (that is assumed proportional to the damage) per damaged gear mesh contact cycle. Figure 3 shows the stiffness profile used for the current modeling effort. Additional information regarding this model and the simulations performed can be found in [14].

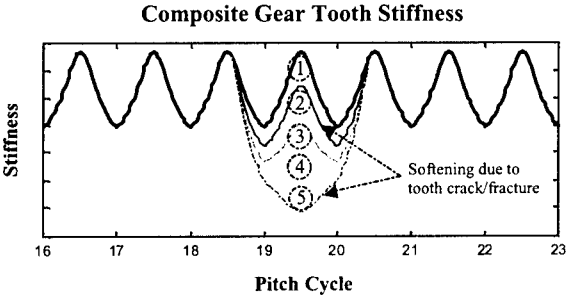


Figure 3. Nominal Gear Tooth Stiffness Variation for a Cracked Tooth

Diagnostic Figures of Merit: Many types of defects or damage in rotating machinery are manifested as increases in the vibration levels. These vibration levels can be converted to electrical signals for data measurement recording through the use of accelerometers. In principle, information concerning the relative condition of the monitored machine can be extracted from this vibration signature, and health assessments can be made through the comparison of the vibration response with prior responses to identify any anomalous

conditions. In practice, however, such direct comparisons are not effective mainly due to the large variations between subsequent signals. Instead, several more useful techniques have been developed over the years that involve feature extraction from the vibration signature. [15] Generally these figures of merit, or "features", are more stable and well behaved than the raw signature data itself. In addition, the features constitute a reduced data set since one feature value may represent an entire snapshot of data, thus facilitating additional analysis such as pattern recognition for diagnostics and feature tracking for prognostics. Moreover, the use of feature values instead of raw vibration data will become extremely important as wireless applications, with greater bandwidth restrictions, become more widely used.

The feature extraction method may require several steps, depending on the type of feature being calculated. Some features are calculated using the "conditioned" raw signal, while others may use a time-synchronous averaged signal that has been filtered to remove the "common" spectral components. A detailed discussion of a variety of feature processing methods is provided in [16].

Many features have been developed and are discussed in the literature. [15] The results presented in this paper will focus on only a few of the common features, namely FM4, FM0, NA4, M6A, and M8A. A detailed discussion of each of these features can be found in [16].

Feature Results: Simulations were performed for several degrees of tooth softening, as illustrated in Figure 3, to generate torsional acceleration predictions. Selected features were then applied to these signals and compared to the features obtained from the MDTB acceleration measurements. The comparisons focused on data obtained during MDTB Run 14 since this run resulted in gear tooth breakage and has a ground truth capability via borescope images taken during the run. These images, albeit limited in their ability to show tooth crack lengths, will facilitate comparisons between the simulated and empirical results. A plot of one of the common diagnostic features, FM4, is provided in Figure 4 for Run 14 during the time period several hours prior to the fault initiation and through the end of the run.

As shown in this figure, FM4 begins to react prior to any visible damage. This area is highlighted in Figure 4 and is labeled "incipient tooth crack". All comparisons in this paper will focus on this region since the simulations were performed for the degradation of one gear tooth. While not the intent of the present work, multiple gear tooth faults could be simulated by modifying the tooth stiffness profile shown in Figure 3. However, the number of faulty consecutive gear teeth that can be modeled using this method will be limited by the contact ratio of the gear set.

One difficulty in comparing the theoretical and empirical results involves relating the tooth stiffness to a particular point in the test run. The limited capability to ground truth the test data to an actual crack length precludes an accurate estimation of the stiffness parameter. Moreover, the complex geometry of the helical gear set further obscures the correlation between tooth damage and stiffness change. Therefore, the results presented below should be interpreted with the understanding that the location of the simulated results are not necessarily tied to the abscissa in each plot. In fact, one method for determining a system's degradation level would be to match measured feature values with simulated features based on a specific degradation

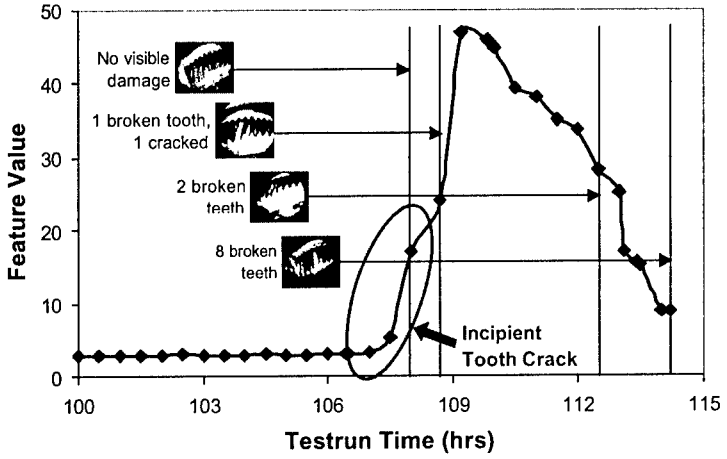


Figure 4. FM4 Applied to MDTB Run 14 Accelerometer Data

using a validated model. *In other words, a validated model could be used to infer the actual degree of damage by simply correlating and aligning the relevant diagnostic features.*

Plots of feature values for both the measured MDTB data and the simulated data are provided below for FM4, NA4, FM0, M6A, and M8A. The simulated data points are numbered such that they can be related to the stiffness profile shown in Figure 3.

Except for FM0, each of these plots show increased feature levels at around 107.5 hours into the test run for the MDTB data. The simulated data for tooth profile number 4 closely matches the results of the features at this point, and thus were plotted accordingly. The results for the first 3 profiles did not show any significant increase in value, and some actually decreased. The results for the 5th profile were plotted at 107.7 hours into the test run. Assuming the model provides an accurate representation of the system, these results could be used to infer the actual state of tooth degradation. However, note that in addition to the modeling approximations these results assume damage is limited to a single gear tooth. It is unclear at this point what caused the discrepancies between FM0 for the data sets. Further investigation is required to resolve these differences.

A descriptive overview and the respective fault sensitivity for these figures of merit with supporting references is provided in [17]. FM4 is a bootstrap recognition figure of merit and depends upon the normalized kurtosis value. Since the simulation includes the parametric excitation forces due to stiffness change, there is an expected increase in kurtosis (4th moment) energy that can be correlated with the experimental data. Clearly, the simulated FM0 feature, which should vary with the amplitude of the mesh tones in the gear average, is affected by the change in compliance. Thus, the traditional features dependent on higher order moments (4th, 6th, 8th) seem to be sensitive to the parametric excitation produced by stiffness profile changes in

the simulation. This excitation produces the amplitude and phase modulations that typically occur in geared systems. Amplitude modulation produces sidebands around the carrier (gear meshing and harmonics) frequencies and in non-faulted components are often associated

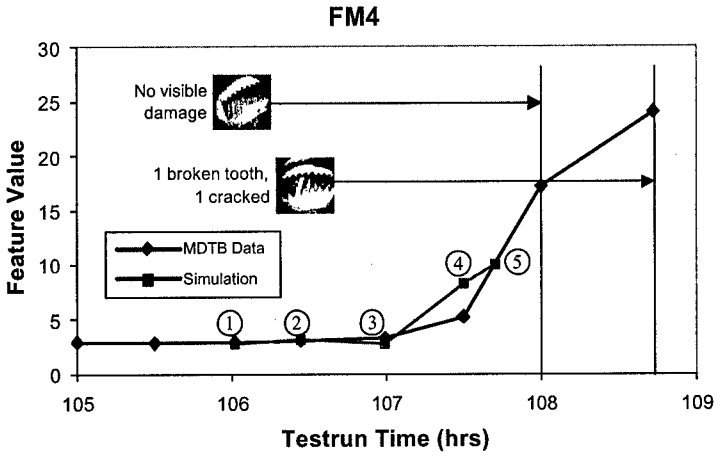


Figure 5. FM4 Diagnostic Feature

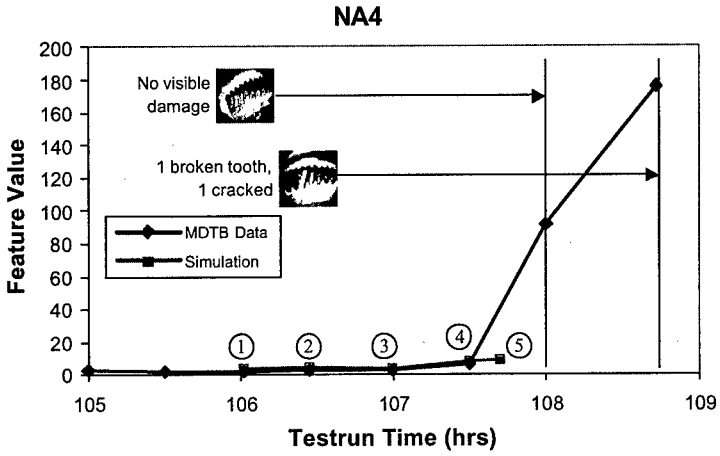


Figure 6. NA4 Diagnostic Feature

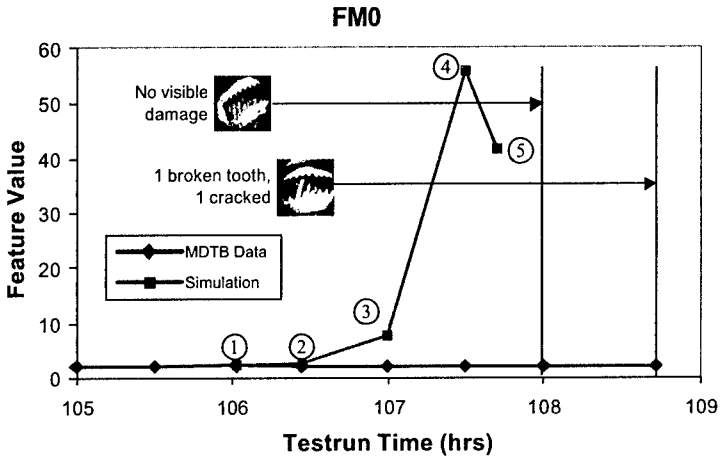


Figure 7. FM0 Diagnostic Feature

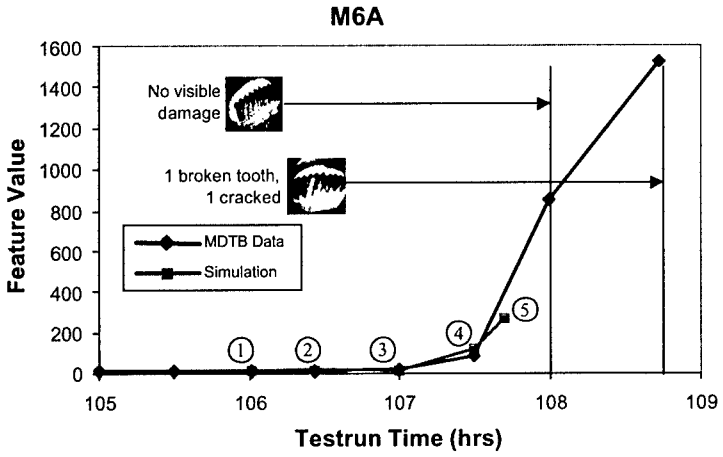


Figure 8. M6A Diagnostic Feature

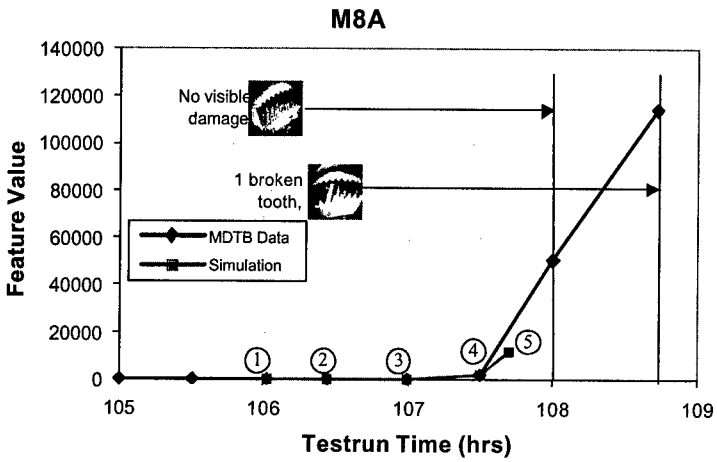


Figure 9. M8A Diagnostic Feature

with eccentricity, uneven wear, or profile errors. Phase or frequency modulation will produce a family of sidebands. How these occur in real systems will either add or subtract to produce an asymmetrical family of sidebands.

The lower order figure, FM0, and modified 4th order moment, NA4, do not provide correlation between the experimental and simulated cases. The reasons are not clear. Perhaps the mesh tone energy level is not suitably impacted by the simplified stiffness profile, thus causing the large values for FM0. The apparent correlation of FM4 but not NA4 leads one to conjecture that some artifact of the processing has produced this effect. Clearly, this is an area of future investigation among others.

Future Work: There are several aspects of this damage modeling effort that can be extended in the future. Accurate methods for relating gear tooth crack length to the composite mesh stiffness is one area for investigation. Better estimates may be obtained based on a finite element model of the gear mesh. The ability to simulate damage to multiple, juxtaposed gear teeth represents another fruitful area for future work. This capability would be helpful for identifying advanced damage conditions. Another extension of this effort would be to expand the model to include the gearbox casing and compare features taken from vibrations measured on the MDTB gearbox casing with the simulated results. Such a model would allow a more accurate representation of the system and would facilitate simulations of other fault types.

Acknowledgment: This work was supported by the Office of Naval Research through the *Multidisciplinary University Research Initiative for Integrated Predictive Diagnostics*

(Grant Number N00014-95-1-0461). The content of the information does not necessarily reflect the position or policy of the Government, and no official endorsement should be inferred.

References:

1. Kozlowski, J.D., and Byington, C.S., 1996, *Mechanical Diagnostics Test Bed for Condition-Based Maintenance,* ASNE Intelligent Ships Symposium II, November 25-26, 1996.
2. Byington, C.S., and Kozlowski, J.D, 1997, *Transitional Data for Estimation of Gearbox Remaining Useful Life*, 51st Meeting of the MFPT, April 1997.
3. McClintic K. T., et al, *Residual and Difference Feature Analysis with Transitional Gearbox Data*, Proceedings of the 54th Meeting of the MFPT, Virginia, May 2000.
4. Begg, C. D., Byington, C. S., and Maynard, K. P., *Dynamic Simulation of Mechanical Fault Transition*, Proceedings of the 54th Meeting of the MFPT, Virginia, May 2000.
5. Natke, H. G., and Cempel, C., *Model-Based Diagnosis – Methods and Experience*, 51st MFPT Proceedings, April 1997.
6. LaLanne, M. and Ferraris, G., **Rotordynamics Prediction in Engineering**, 2nd Ed, John Wiley and Sons, Chichester, England, 1998.
7. Choy, F.K., et al., 1992, *Modal Analysis of Multistage Gear Systems Coupled with Gearbox Vibrations*, Journal of Mechanical Design, Vol. 114, pp. 486-497.
8. Vinayak, H., et al., 1995, *Linear Dynamic Analysis of Multi-Mesh Transmissions Containing External, Rigid Gears*, Journal of Sound and Vibration, Vol. 185, No. 1, pp.1-32.
9. Mark, W.D., 1989, *The Generalized Transmission Error of Parallel-Axis Gears*, Journal of Mechanisms, Transmissions, and Automation in Design, Vol. 111, pp. 414-423.
10. August, R., and Kasuba, R., 1986, *Torsional Vibrations and Dynamic Loads in a Basic Planetary Gear System*, Journal of Vibration, Acoustics, Stress, and Reliability in Design, Vol. 108, pp. 348-353.
11. Przemieniecki, J. S., **Theory of Matrix Structural Analysis**, McGraw-Hill, 1968.
12. Kahraman, A., *Effect of Axial Vibrations on the Dynamics of a Helical Gear Pair*, Journal of Vibrations and Acoustics, V 115, p. 33, 1993.
13. Choy, F.K., Polyshchuk, V., Zakrajsek, J.J., Handschuh, R.F., and Townsend, D.P., 1996, *Analysis of the Effects of Surface Pitting and Wear on the Vibrations of a Gear Transmission System*, Tribology International, Vol. 29, No. 1, pp. 77-83.
14. Begg, C. D., and Byington, C. S., *Gear Fault Transition and Ideal Torsional Structural Excitation*, Proceedings of COMADEM 2000, Houston, TX, December 2000.
15. McClintic, K., *et al*, *Residual and Difference Feature Analysis with Transitional Gearbox Data*, 54th Meeting of the MFPT, Virginia, May 2000.
16. Lebold, M., *et al*, *Review of Vibration Analysis Methods for Gearbox Diagnostics and Prognostics*, 54th Meeting of the MFPT, Virginia, May 2000.
17. Byington, C. S., et al., *Vibration and Oil Debris Feature Fusion in Gearbox Failures*, 53rd Meeting of the MFPT, April 1998.

REAL-TIME CONDITION BASED MAINTENANCE FOR HIGH VALUE SYSTEMS

William W. Matzelevich

Arete Associates
Crystal Square Two, Suite 703
Arlington, VA 22215
matzelevich@arete-dc.com

Abstract: Many industries operate high value equipment – often remotely -- that requires reliable performance in severe environments. Similarly, the U.S. Navy's submarine Towed Array Systems (TASs) stress conventional approaches to operating and maintaining this system level capability comprised of integrated hydraulic, mechanical, electronic and acoustic sub-systems.

The Navy invested in a Condition Based Maintenance (CBM) proof of concept for an individual ship TAS by developing the Thinline Health Monitoring System (THMS). THMS collects real-time discrete reliability data and synchronizes this data with other historical information and the TAS's current condition assessment. As a predictive "intelligent code" it uses Bayesian Belief Networks (BBNs) to extract the full value of real-time data and provide a complete range of system performance evaluations -- from diagnosis to prediction.

Drawing upon THMS' success, the U.S. Navy supported expanding this capability fleet-wide to encompass health assessments of the entire submarine TASs population. Plans have been developed to build a relational database that is accessible to a geographically separated towed systems community via the Internet for interactive analysis and diagnostics.

These system level analyses and first principal processes are directly translatable to other government and commercial critical systems that cannot afford unscheduled -- or unnecessary -- maintenance.

Key Words: Condition based maintenance; Maintenance; Prognostics; Reliability

INTRODUCTION: Submarine Towed Array Systems (TAS) stress conventional approaches to operating and maintaining a system level capability necessary for ships' operations. TASs are mission essential for obtaining acoustic information in support of a high percentage of critical submarine deployments. By itself, a TAS is a complex configuration that requires integrated remote operation of hydraulic, mechanical,

electrical, electronic and acoustic subsystems in a severe ocean environment. It is nearly impossible to observe the full functioning of each TAS component during operation. Further if malfunctions occur at sea, most failures require repairs to be deferred until return to port. Repairs are frequently costly, and when performed waterborne, have the potential to result in negative work through repair activity induced failures associated with poor accessibility and an adverse repair environment. Adopting a prognostic Condition Based Maintenance (CBM) capability completely alters the TASs' maintenance landscape by monitoring current system conditions and predicting failures so that necessary repairs can be completed in advance under favorable conditions.

At the individual ship system level, the U.S. Navy invested in a proof-of-concept Thinline Health Monitoring System (THMS) for an OA-9070/TB-23 thinline towed array baseline configured SSN 688 Class submarine [1]. THMS provides a real-time method for assessing the current condition of the TAS and demonstrates the ability to dynamically predict future system health. The principal elements that support this capability are real-time sensor inputs, a mature Reliability Centered Maintenance program, and an embedded Bayesian Belief Network (BBN) "intelligent" code. Having demonstrated a prognostic, next generation maintenance capability for individual systems, the U.S. Navy funded CBM concept development for the fleet population of towed arrays. Integral to this capability is a comprehensive discrete historical and real-time web-based relational database and a powerful software toolbox to permit diagnostic and prognostic information mining simultaneously to geographically separated users (including operators, design engineers, vendors and logisticians). The inherent object oriented BBN tree framework permits the extension of the THMS health assessment output to serve as an input to the overarching BBN population model. A similar approach is directly applicable to other systems and industries.

BACKGROUND: The Navy's core maintenance efforts reside in a Reliability Centered Maintenance (RCM) program. RCM uses statistical information derived from historical data in order to develop maintenance practices for satisfactory system performance. The shortfalls associated with a stand-alone RCM program have been well-articulated [2][3]. Up front, its usefulness is dependent upon the quality of the database supporting the program. Many factors influence the nature of the database including the quality of data inputs, available validated test data and system level cause-and-effect data, and scalability [4]. Second, considerable system level understanding is required to select a suitable mathematical performance model. The modeler must inevitably make assumptions to characterize failure behavior. Frequently these assumptions are over-simplified and at best are based upon statistical formulations. Finally, there are often real-life events that introduce inaccuracies in the selected model. What effects do improved replacement components have on system reliability? Is it true that following maintenance, the component is returned to "as new" condition? Is it reasonable to assume that no degradation in system or component performance occurs when component inspections are performed as part of preventive maintenance? Can the model accommodate unrealized conditions characteristic of real systems such as localized hot bearings, peak starting currents, clogged lubrication ports, prolonged operational periods

in extreme environments or downstream effects when failures to non-critical components occur?

When RCM is applied to a complex system, it often practically leads to an exhaustive and costly maintenance program that requires a number of compromises affecting system operations and base-line program execution. Some of these compromises are (1) frequent system shutdowns to conduct maintenance inspections, (2) premature component replacement to hedge against system failures, and (3) high logistics costs – and hence total ownership costs -- to maintain a spare components stockpile. Submarine TASs typify the compromises and excessive costs that have become associated with maintaining a complex, high value system.

For submarine TAS, a voluminous database has been constructed from a number of sources to attempt to track and statistically gain performance insight. Beyond the shear exhaustive effort required to collect this data, because of the universe of reporting sources, data has frequently been incomplete or has been inadequately characterized to identify failure modes. The net result has been an amalgamation of data that has not provided the necessary insight to deliver the best possible RCM program, much less lead to a next generation CBM capability. Only through the goal to deliver a TAS CBM program has an initiative to make sense of years worth of TASs data taken shape and meaningful distribution models for selected critical components been developed.

Submarine TAS also labor under the real-world realities that can frustrate achieving desired system reliability. While component and subsystem design processes account for reliability (frequently through an expression of operational availability Ao), when these reliability factors are aggregated to reflect a composite system reliability design factor (if they can) the desired multi-system reliability either falls short of the goal or is strained over system life. For complex systems such as towed arrays, reliability must be addressed up-front through a separately considered maintenance program. TAS maintenance has attempted to improve reliability at great operational and fiscal cost through a variety of hedges that include frequent intrusive inspections and wholesale replacement of major components (such as the entire $1 million array itself) prior to submarine deployments.

THINLINE HEALTH MONITORING SYSTEM (THMS): THMS was built to provide a data collection, storage, condition health assessment and dynamic prognostic capability for a single submarine's TAS, which is comprised of four principal sub-systems; mechanical, hydraulic, electrical and acoustic signal path. Up front, the power of THMS resides in its ability to collect 39 reliability relevant signals real-time, analyze, and then integrate the incoming data stream with other logged discrete reliability data (including previous maintenance and repair or replacement actions). Synchronizing this incoming data stream with historical data results in a current condition assessment. The assessment is at once updated and passed to THMS system memory, and when combined with THMS' Bayesian Belief Network (BBN) "intelligent code," provides an empirical foundation to reliably predict future TAS performance at three operator selected future periods. The end result is an effective risk management system that avoids

"unexplained" reliability failures, provides updated analyses, and uses feedback to improve TAS performance over time [1].

Hardware: THMS hardware is a rugged stand-alone PC package (7.25" x 17.5" x 16") that provides the interface between the analog shipboard Thinline TAS equipment sensors and its embedded software. It is comprised of the following components:

- PXI Chassis with Back-plan Bus (National Instruments based)
- Embedded Computer with 4 GB Hard Drive
- 32 Channel 1.25 MB Analog-to-Digital (A/D) Converter
- Terminal Blocks
- Interconnecting Cables
- Electronic Components for Signal Conditioning and Attenuation
- Flat Screen Monitor
- Keyboard and Mouse

THMS taps the input signals through a high impedance multifunction Input-Output (I/O) card. The architecture of these cards include:

- NI-PGIA gain independent, fast-setting-time instrumentation amplifier
- DAQ-STC counter/timer
- RTSI multi-board/multi-function synchronization bus
- MITE PCI bus transfer interface, and
- Shielded, latching metal connectors [5]

The various input signals are paralleled into the THMS system through a breakout connection and terminal block. These signals are fed into the THMS to perform both single and multiple A/D conversions of a large number of samples. The THMS can perform multiple A/D conversions operations with programmed I/O, interrupts, or Direct Memory Access. The THMS interface software provides for operator interface and governs all aspects of interaction between the hardware and embedded BBN software.

Software: THMS uses a tailored version of two off-the-shelf software products, LabVIEW and Microsoft Visual C++. The software interface module collects and stores real-time input data at a rate of 4 Hertz from selected OA-9070/TB-23 sensors through the THMS Breakout Box. Additionally, the interface module facilitates the display of real-time data, provides data to the BBN decision tree for analysis, and receives analyzed and resulting metric data from the BBN for display. THMS provides multiple user screen variations for input and output displays, and at the functionally highest levels, data trending and system health

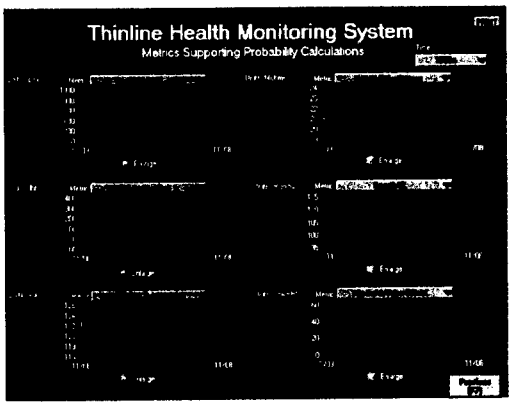


Figure 1. THMS Metrics Display Screen

assessments. Figure 1 illustrates the THMS Metrics Display Screen that is used to show reliability relevant metrics and provides a trend indication of key operational parameters critical to subsystem performance. Figure 2 illustrates the THMS Total and Sub-Group Health User Screen that provides the probability that the TAS will operate successfully. The probabilities are calculated for the total system and the four major system sub-groups using the embedded BBN software code [5].

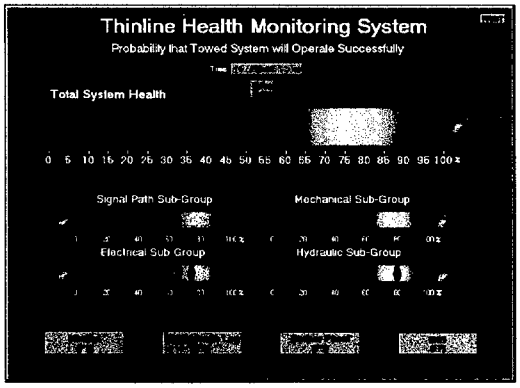


Figure 2. THMS Total and Sub-Group Health User Screen

Bayesian Belief Networks (BBN): At the core of THMS is a BBN structure. It serves as the analytical vehicle to interpret the TAS physical model through reliability analysis and directly provides failure likelihood estimates. Although there were other methods that could have been used – including neural net algorithm training -- BBNs were particularly well suited for TASs because they are tailored to a system-level physical model and are naturally object-oriented. Training an algorithm alone without regard to the physical cause for failures was limited by insufficient data. The data did not include all failure contingencies nor did it represent the actual operating environment. Further, algorithm training did not include a physical understanding that related to predictions and therefore, provided little diagnostic benefit. On the other hand, all that was required to construct the THMS BBN was a connection between an input measurement value and the likelihood that a hypothesis related to that measurement was true. A THMS BBN could be tested and constructed in a qualitative manner independent of data with data being needed to refine the model, not to define it. Finally, if new failure modes became known after model construction, they could easily be added without starting over again.

In general, each BBN component -- represented by a node -- receives input in the form of other measurements or likelihood ratios and produces output linked to other nodes in the form of likelihood ratios. The BBN structure, including the links constructed between nodes, provides a knowledge fusion that relates to the causes of each failure mode (or component) to their consequences (measurement). The BBN fuses system component performance specifications and normal operating parameters or failure probability distributions to form a relationship between each measurement and the likelihood for each possible hypothesis regarding healthy system operation [6]. Although BBNs are typically designed to convert measurements into likelihoods, these "measurements" can take on almost any form from a data concentrator to the system response function of component failure, or even alternative Artificial Intelligent sources. BBNs can even

calculate the posterior probability of a node failure given the evidence of all other nodes. In this way the probabilities are revised as our uncertainty and knowledge changes [7].

As applied to THMS, the top-level BBN nodes for a TAS are provided in Figure 3. The figure shows the connection between the probability of operation of each TAS subsystem and the probability of operation of the TAS as a whole. Each ellipse represents a node in the decision tree. Each sub-system (node) is further modeled and coded in its own complex decision tree. Having been constructed from the physical model and an understanding of the TAS, the embedded BBN or "intelligent code" takes the conditions of a single TAS as provided by the sub-system level sensors and turns them into predictions of successful operation.

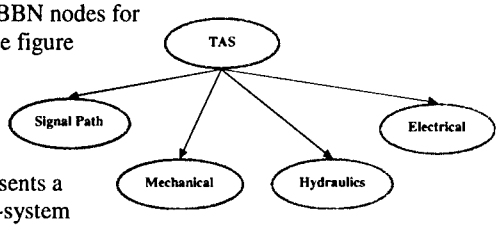


Figure 3. THMS Towed Array System BBN

PROGNOSTIC CAPABILITY FOR FLEET TASs POPULATIONS: Because the BBN schema identifies the probability for a component or sub-system as part of the conditional probability for the entire system, it followed that the THMS BBN tree could be extended to the program level. Figure 4 illustrates this top-level BBN structure. Although yet to be analyzed to this top program level, the U.S. Navy has studied the logical extension of the THMS BBN to the second order Fleet-wide tier [8].

Developing a Fleet-wide prognostic CBM TAS capability made sense from a number of

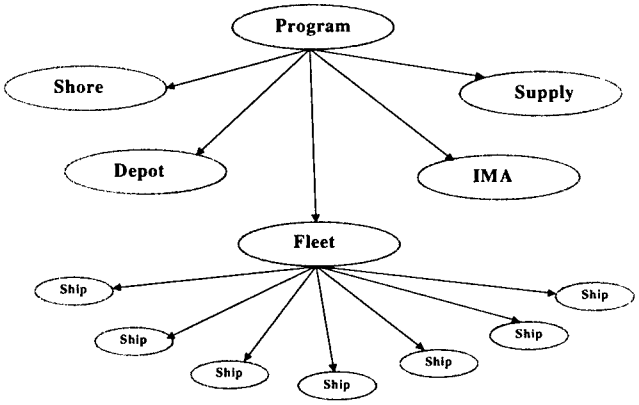


Figure 4. Top-Level TAS BBN Structure

perspectives. By assessing the health of many TASs simultaneously, a feedback process could be established for individual TASs to define a more accurate normal operating condition. This feedback is essential to ensure that individual systems are compared equally to a single standard, and that the standard is fully representative of composite operating system behavior. Second, a comprehensive discrete and real time database could be built that is updated real-time by established data format, validation, and consistency procedures. This database will support the highest quality answers to traditional RCM maintenance failure questions, real-time diagnostic evaluations, and will establish a baseline TAS prognostic CBM capability. Third, as a result of this meaningful, population-wide database, a powerful software toolbox can be fully utilized. Tools can be provided that go far beyond the straightforward cataloguing of Out of Commission (OOC) counts, failures by type and description, and graphing trends over time. Finally, making the database Internet accessible and using web-based formats allows: comparison checks between past performance and current trends using defined metrics, statistical analysis of data referred by specific categories through web browsers, and longitudinal trends and cross-correlation examinations. These are just some of the more meaningful capabilities that will be possible to a geographically separated towed systems community.

Database: Data is entered from a number of sources to an automated centralized database connected to the Internet. Entries can occur either manually through the THMS data logger feature or by Personal Digital Assistants (PDAs) carried by field personnel, or automatically through the automatic THMS function. The data is warehoused on a regular schedule by using Data Transformation Services (DTS) and is saved as a package that includes tasks executed in a coordinated sequence of steps. The DTS units also set aside data entries that have errors or inconsistencies with previously collected data in an Exception Log that is available to the Database Administrator for data reconciliation. Based upon the type of data errors and inconsistencies observed, a subsequent rule-based set of DTS units can subsequently be constructed to accept certain data characteristics as valid.

After the build has been created and saved, it is completely self-contained and can be retrieved by using the database server's Enterprise manager or a DTS utility. In this fashion, the data warehouse provides users with rapid analysis of large data volumes – including investigations of hypotheses for comparisons across and between data groups -- through a variety of network management protocols including indexing, data base de-normalization, star or snowflake schema, and aggregations.

Finally, as the database grows over time, overall performance may be affected by a degraded system response time. Assuming 570 MB of data per month is accumulated per submarine, the database will grow at 340 GB per year with a fleet of 50 submarines. In order to retain a historical database for long-term analysis while maintaining system performance, the database can be archived annually on discs that are retained indefinitely and later disposed when no longer needed [9].

Toolbox: Because of past processing limitations, simplistic analytical assumptions were often made in order to address more complex reliability issues. In the end the resultant analytical conclusions were contradicted by experience. Today's processors permit more powerful survival analysis techniques similar to those pioneered for medical research. These techniques permit statistical inference and data extrapolation within error bounds which in turn, provides reasonable tools to predict the degree of difference between similar articles, the peril rate of articles following from identified failure modes, and the effects of time on future failure events. To cite one example, a recent reliability study on towed array modules showed that the previous simple exponential failure distribution fell far short of accounting for the effects of module aging. Through proportional hazards modeling and competing risk formulations, a two parameter Weibull distribution for module life has since been highly correlated with historical data. While a marked improvement, this was done acknowledging that the time basis for analysis was from available data. With the construct of a rigorous database, it will be possible to provide further time-based granularity for specific times of operation, handling, power-on and other periods of interest. This example points to the feasibility and need for a TASs Toolbox that provides a family of statistical techniques to (1) validate a specific mathematical model in order to describe failure and repair times, and (2) describe specific systems and deliver comparisons between groups within a population [8].

In the first case, the Toolbox can be equipped to investigate situational questions such as how to understand the interaction between external hydraulic oil and TAS performance. This can be accomplished by executing Analysis of Variance (ANOVA) routines that rely upon conducting comparisons or groupings. Other tools used to describe failure and repair times could include Hazard Functions, serial Correlation Tests, Trend Tests, and Post Maintenance or Inspection Differential Failure Rate Analysis.

In the second case, an application would be to expand the use of Fleet TAS operational availability (Ao). Currently, Ao is determined through steady state Markov chain probabilities for towed array and handler type only for attack submarines. With daily TAS data from all submarines, large-scale matrix multiplications for each category of Ao calculation will be required at the minimum once per month information interval. The Toolbox will not only be equipped to accommodate the additional data, but will be able to extend this Ao methodology to cover a variety of other meaningful subgroups of ships and systems including ballistic missile submarines, submarines readying for deployment and those actually deployed. Another useful comparison between groups would be a comprehensive survival analysis for array signal path components. The fidelity of component reliability distributions has already been improved through the comparison between the types of components within populations as newer components are manufactured and placed into service. Improved survival probabilities – beyond the 17 components of interest in the TAS signal path – will be possible through advanced competing risk and proportional hazards modeling that is based upon failure causes and sequential failure history. Significantly, this approach will allow the unique TAS configuration on any specific submarine to be analyzed for life predictions based upon individual components.

It's clear that by putting these tools in the Toolbox, significant diagnostic improvements are possible for TAS RCM. Most significantly, however, these tools will at the same time create the conditions for a BBN based Fleet-wide prognostic CBM capability. This results from the ability to include in the Toolbox the ability to formulate conditional probability density functions for all hypotheses being tested or by using inference. The former technique relies upon a fundamental understanding of the physical model while the later requires a statistical relationship between a measurement such as stress, and a key characteristic such as strength in order to provide the terms for the probability of failure given the measurement. These reliability connections appear in the BBN structure and fuse knowledge of the system with the expected behavior of each failure mode over time in order to predict the future time to the next fault.

Displays: Integral to the Toolbox's utility is its associated information displays. To achieve the highest possible value, displays will be provided using an Internet web page format with the capability to interface to the database from queries initiated from any user. Beginning with a Home Page, obvious trends will be highlighted in one section and the format will provide a convenient and consistent location for the metrics and details of daily importance. One-liner links to news stories with a somewhat larger description of events will be available that will provide follow-on links to the full story. Also included will be a search box that will allow Boolean operators to make specific searches. The Home Page will also include box scores and program indicators with recent changes, and links to categories of information depending upon the specific user's interest at the time.

Similar to financial web sites, beyond the Home Page will be an ever-expanded view of the data lying behind the information. By clicking on an area of interest in a data table or graph, a complete palette of selections will become available to display additional charts, tables, or graphs to permit analysis through comparison, correlation, or statistical review. Based upon user request, information and data can be displayed into numerous population sorts to include time dependent or independent displays, submarine class, geographic location, operational status, and array type and revision.

In the background, the Toolbox will continue to provide data analysis for formal routine reports. Not so obvious correlations, cross-correlations and partial correlations that exceed a specific coefficient threshold can be displayed as an optional page linked to the Home Page. Alert and focus pages tied to subject matter specific to a personalized user page can also be provided. For example an array engineer could add towed array and module subjects to his personalized page. Then, any time there are specific alerts to arrays or modules, the engineer would find an alert on his personal Home Page.

Finally, this web page structure significantly facilitates top-level TASs population management. End-of-year predications can accompany any Home Page linked item in order to provide managers useful indicators for what to expect from available data. Together with their accompanying error bounds, these predictions will put the data in perspective and will highlight the importance of taking action if the predicted outcome is undesirable [8].

CONCLUSIONS: The same characteristics of submarine TASs -- system complexity, significant impact to operations should failure occur, high repair costs, and adverse operating environments – are common to numerous other military and commercial projects. The costly inefficiencies of companion maintenance programs geared to attempt to achieve desired system reliability goals are also common.

There is a process that leads to the next generation maintenance solution for these complex systems. This process is defined by constructing a physical model that is based upon a system understanding and an analysis of existing reliability and maintenance data. As part of this development, additional data reduction or information is typically identified to create and validate the final model. Once the model is built, a solution can be engineered to meet reliability goals. Three tools in particular have proven useful to develop the solution; (1) a flexible architecture that permits predictive performance, (2) optimized processing power, and (3) web-based Internet accessibility.

Although other flexible architectures exist, BBNs have proven a particularly good fit for this system level approach. This is because of the BBNs' object-oriented structure, physical model based estimations, ability to fuse data with system knowledge, and tree structure. BBNs permit both externally and internally acquired system level knowledge to be included in prognosis. Finally, BBNs allow for a scalable capability that can be readily adjusted as system level understanding and growth both occurs.

Today's technology permits real-time processing of large data quantities and using sophisticated statistical models that go far beyond the simple assumptions that were common in the past. Because processors can be equipped with powerful analytical toolboxes, improved statistical reliability products and entirely new prognostic capabilities are possible.

Leveraging Internet technical and functional architectures provides the final piece for this next generation prognostic CBM. By creating relational databases founded upon Data Transformation Services, significant improvements can be made to ensure data validation and consistency and to conduct rapid analyses of data warehouses. Finally, simultaneous access to multiple geographically separated users qualitatively improves data analysis, focuses time critical diagnostic action, and creates shared efficiencies to program managers and site supervisors.

Following this approach used by the U.S. Navy for submarine towed arrays has the potential to significantly benefit any military or commercial sector that deals with costly time related failures that can be physically modeled. Particular cost-benefits may be possible for those high value systems that have significant maintenance costs and for which reliable system operation is important.

ACKNOWLEDGEMENTS: This work was supported by a Phase I SBIR Award to Life Cycle Engineering, Inc. and Arete Associates under NAVSEA contract N00024-00-C-4094.

The author thanks Harry Bishop for his encouragement and review, and Randy Potter, Danny White, John Holdsworth, and Gary Chamberlain for their reviews.

REFERENCES:

[1] Thinline Health Monitoring System (THMS) Technical Data Package Revision 1, prepared for the Naval Sea Systems Command PMS 425, by Life Cycle Engineering, Inc. and Arete Associates, 4 January 2000.
[2] Magdi Essawy and Saleh Zein-sabatto. "Measures of Effectiveness and Measures of Performance for Machine Monitoring and Diagnosis Systems," Proceedings of University of Tennessee on Maintenance and Reliability Conference, 12 May 1999, p. 42.02.
[3] Von Alven, William H. Reliability Engineering, New Jersey: Prentice Hall, 1964, pp. 389-396.
[4] Nickerson, G. William, Michael Van Dyke, and Carl S. Byington. "Quantification and Validation of Diagnostic Techniques Using Fleet Data," Proceedings of American Society of Naval Engineers on Condition-Based Maintenance Symposium, (Ed.) James A. MacStravic, July, 1998, pp. 177-179.
[5] Bishop, Harry, Brian Domrese, and Ronald D. Thomas. "Complex Government System Health Monitor and Predictive Tool," presented at National Instruments Best Applications Conference, August 2000.
[6] Littlewood, Martin Neil Bev and Norman Fenton. "Applying Bayesian Belief Networks to System Dependability Assessment," Proceedings of Safety Critical Systems Club Symposium, Leeds, 6-8 February 1996.
[7] Ellsworth, Kirk and Harry Bishop. "Using Bayesian Belief Networks to Predict System Reliability for Condition-Based Maintenance," April, 2000.
[8] SBIR N00-048 Phase I Final Report: Design and Develop a Real-time On-line RMA Trends and analysis Reporting assessment Database for Towed Systems, prepared for the Naval Sea Systems Command PMS 435, by Life Cycle Engineering, Inc. and Arete Associates, 29 November 2000.
[9] Hsu, Meeiyun, Towed Array System Architecture Design Document, November 2000.

REVIEW OF VIBRATION-BASED HELICOPTERS HEALTH AND USAGE MONITORING METHODS

Victor Giurgiutiu, Adrian Cuc, Paulette Goodman

University of South Carolina, Department of Mechanical Engineering

Columbia, SC 29208

Abstract: The purpose of this paper is to review the work that has been done in the past years by various researchers in vibration based health and usage monitoring and to identify the principal features and signal-processing algorithm used to this purpose. The damage detection concepts and the signal analysis techniques will be reviewed and categorized. Latest advances in signal processing methodologies that are of relevance to vibration based damage detection (e.g., Wavelet Transform and Wigner-Ville distribution) will be highlighted. These vibration signal-processing methods play an important role in early identification of incipient damage that can later develop in a potential threat to the system functionality, and even a flight accident. In aerospace applications, HUMS capabilities are to minimize aircraft operation cost, reduce maintenance flights, and increase flight safety.

Key Words: HUMS; Wavelet Transform; Wigner-Ville distribution; O&S; Machinery Failure Prevention; Neural Networks; Helicopters

INTRODUCTION

Machinery Failure Prevention is an important component of the maintenance activity for most engineering systems. Helicopters are continuously subjected to periodic loads and vibration environments that initiate and propagate fatigue damage in many components. Current helicopter maintenance practice requires a large number of parts to be monitored and replaced at fixed intervals. This constitutes an expensive procedure that adds considerably to the helicopter Operational and Support (O&S) costs. Health and Usage Monitoring Systems (HUMS) have been developed in recent years to detect incipient damage in helicopter components, predict remaining life, and create the premises for moving over from scheduled based maintenance to condition based maintenance. Of prime importance in such a HUMS system is the capability to evaluate the damage or undamaged state of a critical component using only the vibration data signals recorded during flight and ground operation. With such capability, the need for frequent disassembly and bench checking of certain critical components can be reduced and ultimately eliminated, with important ancillary savings in the O&S costs. However, to achieve such capability, advance vibration signal processing algorithms are necessary that can distinguish the damage related features from the background and system noise perturbations. Enhancements such as signal averaging, cepstrum and kurtosis analysis, time-frequency domain analysis, Wavelet transform, neural network systems have shown promising results (McFadden, 1985; Monsena, 1994). However, the challenge remains to translate this knowledge into fault prediction.

The earliest work toward detecting defects in helicopter gearboxes used a high-resonance technique and was a *off-line* monitoring tool focused on finding the exact location of the defect. Based on this technique, the frequency spectrum of the vibration signal is search to find a change in the spectral line at one particular frequency associated with any particular faults. This method has proved good results for simple bearing configurations but is not satisfactory for complex configuration or for a large damage (McFadden, 1985)

Fraser and King (1990) using kurtosis observed that when a gearbox component is severely damaged multiple impulses will appear in the frequency domain, resulting in a cumulative response that tend to reduce the kurtosis value. Another approach is to apply cepstrum based techniques, in an attempt to condense the frequency domain information into an easier to interpret form, thus providing a practical system for routine prognostic monitoring.

Forester (1990) demonstrates that time-frequency techniques, such as Wigner-Ville distribution can describe how the spectral content of the signal is changing with time and provided a framework for developing robust detection and classification schemes for helicopter gearbox faults.

WAVELET TRANSFORM METHOD FOR HELICOPTER HEALTH MONITORING

The wavelet transform is a signal-processing tool, which allows both the time domain and frequency domain properties of a signal to be viewed simultaneously. Performing a wavelet transform consist of convolving the signal with time shifted and dilated. The result of wavelet transform will be a set of coefficients, which are function of time and frequency, also called scale. These coefficients can be used to form a unique mean square wavelet map, a time-frequency representation of the signal.

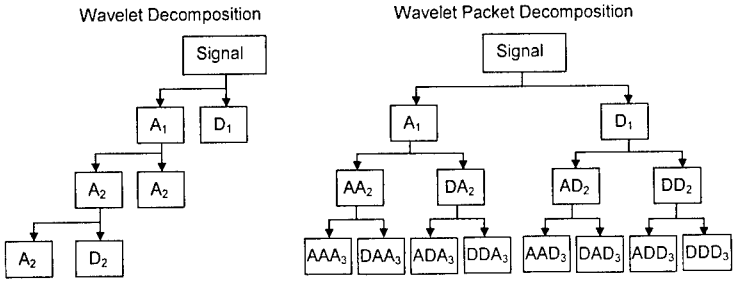


Figure 1 A comparison of the DWT algorithm and the DWPT algorithm. A_i is approximation at level i (low frequency) and D_i is detail at level i (high frequency). (Samuel et al., 1998)

Mallat (1988) discovered a recursive algorithm to compute the DWT consisting in basic wavelet function to form sets of filters, each one consisting of lowpass and highpass filter. The signal is pass through the first set of filters and the result will be two signal each with half of number of coefficients as the original signal. The signal formed using the lowpass filter and thus containing the low frequency information is known an the approximation, and the second signal formed using the high pass filter and thus

containing the high frequency information is known as the detail. For the second recursion the approximation is passed through the next set of filters and so on until an approximation and detail each consisting of one coefficient are formed.

Newland (1993) presented the harmonic wavelet basic function, which is a smooth wavelet providing high numerically accuracy in signal reconstruction and it is completely band limited in the frequency domain. A consequence of the above is that DHWT need not be restricted to octave frequency bands. This form of DWT is known as the discrete harmonic wavelet packet transformation (DHWPT). The algorithm to compute DHWPT is the Mallat recursive algorithm and a comparison between the two algorithms is presented in Figure 1

Samuel et al. (1998) collected and analyzed data from an OH-58A main rotor transmission. The test was run at 6060 rpm (100% of the maximum speed), which resulted in a mesh frequency of 573 Hz, for nine days, eight hours per day at a 117% design torque as part of an accelerated fatigue test. The results were represented in mean square DHWPT maps. The mean square wavelet maps clearly shown the presence of the fault in day nine. Using the normalized power computed for the mash frequency and its accompanying frequency bands the evolution of the fault from day one to day nine was identified (Figure 2a).

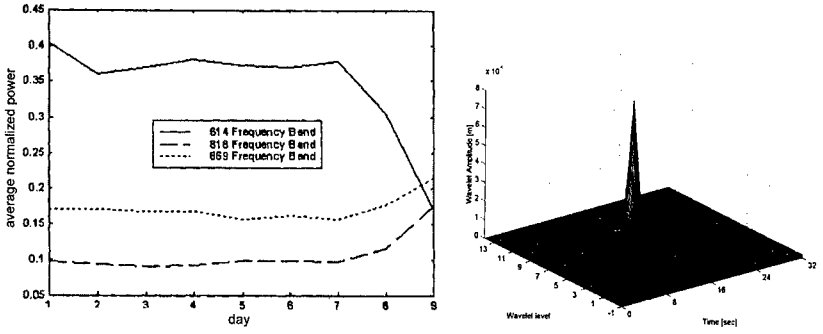


Figure 2 (a) Average normalized power index (Samuel et al., 1998); (b) Directional harmonic wavelet map: 10% crack with 5% noise (Kim and Ewins, 1999)

Kim and Ewins (1999) applied the directional harmonic wavelet transform (DHWT) to investigate the transient vibration response using a numerical simulation of a rotor system during the run-up period with a speed ramp rate of 100[rpm/sec] for the case of a rotor with 10% crack (transverse) depth relative to its diameter. To validate the advantages of DHWT, short time Fourier transform (STFT) was applied to the same set of data (noise contaminated signal). The results presented in Figure 2b reveal the advantage of DHWT over the STFT, because the results from DHWT are insensitive to the random noise while the STFT provides noise-contaminated results (Kim and Ewins, 1998).

NEURAL NETWORK-BASED AND NEURO-FUZZY HELICOPTER HEALTH MONITORING

A common trend in the recent technology and applied research efforts is to create smart systems that are capable of self diagnosis, assessment of self efficiency and operating condition, adaptation, and may be even self remedy. One source of inspiration for researchers is the human body. Studies of the microstructure of the nervous system and the decision making process of the humans have lead to the new concepts as neural network and neuro-fuzzy logic.

Monsena and Dzwonczyc (1995) proposed a hybrid (digital/analog) neural system as an accurate off-line monitoring tool used to reduce helicopter maintenance costs, and an all analog neural network as a real time helicopter gearbox fault monitoring. The hardware platform used is an analog neural network platform, Integrated Neural Computing Architecture (INCA/1). The vibration data were generated using an intermediate gearbox known as the Hollins TH-1L 42-deg tail rotor. Vibration data was recorded from two Endeveco 2220C accelerometers. A low pass filter with a cutoff frequency of 10 kHz was applied when data was generated. The main objectives of the hybrid and analog neural network are: fault detection, fault classification and fault identification. A comparison between the capability of the hybrid neural network to detect faults and the analog neural network is presented in Figure 3. The results indicate that a system employing 60-point DFT was capable of solving the fault detection problem. For the fault classification and identification problem, a 256-point DFT was required for perfect system performances. The performance results by using the all-analog neural network suggest that it is possible to achieve 100% fault detection with 0% false alarm rate.

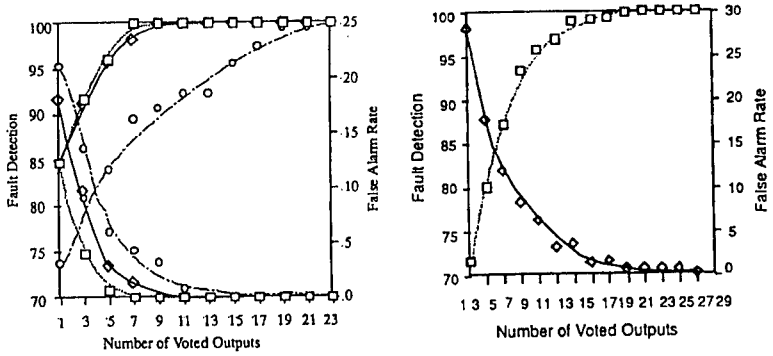


Figure 3 (a) Hybrid neural networks system performance, o = 60-point DFT, ◊ = 128, = 256; (b) All analog system performance, fault detection, ◊ = false alarm (Monsena and Dzwonczyc, 1995)

Essaway et al. (1998) presented an automated predictive diagnosis (IPD) techniques for monitoring the health of helicopter gearboxes. This technique is based on neuro-fuzzy algorithms for pattern clustering, pattern classification, and sensor fusion. The vibration data used was obtained from an aft main power transmission of a CH-46E helicopter. Frequency domain and wavelet analysis techniques were used to analyze the data and

prepare them for the neural network inputs. To train the neural network, a non-supervised learning algorithm known as Self Organizing Maps (SOM) was used. A feedforward backpropagation neural network was used to classify the different faults. In the preprocess part of the vibration data, both auto power spectral density (APSD) and wavelet coefficients were used. A list of the fault types is presented in Table 1. The results obtained using the first feature extraction method (129 points APSD) and second extraction method (wavelet transform) are promising and they are presented in Table 2 and Table 3. These results show that the neuro-fuzzy technique using both APSD and wavelet transform, even though classification results was not perfect for all sensors, produced 100% classification for all cases.

Table 1 List of the fault types created in the test gearbox (Essaway et al., 1998)

Fault #	Fault type	Fault #	Fault type
Fault 2	Planetary bearing corrosion	Fault 6	Helical idler gear crack propagation
Fault 3	Input pinion bearing corrosion	Fault 7	Collector gear crack propagation
Fault 4	Spiral bevel input pinion spalling	Fault 8	Quill shaft crack propagation
Fault 5	Helical input pinion chipping	Fault 9	No defect

Table 2 Neural Network Classification results using APSD features at 100% load (Essaway et al., 1998)

Fault	Load		Acc1	Acc2	Acc3	Acc4	Acc5	Acc6	Acc7	Acc8
Fault 2	100%	Train	100%	100%	100%	100%	100%	100%	100%	100%
		Test	100%	100%	100%	100%	100%	100%	100%	100%
Fault 3	100%	Train	100%	100%	100%	100%	93.3%	90%	100%	100%
		Test	100%	100%	100%	100%	90.9%	100%	100%	100%
Fault 4	100%	Train	100%	100%	100%	100%	86.7%	100%	100%	100%
		Test	100%	100%	100%	100%	81.8%	100%	100%	90.9%
Fault 5	100%	Train	100%	100%	100%	100%	70%	93.3%	100%	96.7%
		Test	100%	100%	100%	100%	72.7%	100%	100%	100%
Fault 6	100%	Train	100%	100%	100%	100%	80%	76.6%	100%	100%
		Test	100%	100%	100%	100%	27.3%	90.9%	100%	100%
Fault 7	100%	Train	100%	100%	100%	100%	100%	93.3%	100%	100%
		Test	100%	100%	100%	100%	100%	90.9%	100%	90.9%
Fault 8	100%	Train	100%	100%	100%	100%	96.7%	83.3%	100%	100%
		Test	100%	100%	100%	100%	81.8%	63.6%	100%	100%
Fault 9	100%	Train	100%	100%	100%	100%	90%	93.3%	100%	100%
		Test	100%	100%	100%	100%	63.6%	100%	100%	81.8%

Table 3 Nural network classification results using wavelet features (Essaway et al., 1998)

Accelerometer#	Fault2	Fault3	Fault4	Fault5	Fault6	Fault7	Fault8	Fault9
Acc 1	100%	100%	66.67%	100%	100%	100%	100%	88.89%
Acc 7 (14x14 SOM)	100%	88.89%	88.89%	100%	100%	100%	100%	100%
Acc 7 (9x9 SOM)	100%	88.89%	44.44%	100%	100%	100%	100%	100%

JOINT TIME-FREQUENCY WIGNER-VILLE DISTRIBUTION ANALYSIS TECHNIQUES

The signal processing methods for machine-health monitoring can be classified into the time-domain analysis, the frequency-domain analysis, and joint time-frequency domain analysis. Some of the parameters used in those methods are FM0, FM4, NA4, NA4*, NB4 and NB48* (Polyshchuk *et al.*, 1998). The Wigner-Ville distribution (WVD) is a joint time-frequency signal analysis. The WVD is one of the most general time-frequency analysis techniques, as it provides excellent resolution for accurate examination in both time and frequencies domains. Some of the problems encounters in using WVD are

related to the aliasing arising in the computation of WVD and to its the nonlinear behavior. To avoid the aliasing problem, the original real signal is transformed into the complex analytical signal. Polyshchuk *et al.* (1998) used the WVD techniques, introducing a new parameter, NP4, to experimental data obtained from a helicopter-tail gear transmission. The damage introduced is single-tooth damage in the tail gear. The WVD and the instantaneous power plot for an undamaged gear were analyzed. Two large components, the first and second harmonics of the gear-mesh frequency, were identified. For a damaged gear, the WVD and the instantaneous power plot were found to be quite different (Figure 4). A clear peak was identified in the WVD and power spectrum density (PSD) plots. This peak was not present in the WVD and PSD plots of the undamaged case. These results proved that the use of WVD could be a good tool in fault detection and failure prevention.

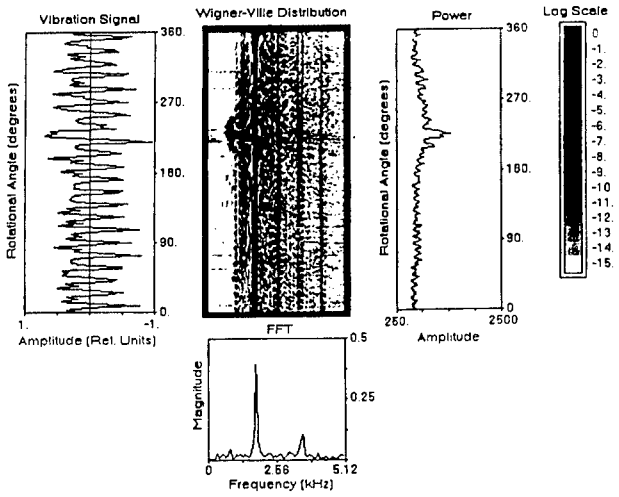


Figure 4. Accelerometer data for damaged gear (Polyshchuk et al., 1998)

VIBRATION ANALYSIS EDUCATION SOFTWARE

Successful predictive maintenance programs will enable an organization to reduce maintenance cost, reduce unscheduled down time, reduce catastrophic failures, improve safety and decrease maintenance man-hours. The keys to a successful condition-monitoring program are accurate prediction of machine faults, accurate repair recommendations, automation (to reduce human error), a refined reporting system and ease of operation and use. Commercial vendors, realizing the importance of creating sound maintenance programs, have

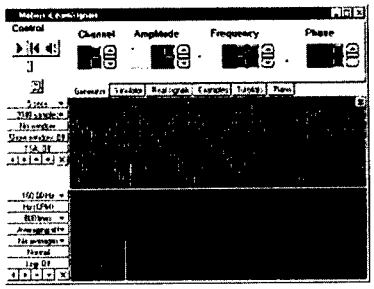


Figure 5 Mobius iLearn Interactive vibration signals.

introduced software programs and products to combat the high cost and complexities involved with equipment maintenance. In recent years due to the changes in technology, software now exists that is capable of exploiting vibration analysis algorithms and data analysis functions, presenting data with a user-friendly graphical user interface. Broadband and narrowband statistical analysis, complete spectrum normalization, trending, automatic peak identification, user selectable warning/alarm level functions are only some of the features offered in vibration analysis software today. This section will comparatively examine three different packages that allow vibration analysis, as well as a computer based training (CBT) programs.

MOBIUS ILEARN INTERACTIVE

In order to be more effective when monitoring a maintenance program, the end user should understand the actual signals being collected to properly interpret the results. A thorough understanding of the relationship between the machine and the characteristics observed will enable the end user to make educated interpretations. Mobius iLearn Interactive is a practical and theoretical training tool, useful to anyone interested in condition monitoring and vibration analysis, regardless of his or her experience in the field. The CBT is self-paced, incorporating simulations, animations, and samples of real data and diagnostic requirements – creating a realistic, interactive, and valuable learning tool. The complete curriculum is split into five, non-progressive, separate modules:

- iLearn Vibration — condition monitoring and vibration analysis learning is narrated while the end user interacts with diagrams and simulations.
- iLearn Hands-On – is a set of vibration measurements taken from a fault test rig. Over two hundred tests covering dozens of fault conditions are analyzed. Analysis can be done on the screen or downloaded into the end users data collector.
- iLearn Case Histories – is a library of spectra and waveforms taken from real machines. Digital recordings enhance the vibration analysis experience. Analysis can be done on the screen or downloaded into the end users data collector.
- iLearn Signals – is a virtual signal generator and spectrum analyzer. This software program will generate simple signals to teach waveforms and spectra. Advanced capabilities delve into signal processing.
- iLearn Machine Faults – allows the end user to model a machine to understand frequencies. The ease of drag and drop technology enable the end user to create a virtual machine and view simulated frequencies, waveforms and spectrum.

SPECTRAL VISUALIZATION AND DEVELOPMENT, SVD INC.

SVD Inc. provides free online courses on their web site. The Canadian based company manufactures vibration analysis software and devices. The courses range from introductory maintenance philosophies to diagnostic methods. The only requirement for the end users is a web browser and Macromedia's Shockwave plug-ins (available as a free Internet download). Twelve courses are offered; four are described below:

- Introduction to Mechanical Vibrations — the basic concepts of mechanical vibrations are presented using a mass-spring-damper example of a vehicle suspension. End users

gain an understanding of mechanical vibration, linear systems and system resonance while analyzing mechanical vibrations.

- Introduction to Machinery Signals – the end user is taught the basics of data acquisition, such as when and how to take measurements, aliasing and the alias foldover effect, and identifying time and stationary signals. Deterministic stationary signals and the processes that generate them are covered as well.
- Introduction to DSP (Data Signal Processing): Time and Frequency Domain – are two separate courses that are offered by SVD Inc. Data acquisition issues such as single channel and multi-channel analysis are taught along with unit of measurement. The concepts of mean, average and correlation, and how they relate to stationary signals are presented. The basic types of spectral plots and spectral analysis and their units of measurement are covered in frequency domain along with spectral estimators, parametric and non-parametric.

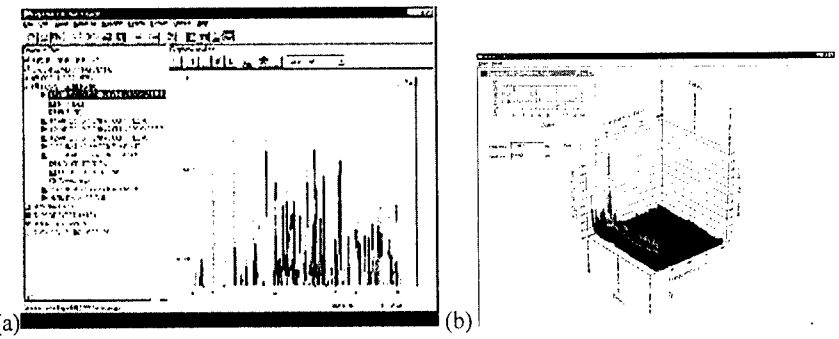


Figure 6 Vibration analysis software monitoring vibration signals for analysis, training and preventive maintenance: (a) ExpertALERT from Predict-DLI; (b) SpectraScope CAF from Spectral Visualization and Development, SVD Inc.

PREDICT-DLI

Predict DLI provides onsite training at their Cleveland or Seattle training center or customized training classes for companies interested in setting up their own onsite training classes. Predict DLI offers training in vibration analysis, lubricant analysis, thermography with digital imagery and visual inspections with digital imagery. The vibration curriculum currently consists of three courses described below:

- Vibration Analysis I and Machine Balancing – is a beginner's course for end users who have little or no experience in vibration data analysis. The course emphasizes vibration sources and measurement techniques as well as fundamental machine fault recognition. This course is designed for a practical focus on vibration analysis to detect major problems.
- Vibration Analysis II and Laser Alignment – this course is a follow-up to the basic course described up above. Here emphasis is placed on single channel analysis of vibration spectra. Problems found in gearboxes and belt driven machines are used for

examples. Alignment tools and techniques are covered to include laser pre-alignment checks.

- Expert ALERT for Voyager – is designed for individuals who have purchased vibration analysis equipment from Predict DLI. Software commands and functions are discussed as well as analyzing, fine tuning data and manipulation of various plotting and display functions. Emphasis is placed on setting up the database and data collection communications and software interface.

CONCLUSIONS

This paper has partially reviewed previous work on vibration-based helicopter health and usage monitoring methods. Machinery failure prevention was shown to be an important component of the maintenance activity for most engineering systems, especially in aerospace. Due to specific continuous vibrations induced by the rotors, helicopters are particularly exposed to operations-induced fatigue damage, and their failure prevention increasingly relies on vibrations, health, and usage monitoring systems., helicopter vibration monitoring has evolved considerably over years, a special focus point being the signal analysis and interpretation algorithms. In recent years, new and more powerful signal processing methods have been developed. Three major directions have been identified and discussed in this paper:

a) Wavelet transform

b) Joint time-frequency Wigner-Ville distribution

c) Neural network and neuro-fuzzy

These vibration signal-processing methods have a generic origin, but their application to helicopter health-monitoring methodology is quite recent. These methods play an important role in early identification of incipient damage that can later develop in a potential threat to the system functionality, and even a flight accident. In aerospace applications, HUMS capabilities are to minimize aircraft operation cost, reduce maintenance flights, and increase flight safety.

Another important area identified in this paper is that of vibration-analysis education software. Several industries are currently utilizing predictive maintenance programs that utilize hardware and software capable of vibration-based diagnostic and prognostics. Though these capabilities have been developed for on-the-ground plants and equipment, the methodology adopted in their development could be transitioned to airborne equipment and helicopters. These predictive maintenance systems vendors offer vibration-analysis educational software that presents considerable opportunities.

The present study is neither final nor exhaustive. During the literature search, it was found that considerable effort is being currently invested in this field, far beyond the limited space available in this short paper. The authors are dedicated to continuing their search, revisiting the subject, and coming back with a new publication in continuation of our present efforts.

REFERENCES

Essaway, M.A.; Diwakar, S.; Zein-Sabatto, S.; Garga, A.K., "Fault Diagnosis of Helicopter Gearboxes Using Neuro-Fuzzy Techniques", Proceedings of the 52nd meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, Virginia, March 30-April 3, 1998, pp. 293-303

Forrester, B.D., "Analysis of gear vibration in the time-frequency domain", in Current Practices and Trends in Mechanical Failure Prevention, edited by H.C. Pussy and S.C. Pussy (Vibration Institute, Willowbrook, IL), pp. 225-234, 1990

Fraser, K.F., and King, C.N., "Helicopter gearbox condition monitoring for the Australian Navy", Detection, Diagnosis, and Prognosis of Rotating Machinery to Improve Reliability, Maintainability, and Readiness Through the Application of New and Innovative Techniques, edited by T.R. Shives and L. J. Mertaugh (Woodbury, New York), pp. 49-58, 1986

iLearn Interactive (2000). Mobius iLearn Interactive, "Practical and theoretical training for Machinery vibration analysis", http://www.mobius-online.com.

Kim, J; Ewins, D.J. "Monitoring Transient Vibration in Rotating Machinery Using Wavelet Analysis", Proceedings of the 2nd International Workshop on Structural Health Monitoring, Stanford University, Stanford, CA September 8-10, 1999, pp. 287-297

Mallat, S.G., "A theory for multiresolution signal decomposition: the wavlet representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, pp. 674-693, July 1989

McFadden, P.D., Smith, J.D., "A signal processing technique for detecting local defects in gear from the signal average of the vibration", Proc. Institute of Mechanical Engineers, Vol.199, No. C4, pp.287-292, 1985.

Monsena, P.T., Dzwonczyk, M., "Analog neural network-based helicopter gearbox health monitoring system", Journal of Acoustical Society of America, Vol.98, No.6, pp.3235-3249, December 1995.

Newland, D.E., "An introduction to random vibrations, spectral and wavelet analysis", Third Edition, Longman Group Ltdm, Essex, pp. 359-370, 1993

Newland, D.E., "Wavelet Analysis of Vibration, Part I: Theory", Journal of Vibration and Acoustics, Vol. 116, pp. 409-416, 1994

Newland, D.E., "Wavelet Analysis of Vibration, Part II: Wavlet Maps", Journal of Vibration and Acoustics, Vol. 116, pp. 417-425, 1994

Predict DLI (2000). Predict/DLI – Course Descriptions, http://www.predict-dli.com.

SVD (2001). Spectral Visualization and Development Inc., http://www.svdinc.com

FAILURE MODES AND ANALYSIS I

Chair: **Ms. Debbie Aliya**
Aliya Analytical

AFTERMARKET PARTS: ARE THEY ALL THEY ARE "CRACKED" UP TO BE?

Victor K. Champagne
U.S. Army Research Laboratory (ARL)
Weapons and Materials Research Directorate
(AMSRL-WM-MD)
Aberdeen Proving Ground, Aberdeen, Md. 21005-5059
Email: vchampag@arl.army.mil
(410)306-0822
(410)306-0806

Abstract: This report contains the results of a failure analysis investigation of a fractured main support bridge from an army helicopter. The part failed component fatigue testing while those of the original equipment manufacturer (OEM) passed. Even though the same technical data package was used by both manufacturers and there were no material discrepancies found, a great disparity existed in the fatigue test data. This has been a recurring problem within the Army and the intent of this paper is to provide some insight as to the technical reasons why this can occur. Emphasis will be placed on the effects of manufacturing processes on fatigue. Other failure analyses will be discussed in relationship to this topic.

Objective: To perform a metallurgical examination comparing components fabricated by "Contractor IT" to those of the OEM, with the intent of determining the disparity in fatigue life.

Conclusion: The metallurgical data collected during this investigation indicated that the difference in fatigue life between the components fabricated by IT and the OEM may be attributable to a difference in dimensions at the web where fatigue crack initiation occurred. The webs of the two OEM parts examined contained cross-sectional thicknesses that measured significantly larger than the IT components.

Recommendations:

1. Change the web reference dimension of 0.38 inch to include a tolerance range based upon a fracture mechanics model.
2. Control the shot peening process especially at the critical areas of the web, to assure complete coverage and proper compressive residual stresses.
3. The engineering drawing includes only a shot peening intensity. No direction is given with respect to type of shot, shot size or coverage.

Background: The aforementioned four helicopter main support bridges were shipped to ARL for analysis. The parts had been subjected to fatigue testing (results listed above), and had shown a difference, by an order of magnitude, in fatigue resistance between the IT components and the two OEM parts. An independent laboratory (IL) analyzed IT0011 initially, and concluded that the shot peening intensity for the IT part was most likely excessive, which produced surface microcracks, leading to premature failure. IL also

stated that the microcracks may have relieved some of the residual stress on the surface of the part. ARL performed a comprehensive investigation in order to identify the cause of decreased fatigue life within the IT parts, including dimensional verification, visual examination, chemical analysis, surface roughness, hardness, conductivity, metallography, tensile and fatigue testing, fractography, shot peening analysis, residual stress, and TEM analysis.

Dimensional Verification: The thickness of the web at the fatigue crack initiation site was measured and compared for each of the four components. As shown in Table 1, a trend was noted. The thickness measurements of the IT webs were appreciably lower. The requirement of 0.38-inch was a reference dimension only.

Table 1
Dimensional Measurement of Web
At Fatigue Crack Initiation Site

Component	Cycles To Failure	Thickness (inch)
IT0011	38,373	0.370
IT0067	36,256	0.399
01344AI	356,942	0.421
1316HMW	435,000	0.421
Requirement		0.38 (ref.)

Visual Examination: Component IT0011 was examined in the as-received condition. The part number and serial number of the failed bridge were visible. The location of fracture, as well as the material prepared for metallographic examination by IL during their analysis was also examined. Oblique lighting was used to highlight the river patterns which corresponded to a fracture origin at the edge of the cross section. Two distinct origins were observed. No gross defect was noted at the origin sites.

Chemical Analysis: A section of the IT0011 component was analyzed to verify conformance to the required chemical composition. The results compared favorably to the nominal composition of 7075 aluminum alloy as shown in Table 2.

Table 2
Chemical Composition Results (Weight Percent)

Element	UH-60 Bridge	Typical 7075 Aluminum
Copper	1.57%	1.0 - 2.0%
Silicon	0.049	0.40 max.
Iron	0.19	0.50 max.
Manganese	0.009	0.30 max.
Magnesium	2.48	2.1 - 2.9
Zinc	5.49	5.1 - 6.1
Chromium	0.22	0.18 - 0.28
Titanium	0.028	0.20 max.
Zirconium	0.021	Other elements: 0.05 max. each, Other elements: 0.15 max. total
Vanadium	0.012	
Nickel	0.006	
Aluminum	remainder	remainder

Surface Roughness: Drawing 70400-08116 requires a surface roughness of 125 maximum all over. Conversation with representatives of the US Army Aviation and Missile Command indicated that this requirement applied to the part before it was shot peened. Since all parts were received by ARL already shot peened, it was impossible to verify conformance to this requirement. However, surface roughness measurements were taken on all four components using the stylus technique, for comparative purposes. Data were measured using a Mitutoyo Surftest Analyzer 401 stylus apparatus, and were taken on surfaces that had the paint removed with methylene chloride. A total of ten readings were taken on each sample, with the first five readings oriented perpendicular to the remaining five readings. As Table 3 shows, the average values were similar in magnitude, and no deleterious trends were noted.

Table 3
Surface Roughness (R_a) Results

IT0011		IT0067		01344AI		1316HMW	
Reading	Ra	Reading	Ra	Reading	Ra	Reading	Ra
1	260	1	180	1	240	1	220
2	240	2	220	2	210	2	200
3	190	3	200	3	230	3	200
4	220	4	220	4	200	4	140
5	180	5	220	5	160	5	210
6	180	6	220	6	190	6	180
7	240	7	190	7	200	7	200
8	180	8	230	8	200	8	170
9	190	9	200	9	180	9	220
10	200	10	200	10	220	10	170
Average	208	Average	186	Average	203	Average	191

Hardness: The hardness of the components was measured and compared. The governing specification, MIL-H-6088 requires a hardness of 78 HRB minimum, for 7075 aluminum in the T73 condition. Readings were made in the grip region of the dogbone specimens used for tensile testing. The following table summarizes the results of five readings on each component. Each component met the governing requirements, and no deleterious trends were noted.

Table 4
Hardness Measurement Results (HRB Scale)

IT0011		IT0067		01344AI		1316HMW	
Reading	Ra	Reading	Ra	Reading	Ra	Reading	Ra
1	82.9	1	81.3	1	81.3	1	81.6
2	83.6	2	81.0	2	81.3	2	82.7
3	84.2	3	80.9	3	81.9	3	82.1
4	84.4	4	81.6	4	81.6	4	82.6
5	84.7	5	81.7	5	81.9	5	82.6
Average	83.4	Average	81.3	Average	81.6	Average	82.3
Requirement	78 min.	Requirement	78 min.	Requirement	78 min.	Requirement	78 min.

Conductivity: Conductivity testing was performed to determine whether the components were properly aged. The governing specification (MIL-H-6088) lists a typical conductivity range of 40.0 – 43.0 %IACS. A total of five readings were taken on similar cross sections representative of each bridge. As shown in Table 5, the results conformed to the governing requirement. No significant difference was noted between the IT and OEM parts.

Table 5
Conductivity Measurement Results
%IACS

IT0011		IT0067		01344AI		1316HMW	
Reading	%IACS	Reading	%IACS	Reading	%IACS	Reading	%IACS
1	40.63	1	42.46	1	42.14	1	41.19
2	41.72	2	42.49	2	41.00	2	41.21
3	41.52	3	42.56	3	41.03	3	41.41
4	41.95	4	42.67	4	41.24	4	41.25
5	41.43	5	42.30	5	41.01	5	41.08
Average	41.45	Average	42.50	Average	41.28	Average	41.23
Requirement	40 - 43	Requirement	40 - 43	Requirement	40 - 43	Requirement	40 - 43

Metallography: Samples were metallographically prepared representing transverse and longitudinal orientations of part IT0011. The intent was twofold; to observe the presence (if any) of gross internal defects that may have led to premature failure, and to determine whether the part had been aged properly. The samples were rough polished utilizing silicon carbide papers of increasing grit number, followed by fine polishing consisting of diamond paste and alumina. The samples were examined in the as-polished condition. No gross defects of inclusions were observed within the samples, with the exception of "foldovers" which resulted from the shot peening process. The surface of each of the samples was examined. Metal foldover was observed within each sample. The samples were subsequently etched with Keller's reagent, and examined. The structure was examined at both low and high magnifications, and appeared consistent with the typical structure of this alloy in this condition. Further microstructural characterization was conducted on all four components and is included in the TEM Analysis section of this report.

Tensile Testing: Tensile testing was conducted on a total of four specimens from each bridge. MIL-H-6088 does not list tensile property requirements, and indicates that the properties are governed by the engineering drawing. Since no requirement was noted on Drawing 70400-08116, typical values for this alloy are listed in Table 6. These typical values were referenced from the textbook, Aluminum: Properties and Physical Metallurgy [1]. No significant differences were noted between the IT and OEM parts, as the IT specimens exhibited both the lowest and highest strength. With respect to strength, it appeared only the IT0011 specimens met the typical values of this alloy and temper.

Table 6
Tensile Testing Results

Specimen	Area (sq. in.)	0.2% Y.S. (ksi)	UTS (ksi)	%Elongation	Modulus ($x10^6$ psi)
IT0011 #1	0.0313	65.7	74.9	14.2	10.3
#2	0.0314	64.8	74.5	15.0	10.1
#3	0.0313	65.3	75.3	13.6	12.6
#4	0.0313	65.7	74.8	14.2	10.5
Average		**65.4**	**74.9**	**14.3**	
IT0067 #1	0.0313	60.0	70.6	12.6	12.9
#2	0.0314	59.1	70.9	13.0	14.6
#3	0.0312	60.4	71.1	13.1	10.6
#4	0.0314	60.2	71.0	13.6	10.7
Average		**59.9**	**70.9**	**13.1**	
01344AI #1	0.0314	58.8	69.7	13.4	10.3
#2	0.0313	57.0	68.7	16.5	10.4
#3	0.0313	61.5	71.9	16.3	10.3
#4	0.0313	61.6	72.0	14.3	10.1
Average		**59.7**	**70.6**	**15.1**	
1316HMW #1	0.0313	61.6	72.9	13.7	11.4
#2	0.0314	61.3	72.3	13.5	10.9
#3	0.0313	59.6	72.3	14.8	10.4
#4	0.0315	62.2	72.9	12.9	13.7
Average		**61.2**	**72.6**	**13.7**	
Typical 7075-T73		63.1	73.3	13	

Fatigue Testing: Fatigue testing was conducted on a total of five to six specimens from each bridge. The specimens were sectioned from the flanges. The testing was conducted at a frequency of 25 Hz, and an R-value of 0.1. A stress level of 45 ksi was utilized. The objective of this testing was to determine whether the base material of the IT parts had a similar fatigue resistance as the OEM parts. Since all specimens were fabricated similarly, this laboratory testing would eliminate such factors as surface asperities, and dimensional irregularities, and compare the actual base material of each component. There was considerable scatter in the results (as shown in Table 7), and no concrete conclusions could be drawn. The "inner" and "outer" in Table 7 refer to the location of the flanges.

Table 7
Fatigue Testing Results

Specimen	Diameter (inch)	Frequency (Hz)	R Value	Stress (ksi)	Cycles
IT0011 Inner #1	0.1495	25	0.1	45	Setup
#2	0.1495	25	0.1	45	69,837
#3	0.1495	25	0.1	45	74,676
IT0011 Outer #1	0.1500	25	0.1	45	77825
#2	0.1500	25	0.1	45	143,180
#3	0.1490	25	0.1	45	87,342
IT0067 Inner #1	0.1500	25	0.1	45	105,847
#2	0.1500	25	0.1	45	239,024
#3	0.1500	25	0.1	45	1,500,000+
IT0067 Outer #1	0.1500	25	0.1	45	707,433
#2	0.1500	25	0.1	45	334,010
#3	0.1500	25	0.1	45	105,858
01344AI Inner #1	0.1500	25	0.1	45	Setup
#2	0.1500	25	0.1	45	89,248
#3	0.1500	25	0.1	45	127,885
01344AI Outer #1	0.1500	25	0.1	45	1,000,000+
#2	0.1495	25	0.1	45	1,000,000+
#3	0.1505	25	0.1	45	72,130
1316HMW Inner #1	0.1495	25	0.1	45	314,130
#2	0.1500	25	0.1	45	162,292
#3	0.1495	25	0.1	45	475,112
1316HMW Outer #1	0.1495	25	0.1	45	299,996
#2	0.1495	25	0.1	45	66,374
#3	0.1500	25	0.1	45	69,280

Fractography: The morphology of the fracture surface of IT0011 was mapped, and SEM micrographs were taken to document the findings. The objective was to determine whether a surface or internal anomaly caused the premature failure of the IT0011 gear during fatigue testing. SEM photomacro- and micrographs were taken of the fracture surface containing the origin. River patterns were clearly discernable, leading directly to the origin, which was located on the surface (versus a subsurface origin). No gross defects were noted at the origin. The smearing below the origin was most likely a post-fracture occurrence. These findings were consistent with those of IL The fracture morphology was transgranular at the location, which is to be expected of this alloy under fatigue conditions. Fatigue striations were noted. These striations were approximately 0.323-inch from the origin. Striations were observed as close as 0.0375-inch from the origin. A transition between transgranular morphology and the ductile region characteristic of tensile overload was observed. A ductile dimpled morphology was noted in the tensile overload region. There were no gross internal or surface defects observed during SEM analysis.

Scanning electron microscopy was also beneficial in characterizing the shot peened surface of the failed bridge. The size of the dimples is an indication of shot peening intensity, which was later compared to those in the "Shot Peening Analysis" section.

TEM Analysis: To further investigate the possibility that the material structure varied between the IT and OEM components, a representative sample of each component was analyzed using transmission electron microscopy (TEM). The second phase precipitates within the matrix and grain boundaries of each component was analyzed. TEM specimens were prepared by cutting thin slices from the sample sections, followed by grinding to a thickness of 200μm. Discs 3mm in diameter were subsequently punched from this material, and electropolished in a 20% nitric-methanol electrolyte at –30°C. The specimens were examined using a Philips CM-12 electron microscope fitted with a Princeton Gamma Technologies (PGT) Energy Dispersive Spectroscopy (EDS) system. Table 8 summarizes the types of secondary phases noted within the samples.

Table 8
Secondary Precipitates

Secondary Precipitate and Comments	Sample Found Within
Coarse stringers of Al_7Cu_2Fe, evident in optical and SEM	IT0011, IT0067, 01344AI, 1316HMW
Coarse Al-Si oxide particles, evident in optical and SEM	IT0011, IT0067, 01344AI, 1316HMW
Fine (E-phase) dispersoids ($Al_{18}Mg_3Cr_2$), evident only in the TEM	IT0011, IT0067, 01344AI, 1316HMW
Strengthening precipitates, evident only in the TEM	IT0067, 01344AI
Ultrafine, matrix strengthening precipitates, evident only in the TEM	IT0011, 1316HMW

Of these, it was not possible to determine differences in the size, density and distribution of the Al_7Cu_2Fe and Al-Si oxide particles in the TEM due to their coarse size. However, examination of the electro-polished TEM discs in an optical microscope did not reveal any significant difference between the samples.

The grain size varied from 1 to 5 μm, for the IT0011 and 1316HWM material, but was larger for the IT0067 and 01344AI material (2 to 10 μm). EDS was performed to characterize the chemical composition of the dispersoid and strengthening precipitates within each sample. It was determined that the median size of these particles was 450 Angstroms for sample IT0011, 750A for sample IT0067, 800A for sample 01344AI, and 800A for 1316HWM. These measurements should be considered only estimates based upon the different shapes of the dispersoids. A possible difference in the size of dispersoids could be due to differences in solutionizing treatment temperatures. A larger dispersoid would be associated with a higher solutionizing temperature. Since the dispersoid particles remain undissolved during the solutionizing treatment, they would coarsen at higher solutionizing temperatures. Attempts to estimate the volume fraction of the dispersoids from the TEM images did not provide reproducible results, presumably due to variations in specimen thickness and image shapes.

In short, sample IT0011 had an E-phase size about 50% smaller than the other three samples, most likely a factor of the prior solutionizing temperature. Other microstructural features such as size of the strengthening precipitates, width of the

precipitate free zone and the size and distribution of the coarse (Al_7Cu_2Fe, Al-Si oxide) particles were comparable.

Shot Peening Analysis: Note 4 of Engineering Drawing 70400-08116 states, "After final machining, shotpeen all over per SS8767 to 0.008A minimum intensity. Complete shotpeen coverage not necessary in areas noted. Overspray of these areas is permissible". A review of the shot peening invoices for each of these components, revealed that each of the parts had been shot peened by one company, but at three separate locations. Both IT0011 and IT0067 were peened at a plant that was not identified on the Purchase Order, while 01344AI was peened at a plant in West Babylon, NY. The 1316HMW part was peened at a plant in Wyandanch, NY. Table 9 summarizes the parameters used by the shot peen vendor at each of their plants.

Table 9
Shot Peening Parameters

Plant	Shot Size	Intensity	Coverage
Unidentified (IT0011, IT0067)	CS-230	0.009A	100%
West Babylon, NY (01344AI)	CS 330R*/230R*	0.008-0.012A	200%
Wyandanch, NY (1316HMW)	CS 330/230	0.008-0.012A	200%
Requirement (Dwg. 70400-08116)		**0.008A minimum**	

* - R=Regular shot, 45-52 HRC

As shown in Table 9, variation existed as to shot size used, as well as the coverage. The IT parts were subjected to only the 230 sieve size cast steel shot, while the remaining parts were shot with 330, and 230 (as listed in Table A of SS8767, Rev. 5, a cast shot size of 330 has a sieve opening of 0.0331-inch, while a cast shot size of 230 has a sieve opening of 0.0234-inch). Therefore, the OEM components were peened with a coarse shot, followed by peening with the finer shot. This explains the 100% coverage for the IT parts, and the 200% coverage for the OEM components. It should be noted that 200% may be beneficial for compressive depth for this material [2].

IL believed that the intensity was excessive for part IT0011, since they reported that microcracking was prevalent on the surface of the part. However, generally speaking, the diameter of a peening "dimple" should be equal in magnitude to the intensity used to peen. With this in mind, several dimples on the IT0011 were measured. The resulting average of 0.0059-inch, was well below the 0.008 minimum intensity requirement. It appeared as if the intensity for the IT part was less than nominal, rather than excessive. Surface residual stress measurements were also taken within this area and the resultant values were lower than anticipated (refer to "Residual Stress" section). The same trend was noted for the OEM parts as well.

Additionally, a piece of material taken from part 01344AI was sent to a reputable vendor to shot peen under the following conditions: CS 230 shot size, 0.008 A minimum intensity and 100% coverage. This was used as a standard for comparative purposes. The piece was milled prior to shot peening. Subsequent to peening, there was no evidence of the "foldover". The dimples on this piece had a larger diameter than those of

the IT parts, indicative of a higher intensity . A residual stress profile was also performed on this piece, and revealed that the compressive stress was nearly double that of the bridges. These results are located in the "Residual Stress" section.

Residual Stress: A technology for Energy Corp. (TEC) Model 1610 X-Ray Residual Stress Analysis System was used to characterize shot peened induced surface and subsurface residual stresses. All data were obtained utilizing the $\sin^2\psi$ stress-measuring technique with chromium Kα radiation diffracted from the (311) crystallographic planes at a zero-strain peak position of 139° 2θ. Surface measurements were performed on each component; subsurface measurements were performed on components IT0067 and 1316HMW and on test section 01344AI. Layer removal and stress gradient corrections were applied to the subsurface data per SAE J784a [3]. The longitudinal stress direction was arbitrarily chosen (the transverse direction was 90° clockwise from longitudinal). The area of measurement was as close to the fatigue crack initiation site as geometry would allow. Initial surface residual stress data from component IT0011 was observed to be approximately half the value of the other components (see Table 10). However, the other IT part (IT0067) exhibited the highest readings, suggesting that surface residual stress may not have played a part in the vastly different fatigue lives.

Table 10
Results of Surface Residual Stress Measurements
Actual Components

Component	Residual Stress ksi
IT0011 - Trans	-15.1
IT0011 - Long	-16.9
IT0067 - Trans	-27.9
IT0067 - Long	-27.4
01344 - Trans	-24.3
01344 - Long	-23.2
1316HMW - Trans	-26.4
1316HMW - Long	-25.0

Subsurface measurements were performed on a representative IT component (IT0067) and a representative OEM component (1316HMW). The purpose of this testing was to compare the residual stress values at increasing depth below the surface. Again, measurements were taken in the longitudinal and transverse directions. The results (Table 11) showed that the OEM component had a compressive layer of higher magnitude than the IT component. The IT component exhibited its highest compressive stress at 0.004 – 0.0045 inches in depth, while the OEM component exhibited its highest compressive stress at 0.007 – 0.0075 inches in depth.

Table 11
Results of Subsurface Residual Stress Measurements
Actual Components

Depth (inch)	IT0067		1316HMW	
	Longitudinal Residual Stress (ksi)	Transverse Residual Stress (ksi)	Longitudinal Residual Stress (ksi)	Transverse Residual Stress (ksi)
Surface	-11.3 ± 0.6	-8.7 ± 1.0	-8.6 ± 0.9	-7.0 ± 1.1
0.001 – 0.0015	-16.5 ± 0.5	-18.0 ± 1.7	-20.6 ± 0.5	-16.9 ± 3.5
0.002 – 0.0025	-22.1 ± 1.0	-20.4 ± 2.3	-22.0 ± 1.2	-24.8 ± 5.6
0.004 – 0.0045	-26.3 ± 1.4	-23.3 ± 2.8	-28.6 ± 3.5	-25.5 ± 5.2
0.007 – 0.0075	-20.4 ± 2.0	-16.7 ± 4.1	-30.3 ± 4.1	-29.9 ± 2.7
0.012 – 0.0125	-1.4 ± 1.4	-5.3 ± 0.9	-3.0 ± 1.6	-5.1 ± 2.4
0.018	0.0 ± 1.1	-1.0 ± 2.0	-0.8 ± 2.8	-1.3 ± 2.9
0.0235 – 0.024	0.3 ± 1.0	-2.4 ± 1.8	-2.5 ± 0.5	-1.4 ± 4.7
0.0295 – 0.030	-4.1 ± 1.0	-3.4 ± 1.6	-0.1 ± 0.5	0.1 ± 1.6

Generally, the magnitude of the compressive stress should equal approximately 60% of the UTS for this alloy [2]. Appropriately, this material should have had a compressive stress approaching 44 ksi. The highest subsurface stress measured by x-ray analysis was –30.3 ksi, which was approximately 30% lower than nominal.

The residual stress of the "standard" shot peened by Metal Improvements was also determined. The readings were measured to a depth of 0.0150-inch. As shown in Table 12, the magnitude of the compressive stress throughout the standard was greater than both the IT and OEM components. The "60% of the UTS" maximum residual stress observed was comparible to the rule of thumb.

Table 12
Results of Subsurface Residual Stress Measurements
Test Section from Component 01344AI

Depth (inch)	Residual Stress (ksi)
Surface	-38.10
0.0005	-46.21
0.0015	-46.30
0.0025	-47.88
0.0050	-38.06
0.0095	-8.08
0.0150	0.97

Discussion:

Preload During Fatigue Testing: It was reported to ARL from Wayne Rainey of AMCOM that the pre-loads were "very much" higher in the IT components that were fatigue tested. This was detected in strain gages located in the area of failure upon loading of the parts into the test fixture during a static survey. The fact that the preloads were higher in the IT components served as an indicator that a dimensional discrepancy may have been present somewhere on these parts. This was verified when the thickness of each component was measured at the location of the fracture and the IT components were found to be significantly lower. The extent to which the difference in thickness affected the fatigue results should be investigated through a stress analysis of the area of concern.

Shot Peening: A concern about the integrity of the peened surface of the IT components was raised byIL. It was reported by ILthat evidence of broken shot was detected along with microcracks on the surface which in turn suggested a higher than acceptable shot peening intensity. This prompted ARL to research these claims in depth. The results indicated that the intensity may have been too low as substantiated by low residual stress values of the shot peened surface of all components, IT and the OEM. The values obtained were compared to that of a shot peened "standard" provided by Metal Improvements, the vendor that shot peened both the OEM and IT components utilizing the same parameters that were used on the components. The residual stress results clearly indicate that the shot peening operation performed on the IT and OEM components resulted in surface residual stresses that were below the standard. It is important to note that the standard was fabricated from material taken from OEM component 01344AI in which the surface was milled prior to shot peening to remove any previous effects of fabrication. Microcracks and broken bits of shot were not observed by ARL on either the OEM or IT parts. What was detected was significant evidence of "foldover" on both the OEM and IT components, which may have been misinterpreted as microcracks. Foldover can be caused by directing the shot at an angle, but since it was observed on all four components, it was not believed to have contributed to the difference in fatigue results.

Another important observation made on all four parts was the non-uniformity of the shot peened surface through visual examination. The extent to which this may have affected the residual stress pattern of each component was not investigated because of the time and expense involved and due to the fact that the residual stresses were measured adjacent to the fracture zones on all four components. The fracture zone is the area of concern, and the remaining surfaces would not be involved in reducing the fatigue life, but the issue should be raised that uniformity of the shot peened surface could play a role in fatigue life.

References:

[1] Aluminum: Properties and Physical Metallurgy, Edited by John E. Hatch, American Society for Metals, Metals Park, Ohio, 1984, p. 364.

[2] Personal conversation with Win Welsch, 9/15/99, during meeting at ARL, APG, MD

[3] Residual Stress Measurement by X-Ray Diffraction – SAE J784a, Society of Automotive Engineers, Warrendale, PA, 1971.

Acknowledgement:

The author wish to thank STEM, Inc. for their assistance with TEM characterization.

FAILURE MODES AND PREDICTIVE DIAGNOSTICS CONSIDERATIONS FOR DIESEL ENGINES

Jeffrey Banks, Jason Hines, Mitchell Lebold, Robert Campbell,
Colin Begg and Carl Byington

The Pennsylvania State University/Applied Research Lab
Condition-Based Maintenance Department
University Park, PA, 16804-0030

Abstract: Diesel engines are well known for their operational robustness and efficient performance. These attributes make them a leading choice for prime movers in critical DoD, industrial, and mobility applications. Despite the diesel engine's known reliability, there are some operational issues that justify monitoring critical engine components and subsystems in order to increase the overall availability and readiness of diesel-powered systems. Moreover, engines typically constitute a significant fraction (1/10-1/5) of the acquisition cost and a comparable fraction of the life cycle cost for mobility applications (trucks, armored vehicles), thereby providing the motivation for engine condition monitoring on the basis of reducing life cycle costs. Review of the available literature indicates that the fuel injection and cooling subsystems are among the most problematic on diesel engines contributing to reduced readiness and increased maintenance costs. These faults can be addressed and studied using scaled testing to build the necessary knowledge base to quickly transition the methods to full-scale, more costly diesel engines.

Towards this goal, a Diesel Enhanced Mechanical Diagnostics Test Bed (DEMDTB) has been developed that uses an array of sensors to measure pressure, temperature, vibration, and displacement. The test bed is used for experimental collection of healthy, seeded fault, and transitional fault test data from the diesel engine and driveline components. The data is analyzed with time and frequency based analysis methods to characterize 'healthy' and 'faulty' operation.

The purpose of this paper is to present an overview of previous research conducted for diesel engine diagnostics, discuss recent diesel engine diagnostics developments, and to lay the basis for straightforward concept designs for practical diesel engine monitoring/diagnostics systems that will enable system prognostics.

Key Words: Condition-Based Maintenance (CBM); diesel engines; Diesel Enhanced Mechanical Diagnostics Test Bed (DEMDTB); FMEA.

Introduction: Diesel engines are widely used as generators and prime movers in industry and the military for their durability and efficient performance and they are often used in applications where reliability is a crucial operating requirement. Large and

medium size diesel engines can be found in electrical power plants as the prime movers of large oceanic vessels. Meanwhile smaller, high-speed engines have been found in tractors, trucks, cars and small marine vessels. Diesel engines are also used for a wide variety of military applications. For example, the US Navy uses diesel engines in a variety of roles in the fleet. Numbering in the thousands, applications for these engines range from main propulsion and service power generation down to fire hose pumps. These engines range in power from less than 50 hp to above 12,000 hp. Diesels are used on roughly 30 classes of ships across the Navy. Currently over 200 diesel engines that are greater than 2,000 hp are being used for main propulsion on the LSD, LST, and PC class ships [1]. The U.S. Army and Marine Corps are also heavily reliant upon diesel engines as prime movers where the majority of combat and transport vehicles in use (with the exception of main battle tanks) are powered by diesel engines. The Advanced Amphibious Assault Vehicle, currently in the acquisition process, uses a high-powered, MTU diesel engine. Some of these systems have a good deal of performance monitoring and limited diagnostics. However, the authors have seen no commercial system with true prognostic capability.

Considering the manning reduction issues that the military and many industries are faced with coupled with the need for diesel engine maintainability, it is logical that a Condition-based Maintenance (CBM) system be developed for monitoring diesel engine operation. The justification for implementing a CBM program should be evaluated on a case-by-case basis but when diesel engine dependability is crucial to mission effectiveness, then that system is an excellent candidate for the application of a CBM program. Financially feasible applications for such advanced maintenance systems are in nuclear power plants, offshore oil rigs, hospitals, and in various remote unmanned facilities. The need for cost-efficient maintenance programs in the military is evident by the overwhelming size and age of the armed forces vehicular fleet. The average age of the U.S. military's 850,000 vehicles is twelve years, which makes maintaining operational readiness a paramount concern and also a costly expenditure of more than five billion dollars annually [2].

Operational Characteristics of Diesel Engines: Diesel engines are comparable to spark ignited (SI) engines in many respects, with the exception that they use the heat produced from the compression stroke for ignition rather than spark plugs. Diesel fuel is injected at high pressure into the cylinder after the air has been compressed to such a point where auto-ignition occurs. The compression ratio of diesel engines can be greater than twice that of SI engines, which translates into greater efficiency.

The combustion process of the diesel engine is highly dependant upon the precise injection of atomized fuel into the cylinder or swirl chamber. The fuel injection system controls the injection pressure (necessary for atomization and mixture) and dispenses a metered amount of fuel for specified speed and load conditions, and has effects upon emissions and overall engine noise. The reliable operation of the fuel injection/delivery system is thus a paramount concern for both engine manufacturers and maintenance personnel. The development of the distributor-type injection pump with automatic timing (introduced in the early 1960's) and Electronic Diesel Control (developed during the

1970's) have contributed greatly to the increased power output and lower emissions of modern engines [3].

Diesel Engine Fault Analysis: Research efforts related to the development of diesel engine diagnostic systems have typically been guided by a thorough knowledge of component failure modes. A reliability-based engineering method that is commonly employed as an evaluation tool is Failure Mode and Effects Analysis (FMEA). FMEA charts describe the function of a component, potential failure modes, possible causes of failure, and the effects such failures would have upon the system's operation. Other versions of how to capture this information are employed in RCM II analysis and FMECA (Failure Modes, Effects and Criticality Analysis). The FMEA chart shown in Figure 1 is for large, medium-speed marine diesel engines typically employed in large commercial vessels.

FMEA Chart for Fuel Oil Supply System

Component	Function	Failure Mode	Possible Cause of Failure	Effects on System
Fuel Oil Injection Pumps	provide engine with fuel in quantities corresponding to power required and timed correctly	broken delivery valve springs		poor atomization fouling misfiring of cylinders
		choked fuel valves	contaminated fuel	loss in power
		cavitation	local pressure falls below saturated vapor pressure of fuel	pump erosion
Fuel Oil Injectors	atomize fuel in combustion chamber and to ensure that it mixes with sufficient air for complete combustion in cycle time available	incorrect atomization	choked atomizer due to contaminated fuel debris and hot gas from cylinder forming carbon	incorrect combustion
		cavitation	low pressure caused by pressure waves that move between injector and fuel pump at end of fuel injection; delivery valve breakage also aggravates cavitation	injector erosion
High pressure Fuel Lines		cavitation	same as item 2	erosion in high-pressure fuel lines, ultimately resulting in rupture of main fuel line

Figure 1 - FMEA Chart of Fuel Oil Supply Systems [3]

It is also important to understand which of the failures described in the FMEA have the highest rate of occurrence during operation. Comparisons of component failure rates obtained from studies of twenty similar marine diesel engines are shown in Figure 2. This information allows diagnostic research to be tailored to address failure mechanisms that occur on a significant basis under typical operating conditions.

Failures of the fuel oil valve represent greater than 30% of the recorded failures, while twelve components account for roughly 90%. A review of other studies corroborates the

high failure rates of the fuel delivery system, cylinder head, valves and cooling system shown in Figure 2. A summarized list of fault areas obtained from the reviewed studies is shown below (listed in order of priority).

(1) Fuel Injection System
(2) Cylinder Head and Valves
(3) Charging and Exhaust System
(4) Cooling System
(5) Bearings, Pistons, Liners, Timing Gears, etc.

The reviewed information clearly indicates that the fuel injection system has been the most prevalent source of problems for diesel engines. Meanwhile, engine components subjected to high levels of wear, such as pistons, liners and bearings rank near the bottom.

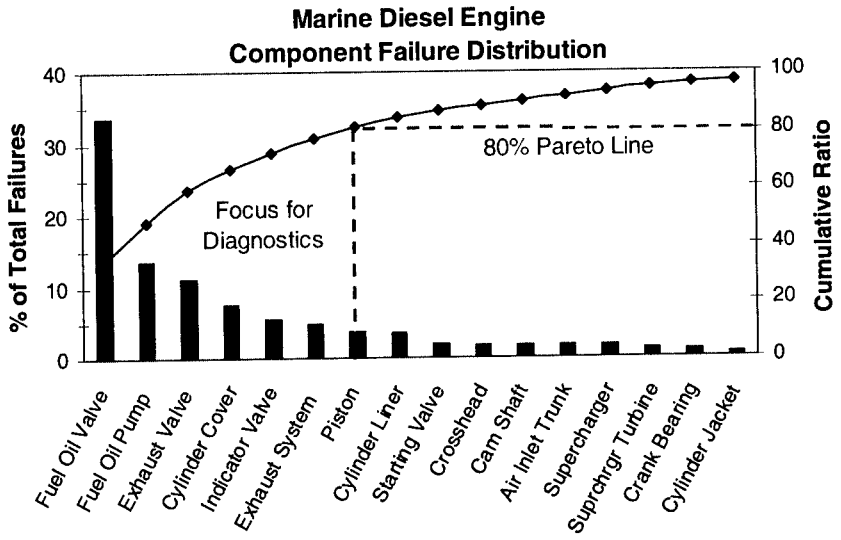


Figure 2 - Marine Diesel Engine Component Failure Distribution [4]

From the study on marine diesel engines, a comparison was made between the amount of maintenance attention given to engine components versus their corresponding failure rates. Figure 3 shows that the high failure rates of individual components described in Figure 2 are not caused by lack of proper maintenance attention. It also shows that preventative maintenance (PM) is properly distributed to those parts that are most prone to failure or are critical to engine performance.

It should be noted that due to the nature of the preventive maintenance techniques employed, it would be expected that frequent component replacements, regularly scheduled maintenance, and yearly overhauls should be reflected in these numbers as components are discarded prior to the end of their useful life. Unlike a preventative maintenance program, a CBM system would allow these components to operate

throughout their extended design life cycle. A shift might therefore be expected in fault distributions once the engine is being maintained with a CBM approach rather than by PM.

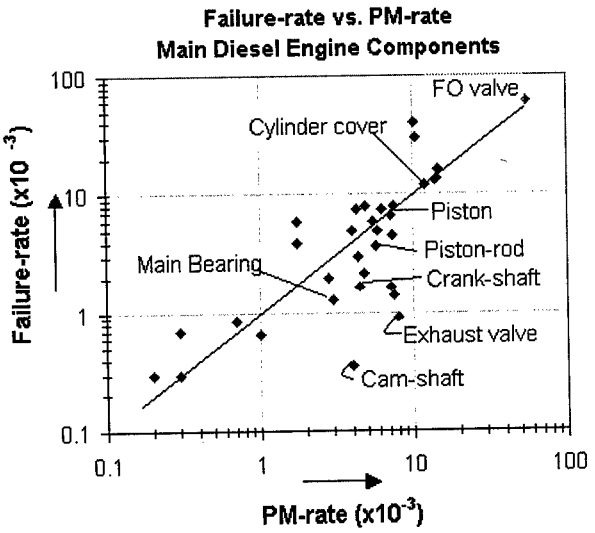


Figure 3 - Main Diesel Engine Components, Failure-rate vs. PM-rate [4]

Diesel Engine Diagnostics: Condition monitoring systems and fault diagnostics techniques have been developed to reduce maintenance costs and to increase machinery availability for critical mechanical systems. This is accomplished by comparing measured operational parameters to normal machine condition baseline levels, with the residual representing an indication of a possible fault condition. Understanding the correlation between the parameter and the components or mechanical functions that they represent provides insight into the root cause of machinery faults.

There are a number of parameters that can be measured and monitored on diesel engines including: pressure, temperature, flow rates and vibration. Currently, the most prevalent diagnostic evaluation technique is cylinder pressure analysis, which has been used extensively to monitor the engine combustion process. By evaluating deviations from pre-established, 'healthy' pressure-time curves of each of the cylinders it is possible to detect a variety of abnormal operating conditions. Peak firing pressure, peak firing pressure crank angle, maximum pressure rise rate, start of injection, and start of combustion are key reference points used in diesel engine cylinder pressure analysis [5,6].

Additionally, vibration and acoustic analysis of diesel engine operation has been increasingly employed in diagnostic systems and holds great potential for predictive diagnostics research. Spectrum analysis of engine vibration data has typically been a

subjective process in which observed patterns are compared to reference or baseline conditions in order to identify operating anomalies. Liner scuffing, blow-by and improper fuel injection are a few specific faults that have been detected by this method. The basis for acoustic (ultrasound) analysis of engine noise stems from the capability of experienced maintenance personnel to diagnose faults through observed sound qualities. This process involves using either high frequency vibration transducers (35-45kHz range) or acoustic ultrasonic detection equipment [5].

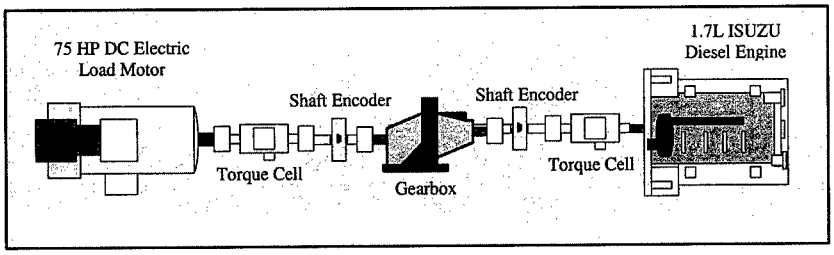
The System Material Analysis Department of the Commonwealth Edison Company illustrate a method for diagnosing diesel engine blow-by using ultrasonic and pressure analysis [7]. Blow-by is fault condition usually indicative of piston ring or cylinder wear. The most salient symptom for this condition is an increase in the vibration level that coincides with high cylinder pressures. The large vibrations occur due to the leakage of combustion gases past the piston rings.

Condition monitoring and wear debris analysis of lubricating oil are frequently used to complement data obtained from the cylinder pressure analysis. Assessing the health of lubrication oil is especially important for industrial and marine applications of large and medium size diesel engines. Lubricant contamination occurs primarily in the form of metallic debris produced from the mechanical wear of engine components and from leakages into the system. Various methods of wear debris analysis, typically involving forms of spectroscopy or ferrography, offer valuable insight into component wear rates and thus provide a means to detect rapidly deteriorating engine components prior to failure. Spectroscopy assesses the elemental content of oil by measuring the frequency and intensity of light emitted from electrically excited particles, whereas ferrography entails the separation of ferromagnetic fluid particles from the lubricant through contact with a magnetic field [8].

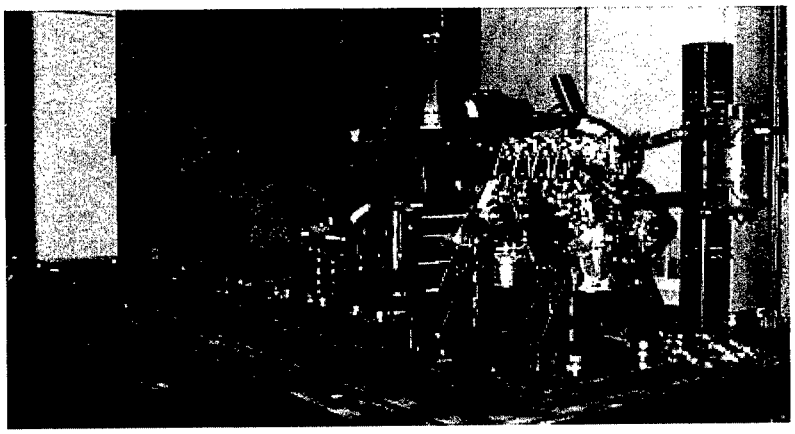
Diesel Enhanced Mechanical Diagnostic Test Bed (DEMDTB): A test cell for performing transitional and seeded fault research on a reciprocating engine and mechanical drive train parts is a key enabler towards the development of diagnostic and prognostic capability. The Diesel Enhanced Mechanical Diagnostic Test Bed (DEMDTB) system is capable of generating diesel engine operational data at the component and subsystem level under variable speed and load conditions. It provides the opportunity for conventional and advanced sensing techniques on real machinery in a controlled environment.

The DEMDTB provides two methods of driving forces for testing: electric motor or diesel engine. A block diagram layout of the DEMDTB in the diesel engine drive configuration is shown in Figure 4. The DEMDTB provides accurate torque and speed information via torque cells and shaft encoders mounted on the drive and load shafts. Secondary torque and speed measurements are also provided from the electric motor controllers.

The 1.7-liter 4-cylinder Isuzu diesel engine shown on the test bed in Figure 4, provides a continuous output of 36.1 bhp @ 3000 rpm with a maximum rating of 80.0 ft-lbs @ 1800 rpm. This test bed provides an effective means for studying health indication parameters for a representative diesel engine. Seeded faults in the diesel engine may



(a)



(b)

Figure 4 - Schematic and Picture of Diesel Enhanced Mechanical Diagnostic Test Bed

include excessive wear on piston rings and valves or a cracked crankshaft and lifter rods. While the DEMDTB provides a new mechanism for developing diesel diagnostics, the test bed still provides a means for testing different types of gearboxes and other mechanical devices.

Data Acquisition and Control System: A C-sized, VXI rack-mount data acquisition system connected to a Pentium based rack mount computer via an IEEE-1394 fire-wire interface is implement on the DEMDTB. This rack mount system is capable of housing 13 different types of VXI boards. Currently, the system contains one Agilent E1433 and one Agilent E1437.

The Agilent E1433 digitizes 8-channels at a rate of 196,000 samples/sec with 16-bits of amplitude resolution. The module provides transducer signal conditioning, anti-alias protection, digitization and measurement computation. The onboard digital signal processor and 32 Mbytes of RAM maximizes total system performance and simplifies

system integration. Using separate analog-to-digital converters (ADC) for each channel, the E1433 provides simultaneous sampling across all channels. Simultaneous sampling guarantees accurate channel-to-channel comparisons, both in the time and frequency domains and is required for phase analysis and order resampling analysis. The E1433 uses sigma-delta ADCs with 64X over-sampling which allows for low-order analog anti-alias filters and permits all filtering and sample-rate conversions to be performed digitally thereby providing stable and drift-free filtering.

The Agilent E1437 digitizer provides one channel at 20Msamples/sec sample rate with an amplitude resolution of 23-bits. This module is capable of recording reliable high frequency data up to a 10MHz bandwidth for torsional vibration analysis and high frequency analysis. The E1437 module includes input signal conditioning and an 8 MHz anti-alias filter that guarantees that signals outside the analysis bandwidth do not corrupt data samples.

A benefit of a VXI data acquisition system is its ability to trigger all the DAQ boards installed in the mainframe, and also across multiple mainframes, at the same time to provide simultaneous sampling. Besides the capability of recording accelerometer, speed, torque and general voltage measurements such as pressure and strain, the DEMDTB data acquisition system is also capable of recording up to 32 channels of temperature data utilizing an Omega TempScan unit. A National Instrument input/output (I/O) board (VXI-AO-48XDC) is also provided in the VXI mainframe for basic I/O monitoring and control. This high-precision analog source module has 48 voltage (±10.1v) and 48 current outputs (0-20.2ma) for the generation 96 analog signals with 18-bits of resolution. The VXI-AO-48XDC also provides 32 bi-directional TTL compatible digital I/O lines for control and sensing.

The DEMDTB utilizes a second computer whose primary purpose is control and automation of the test bed. The control computer system consists of a Pentium based rack mount computer that utilizes a National Instruments (NI) PCI based data acquisition board (PCI-6025E) for monitoring and control of the diesel test bed operations and auxiliary support systems. The NI card has the capability of simultaneously sampling up to 16 analog channels with a 200Ksamples/sec sample rate. This card also includes 12-bit analog output channels and 32 digital input/output lines for control and sensing signals.

Custom software was developed for both the DAQ and control computers with open systems architecture (OSA) and the Internet in mind. Each of the programs can be setup and controlled remotely via the Internet using a TCP/IP protocol. During testing operations, the control computer will monitor and control the diesel, electric motors, support systems and the VXI data acquisition computer. The DAQ computer will record data and process diagnostic features between snapshots. These diagnostic features will provide warnings and alarms to the operator along with emergency shutdown signals to the control computer. The distributed network system described above is shown in Figure 5.

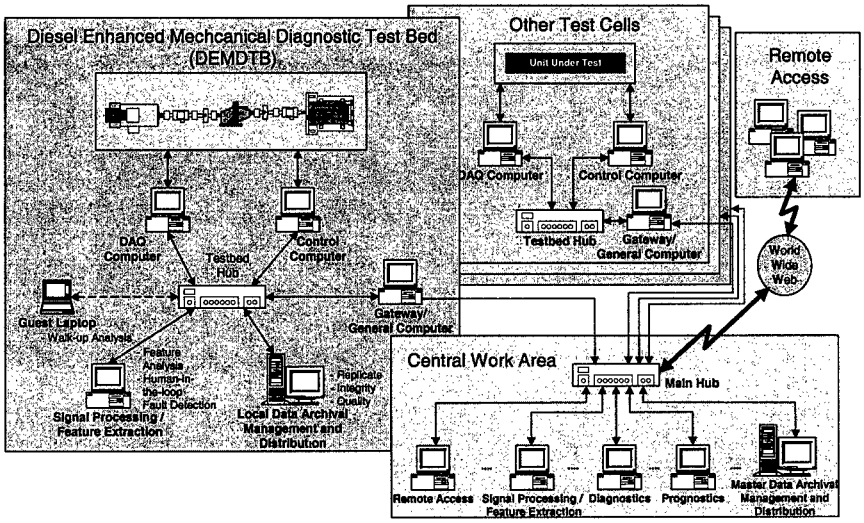


Figure 5 - Global Data Management Network System Architecture

Various computers will be attached locally to the DEMDTB network to process and manage the data generated from the test bed. The DAQ computer will instruct the data archival computer to generate a local mirror image of the collected data and then prompt the processing computer to perform analysis on the latest snapshot of data. A gateway computer on the test bed network will allow remote control via the World Wide Web along with remote data analysis and management.

The DEMDTB as well as several other test cells are centrally supported by the central network system. This provides the capability to conduct data archival/retrieval, signal processing, diagnostic analysis and prognostic analysis from remote locations as well as provide the general public access to the research being conducted through the World Wide Web.

Advanced Diagnostics and Prognostics: Previous diagnostics research has offered insight into the primary failure modes and component failure rates for representative diesel engines. Work has also been conducted to provide measurement parameters as indicators for typical engine faults. Although the state of diesel engine diagnostics has been well developed, there is still no mechanism for predicting the remaining useful life (RUL), which is the ultimate goal of any CBM plan. In an effort to develop the capability to predict the engine component or system RUL, the ARL/Penn State Condition Based Maintenance department proposes to evaluate the use of advanced diagnostic/prognostic techniques. In addition to measuring the standard parameters used for diesel engine diagnostics such as pressure, temperature, flow rate, displacement and vibration, the use of torsional vibration analysis [9] and structural surface intensity [10] will be monitored to evaluate their ability to reliably diagnose and track diesel engine fault conditions. These parameters can be further enhanced to provide a possible prognostic metric by the use of statistical feature extraction [11], which can accentuate

the reaction effect of the measured indication parameter to progressing engine component failure. Feature analysis also lends itself well to the application of pattern recognition and tracking techniques, which are necessary for the implementation of a real time condition monitoring system. Increasing the reliability and effectiveness of diesel engines will require the ability to constantly monitor the engine's operational parameters for the purpose of extracting and processing the useful information that reveals the health condition of the engine and future effectiveness of the prime mover.

Acknowledgment: This DEMTB design and development was supported through the Office of Naval Research by the *Defense University Research Instrumentation Program* (Grant Number N00014-99-1-0648). The content of the information does not necessarily reflect the position or policy of the Government, and no official endorsement should be inferred.

References:

1. Sharpe, R., Jane's Fighting Ships. 102 ed. Directory and Database Publishers Association, 1999.
2. Tank Automotive Research, Development and Engineering Center (TARDEC), **http://www.tacom.army.mil/tardec/engmain.htm**
3. Perakis, A. N. and I. Bahadir, *Reliability Analysis of Great Lakes Marine Diesels: State of the Art and Current Modeling*, Marine Technology, Vol. 27, No.4, pp. 237-249, July 1990.
4. Kawasaki, Y., *The Marine Diesel Engine and its Reliability Problems,* Bulletin of the Marine Engineers Society of Japan, Vol. 8, No. 1, pp. 3-13.
5. Challen, B., Editor, Diesel Engine Reference Book. SAE International: Warrendale, PA, 1998.
6. Hunt, G. A., *Diesel Engine Analysis Review*, ASME ICE Div., Vol. 27-1, pp. 109-117, 1997.
7. Long, B. R. and K. D. Boutin, *Enhancing the Process of Diesel Engine Condition Monitoring,* ASME ICE Div. 1996 Fall Technical Conference, Vol. 27-1, pp. 61-68, 1996.
8. Collacott, R. A., Mechanical Fault Diagnosis and Condition Monitoring. Halstead Press: New York, 1977.
9. Maynard, K. P., et al, *Application of Double Resampling to Shaft Torsional Vibration Measurement for the Detection of Blade Natural Frequencies*, Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA, pp. 87-94.
10. Banks, J. and S. Hambric, *Structural Surface Intensity as a Diagnostic Indicator of machinery Condition*; Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA, pp. 551-558.
11. McClintic, K., et al, *Residual and Difference Feature Analysis with Transitional Gearbox Data*, Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA, pp. 635-645.

THE ROLE OF MANUFACTURING DEFECTS IN MUNITION COMPONENT FAILURES

Marc Pepi
US Army Research Laboratory
Weapons and Materials Research Directorate
AMSRL-WM-MD, Building 4600
Aberdeen Proving Ground, Maryland 21005-5069
mpepi@arl.army.mil

Abstract: The US Army Research Laboratory performs numerous failure analysis investigations on munition-related components. Many of these failures are attributable to defects that can be traced back to the manufacturing process. This paper will discuss the impact of these defective parts making their way into service. Munition component failures are very costly, and may seriously affect the safety and readiness of the fleet, as well as leading to a system grounding depending on the severity of the problem. Typical defects included those associated with the material, forging, casting, welding, and heat treatment processes. Also, dimensional anomalies have been noted. Specific examples of component failures will include bomb fin retaining bands, general-purpose bomb suspension lugs, missile launcher attachment bolts, cluster bomb tailcones, general-purpose bomb fins, and Gatling gun breech bolt assemblies. In addition, this paper will focus on the importance of proper manufacturing techniques in this industry.

Key Words: Failure Analysis; Metallurgical Investigation; Flight Safety Critical Components, Manufacturing Defects

COMPONENT: MK 15/Mod 6 Snakeye Bomb Fin Retaining Band
MANUFACTURING DEFECTS: Improper heat treatment / Improper dimensions

Background: A retaining band from the MK15 Mod 6 Snakeye bomb fin unwrapped during a practice flight, causing the bomb fins to deploy, as well as triggering an adjacent retaining band to become unraveled. The pilot was able to land without incident, and upon inspection of the bomb fin, it was noticed that the retaining band had not actually broken, but had simply loosened from its original tightened position.

ARL Investigation: The Naval Air Warfare Center (NAWC) sent the "failed" as well as, an intact retaining band to ARL for inspection and analysis. Chemical analysis, dimensional verification, hardness testing, metallography and tensile testing were performed in order to determine the cause for premature failure.

Results of Investigation: The chemical composition of the components compared favorably to Type 302 stainless steel, which conformed to the governing requirement (Type 301 or 302). Dimensional inspection revealed that the band was thinner than required. The results of hardness testing were lower than required, and compared more favorably to the material in the annealed condition, rather than the ¼-hard condition that was required. Metallography results were in agreement with the hardness results, as the grains of the Type 302 stainless steel were equiaxed, rather than flattened, or "pancaked" (see Figure 1). Tensile testing confirmed that the component was annealed, as the results did not compare favorably with those for the ¼-hard condition.

Effect of Manufacturing Defect on Performance: The retaining band was able to unwrap itself from the clamp tightener because it was thinner than required, and softer (less stiff) than intended.

Outcome: The NAWC is going to scrap the retaining band kits fabricated by the manufacturer of the suspect kits, and procure new components. They will oversee the manufacturers procedures and perform first article testing to ensure this type of situation does not occur again.

Figure 1 Micrograph showing the equiaxed grains of the Type 302 stainless steel, typical of the annealed condition. Mag. 400x.

COMPONENT: MS3314 General-Purpose Bomb 1,000-Pound Suspension Lug
MANUFACTURING DEFECTS: Forging Laps and Seams

Background: Two MS3314 suspension lugs are threaded into each general-purpose bomb, such that the munitions can be loaded onto the underside of Navy aircraft. A total of three AISI 4340 MS3314 suspension lugs failed during routine proof load testing. The proof load testing required the part to sustain a tensile load of 35,000-pounds for one-minute at a 6-degree angle, as well as 24,000-pounds at a 35-degree angle for one-minute. These three lugs failed to achieve the one-minute duration before failure occurred.

ARL Investigation: Two of the three failed lugs were sent to ARL from the Naval Air Warfare Center (NAWC) for failure analysis. The failure investigation included visual examination, chemical analysis, metallography, hardness testing, scanning electron microscopy, and energy dispersive spectroscopy.

Results of Investigation: Visual examination revealed a blackened region at the crack origin of each failure. In addition, a forming lap was found on the external bail surface of one of the failed lugs. Material sectioned from the failed lugs and subjected to chemical analysis conformed to governing specification. Metallographic examination adjacent to the blackened regions at the crack initiation sites showed slight carburization upon etching. This indicated that the regions were exposed to the high temperatures associated with the heat treatment. The hardness of the component was acceptable. Electron microscopy of the blackened surface revealed a featureless condition associated with oxide formation. It was concluded that the lugs failed due to overload conditions, as determined by the predominantly ductile dimpled fracture surface. Energy dispersive spectroscopy of the blackened regions revealed evidence of a corrosion product or heat treat scale.

Additional Testing: ARL performed the required proof testing on a number of lugs in inventory in order to determine the extent of the manufacturing defects. When many components failed the proof testing, it necessitated a magnetic particle inspection (MPI) screening of the hundreds of thousands of lugs in inventory. Figure 2 is a blacklight photomacrograph showing an example of a forging lap contained within the bail of a lug subjected to this screening. Concurrently, ARL and NAWC representatives visited the manufacturing facility in order to determine how the defective parts had made their way into inventory. It was determined that the contractor was not using an authorized written procedure for MPI. In addition, the contractor was using a system that was not capable of detecting defects in certain orientations. Further, poor lug handling practice was observed during the MPI process. This combination of factors allowed defective components to leave the facility undetected. As for the forging, the manufacturer took steps to minimize the amount of defects, including the use of a lubricant and decreased impact energy.

Effect of Manufacturing Defects on Performance: A lap is caused by the folding over of metal into the surface of the part during forming [1], while a seam is a discontinuity in a part caused by an incomplete joining of material during forming [2]. As shown in proof testing, the lugs were very sensitive to these surface anomalies. It was fortunate that defective lugs were revealed as a result of this proof testing (which is performed on a sampling basis), rather than in service.

Outcome: An extensive lug screening process was undertaken, whereby the parts that were previously magnetic particle inspected were subjected to an additional inspection consisting of a central conductor shot and a head-shot. The handling of the lugs subsequent to inspection was also improved, in an effort to reduce the masking of defective parts. Thousands of lugs were scrapped as a result of this re-inspection, and the warranty clause was invoked by the NAWC.

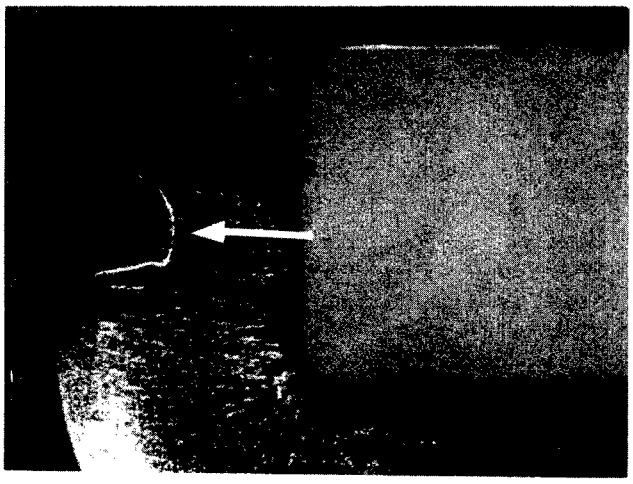


Figure 2 Blacklight macrograph showing typical lap defect within the bail region. Mag. 3x

COMPONENT: LAU-7 Missile Launcher Attachment Bolts
MANUFACTURING DEFECTS: Machining Rather Than Forging, Inadvertent Carburization

Background: Two LAU-7 missile launcher attachment bolts were found broken at Oceana, Virginia during pre-flight inspection. The bolts were installed on the aircraft for a total of two months before the failure was noted. The component is used to attach the missile launcher rails to the underside of Navy fighter aircraft, and is fabricated from Hy-Tuf® steel (AISI 4340 derivative). The bolts were required to be vacuum cadmium coated.

ARL Investigation: The failed bolts were sent to ARL from the Naval Air Warfare Center (NAWC) for failure analysis. The failure investigation included visual examination, metallography, chemical analysis, fractography, hardness testing, and stress durability testing.

Results of Investigation: Metallography and hardness testing revealed that the bolts were inadvertently carburized, which was not in conformance with the governing requirements. Additionally, macroetching revealed that the parts were machined from stock (rather than forged) and the threads were cut (rather than rolled). Figure 3 shows the etched grain flow within the threads. The rolling process would have produced a grain flow that followed the contour of the threads, however, the grain flow in Figure 3 does not follow the contour. It was determined that hydrogen-assisted stress corrosion cracking (SCC) was the probable cause of failure in both bolts. Hydrogen charging resulted from the surface corrosion. Contributing factors to SCC included surface carburization and the unacceptable grain flow pattern. Carburization resulted in a much harder (less tough) surface, while the stress distribution within the bolt head was adversely affected by the improper grain flow.

Additional Testing: As previously mentioned, stress durability testing was conducted on bolts from inventory to verify that the parts did not fail due to hydrogen charging from the plating process (in the case that the parts were electroplated rather than the required vacuum coating). The bolts were loaded to 80% of the UTS, and sustained for 200 hours. No failures occurred as a result of this testing. Also, ARL examined a number of bolts from different manufacturing lots for carburization and grain flow, in an effort to verify the extent of the problem. ARL was able to identify specific heat lots that were affected, and recommend others for continued use.

Effect of Manufacturing Defects on Performance: The forging process results in a grain flow that follows the contour of the part, and offers three distinct advantages compared to a part that was machined; enhanced directional strength, structural integrity and dynamic properties [3]. By refining the grain structure and developing optimum grain flow, forging promotes desirable directional properties such as tensile strength and ductility, and dynamic

properties such as impact toughness, fracture toughness and fatigue strength. With respect to structural integrity, forged parts are generally free from voids and porosity. All of these properties for the bolts under investigation were compromised as a result of the parts being machined rather than forged. The same advantages apply to threads that are rolled rather than cut. Carburization raises the surface hardness of the part, and is usually beneficial with respect to surface wear and fatigue resistance. However, the increased surface hardness made these components more susceptible to hydrogen attack.

Outcome: As mentioned, ARL offered a short-term recommendation concerning which bolts to continue using, and a long-term recommendation to consider changing the bolt to a lower strength (higher ductility, fracture toughness) material.

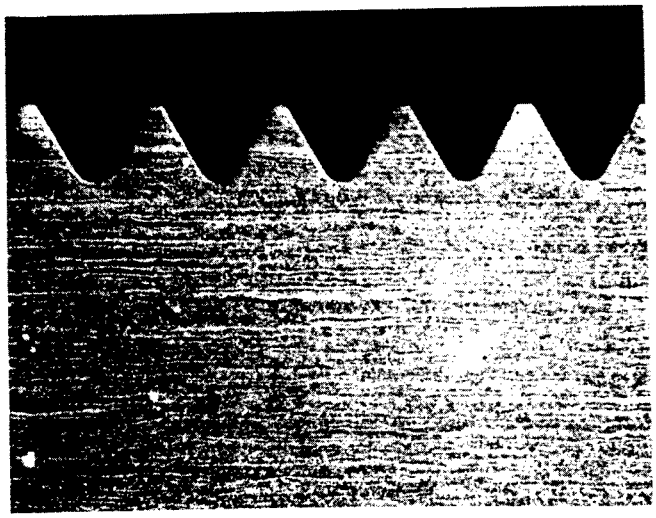


Figure 3 Macrograph showing the grain flow within the threads of the LAU-7 bolt. Note the flow does not follow the contour of the threads, indicating the threads were cut (machined) rather than rolled. Mag. 12.5x.

COMPONENT: **Rockeye XVI Cluster Bomb Tailcone Assemblies**
MANUFACTURING DEFECTS: **Casting Heat Checks, Inclusions, Porosity and Shrinkage**

Background: ARL conducted an analysis of two semicircular aluminum die-castings (alloy A356) that are components of the tailcone assembly of the Rockeye XVI Cluster Bomb. As the name implies, these components are located in the aft section of the bomb. The parts were rejected as unserviceable but repairable by the NAWC based upon a surface condition noted during visual examination of the tailcones in inventory, and sent to ARL.

ARL Investigation: At ARL, the two tailcones were subject to visual and radiographic inspection, chemical analysis, metallographic examination, hardness testing, tensile testing and scanning electron microscopy with energy dispersive spectroscopy.

Results of Investigation: Visual examination revealed the presence of "heat checks" (Figure 5) on the surface of the tailcones, while radiography showed indications of foreign material (both more and less dense than the casting material), gas holes, and shrinkage defects. Tensile testing showed that the specimens fabricated from the components did not meet the required mechanical properties for this alloy. In some cases, the tensile failures initiated at large inclusions.

Additional Tasks: As a result of these findings, an inspection was performed at the component manufacturing facility. A tour was given of the entire production process, whereby it was witnessed that hardened slag products along the outside of the crucible were inadvertently being poured into the casting die with the molten metal. It was concluded that this may have attributed to the inclusions noted in the two tailcones that were examined. Recommendations were provided concerning this condition, as well as the presence of heat checks. To determine the extent of this casting problem, twelve tailcone sections from inventory were sent to ARL for analysis, similar to that performed on the two original tailcones. Several of these components exhibited shrinkage cavities, gas holes, and most failed to meet the required mechanical properties.

Effect of Manufacturing Defects on Performance: Heat checks are most often located on surfaces that correspond to areas in the die which are subject to high thermal stresses or where the liquid metal flows at high speeds causing erosion of the mold or die cavity [4]. These defects can also be caused by extended mold usage or die wear, and are open to the surface, resulting in decreased structural integrity. Gas holes (porosity) are generally formed by an excessive amount of gas in the metal bath, which is released during solidification. This defect reduces the cross sectional area of the component. Shrinkage typically occurs in the last areas to solidify, or areas in contact with gates. This defect reduces the cross sectional area of the component to a greater extent than porosity. In

aluminum die casting, especially Al-Si-Cu alloys containing iron (alloy of parts under investigation) intermetallic compounds (Fe-Al-Mn-Si combinations) which form locally or throughout the melt in the forms of grains or needlelike crystals because of excessively low temperature in the crucible holding furnace [5]. This was the situation noted at the manufacturer. As shown, these inclusions were present on the fracture surface of the tensile specimens, and most likely played a role in failure location.

Outcome: Not only did the components fail to meet the required mechanical properties, but the workmanship of the parts was less than acceptable. The extended analysis performed by ARL indicated that the problem was rather widespread. It was concluded that the serviceability of the casting was adversely affected as a result of these findings.

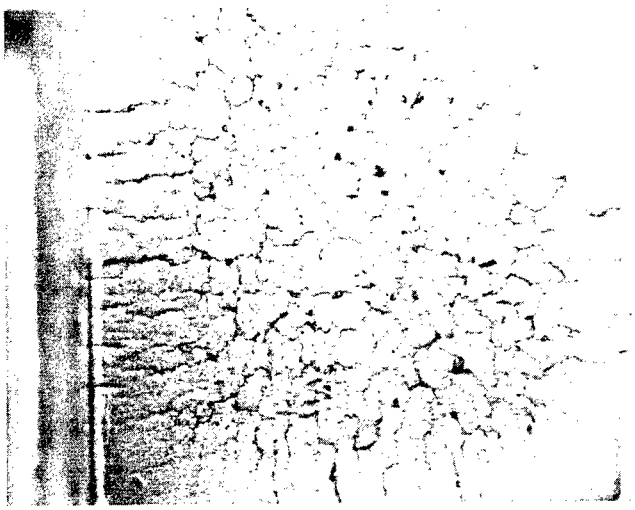


Figure 4 Macrograph showing the heat checks noted on the tailcone sections. Mag. 1.8x.

COMPONENT: MK 83 and MK84 General Purpose Bomb Fins
MANUFACTURING DEFECT: Non-Penetrating Spot Welds

Background: A First Article Inspection was performed for the MK83 and MK84 conical bomb fins. It was noted upon visual examination at the manufacturing plant that the chamfered butt welds (fusion weld) and the plug welds (resistance weld) showed less than optimal workmanship. The plug welds attach the outer skin to the inner spar, while the chamfered butt weld attaches the skin to the ring that is used to secure the fin to the bomb. The skin and associated components are made of low carbon steel.

ARL Investigation: Several of each type of conical fins was shipped to ARL for examination. Among other characteristics, the integrity of both the chamfered butt weld and the plug welds was examined. ARL performed radiography, tensile testing, visual examination and metallography of the failed test specimens in order to characterize the weldments.

Results of Investigation: Radiographic examination did not reveal significant nonconformities within the welds. Initial peel tests resulted in sheared spot welds, and unacceptably low loads. Figure 5 shows an example of a sheared plug weld. Note the burning that occurred on the underside of the parent material, indicating a lack of control. The process was improved, however, subsequent peel testing revealed a lower than nominal nugget size and corresponding lower than nominal pull loads for the plug welds. The tensile data from the butt welds conformed to the governing requirements. Visual and metallographic examination confirmed inadequate plug weld penetration. Visual examination of the sectioned fins also revealed cracked welds prior to tensile testing, plug welds that were misaligned with the intended positioning, and only partially attached plug welds to the spar. The data and photos were presented to the contractor, and the plug welding process was further improved. Not only were the through-holes for the plug welds enlarged, welding parameters were altered to allow for increased depth of penetration (i.e. increased dwell time, and increased current). This resulted in acceptable welds.

Effect of Manufacturing Defects on Performance: In general, the minimum depth of fusion is generally accepted as 20 percent of the thickness of the thinner piece [6]. The depth of fusion of the parts in question was much less than this amount, indicating less than adequate heating during welding. Since these welds were of inferior quality, the peel strength and the nugget diameters were well below the specified requirements. This may have led to eventual failures during service or storage.

Outcome: As mentioned previously, the spot weld dwell time was increased, as well as the current used to perform the welding. Since this anomaly was noted during a First Article Inspection of the manufacturing plant, none of the defective parts made their way into service.

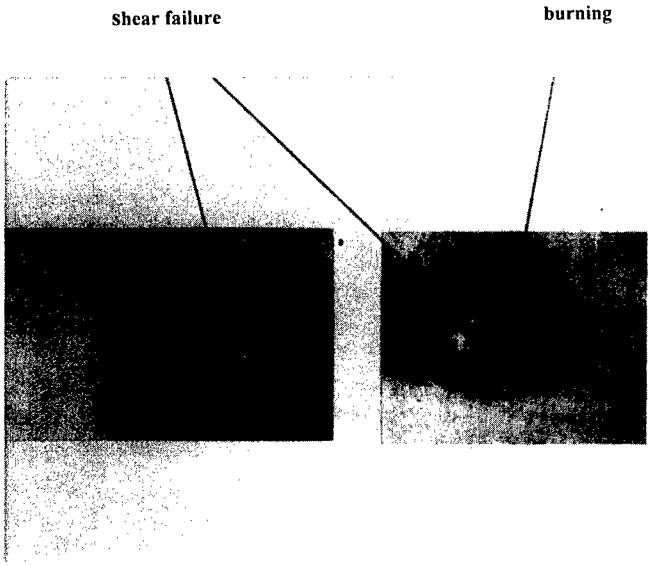


Figure 5 Micrograph showing shear failure of a resistance plug weld. Mag. 1x.

COMPONENT: M61A1 Breech Bolt Assembly
MANUFACTURING DEFECT: Improper Chemistry, Heat Treatment

Background: ARL characterized unused and failed "after-market" breech bolt assemblies from the M61A1 Gatling gun used on F-14 and F-18 Navy fighter jets. In addition, an individual locking block was examined. Similar components had exhibited accelerated wear during F-14 gun mount firing tests conducted at NAWCAD, Patuxent River.

ARL Investigation: A multitude of tests were performed at ARL to characterize the aforementioned components. The entire breech bolt assembly was subjected to a continuity test and a high voltage test, while the firing pin protrusion was measured in the locked and unlocked position. The component was subsequently disassembled, such that the individual components could be examined. The following analyses were conducted on the individual components; visual examination, surface finish, dimensional verification, magnetic particle inspection, metallography, chemical analysis, microhardness testing, macrohardness testing, coating thickness (where applicable), decarburization measurement (where applicable), case depth (where applicable) and double shear testing (where applicable).

Results of Investigation: Visual examination of the individual locking block revealed corrosion and a rougher than nominal surface finish. Metallography revealed a *complete* layer of decarburization along the periphery of the component, which was prohibited according to the governing specification. The depth of decarburization was greater than specified (0.006 vs. 0.003-inch). Hardness testing showed that the entire component (not just the area above the bail, as specified) was hardened to 50 – 55 HRC, not 38 – 43. Hardness testing also showed a loss of surface hardness due to the presence of the decarburization (~20 HRC at the surface, as opposed to 50 – 55 HRC elsewhere). Figure 6 shows a Knoop microhardness profile through the decarburized layer. Note the larger surface readings corresponding to a lower hardness. In addition, the unused top and bottom bolt shafts were not nitrided, as confirmed through metallography and microhardness testing. These components also had a higher than specified silicon content by almost double. Finally, both of the spiral spring pins failed to achieve the 3,900-pound double shear load. One of these pins contained carbon approximately 10 times the requirement, which could have attributed to the poor double shear results.

Additional Tasks: ARL also examined a bolt shaft that had failed while in service. The results showed that the component failed due to fatigue. In addition, the part was not nitrided as required, which most likely led to the premature failure.

Effect of Manufacturing Defects on Performance: With respect to chemistry, the silicon content of maraging steels should not exceed the maximum limit, since the notch tensile strength [7], as well as the toughness [8] have been shown to be detrimentally affected. Concerning heat treatment, decarburization is a loss of carbon from the surface of a hardened steel part usually caused by an excessively high dew point or low carbon potential during the diffusion portion of a carbide-diffuse cycle, or of prolonged reheating in moist air or other decarburizing gas [9]. A complete decarburized layer consists of ferrite, which transitions to a layer of ferrite plus a low-carbon martensite towards the core of the component, followed by the normal tempered martensitic structure. The presence of decarburization acts to lower the surface hardness of a component, as was shown during the ARL analysis. Decarburization also affects the wear and fatigue resistance of a component. Finally, the effect of a lack of nitriding on a maraged steel component was shown by the in-service fatigue failure. Nitriding of maraging steels is performed to provide resistance to wear and fatigue [10], and should not have been neglected given the application.

Outcome: The quality and workmanship of the "after-market" components were poor, and these components should never have made there way into service. A First Article Inspection at the "after-market" manufacturing facility should have revealed these detrimental anomalies. The ARL results were presented to both the Air Force (procuring activity) and the NAWC (receiving activity).

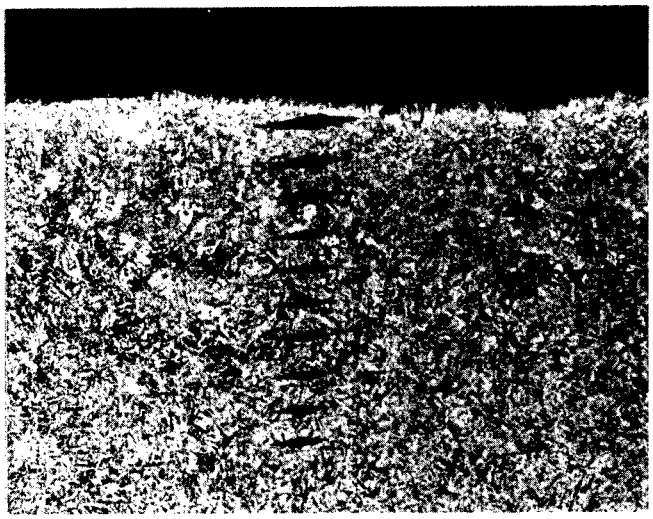


Figure 6 Microstructure of a locking block that contained decarburization. Note the softer surface indicated by the larger Knoop microhardness readings. Mag. 200x.

Conclusion: As shown in these few examples, manufacturing defects have the potential to negatively impact parts that are able to make their way into service. It is the responsibility not only of the manufacturer to adhere to quality workmanship practices, but also of the procuring activity, to ensure such defects are noticed prior to purchase and fielding of the parts. For the Department of Defense, this screening process is known as the First Article Inspection, where manufacturing processes (such as forging, casting, heat treatment, welding, etc.) and fabricated components are scrutinized prior to the start of final production. This inspection is intended to ensure conformance to the governing engineering drawings and specifications.

References

[1] *Product Design Guide for Forging,* Published by the Forging Industry Association, Cleveland, OH, 1997, pp. 148.

[2] *ASM Materials Engineering Dictionary*, Edited by Davis, J.R., ASM International, Materials Park, OH, 1992, p. 398.

[3] *Product Design Guide for Forging*, Published by the Forging Industry Association, Cleveland, OH, 1997, pp. 6-7.

[4] *International Atlas of Casting Defects*, Edited by Rowley, M.T., American Foundrymen's Society, Des Plaines, IL, pp. 46-47.

[5] *International Atlas of Casting Defects*, Edited by Rowley, M.T., American Foundrymen's Society, Des Plaines, IL, p. 269.

[6] *Welding Handbook*, Eight Edition, Volume 1, Welding Technology, Edited by Connor, L.P., American Welding Society, Miami, FL, 1987, p. 371.

[7] Decker, R.F., Eash, J.T., Goldman, A.J., *18% Nickel Maraging Steel,* Source Book on Maraging Steels, American Society for Metals, Materials Park, OH, 1979, p. 370.

[8] Vishnevsky, C., *Literature Survey on the Influence of Alloying Elements on the Fracture Toughness of High Alloy Steels*, Interim Report – Contract DAAG 46-69-C-0060, Army Materials and Mechanics Research Center, February 1970, p. 34.

[9] Carburizing and Carbonitriding, Edited by Gray, A.G., American Society for Metals, Materials Park, OH, 1977, p. 42.

[10] Appendix, Source Book on Maraging Steels, American Society for Metals, 1979, p. 370.

EVALUATING THE IMPACT OF ENVIRONMENTALLY FRIENDLY CLEANERS ON SYSTEM READINESS

Wayne Ziegler
Military Environmental Technology Demonstration Center
Aberdeen Proving Ground, Maryland 21005-5059
wziegler@arl.army.mil

Albert J. Walker, Jr.
U.S. Army Environmental Center
Aberdeen Proving Ground, Maryland 21005-5401
Albert.Walker@aec.apgea.army.mil

Abstract: Federal, state and local regulations limiting the use, storage and disposal of hydrocarbon-based cleaning solvents have led to the uncontrolled replacement of solvents with environmentally friendly products. The Army and other defense agencies rely on these solvents to maintain unique, mission critical systems and materiel and the replacement of hydrocarbon solvents has resulted in use, approval and compatibility issues. The U.S. Army Environmental Center (USAEC) and Aberdeen Test Center (ATC) have developed an Alternative Cleaner Compatibility and Performance Evaluation Program to survey user needs, validate alternative cleaner performance, consolidate lessons learned and provide a web based dissemination, review and evaluation tool. The test criteria were developed based on input from the technical community, the test community and the user community. A cooperative program between cleaner manufacturers and USAEC/ATC is currently in progress to evaluate currently available cleaner technology. The objective of this paper is to discuss the current status of the alternative cleaner testing and efforts to develop a universal test protocol that will provide the DOD community with the data necessary to make wise decisions concerning the replacement of hydrocarbon based cleaners. The culmination of the process will provide a user-friendly mechanism to facilitate implementation of environmentally friendly replacement products and technologies.

Key Words: Aqueous cleaners, cleaning, environmentally friendly products, hydrocarbon-based solvents, material compatibility, product validation, system readiness.

Introduction

Background: The U.S. Army Environmental Center (USAEC) and the U.S. Army Aberdeen Test Center (ATC) are currently leading an effort to investigate the appropriateness of using aqueous based cleaners during general maintenance and repair operations. These efforts were prompted due to complaints from field components that alleged corrosion of equipment after continuous operational use of aqueous based cleaning systems. USAEC was given the task of executing a project for the purpose of

substantiating or disproving these performance claims. There are several issues driving the search by installation and field components to replace hydrocarbon-based cleaning solvents including many federal, state and local regulations that limit the use, storage and disposal of hydrocarbon-based cleaning solvents and in many cases significant cost savings. Unfortunately, the Army and other defense agencies rely on these solvents to maintain unique, mission-critical systems and materiel and these systems may be compromised by indiscriminate use of unqualified cleaners.

Problem: To put the problem into context, in 1998 more than 40 Army installations sought money for alternative cleaning systems through the Pollution Prevention Investment Fund (P2IF). The total FY99 funding request for these projects was $1.1 M for a total net annual cost avoidance for $642 K (payback 2.74 years). While this shows initiative and a commitment to stewardship, many of the installations have bought (or are trying to buy) products that have not been fully qualified for use on Army equipment. It also must be kept in mind that this is only one funding source and only in the Army. The true magnitude of the problem throughout the various branches of the armed forces is not well documented. The problem is compounded by the fact that many of these products have GSA contract numbers and are listed as "environmentally friendly" replacements in Defense Logistics Agency (DLA) catalogs. Many purchasing organizations are unaware of the requirement to request approval for changes in cleaner systems from the respective commodity manager.

Solution: The purpose of the effort initiated by USAEC was to provide a mechanism to evaluate aqueous-based cleaners for applicability to U.S. Army and DOD maintenance and repair activities. To achieve this goal USAEC and ATC coordinated the development of a comprehensive aqueous based cleaner test protocol. The protocol is unique because it is the first comprehensive test protocol known to have been developed for this purpose with input by stakeholders from the aviation, small arms, and tank, automotive and armaments communities. The initial test protocol development included Army stakeholders, however, ongoing efforts have included input from stakeholders in the Navy, Air Force and Marines. The goal of the effort has expanded to include the development of DOD test protocol for aqueous cleaners.

Current Status: The protocol has been implemented on a limited basis to test several cleaners to specific requirements of specific Army activities. The lessons learned from these small-scale applications have been incorporated into the final draft protocol currently being circulated to stakeholders within all branches of the armed services. USAEC and ATC are currently leading a multi-agency initiative to comprehensively test several cleaning products and gather data that can be used to make procurement and usage decisions. The agencies involved will use a through screening process to decide which products to put through the full range of performance tests. Testing will be jointly funded; solvent manufacturers will pay for the test on their specific products, while the Army will maintain overall test capabilities and purchase materials needed to conduct the test.

Test protocol

General: The cleaner performance test protocol includes three sub-test areas: Cleaner Evaluation; Material Compatibility; and Service Test. The issues of greatest concern to the technical user community were the material compatibility of the aqueous based cleaners with the materiel being maintained and the performance characteristics of the cleaners. Other areas of concern during protocol development were product characteristics, worker health and safety issues and environmental impacts. The criterions were developed based on military objectives and materiel. The evaluation methodologies were, however, based on national and international standards. Standard test methods were used wherever possible to promote broader acceptance and applicability of the test results for both the DOD maintenance and manufacturing communities.

Protocol Development: The development of the test protocol addressed three tasks in the following order: criteria development, selection of materials, and selection of test methods. The cleaner protocol was developed to satisfy a diverse set of criteria from the user community, the materials developer community and the scientific community. This set of criteria provided the basis for developing a protocol that would address the issues of all concerned parties. The criteria fell into three categories of; general product characteristics related to worker safety and environmental impact, performance of the product and material compatibility issues with regard to current solvent applications. The interested parties also identified those materials for evaluations that they felt were the most problematic in their operations. The final step of developing the protocol was to identify appropriate test methods to access the performance of the cleaners against the established criteria. The bulk of the test methods selected were standard test methods from recognized organizations like the American Society of Testing and Materials (ASTM) or Society of Automotive Engineers (SAE). In a few instances military standards were used to address issues specifically related to military materiel or missions. The protocol was reviewed and approved by all interested parties. Table I shows the resulting test matrix of test methods. Table II shows a list of the materials currently incorporated into the various sub-tests. These tables are attached at the end of the paper. Both of the tables represent the methods and materials that cover the concerns of the Army stakeholders who were the primary parties involved in the current draft of the protocol.

Materials Compatibility

Overview of Testing to Date: The products examined to date have excelled in areas of worker health and safety. They also performed well against the environmental impact and characteristic criteria that support the operational cost benefits of the product. These results confirmed the expected results since these are the primary advantages of environmentally friendly products. There have been some issues with cleaner performance. In some cases even though the cleaners cleaned off the contaminant materials, they left a residue on the materiel that adversely impacted the test results and maintenance procedures. In the area of material compatibility problems were identified

during corrosion testing of some of the metals as well as degradation of some plastics and coatings.

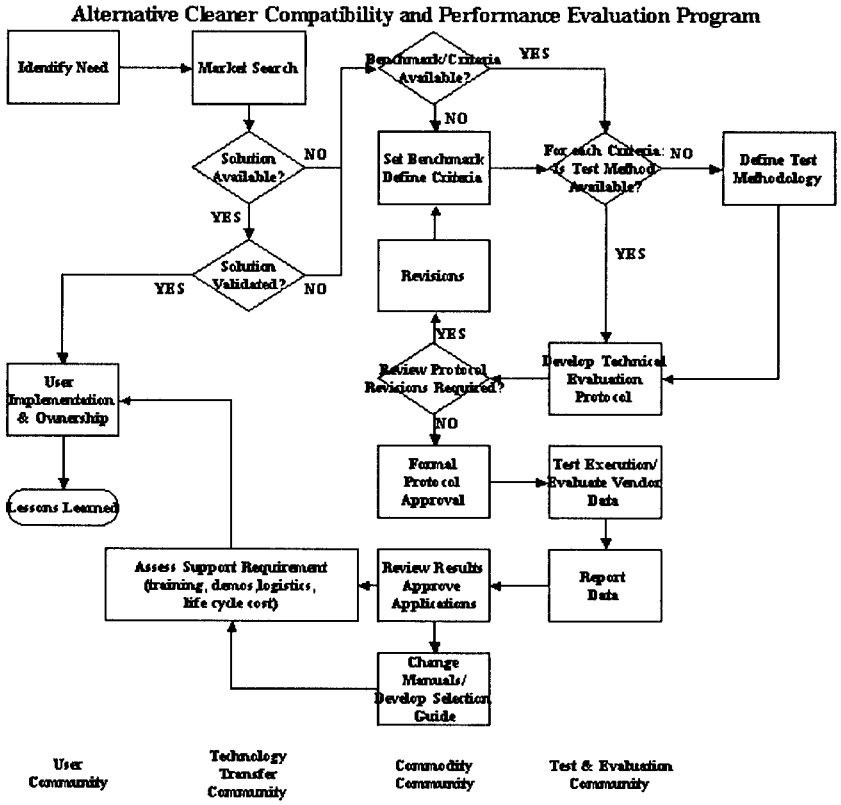


Figure 1. Flowchart representation of environmentally friendly product evaluation protocol development process from identification of need to user implementation and ownership.

Problems & test modifications: The objective of the protocol is to provide technical data on aqueous cleaners, which can be used to determine the cleaner's applicability to U.S. Army maintenance and repair activities. During test method selection every effort was made to use test methods that would simulate actual operation environment conditions for the tested cleaners. Towards this end some of the standard test methods were modified to provide test conditions that more accurately reflected operating environments. During the initial test program several problems, shortcomings and improvements were identified. In some cases the criteria were vague and difficult to use in evaluation or did not allow for a full indication of the test results. An example of this

was the total immersion corrosion criteria that stated a weight loss criterion and referenced a visual inspection but provided no pass/fail criteria for the visual inspection. As a result instances occurred where a specific combination of cleaner and material resulted in significant general corrosion, however, the weight loss criteria was not exceeded. Another challenge was the procurement of the materials. Some of the material alloys identified are costly and difficult or impossible to procure in the relatively small quantities required for testing. In some cases it was difficult to identify an appropriate test method to evaluate properties such as cleanliness or odor. All these issues need to be addressed keeping in mind the requirement of not only developing a technically sound protocol but also a test matrix that can be completed in a reasonable period of time for a reasonable investment of funds. One improvement that has already been incorporated into the protocol is a phasing of the test matrix. The test methods were grouped to provide maximum return on the dollar early in testing. Each phase consists of group of test methods that are conducted simultaneously and the performance of each cleaner is evaluated at the completion of each phase. This protects the investment of the manufacturer by identifying potential problems quickly utilizing relatively inexpensive test methods. A cleaner that fails a critical sub-test during a given phase of testing is not required to continue and thereby saves the cost of the later phases of testing.

Following is a discussion of several of the test methods utilized and any modifications applied as well as representative results.

Total Immersion: The total immersion corrosion caused by the manufacturer's suggested working concentration of the cleaner is determined using ASTM F-483-90, Standard Test Method for Total Immersion Corrosion Test for Aircraft Maintenance Chemicals. In addition to the requirements of ASTM F-483-90 the testing is conducted at the operating temperature for the cleaner. In many cases for aqueous cleaners the operating temperature is 100-105 °F. Problems in past testing have included excessive weight loss for Mg and Cd-plated 4340 samples. In some cases significant general corrosion was noted in test specimens that met the weight loss criteria. This is an example of an area where inspection criteria need to be better defined.

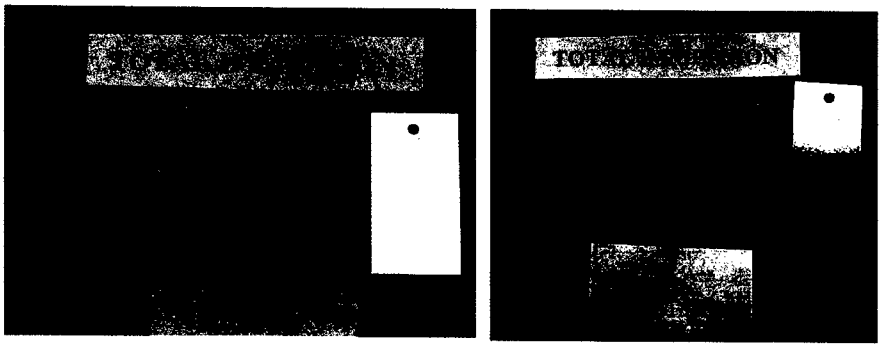


Figure 2. Total immersion test specimens of maraging C-250 and Cu UNS 36000 after 168 hours of exposure in an aqueous based alternative cleaner.

Sandwich Corrosion: The sandwich corrosion caused by the manufacturer's suggested working concentration a cleaner product was determined using ASTM-F-1110-90, Standard Test Method For Sandwich Corrosion Test. The criterion in the Test Protocol states, "The manufacturer's suggested working concentration shall not cause a corrosion rating greater that two (2) on any test panel and the manufacturer's suggested working concentration shall not cause a corrosion rating greater than P-D-680 (II)" (for zinc-phosphated 4340 coupons only). ASTM F-1110-90 states, "Any corrosion in excess of that shown by 'reagent water group' shall be cause for rejection." Any panel with pitting was given a severity rating of 4. Some cleaning products tested to date have had difficulty meeting the criterion for sandwich corrosion for the following materials: PH 13-8 Mo stainless steel, maraging C-250 steel, AISI/SAE 4340 steel, magnesium AMS 4377, and zinc-phosphated 4340 steel alloy (figure 2).

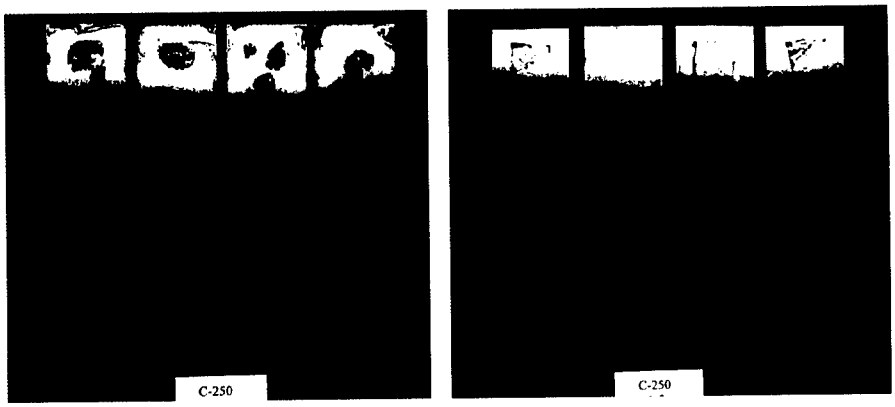


Figure 3. Sandwich corrosion maraging C-250 steel test specimens exposed to the test cleaner (left) and reagent water (right).

Effects on Painted Surfaces: The criteria is that the manufacturer's suggested working concentration of the cleaning compound shall not cause streaking, discoloration, blistering or a permanent decrease in film hardness of more than one (1) pencil hardness level on any painted surfaces. The effect of the manufacturer's suggested working concentration of the cleaning compound on the painted surfaces is determined using ASTM F-502-93 (app C, ref 15), Standard Test Method for Effects of Cleaning and Chemical Maintenance Materials on Painted Aircraft Surfaces, modified by the Test Protocol. One of the previously tested products did not meet the criterion for effects on painted surfaces for the MIL-P-14105 heat-resistant paint. There was also a slight color change on the exposed end of the MIL-C-46168 aliphatic polyurethane, single-component topcoat panels, which indicated marginal compatibility with this coating. One of the topcoat products specified was unavailable and it has been recommended that a replacement be selected by the interested technical POC.

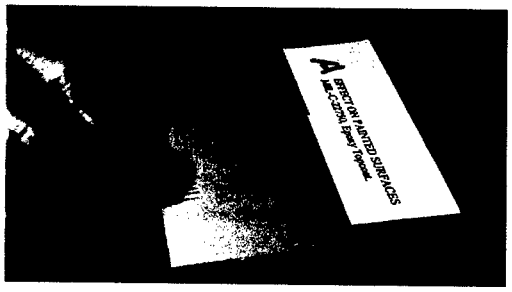


Figure 4. Testing for hardness changes in paints as a result of exposure to aqueous based cleaners.

Effects on Acrylic Plastics and Polycarbonate Plastics: The stress-crazing effect that the manufacturer's suggested working concentration of the test cleaner product has on acrylic plastics and polycarbonate plastics is determined using ASTM F-484-83, Standard Test Method for Stress Crazing of Acrylic Plastics in Contact with Liquid or Semi-Liquid Compounds. The criteria states that the manufacturer's suggested working concentration shall not cause stress crazing or staining of polycarbonate plastics. This test has been a challenge to set up for since it is difficult finding suppliers of these specific plastics in reasonable amounts for a single test. These materials are normally sold in large sheets and are relatively expensive. It has proved to be a valuable screening test since crazing of the polycarbonate plastic has been observed within 30 minutes of exposure to a subject cleaner.

Next Steps

Alternative Compatibility and Performance Evaluation Program: The program efforts at this stage are proceeding along two paths. USAEC and ATC are preparing to test a number of cleaner candidates based on the current protocol. The funding to purchase test materials and specimens has been provided through government sources, primarily through P2IF. The participating manufacturers will pay for each phase of testing as a lump sum and based on the results of each phase they may or may not continue to participate in subsequent phases. The methods included in each phase were chosen to provide maximum return on investment earlier in the testing. A program kick-off meeting to be attended by all interested parties is planned for February 2001. The data results produced by this testing will be passed along to the commodity managers and materiel developers so that they can make decisions regarding the use of a given cleaner. ATC's role is that of an independent test organization. ATC will provide the test services and technical evaluation of each cleaner relative to the test criteria. Commodity managers and materiel developers who have both the expertise and authority to make these decisions will make decisions regarding product use.

The second aspect of the ongoing aqueous based solvent project is the continued refinement of the protocol itself. The protocol is a living document at this point and one

of the main thrusts is to involve a maximum number of technical organizations within the DOD system in order to include their input. The ultimate goal is to develop a universal test protocol that will provide all members of the DOD community with the data they need to make wise decisions concerning the replacement of hydrocarbon based cleaners. Currently personnel within the Navy, Air Force and Marines are reviewing the protocol and the Navy has already provided input about additional test methods and materials that they will need added to the protocol to cover their unique requirements.

Related Efforts Concerning Other Categories of Environmentally Friendly Products: One of the lessons learned throughout the process of identifying the requirements and developing the protocol for evaluating aqueous based cleaners relates to the whole arena of environmentally friendly replacement products. There are many products throughout the government procurement system that are being billed as environmentally friendly alternatives to approved products. There is a separate DLA catalog for these products and many of these products are being offered for procurement in the same manner as the hydrocarbon based solvent replacements, without approval for use from the appropriate agencies. There is no standard mechanism for evaluating the claims of these products or the impact of these products on DOD materiel. A proposal has been submitted and is currently being evaluated recommending that the protocol development process used for this project be used as a strawman for the process of evaluating other categories of replacement products. In general, a small test protocol development team would be assembled at ATC and operate under the direction of a DOD working group involving appropriate technical personnel from commodity managers, materiel developers and testing communities. The working group or the funding source would identify a priority list of those product categories that will require evaluation. The working group would solicit and identify the requirements (criteria) for a given product category, review the test methodology (test protocol) and provide final approval of the test protocol. The development team at ATC will develop a draft test protocol for each product category using the working group requirements as the metrics for the product evaluation, and national and international test methods, when available, to ensure product vendor acceptance of the protocol.

Conclusion

The Alternative Solvents Substitutes Performance Validation Test Protocol addresses many of the concerns that both the user community and the material developer communities have identified. Thanks to the Alternative Solvent Substitution Performance Validation Program, the Army and other DOD agencies will be able to better preserve readiness, save money and avoid bad decisions by knowing which alternative cleaning products meet its stringent requirements for performance, soldier safety and environmental compliance. Vendors and manufacturers will have a clearly defined and accepted process for validating their products for possible defense procurement. Using this program as a model, performance validation protocols for other environmentally friendly product replacements can be developed and implemented.

Table I. Test Matrix of Methods Currently Incorporated into the Alternative Cleaner Compatibility and Performance Evaluation Test Protocol.

Test Matrix - Alternative Cleaner Compatibility & Performance Evaluation Program		
Test Title	**Test Method**	**Requesting Agency**
Flash point	ASTM-D-92-90	All
pH	ASTM-E-70-90	All
Heat stability	MIL-C-87937B	AMCOM
Toxicity	AR 40-5	AMCOM/All
Biodegradability	40CFR Part 796.3100	ATC
Non-volatile residue	MIL-C-87937B	AMCOM
Cleaning efficiency	ASTM-F-22-65	AMCOM
Constituents	MIL-C-29602	All
Appearance	MIL-C-29602	AMCOM
Volatile organic chemicals	EPA Method 8206A	All
Water break free	ASTM-F-22-65	AMCOM
Cold stability	MIL-C-87937B	AMCOM
Fluorescent penetration Inspection	Level(IV) Inspection	AMCOM
Drying point	ASTM-D-86-96	TACOM
Relative solvency	TACOM Method	TACOM
Non-volatile residue (TACOM)	ASTM-E-1131-93 Mod	TACOM
Coating adhesion	Fed Std Method 6301.2	AMCOM
Effects on painted surfaces	ASTM-F-502-93	AMCOM
Total immersion corrosion	ASTM-F-483-90	AMCOM
Sandwich corrosion	ASTM-F-1110-90	AMCOM
Hydrogen embrittlement	ASTM-F-519-93	AMCOM
Effects on unpainted surfaces	ASTM-F-485-90	AMCOM
Effects on polymide wire	MIL-C-87937B	AMCOM
Effects on acrylic plastic	ASTM-F-484-83	AMCOM
Rubber compatibility	ASTM-D-2240-95	AMCOM
Effects on polysulfide sealant	MIL-C-87937B	AMCOM
Effects on polycarbonate plastic	ASTM-F-484-83	AMCOM
Effects on bonding	ASTM-D-3167-93	AMCOM
Stress corrosion	ASTM-G-44-94	AMCOM
Effects on sealant peel strength	AMCOM Procedure	AMCOM
Copper corrosion	ASTM-D-130-94	TACOM
Steel corrosion	ASTM-D-130-94 Mod	TACOM
Bimetallic couple corrosion	Fed Std 791C	ARDEC
Effects on storage	ATC Test Method	ARDEC

Table II. Listing of the Materials Currently[1] Incorporated into Alternative Cleaner Compatibility and Performance Evaluation Test Protocol.

Metals

—Aluminum, 2024-T3 (Anodized per MIL-A-8625, Type I)

—Aluminum, 2024-T3 (Conversion Coat per MIL-C-5541)

—High strength steel AM-355 CRT

—High strength steel PH 13-8

—High strength steel Maraging C-250

—Aluminum 7075-T6

—Titanium 6AL-4V

—Steel 4340

—Aluminum 7075-T6 (Alclad)

—Magnesium AMS-4377 (surface treatment MIL-M-3171, Type III)

—Cadmium plated steel (4340), **ASTM-F-519-93 plating method**

—Nickel plated steel (4340)

—Steel 1020

—Inconel 718-Bar

—Ti-6AL-4V-Bar

—Zinc phosphated steel (4340), per DOD-P-16232F

—Manganese phosphated steel (4340), per DOD-P-16232F

—Copper alloy UNS C36000

—Copper, hard tempered, cold-finished, 99.9 % purity

Paints

—Primer coating, MIL-P-23377 epoxy

—Primer coating, MIL-P-85582

—Top coat MIL-C-85285, polyurethane, High solids

—Top coat MIL-C-22750, Epoxy

—Top coat MIL-C-46168, Aliphatic, polyurethane, single component

—Top coat MIL-L-46159, Lacquer, acrylic, low reflective

—Top coat MIL-P-14105, Heat resistant

—Top coat MIL-E-52891B, Enamel, lusterless, zinc phosphate, styrenated alkyd type

Other materials

—Acrylic plastic MIL-P-5425, Finish A

—Acrylic plastic MIL-P-8184, Finish B

—Acrylic plastic MIL-P-25690

—Polycarbonate plastic MIL-P-83310

—Polymide wire

—Rubber, Type SAE 3204

—Rubber, Type SAE 3209

—Polysulfide sealant MIL-S-81733, Type 1

—Polysulfide sealant MIL-S-8802, Type 1

Note: 1. Additional materials will be added as required to support the requirements of the Navy, Air Force and Marines.

ESTIMATION OF RELIABILITY GROWTH DETERMINATION IN CRACKED SPECIMENS UNDER FATIGUE FAILURE

M. Riahi and M. Aslanimanesh

Department of Mechanical Engineering; Iran University of Science and Technology
Riahi@Sun.iust.ac.ir , Aslanimanesh@Yahoo.com

Abstract:

A probabilistic analysis of the fatigue crack growth for reliability growth calculation on the mechanical component is presented on the basis of fracture mechanics and theory of random process. The loading is postulated to be stationary, narrow-band random Gaussian process and consequently, randomized Paris-Erdogan law is applicable. As a specific problem, a thin plate having a central crack is analyzed by two analytical methods "stochastic averaging" and "damaged linear accumulation''. Simultaneously, the aforementioned plate is being analyzed by the utilization of "NISAII", Software. Analysis made by "NISAII" is on the basis of Monte-Carlo simulation random analysis. In the end, all the results are compared with each other and conclusion is drawn.

Key Words:
Fatigue damaged linear accumulation; Monte Carlo simulation; Random fatigue crack growth; Random loading; Reliability growth; Stochastic averaging

Introduction:

One of the main contributors to the structure failure of an aircraft is a phenomenon known as fatigue. Most of the mechanical components, especially aircraft structure, have been designed on the basis of "Fail safe" and "damage tolerance" concepts. Providing assurance of no failure operation is highly important in these components. Reliability evaluation can provide useful information in the fatigue control.

Evaluation of fatigue effect on mechanical components containing initial crack is rather complex. The complicity arising from crack growth in component that depends on essential properties of material under fatigue load also depends on geometry of component as well as environmental conditions. In the laboratory controlled conditions,

if an experiment is repeated many times over and test conditions remain the same, the results are not deterministic, and in most cases are scattered. Critical elements designed on the basis of the above-mentioned points will be very reliable. One of the important concepts in reliability analysis is the definition of failure criteria, which can estimate probability of useful operation. In mechanical components, existence crack is one of major failure causes, hence reaching crack length to a critical length can be defined as the failure criteria. Equations of fracture analysis are in the form of probability, therefore we can't apply deterministic fracture mechanics. Stochastic behavior should be evaluated by statistical and probability theories for reliability analysis which introduce complicity in analysis results. Reliability can be analyzed with obtained results and proper modeling and solution by stochastic method Thus becomes an important point.

In this regard, there are a host of stochastic methods. Moreover, recently presented methods reduce essential complicity of probability calculations. Stochastic averaging method for fatigue crack measurement in a component under random loading presented by Zhu and Lin (1992)[1] is to name one. B.Kececioglu in 1998 analyzed reliability of mechanical component under wide-band and narrow-band loading by definition of damage function [2]. Results obtained from the aforementioned researches are used as comparative criteria in the presented study here.

In this paper, reliability analysis of a mechanical component is evaluated under critical conditions. In addition, failure cause is assumed to be the fatigue fracture that arises from crack growth. Component behavior is evaluated on the basis of probability fracture mechanics and random vibration theories. Critical conditions are presumed to be simpler assumptions for reducing governing probability complicity. On the basis of these assumptions, the problem is designed and solved by governing theories. Property of a dynamic system is light damping and stationary narrow-band Gaussian random loading process. Consequently, the system response would be come close to the system's resonance response, (i.e. high stress fatigue is taken into account.) "Stochastic averaging method" and "damaged linear accumulation method" is used as analytical solutions. Software analysis (by NISAII) is presented on the basis of Monte-Carlo random simulation method.

Methodology:

In this section, analytical and software methods are used for the reliability growth determination in a cracked mechanical component under fatigue failure. Considering the complicity of methods for reliability growth in this field, it becomes necessary to design a problem with simpler conditions in order to make it possible to expand into general conditions.

Problem Assumption:

Dynamic Analysis

- one-degree of freedom and light damping system
- stationary wide band Gaussian random loading process(white noise)

Fracture Analysis

- only one crack in component
- length and location of crack known

- material behavior, brittle, homogeneous, isotropic
- high stress fatigue

Reliability Analysis

- failure case, fatigue fracture
- failure criteria, a=ac

Model Selection:

Model assumptions should be one-degree of freedom, light damping and wide band loading. Test specimen concerned is thin plate with a central crack, however, it can be expanded to other specimens.

Problem:

Consider a thin square plate 1×1, shown in Fig. 1, with an initial central crack of length $2a_o$ and supporting an infinitely rigid heavy mass M at its end. The plate is idealized to be massless, homogeneous, isotropic, and with light damping, and the mass M is subjected to a Gaussian load process perpendicular to crack (Y) with a wide-band one-sided spectral density G.

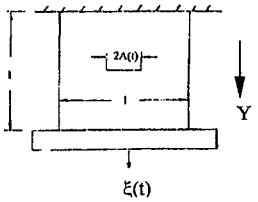


Fig.1 Thin plate containing central crack and mass in its end

Analytical Solution Methods:

For calculation of plate reliability growth presented "stochastic averaging" and "accumulative linear damage" as analytical methods

(1)Stochastic Averaging Method[3]

Model of System Dynamic

For one degree of freedom, the system can be considered as dynamic model of mass, spring and damping system, then the dynamic equation will be:

$$M\ddot{X}(t) + C\dot{X}(t) + g[A(t)]X(t) = \xi(t) \qquad (1)$$

The stiffness function g[A(t)] can be approximated by following polynomial expression[4]

$$g(a) = g(0)(1 - 1.708u^2 + 3.081u^4 - 7.036u^6 + 8.928u^8 - 4.266u^{10}) \qquad (2)$$

Which g(o) is without crack plate stiffness and $g(a_0 = 0) \cong E.A/L$ and $u = \frac{2a}{L}$.

We shall assume that the stress is equal to the displacement multiplied by a constant.

Fracture Analysis

Based on Paris-Erdogan model, the equation for crack growth rate is:

$$\frac{dA}{dt} = \frac{\omega(A)}{2\pi}\alpha(\Delta K)^{\beta} \tag{3}$$

Where:

$$\Delta K = 2\sqrt{\frac{\pi G g(a)}{C}}h(a)R \tag{4}$$

R , stress range, an approximate expression for h(a) has been provided in [4]as following:

$$h(a) = \sqrt{u}(0.467 - 0.514u + 0.960u^2 - 1.421u^3 + 0.782u^4) \tag{5}$$

now eq.(3) can be rewritten as follows:

$$\frac{dA}{dt} = \gamma Q(A)S^{\beta} \tag{6}$$

$$\gamma = \frac{\alpha}{2\pi}(2\sqrt{(\pi G/c)})^{\beta} \tag{7}$$

$$Q(A) = \omega(A)[h(A)\sqrt{(g(A))}]^{\beta} \tag{8}$$

In which S is the stress envelope process and R (t)=2 S(t). Then by solving the above equations, the following transition probability density of crack size A (t) will be yield.

$$q(a,t|a_0,0) = \frac{1}{\sqrt{(2\pi t)}Q(a)\sigma\Phi(m\sqrt{t}|\sigma)}\exp\left\{-\frac{\left[\int_{a0}^{a}\frac{du}{Q(u)} - mt\right]^2}{2\sigma^2 t}\right\}; \quad a \geq a_0 \tag{9}$$

Where:

$$m = \gamma\,\Gamma(1+\beta|2) \tag{10}$$

$$\sigma^2 = \gamma^2\sum_{n=1}^{\infty}c_n^2\tau_n \tag{11}$$

$$c_n = \sum_{k=0}^{n}\frac{(-1)^k n!}{(n-k)!\,(k!)^2}\Gamma(1+k+\beta|2) \tag{12}$$

$$\tau_n = 2\int_0^{\infty}\rho^{2n}(\tau)d\tau \tag{13}$$

$\rho=\exp[-C\,\tau/(2M)]$ correlation coefficient of loading process.
The reliability function follows:

$$R(t,a_{cr}|a_0,0)=\int_{a_0}^{a_{cr}} q(a,t|a_0,0)da=1-\frac{\Phi\left[\dfrac{mt-z_{cr}}{\sigma\sqrt{t}}\right]}{\Phi\left[m\sqrt{t}|\sigma\right]} \tag{14}$$

Where:

$$z_{cr}=\int_{a_0}^{a_{cr}}\frac{da}{Q(a)}$$

(2)*The Linear Accumulation Hypothesis in random fatigue crack growth*
Paris-Erdogan model is given in the form [5]

$$\frac{da}{dn}=C(\Delta k)^m \tag{15}$$

Where:

$$\Delta k=\Delta S\sqrt{\pi a}\,\eta(a)$$

When the crack size is very small in comparison with the component dimensions, the above equation can be rewritten as

$$\frac{da}{dn}=CG(a)(\Delta S)^m \tag{16}$$

Where:

$$G(a)=[\eta(a)\sqrt{\pi a}]^m$$

and Eq.(16) can be written

$$\Psi(a)=\int_{a_0}^{a}\frac{dz}{G(z)} \tag{17}$$

Where

$$\frac{d\Psi(a)}{dn}=C(\Delta S)^m$$

the damage indicator is defined as:

$$D=\frac{\Psi(a)}{\Psi(a_c)} \tag{18}$$

It can be shown that when a=a0, D=1; and when a=ac, D=1. The increase rate of the damage indicator therefore is

$$\frac{dD}{dt} = \frac{2^m f C(\Delta S)^m}{\Psi(a_c)} \tag{19}$$

Where S is the stress amplitude process, and $S = \frac{\Delta s}{2}$.

Definite the coefficient A as

$$A = \frac{\Psi(a_c)}{2^m C} \quad , \quad b = m \tag{20}$$

It can be shown in [5] that the damage accumulation under stationary stress loading process is a normally distributed random variable; i.e.,

$$D(t) \propto N[\bar{D}(t), \sigma_{D(t)}] \tag{21}$$

For narrow band stress process:

$$\bar{D}(t) = \frac{\omega_n}{2\pi} \frac{t}{A} (\sqrt{2}\sigma_x)^b \Gamma(\frac{b}{2}+1) \tag{22}$$

$$\sigma_{D(t)} = \frac{(\sqrt{2}\sigma_x)^b}{A} \Gamma(\frac{b}{2}+1) \sqrt{\frac{\omega_n}{2\pi} \frac{\varphi_1(b)}{\zeta}} \tag{23}$$

$\varphi_1(b)$ which is given in[2] and α and β definite:

$$\alpha = \frac{v}{\sqrt{u}} \quad , \beta = \frac{1}{u} \tag{24}$$

$$u \cong \frac{\omega_n}{2\pi A} (\sqrt{2}\sigma_x)^b \Gamma(\frac{b}{2}+1) \tag{25}$$

$$v \cong \frac{(\sqrt{2}\sigma_x)^b \Gamma(\frac{b}{2}+1)}{A} \sqrt{\frac{\omega_n}{2\pi} \frac{\varphi_1(b)}{\zeta}} \tag{26}$$

The reliability is then given by

$$R(t) = \Phi\{\frac{1}{\alpha}[(\frac{t}{\beta})^{-1/2} - (\frac{t}{\beta})^{1/2}]\} \tag{27}$$

Analysis:

Reliability-Time curve shown in Fig.2 for analytical method.

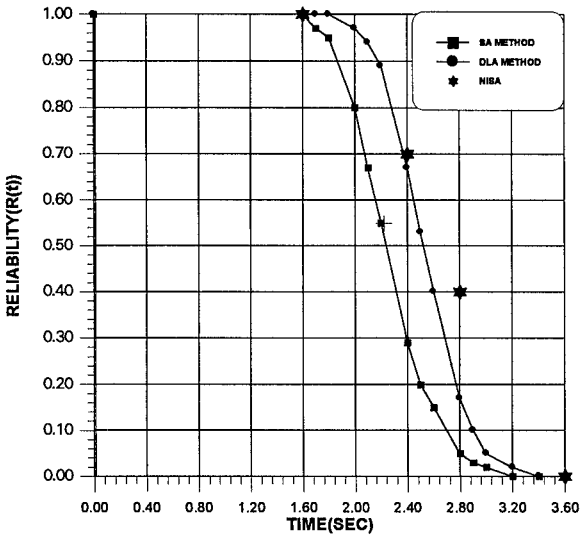


Fig.2 curve of Reliability –Time with two methods

Physical Data of Problem:

$2a_0$=0.00254 [m]
2ac=0.01016 [m]
C=3502.536 [Kg/S]
E=68.95 [GPa]
G=1.243[N^2/HZ]
K_{IC}=91.3 [Mpa $\sqrt{m}$]

L= 0.0254[m]
M= 54.28 [Kg]
Sy=560[Mpa]
th = 0.00254 [m]
α=0.1354E-8
β=2.25

Computer Solution Method:

The first dynamic analysis and former fatigue fractures analysis are carried out on the basis of Monte-Carlo method. Finite element method is used as numerical solution, which is carried out by "NISAII" software. At the beginning, Monte-Carlo analysis method is presented for random solution.

Monte-Carlo Simulation Method:

This method is the simulation of an experiment with a computer. The set of random numbers are generated for random parameters at the beginning, then random numbers are constituted in response equations from which the set of random numbers for identification of random behavior is obtained. This set is analyzed by statistical methods,

nonetheless it is possible to obtain quantitative and qualitative responses. Applications of this method are enormous. In this paper flow chart of method is shown in Fig. 3.

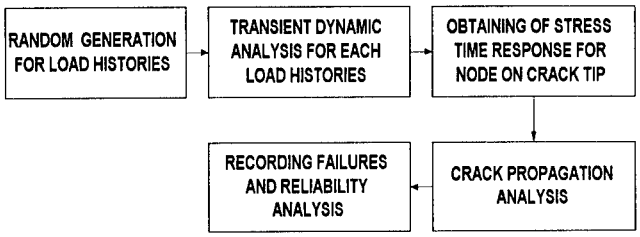


Fig.3 flow chart of reliability analysis by Monte-Carlo method

Dynamic Analysis:

Random vibration analysis is carried out on the basis of given dynamic data and white noise excitation $G_E = 1.24326\ [\frac{N^2}{HZ}]$ for obtained RMS of response that required in "Damage linear accumulation method" and determination of excitation frequency band that required in "Calculation of excitation standard deviation for transient dynamic analysis". Therefore, stress response S_{yy} is evaluated for node on the crack tip. Power Spectral Density (PSD) of S_{yy} stress response is shown in Fig. 4. A RMS value for loading process (force excitation) and stress dynamic response is obtained.

$$(RMS)_{EX.} = 1.6[N] \qquad (RMS)_{Res} = 1.5E2[Pa]$$

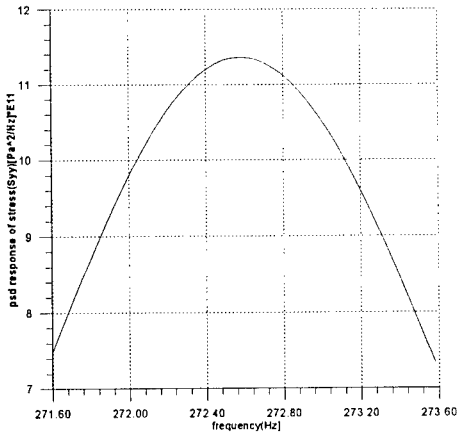


Fig.4 PSD Response of stress S_{yy} in frequency domain related to node on crack tip

Results of Crack Growth and Reliability:

Results of reliability analysis are presented in Tab.I for data of final crack length at time t=1.6 sec, and calculated reliability value is written only for other times.

Tab.I Results of Reliability and Fracture Analysis (a_c = 5.08[mm])

NUMBER OF SIMULATION	1	2	3	4	5	6	7	8	9	10
$a_{f\ [mm]\ at\ Time=1.6}$	3.26	3.6	3.92	2.76	4.3	3.64	3.21	3.48	3.7	3.08
FAILURE($a=a_c$)	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
NO FAILURE TIME(S)	10 t=1.6	7 t=2.4	4 t=2.8	0. t=3.6	0. t=4					
R(t)	10/10 =1	0.7	.4	0	0					

Conclusion:

Estimated reliability is shown in Fig. 2 by three methods "Stochastic Averaging (SA)", "Damage Linear Accumulation (DLA)" and "Software (NISAII)". Proceeding from Fig. 2 one can see that the results of this method are agreeable and very compatible with two methods mentioned previously. **The new proposed method enables us to analyze wider range of specimens in more complicated conditions**.

Reference:

1- J.N. Yang, G.C. Salivar and C.G. Annis, Statistical Modeling of Fatigue Crack Growth in a Nickel-Super Alloy., Engng Fracture Mech. 18, 1983,257-270

2- Dimitri B.kececioglu,A Unified Approch to Random- Fatigue Reliability Quantification under Random Loading,Proceedings Annual Reliability and Maintainability Symposium,California:Institute of Electrical Society, 1998,PP.308-313

3- W.Q.Zhu & Y.K. Lin,on Fatige Crack Growth under Random Loading, Engineering Mechnical Vol 43,No 1, 1992,pp1-12

4-M.Grigoriu,Reliability of Degrading Dynamic Systems.Stract Safety 8,1990,345- 351

5- Paris , P.C and Erdogan,F,"A Critical Analysis of Crack Propagation Llaws",Journal of Basic Engineerlng Trans. ASME,D 85, 1963, PP.528-534

PROGNOSTICS

Chair: Mr. Paul Grabill
Intelligent Automation Corporation

A PROGNOSTIC MODELING APPROACH FOR PREDICTING RECURRING MAINTENANCE FOR SHIPBOARD PROPULSION SYSTEMS

Gregory J. Kacprzynski
Michael Gumina
Michael J. Roemer
Impact Technologies, LLC
125 Tech Park Drive
Rochester, New York, 14445

Daniel E. Caguiat
Thomas R. Galie
Jack J. McGroarty
Naval Surface Warfare Center
Carderock Division
Philadelphia, PA 19112

Abstract: Accurate prognostic models and associated algorithms that are capable of predicting future component failure rates or performance degradation rates for shipboard propulsion systems are critical for optimizing the timing of recurring maintenance actions. As part of the Naval maintenance philosophy on condition based maintenance (CBM), prognostic algorithms are being developed for gas turbine applications that utilize state-of-the-art probabilistic modeling and analysis technologies. NSWCCD-SSES Code 9334 has continued interest in investigating methods for implementing CBM algorithms to modify gas turbine preventative maintenance in such areas as internal crank wash, fuel nozzles and lube oil filter replacement. This paper will discuss a prognostic modeling approach developed for the LM2500 and Allison 501-K17 gas turbines based on the combination of probabilistic analysis and fouling test results obtained from NSWCCD in Philadelphia. In this application, the prognostic module is used to assess and predict compressor performance degradation rates due to salt deposit ingestion. From this information, the optimum time for on-line waterwashing or crank washing from a cost/benefit standpoint is determined.

Keywords: Compressor Cooling, Cost/Benefit, Prognostics

Nomenclature:

C, F – Normal Distributions
N – Speed
P – Pressure
Q – Volumetric Flow
S – Weighted Coefficients
T – Temperature
y - Predicted Value

CIT – Turbine Inlet Temperature
CDT – Turbine Discharge Temperature
CDP – Compressor Static Discharge Pressure
CDP_T – Compressor Discharge Total Pressure
TIT – Turbine Inlet Temperature

Φ - Normalized Cumulative Distribution
α - Weighting Factor
γ - ratio of Specific Heats
σ - Standard Deviations
τ - prediction interval

Introduction: With a growing presence of gas turbine technologies, a stronger focus is being placed on trade-off analysis between performance optimization and O&M costs. As a result, cost/benefit evaluation of performance recovery methods has been at the forefront of these efforts. In both the military and private sectors, reducing the extra costs encountered by degraded performance parameters such as fuel consumption or power loss, have prompted research into prognostic and diagnostic technologies. Kurtz et al (2000) gives an excellent discussion of many performance degradation mechanisms; the majority of which are recoverable either through washing procedures, variable geometry adjustment or component replacement. However, optimization of both compressor crank and on-line washing intervals from the standpoints of fuel consumption and proactive maintenance is of primary interest to the US Navy and the focus of this paper.

The economic benefits associated with an optimized and condition-based on-line and off-line waterwash predictor are significant. Research by Haub et al (1990) showed that as much as 1188 Mw-hrs could be saved through the use of on-line washing. Additional savings of 450 MW-hrs was attributed to reduced maintenance costs and extended operating time between crank-washings. The study performed by Peltier et al (1995) showed that the use of on-line washing decreased average performance degradation from 1% per 100 operating hours without on-line washing to 0.2% per 100 operating hours with on-line washing.

An automated process is desired that is capable of detecting the severity of compressor fouling and relating that to the optimal time to perform maintenance based on O&M costs. When a compressor undergoes fouling, several key performance factors are affected. The most sensitive of these factors is the compressor capacity or referred mass flow Peltier et al (1995). This is because loss of capacity comes from throat blockage and increases in roughness on the suction side of the blading. Unfortunately, in most practical naval applications, compressor capacity is not reliably determinable. The compressor outlet temperature and discharge total pressure can typically be used to find compressor efficiency (Boyce 1995) however CDT, CDP_T are not standard sensors in most Naval platforms.

$$\eta_{adb} = \frac{\left[\left(\frac{CDP_T}{CIP_T}\right)^{\gamma-1/\gamma} - 1\right]}{\left[\left(\frac{CDT}{CIT}\right) - 1\right]} \quad (1)$$

Compressor fouling has also been shown to increase vibration, (Ozgur et al (2000) and Tsalavoutas et al) but there are many drawbacks to using this method to predict fouling severity. First is the complexity of separating out other modes that contribute to vibration increases and secondly is the poor reliability with which performance degradation severity may be assessed. In lieu of these practical issues, a primary goal of this effort was to be capable of predicting the optimal time to waterwash or crankwash using only the most essential parameters that are

currently available on most Naval installations. Specifically, the developed technique utilizes available performance parameters such as fuel flow and CDP, relates them to performance degradation levels utilizing the prognostic model, and then predicts the optimal timing for cleaning procedures.

Accelerated Fouling Testing on the LM2500 and Allison 501: The prognostic model was developed based on data from fouling tests taken at NSWCC in Philadelphia, PA. In order to simulate the amount of salt the typical Navy gas turbine is exposed to on a normal deployment, a 9% salt solution was injected into the engine intake. Over the course of the entire test (3 days) approximately 0.0057m^3 of salt was used to induce compressor degradation at four different load levels (1/3, 2/3, standard and full load levels or "bells"). This method of testing was performed on both Allison 501 and LM2500 Units. Figure 1 shows a borescope image of the salt deposits on the LM2500 1st stage blading.

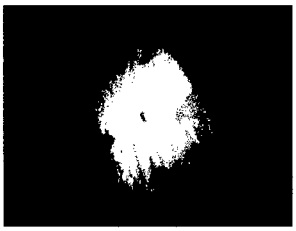


Figure 1 – Borescopic Image of salt deposits on 1st stage blading

In addition to fouling the two engines, testing was also performed on the effects of on-line washing for the Allison 501. The machine was crank washed and fouling was reinitiated. Specifically, at approximately 2% CDP drops, an on-line waterwash was performed using detergent. This cycle was completed 4 times at four different load levels.

During the testing, several of the critical parameters were monitored and their response to degradation was tended. Table 1 contains the measured parameters with their units and ranges (Shaft RPM and Ngg are for the LM2500 testing only).

Table 1 – Recorded Parameters from the DCS

Parameter	Units	Ranges
N_{gg}	RPM	0→9575
Q_{fuel}	GPM	0 → 97
TIT	°F	0 →2000
CDT	°F	0 →1468
CDP	psig	0 →300
CIT	°F	0 →500
Load	k-lbf	0 →300
Shaft RPM	RPM	0 → 274
$P_{barometric}$	inHg	

This list is much smaller than the *Required Instrumentation List* for performance testing purposed by Kurz et al (1999), it represents a much more realistic view of what instrumentation is actually installed on Naval platforms with the exception of CDT. However, through the use of experienced-based correlations, compressor degradation can still be accurately monitored. This

reduced list emphasizes how the methods described in the next section can bypass the need to know all the state variables at all the key gas turbine stations and still be able to track important performance parameters and their trends. The focus is to generate reliable indicators of compressor fouling not necessarily standard thermodynamics features.

Degradation Features: Before any performance features was generated the test data was referred (corrected) to standard day conditions to account for the changing environmental conditions that occurred over the three days of testing. In addition, due to difficulties in holding water brake load constant in the LM2500 test, corrections were also made for speed variations at the various load levels as well. Therefore, correction curves were developed for Q_{fuel} vs. N_{shaft}, CDT vs. N_{shaft} and CDP vs. N_{shaft} to compensate for the fluctuations encountered during testing.

As previously stated, the best features for identifying compressor degradation would be compressor capacity and total pressure ratio changes. The latter feature not only accounts for changes in static pressure drop but also losses in axial velocity due to wake losses and blade exit angle distortions. With total pressure measurements absent there is not enough information to calculate compressor adiabatic efficiency in its strict form (as was shown in eq.1).

Alternately, Eq.2 may be used whose components are shown in Figure 2.

$$\eta_{adb} = \frac{(h_{3t} - h_{1t})_{ideal}}{(h_{3t} - h_{1t})_{actual}} \tag{2}$$

where: $h = Cp(\Delta T)$

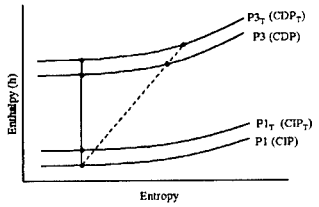


Figure 2 – H-S diagram

With the knowledge of inlet mass flow (and hence velocity) at various load levels and associated bleeds (for an unfouled compressor) obtained from experience, an pseudo-efficiency feature may be calculated which cannot account for axial velocity changes. However, this feature was acceptable considering the degradation feature of interest is the percent change in performance. It should be noted that the Navy intends to place total/static pressure probes on certain K-17s and LM-2500's to improve the accuracy of these fouling features and performance assessment capabilities.

Figure 3 shows the pseudo-efficiency as compared to an unfouled state for the 501–K17 test. Four waterwash events occur in this data set. It is important to note the trend in non-recoverable losses that will require a compressor crank wash, or more detailed overall, to recover.

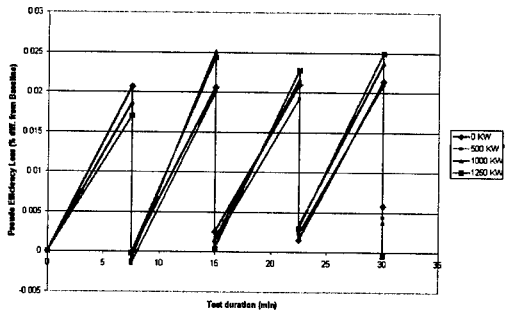


Figure 3 – Fouling / Waterwash Test Results

In addition to this feature, it was found that at higher loads Static Pressure Ratio, CDT and Fuel Flow were all major indicators of degradation due to fouling. The increase in CDT was relatively minor and even risks overlapping thermocouple sensitivity. These results from the LM2500 test are shown in Figure 4.

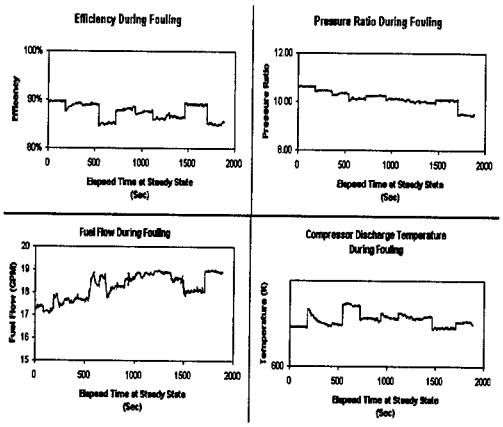


Figure 4 – Parameter Deviation at Full Load (LM2500 test)

Prognostic Model for Predicting Degradation: The compressor performance prognostic module consists of a data preprocessor and specific diagnostic/prognostic algorithms for assessing the current and future conditions of the gas turbine. The data preprocessor algorithms examine the unit's operating data and automatically calculate key corrected performance parameters such as pressure ratios and efficiencies at specific load levels in the fashion already described. As fouling starts to occur in service, probabilistic classifiers match up corresponding parameter shifts to fouling severity levels attained from these tests with corresponding degrees of confidence.

A probabilistic-based technique has been developed that utilizes the known information on how measured parameters degrade over time to assess the current severity of parameter distribution

shifts and project their future state (see Figure 5). The parameter space is populated by two main components. These are the current condition and the expected degradation path. Both are multi-variate Probability Density Function (PDFs) or 3-D statistical distributions. Figure 5 shows a top view of these distributions. The highest degree of overlap between the expected degradation path and the current condition is the most likely level of compressor fouling.

In general, the probability that the current condition (C), may be attributed to a given fault (F) is determined by their joint probability density function. If C and F can be assumed to be normally distributed, the probability of association (Pa) can be found using:

$$p_a = 2\Phi(-\frac{\overline{F-C}}{\sqrt{\sigma_f^2+\sigma_c^2}}) = 2\Phi(-\beta) \tag{2}$$

where:

$\overline{F},\overline{C}$ = the mean of the distributions F and C respectively
σ_f,σ_c = the standard deviation of the F and C distributions

The function $\Phi(\ )$ is the standard normal cumulative distribution. The notation β is defined as the fault index.

Once the current severity level is known with a high degree of confidence, a fault-weighted projection is performed using a modified double-exponential smoothing technique. This approach is a better that a simple multi-variate regression because it weights the most recent performance degradation trends and evolve the current conditions toward the expected degradation path.

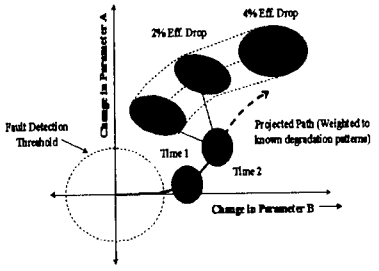


Figure 5 – Prognostic Modeling Approach

To manipulate the data into the form of this model, the time dependency of the test results had to be removed because of the unrealistic fouling rates. This was performed by viewing percent changes in static pressure ratio, fuel flow and CDT in relation to ¼ % pseudo-efficiency drops. This increment was chosen because it was the highest resolution that still permitted statistical analysis. With the assimilation of the data into these discrete bands, the statistical parameters (e.g., mean and standard deviation) can be ascertained for use in the prognostic model.

Figure 6 shows the evolution of the compressor degradation for the LM-2500 test at 1% pseudo-efficiency drops (for visual clarity). The top two plots show the distributions of pressure ratio and fuel flow respectively while the bottom two illustrate the joint probability distributions.

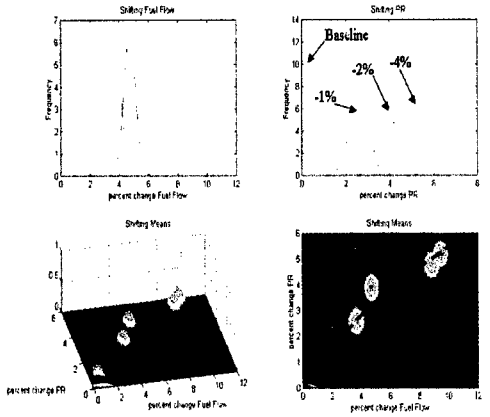


Figure 6 – Prognostic Model Visualization

Once the statistical performance degradation path is realized along with the capability to assess current degradation severity, the final step was to implement the predictive capability.

All compressors will not foul in exactly the same way and certainly not at the same rate as the accelerated tests. Fouling rates may even change between waterwashes or crankwashes for a given compressor. However, the percent changes of parameters relative to each other is still information that should be accounted for when projections of future fouling severity are to be made. The actual unit-specific fouling rate is combined with historical fouling rates with a double exponential smoothing method. This time series technique weights the two most recent data points over past observations. Eq. 3a,b, and c, give the general formulation, Bowerman (1993). Figure 7 shows how this technique can give significantly different results than standard regression.

$$S_T=\alpha y_T+(1-\alpha)S_{T-1} \tag{3a}$$

$$S^{[2]}_T=\alpha S_T+(1-\alpha)S^{[2]}_{T-1} \tag{3b}$$

$$\hat{y}_{T+\tau}(T)=\left(2+\frac{\alpha\tau}{(1-\alpha)}\right)S_T-\left(1+\frac{\alpha\tau}{1-\alpha}\right)S_T^{[2]} \tag{3c}$$

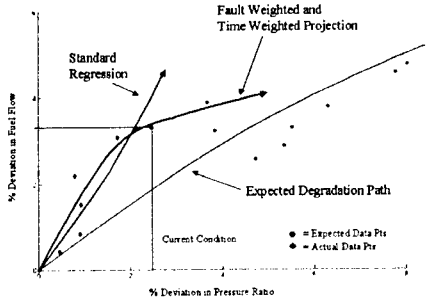


Figure 7 – Prediction of Degradation Rates

Benefits of Test Results: The test data made two essential contributions to the development of this prognostic model. First, they provided a means by which to validate an analytical model of how performance parameters change as a function of compressor fouling. Secondly, they gave insight into the sensitivity and statistical distributions of performance parameters as a function of load. Hence, having been developed and validated on real data, a large amount of knowledge is "built in" to the prognostic model. Along similar lines, the prognostic model may be developed for any particular gas turbine if data is made available on pre and post on-line waterwashing and crank washing.

Optimizing Compressor Wash Intervals: Referring back to the accelerated fouling test results of Figure 3, it is clear that on-line compressor washing was able to recover a majority of the compressor efficiency degradation. Initially, a large portion of these non-recoverable losses are recovered by crank washing but eventually a hot section overall will become necessary to regain performance (Figure 8).

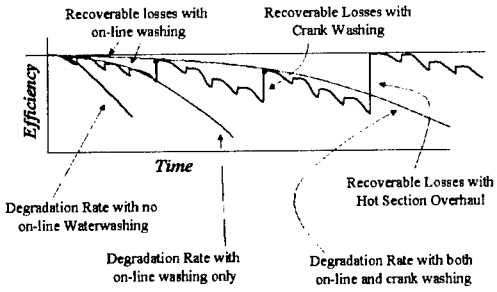


Figure 8 - Types of Degradation Rates

Hence, the question shifts from if washing should be done to when it should be done from an optimal cost/benefit standpoint. To perform this optimization, results from the engineering-based, compressor fouling prognostic model are combined with an economic-based analysis that accounts for the costs associated with efficiency degradation and performing compressor washing.

The compressor washing optimization algorithm developed predicts the optimal time to perform the wash based on the projected efficiency difference between performing the wash (action) to correct the degradation and continuing to run the gas turbine in its current condition (no action). The compressor wash should occur at the point in the future when the benefits of performing it outweigh its costs.

In this process, the engineering projections are merged with the O&M economic information on compressor degradation consequential costs. Factors such as reduced load, downtime, and other replacement "value" costs are all taken into account to quantify the decision to either not perform a type of wash (at the expense of increased degradation) or to perform the wash (but incurring a cost). (Figure 9).

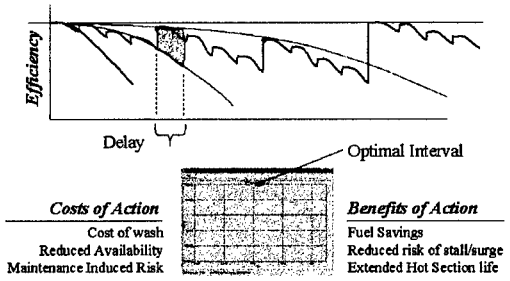


Figure 9 - Optimal Time for waterwashing

The Net Present Value (NPV) is calculated into the future for a specified time period. The NPV is a simple calculation once consequential costs have been determined from the above factors. When the costs and risks associated with keeping the gas turbine operating are thought of as "benefits" the NPV may be thought of as the following mathematical form:

NPV = (Total Expected Cost Associated with Efficiency Degradation) - (Total Expected Cost of Recovering Losses) - (Costs of the Waterwash or Crankwash) (4)

A simplified cost function version of this NPV calculation can be represented as follows:

$C(total) = (\Delta Fuel\ Flow * Fuel\ Cost/Amount\ of\ Fuel) - (Cost_{labor} + Cost_{materials} + PowerLost)_{Wash}$ (5)

In this formulation, a simple minimization problem exists. The prognostic model's gas turbine degradation statistics, forecasting and the probabilistic analysis are used as inputs to the development of this cost minimizing procedure. Eq. (5) will produce a minimum when the costs of performing an online wash equals the amount of extra fuel being consumed due to degradation. Figure 10 shows an actual trend in Δ% fuel flow for the K-17 test and the prediction of future degradation based on the double exponential regression and experience from previous fouling/waterwash/crankwash results. The line in Figure 10 represents the point of the NPV curve beyond which costs outweigh benefits.

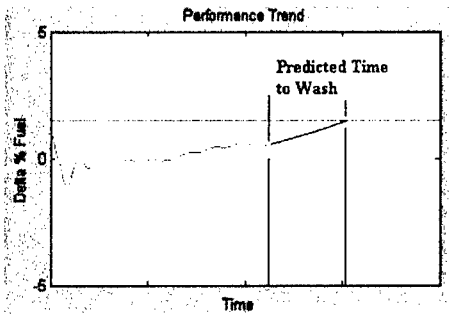


Figure 10 – Wash prediction

Figure 11 illustrates how the cost function changes as a function of fouling severity at "Full" and "Standard" load levels or "bells". This shows that, on a relative basis, a waterwash at "Standard" load need only be performed nearly ½ as frequently than if the unit is always operated at "Full" load. Lower load levels would warrant even less frequent washing. The relative time used in Figure 11 is due to the fact that the actual fouling rate was accelerated. It is assumed that the relative waterwash frequencies will be applicable to the actual operating times of a unit undergoing normal fouling rates. Figure 11 also assumes that the costs associated with performing the online wash are *fixed*. This allows the costs associated with compressor fouling (i.e., excess fuel expenditures) to be the only *variable costs* within the algorithm.

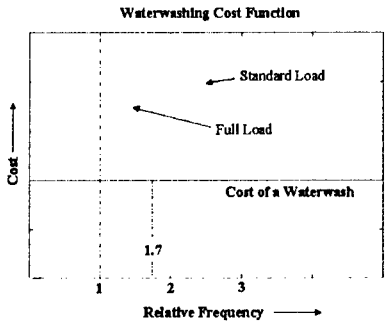


Figure 11 – Water Wash Frequency

Conclusion : A method has been presented that assesses the compressor performance degradation of Naval gas turbines with standard instrumentation and predicts the optimal time for washing processes based on a cost/benefit analysis. The approach utilizes built-in knowledge from accelerated fouling tests for model validation and to predict future performance of an arbitrary unit in-service. With continuous monitoring and cost/benefit analysis the Navy can make informed decisions about incorporating on-line waterwashing and altering crankwash intervals.

Work Cited:

1. Bowerman, Bruce L. and O'Connel, Richard T., Forecasting and Time Series, Duxbury Press, 1993.

2. Boyce, Meherwan, Gas Turbine Engineering Handbook, Gulf Publishing Company, 1995.

3. Haub, Barry L., and Hauhe, William E., "*Field Evaluation of the On-Line Compressor Cleaning in Heavy Duty Industrial Gas Turbines*" International Gas Turbine and Aeroengine Congress and Exposition, Belgium, June 1990.

4. Kurtz, Rainer, Brun, Klaus, and Legrand, Daryl D., "*Field Performance Test of Gas Turbine Driven Compressor Sets*" Proceedings of the 28th Turbomachinery Symposium, Huston, Texas, September 1999.

5. Kurtz, Rainer, and Brun, Klaus, *"Degradation in Gas Turbine Systems"* Proceedings of the ASME TURBO EXPO 2000, May 8-11, 2000, Munich Germany.

6. Neter, John, Kutner, Michael H., Nachtsheim, Christopher J., and Wasserman, William, Applied Linear Statistical Models, IRWIN, Chicago, 1996.

7. Ozgur, Dincer, Morjaria, Mahesh, Rucigay, Richard, Lakshminarasimha, Arkalgud, and Sanborn, S., "*Remote Monitoring & Diagnostics System For GE Heavy Duty Gas Turbines*" International Gas Turbine and Aeroengine Congress and Exposition, Munich, Germany, May 2000.

8. Peltier, Robert V., and Swanekamp, Robert C., "*LM2500 Recoverable and Non-Recoverable Power Loss*" ASME Cogen-Turbo Power Conference, Vienna, Austria, August 1995.

9. Tsalavoutas, A., Aretakis, N., Mathioudakis, K., and Stamatis, A., " *Combining Advanced Data Analysis Methods For The Constitution Of An Integrated Gas Turbine Condition Monitoring and Diagnostic System*" " International Gas Turbine and Aeroengine Congress and Exposition, Munich, Germany, May 2000.

10. Roemer, M.J., and Atkinson, B., "*Real-Time Engine Health Monitoring and Diagnostics for Gas Turbine Engines,*" Proceedings of the International Gas Turbine & Aeroengine.

11. Roemer, M.J. and Ghiocel D.M., "*A Probabilistic Approach to the Diagnosis of Gas Turbine Engine Faults*" 53rd Machinery Prevention Technologies (MFPT) Conference, Virginia Beach, VA, April 1999.

12. Walsh, P.P., and Fletcher, P., Gas Turbine Performance, ASME Press, New York, 1996.

PROGNOSTIC ENHANCEMENTS TO NAVAL CONDITION-BASED MAINTENANCE SYSTEMS

Michael J. Roemer
Gregory J. Kacprzynski
Andrea Palladino
Impact Technologies, LLC
125 Tech Park Drive
Rochester, NY 14623
716-424-1990

Thomas Galie
Naval Surface Warfare Ctr.
Carderock Division
Philadelphia Naval
Business Center,
Philadelphia, PA, 19112-5083

Carl Byington
Mitchell Lebold
Applied Research Lab.
Penn State University
P.O. Box 30
State College, PA 16804-0030

Abstract: In recent years, numerous machinery health monitoring technologies have been developed by the U.S. Navy to aid in the detection and classification of developing machinery faults for various Naval platforms. Existing Naval condition assessment systems such as ICAS (Integrated Condition Assessment System) employ several fault detection and diagnostic technologies ranging from simple thresholding to rule-based algorithms. However, these technologies have not specifically focussed on the ability to predict the future condition (prognostics) of a machine based on the current diagnostic state of the machinery and its available operating and failure history data. Prognostic capability is desired because the ability to forecast this future condition enables a higher level of condition-based maintenance for optimally managing total Life Cycle Costs (LCC). A second issue is that a framework does not exist for "plug 'n play" integration of new diagnostic and prognostic technologies into existing Naval platforms. This paper will outline a generic framework for developing plug 'n play prognostic "modules" as well as examples of specific prognostic modules developed for steam turbine journal bearings and auxiliary gearboxes. The gearbox prognostic module was calibrated and verified using gearbox seeded fault and accelerated failure data taken with the MDTB (Mechanical Diagnostic Test Bed) at the ARL Lab at Penn State University.

Keywords: Prognostics, Condition-based Maintenance, Open System architectures

Introduction: The U.S. Navy has identified the benefits of condition-based maintenance for reducing the life cycle costs of critical shipboard equipment, improving system readiness, and allow more efficient allocation of reduced human resources. Introducing CBM enabling technologies such as advanced diagnostics and prognostics onto Naval platforms that employ SMART and conventional Command, Control, and Communication (C3) systems, appropriate Human-System Interfaces, and various sensor technologies is paramount to achieving these goals. Specifically, these initiatives included:

- Integrating feature-based and model-based prognostics in a real-time environment.
- Developing a "Toolbox" of generic prognostic approaches useful for a wide variety of applications.
- Capitalizing on AI Technologies for fault identification, expert system development, and prediction.
- Designing a Human System Interface concept for knowledge-rich and efficient information access
- Making CBM enabling technologies "Plug 'n Play" in an Open Systems Architecture for ease of data transfer and continuous enhancement of shipboard technologies.

The technology development costs of advanced plug and play diagnostics and prognostics for steam and gas turbine components was validated for the reasons given below. They are aimed at reducing operations and maintenance costs by 50% and predicting component failures and/or degradation with 1-sigma confidence bound of 100 hours.

- Steam and gas turbine component failures and degradation can account for up to 5% of downtime associated with shipboard applications.
- The maintenance and operational costs can be in excess of approximately $5,000 per day for a DDG class ship.
- Prediction of component failure and degradation and maintenance optimization can reduce expected (risk*consequential cost) life cycle costs by up to 30% for steam and gas turbine applications.

Several technologies have been developed or transitioned to help achieve these goals which fundamentally fall in the categories of:

1) Automated sensor/data integrity assessment
2) Improved anomaly detection and feature extraction
3) Data and knowledge fusion processes
4) Feature and model-based prognostics

The capability of each category builds upon the functionality of the previous category with effective prognostics utilizing elements of data validation, anomaly detection, feature extraction and fusion. However, categories 1-3 are outside the scope of this paper. This paper will deal primarily with the design and functionality of prognostic modules, related open system architecture issues, and provide detailed examples of prognostics.

Prognostics Modules: A comprehensive prognostic capability for critical components and/or systems must be capable of integrating existing technologies such advanced features extraction (i.e. vibration, oil analysis, etc.) techniques and empirical/physics-based modeling approaches. In addition, due to the inherent uncertainty involved with predicting future events, prognostic modules should also incorporate a probabilistic framework to directly identify confidence bounds associated with specific component/system time-to-failure predictions. This approach should also be capable of integrating component reliability and inspection results, as well as provide statistical updating methods to accommodate modeling, operational and material property uncertainties known to exist.

To achieve this broad-based inclusion of prognostic technologies into Naval CBM systems, prognostics must utilized and implemented based on inputs from leading feature-based and model-based technologies. It is important to note the intrinsic differences between feature-based prognostics and physics-based prognostics, both in terms of accuracy and applicability, from the operations and maintenance perspectives. In short, the operational perspective relies on more near term predictions of remaining useful life (RUL). The feature-based prognostic approaches address this perspective because they can only make RUL estimates when a particular feature or features associated with a known fault condition has been observed. This characteristic of feature-based prognostics is illustrated in Figure 1. This plot shows how a feature-based RUL prediction becomes accurate only when diagnostic information or features become available. Without these features, no viable prediction can be calculated.

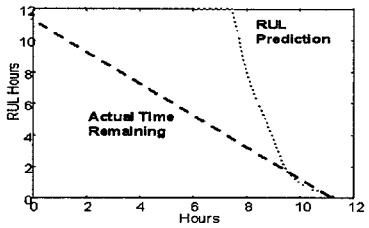


Figure 1 Typical RUL Prediction using Feature-Based Prognostics

Model-based prognostics differ from feature-based prognostics in that they can estimate RUL based only on operational conditions and can be "calibrated" based on any relevant diagnoses that are made. This form of prognostic relies upon high fidelity models (i.e. Finite Element or State Space) that are developed a-priori. Because this form of prognostics can make a RUL estimate in the absence of diagnostic information, it can be used for more long-term predictions as well as short-term ones [6]. It can address questions like what is the failure risk 6 months into the future if the expected future operational profile is known. Figure 2 shows the relationship between diagnostics and a-priori knowledge in the functionality of feature and model-based prognostics from the operations and maintenance perspectives.

One of the key aspects of this integrated prognostic approach is that it is flexible enough to accept input many different sources of information in order to contribute to better fault prediction on remaining useful life. Within this architecture, measured feature data is processed in the diagnostic block, with relevant processed feature information passed to the prognostic block. Next, this information is combined with the model-based estimate to examine the current and future risk associated with a particular failure mode. This block diagram is simplistic in order to highlight the important components of an integrated prognostic module. A more detailed description of the various technologies that have been implemented for particular applications is given in the following sections.

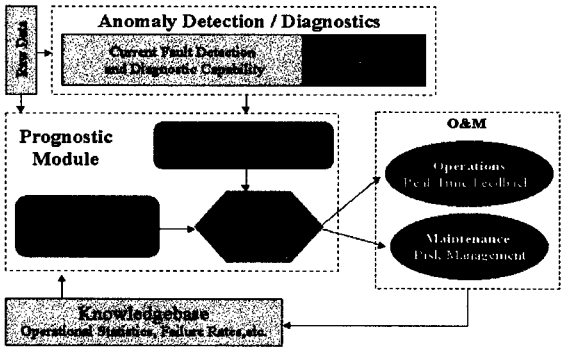


Figure 2 Generic Prognostic Process and Maintenance Integration

Gearbox Prognostic Module: A physics-based model for geartooth failure is the first prognostic model that will be presented. This model was chosen because it could be validated and calibrated on seeded fault / run-to-failure data available with the MDTB (Mechanical Diagnostic Test Bed) at the ARL Lab at Penn State University.

This prognostic module is a near real-time, self-calibrating, physics-based statistical RUL predictor of gear tooth failure due to tooth spalling or low cycle fatigue (LCF) cracking. Figure 3 is a block diagram that illustrates the functionality of this module.

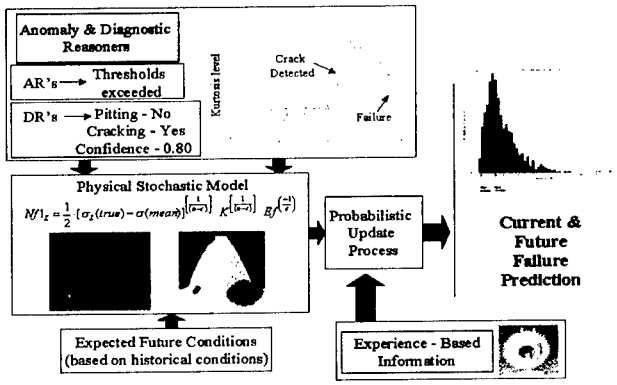


Figure 3 Gear Model-Based Prognostics

A shipboard gearbox of sufficient importance to warrant a dedicated prognostic module would be linked to a on-line data acquisition system capable of extracting vibration, speed and load data. This real-time data would be processed by a pre-developed prognostic module residing on the shipboard CBM system or on a remote server. The prognostic module encapsulates four primary capabilities.

1) Containment of real world calibrated, physic-based algorithms for accumulating the material damage of a gear as a function of operating parameters.
2) The ability to statistically examine past operating condition and extrapolate them into the future or allow for a simulated future operating profile.
3) Containment of algorithms for processing the vibration data and extracting vibration features that are most indicative of gear tooth cracking or pitting.
4) The ability to statistically calibrate the physic-based model results in the presence of a diagnosis of gear wear or with failure rates or inspection results from similar gearboxes.

The output of the prognostic module would be the probability of failure, with confidence bounds, for a specified time into the future.

This model uses American Gear Manufacturer's Association (AGMA) standards for calculation of tooth root stress as a function of transmitted load however sophisticated FE modeling of gear tooth contact could can been employed. The primary failure mode in the Penn State MDTB data was tooth root cracking which is an LCF phenomena. The mean number of cycles to root crack initiation is given in Eq. (1) which relates the LCF damage to localized true stress range.

$$Nf1_L = \frac{1}{2}\cdot[\sigma_L(true)-\sigma(mean)]^{\left[\frac{1}{(n-c)}\right]}\cdot K^{\left[\frac{1}{(n-c)}\right]}\cdot Ef^{\left(\frac{-1}{c}\right)} \quad (1)$$

Nfl_L = the LCF life for the gear (L)
σ_L(true) = localized true plastic stress amplitude at a tooth root

n = cyclic strain hardening exponent, c = fatigue ductility exponent
K = cyclic strength coefficient, Ef= fatigue ductility coefficient

This tooth root stresses fully account for strain hardening and residual compressive stresses by completely modeling the material's hysteresis loop. A Monte Carlo simulation was used to generate a distribution on the time to crack initiation based on uncertainty in mechanical properties and operating conditions. Some examples of this uncertainty include the load application factor, which is a function of manufacturing quality and gear alignment, and the true root notch stress. Having developed a distribution on number of cycles to crack initiation at a given load level, the next step is to find the distribution on total damage level as a function of time.

The damage accumulated due to low-cycle fatigue at a particular time is based on a non-linear Miner's rule Eq. (2). A damage level greater than or equal to 1 would represent an initiated root crack.

$$Damage = \left(\frac{n}{Nf1_L} \right)^{r1} \tag{2}$$

Where: n = number of cycles experienced, r1 = non-linear damage exponent, Nf1 = Number cycles to crack initiation

To be functional as a calibrated prognostic tool, the physics-based model must also consider crack propagation so it can predict the time to gear tooth failure when a diagnostic tool discovers that a crack has initiated. To address crack propagation, a fracture mechanics model was created. The fracture mechanics package used was a 2-D version of Franc-XT. The 2-D analysis yielded the change in stress intensity factor with respect to crack length.

The fundamental differential equation used for the rate of crack growth per cycle (Paris Law) is:

$$\frac{\partial a}{\partial N} = C\Delta Ki^m \tag{3}$$

Where:
C, m = fracture related empirical constants , a = crack length , N = cycle (Low or High)

The total probability of failure is the combination of two independent events; the initiation of a crack and the propagation of that crack to failure. For independent events, the total probability is

$$P_{total} = P(i) * P(p) \tag{4}$$

where:

$$P(p) = \frac{\#Damage > 1}{\#MonteCarlo_pts} \tag{5}$$

Figure 4 shows a screen capture of the notional, plug 'n play prognostic module for gear tooth failure prevention.

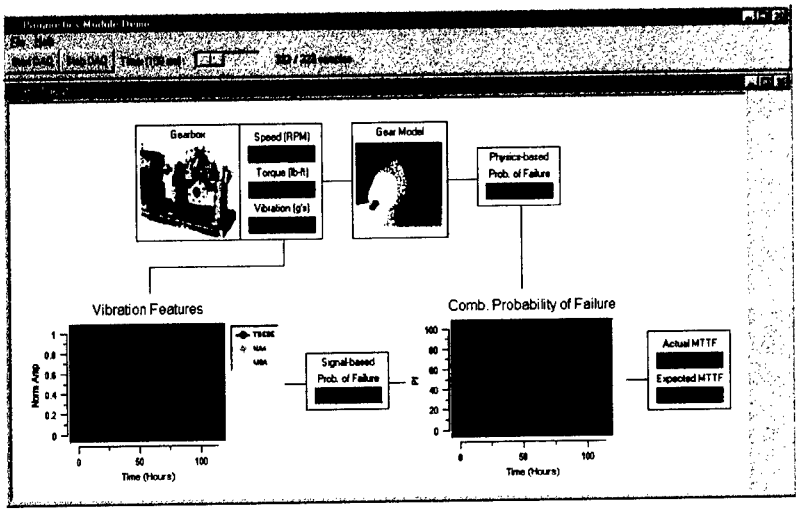


Figure 4 Gearbox Prognostic Module

The layout of this module (Figure 4) is intended to illustrate the knowledge fusion hierarchy that is "behind the scenes". The lower left plot shows three of the 25 vibration features as a function of time. Increases in the normalized amplitude levels have been shown to be indicative of gear tooth cracking [2]. The "Signal-based Prob. of Failure" number is based on the Dempster-Shafer combination of these features [7]. On a parallel path, the raw data gets evaluated by the physical-based prognostic model, which produces its own Prob. of Failure result called "Physics-based Prob. of Failure". A second Dempster-Shafer knowledge fusion process was used to combine the signal-based results with the Physics-based results. The "Actual MTTF" is generated based on the signal information while the Expected MTTF is based on the operational profile (speed and torque) from the physical model.

Actual and Expected MTTF have a higher purpose than just stating that a maintenance event should occur sooner (or later) than expected. The rate of change between actual and expected MTTF is a vital factor in maintenance optimization. Risk, defined as probability of failure multiplied by consequential costs, is always evaluated under two scenarios; 1) what is the risk of failure as a function of time if maintenance is performed in the present vs. 2) if it delayed until some future time. The future probability of failure is performed by extrapolating past speed and loading profile statistics over some future analysis time period.

Babbitted Journal Bearing: Large steam turbine babbitted bearings were identified as high-risk item for the Navy. Therefor a generic steam turbine bearing prognostic module was developed for a two axial groove or pressure bearing design with a tin-based babbitt atop a steel backing. The module is applicable for bearings of approximately 8-10" in length and 12-16" in diameter.

The failure mode of interest for steam turbine bearings is fatigue failure of the babbitt material as a result of fluid film pressure fluctuations [5]. The prognostic module developed would be used as follows:

1) A bearing prognostic module, initialized to specific application and design, would convert data from at least 2 proximity probes near the bearing into rotor eccentricity as a function of time.
2) Via the Reynolds equation and the short bearing model, the magnitude and location of the max fluid pressure will be calculated from the eccentricity
3) Using compiled experimental data relating max. fluid pressure to babbitt life, the mean time to failure (MTTF) with confidence bounds would be determined.

The Bearing Prognostic module is designed to accept two real-time rotor displacement measurements. Processing of the prox. probe data stream yields the eccentricity of the localized rotor motion as a function of time. The eccentricity or "orbit" of the rotor in the journal bearing is an input to a simplification of the Reynolds Equation chosen for this module called the Short Bearing Model. This model was chosen because for most steam turbine bearing designs the Length/Width ratio allows this assumption to be valid.

The Short Bearing Model, which relates non-dimensional fluid pressure to eccentricity is given by:

$$P_{ND} = 6\pi\left(L/D\right)^2\left(\frac{1}{1+er*\cos(\theta_{ar})}\right)\left(er\sin(\theta_{ar})\right)\left(z^2-1\right)+pa \tag{6}$$

This is converted to dimensional pressure via:

$$P = \mu * N * \left(\frac{R}{C}\right)^2 * P_{ND} \tag{7}$$

Where:

P_{ND} - Non Dimensional Pressure, L - Length of Bearing, D - Diameter of Bearing, er - Eccentricity Ratio

θ_{AP} – Angular position, R- Bearing Radius, μ – Fluid Viscosity, z- Non Dimensional axial direction

The solution to the model ultimately yields the fluid pressure distribution as a function of rotor displacement. A fluid pressure distribution is shown in Figure 5a. A sparse but significant set of experimental results have correlated max fluid pressure to number of cycles to babbitt fatigue failure [3]. This run-to-failure data taken from the EPRI Manual of Bearing Failure and Repair of Powerplant Rotating Equipment [4] is shown in Figure 5b. Standard Normal distributions were placed about the linear regression line to capture experimental uncertainty. Hence, with a given fluid pressure, a Mean Time To Failure (MTTF) with confidence bounds can be found.

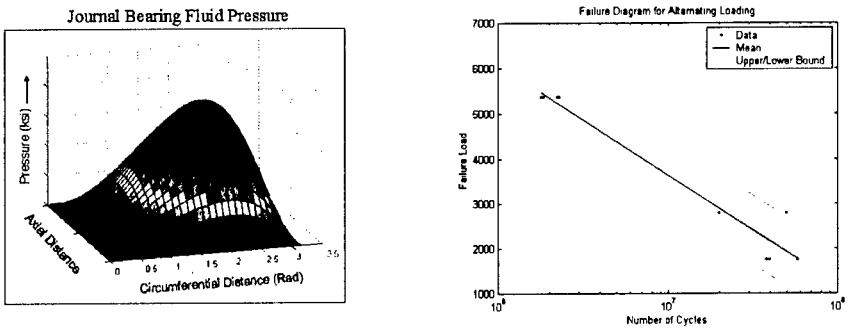


Figure 5a,b Journal Fluid Pressure Distribution and Failure Curve as Function of Load

A simulation was performed were the journal bearing had a normal amount of eccentricity for a period of time and then a rotor misalignment was simulated. The misalignment caused some high fluid pressure fluctuations as shown in Figure 6.

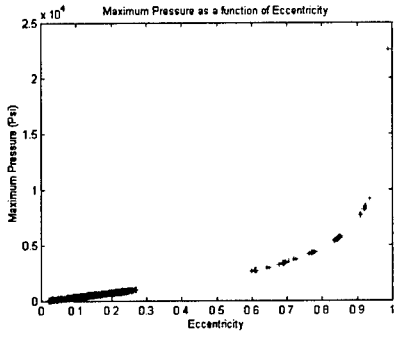


Figure 6 Maximum Pressure as a Function of Eccentricity

Like all the "prognostic" modules, the bearing module contains some components of anomaly detection, diagnostic and prognostic reasoning inherent to its architecture.

In this module an anomalous event that affects bearing life would be rotor misalignment. The model-based prognostic module continuously evaluates the remaining life of the bearing regardless of whether or not a misalignment diagnosis is made. However, the rate of damage accumulation increases dramatically when rotor misalignment is detected.

Figure 7 is a screen capture of a plug 'n play module for the Steam Turbine Bearing Prognostic Module. The real-time orbit of the rotor is shown in the upper left-hand corner along with the raw proximity probe data. The "Bearing Model" is meant to collectively represent FFT capability and the hydrodynamic model. 1 and 2 per rev. features are captured from the vibration spectrum. The severity level (0-100) of two conditions adversely affecting bearing life, rotor unbalance and misalignment are evaluated based on the amplitude levels of the 1 and 2 per rev. respectively. In the case illustrated, misalignment levels where mapped to rotor eccentricities.

Given the rotor eccentricity and Eq. (7), the maximum fluid pressure can be found. Finally, a probability of babbitt fatigue failure is found based on the number of cycles experienced and the known distribution of number of cycles to failure given max. fluid pressure level. Like all modules, a threshold level is placed on the difference between Actual and Expected Mean Time To Failure to alert when maintenance action should be taken.

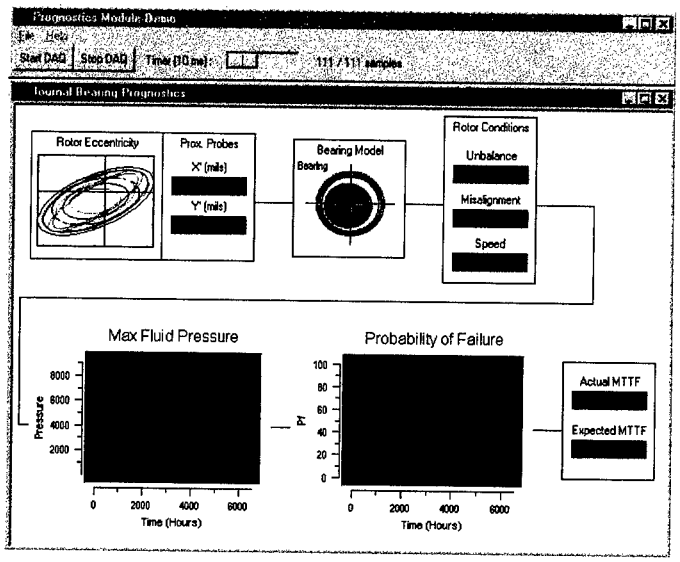


Figure 7 Journal Bearing Plug 'n Play Prognostic Module

Open Systems Architecture: Open systems architecture (OSA) is a design methodology that defines a set of standard publicly known interfaces for specific modules. This published interface standard allows systems to be broken down into independent sub-modules that can be replaced by another party's module as long as it meets the same non-propriety interface format. Each module is viewed as a collection of similar tasks or functions at different levels of abstraction. Figure 8 shows a flow chart of a proposed OSA for a machinery prognostics system. This OSA model has seven sub-components or modules: Human System Interface, Decision Reasoning, Prognostic Processing, Diagnostic Processing, Signal and Feature Processing, Data Acquisition and Sensor. Each module will have a standard input and output interface that enables communication between modules. The hub of the wheel structure represents the communications medium between the modules, which may be accomplished using popular Internet protocols such as TCP/IP or HTTP. This means that modules do not need to reside on the same machine but may reside anywhere on a local, wide area or worldwide network. Open systems architecture design is an essential part to a prognostic's system design to allow maximum flexibility and upgrade ability of the system.

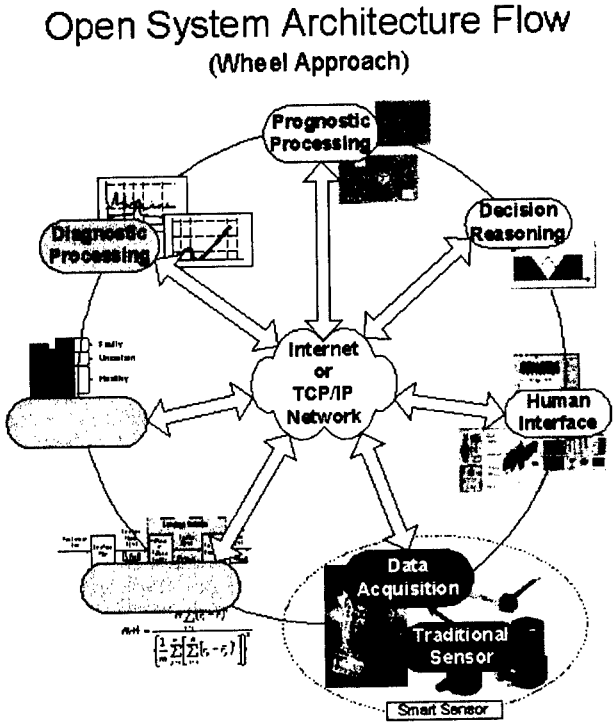


Figure 8 - Data Flow within an Open System Architecture

The requirements of an Open System Architecture (OSA) for prognostic modules such as the ones discussed herein will be further identified by incorporating current OSA formats such as those provided by MIMOSA. Some of the prognostic output protocols that have been considered are (as derived from MIMOSA): Status, State of Health, Rate of Change, Time to Action, Problem Identification, Components Affected, Recommendations, Work Request, Confidence, Remarks/Comments. A detailed discussion of OSAs can be found in [1].

Human System Interface: A proposed human system interface (HSI) concept for incorporating the prognostic modules into a multi-framed, single document interface (SDI) is illustrated in Figure 9. This approach will ensure that critical information is not obstructed or hidden from view by another window of the application. Tab buttons can be utilized to view multiple pages of information with the simple window and hot links will be incorporated to simplify navigation between pages.

The deck navigation frame allows the user to graphically locate compartments within the shipboard platform. This frame has two panels associated with it, allowing the user to graphically pinpoint a specific system or component within the ship. The left portion of the frame will provide a vertical cross-sectional view of the shipboard platform while the right portion of the frame will provide a graphical layout of the selected deck level. The deck level layout will show all the relevant compartments that a user may query. When a user selects the shipboard level and compartment, the

equipment selection frame will then update to show all relevant equipment located in that selected compartment. The colors of the deck and compartment level will be color coded to indicate the current health of the system and components. A gray deck and compartment color may be used to indicate that all machinery within that location is of good health and operating normally, while the color red may indicate a system failure or alert message. A different color, maybe yellow, will be used to indicate a system with a low remaining useful life (RUL) or degraded condition. Blue will be used to indicate a selected region on the layout. If the item was originally red or yellow then the selected item will contain hatch diagonal blue lines in order to still indicate the critical condition of the compartment and deck level.

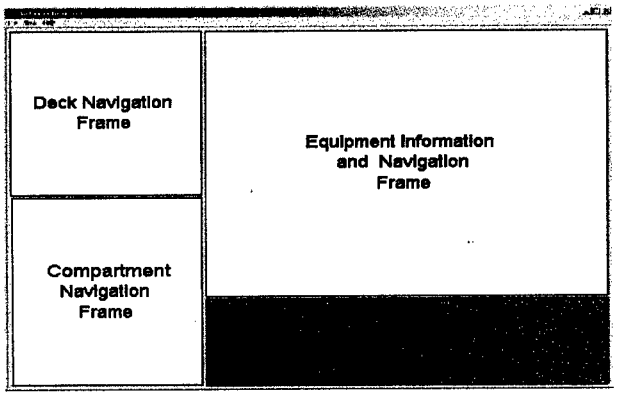


Figure 9 General HSI Layout

The compartment navigation frame will allow the user to visually monitor the heath of all equipment contained within a compartment. The user may then select a piece of equipment for further inspection. The color-coding of the equipment will also have the same methodology as the deck and compartment levels. Additional information may be displayed to the user as the mouse moves over a relevant deck, compartment or component. When a user clicks on a piece of equipment within the compartment, the equipment information and navigation frame will display additional information related to that specific component.

The equipment information and navigation frame allows the user retrieve information about a specific system and sub-components. The top portion of this frame will contain a set of tab controls to view different aspects and information about the system and sub-components. Six tab controls will be utilized for Overview, Detail, Diagnostics, Prognostic, Manuals and Procedures of the system. The "Overview" tab window will contain a picture of the system that will allow mouse-over information and short-cut links to sub-component information. This window will also display the current status and health of the system and sub-components. The current health panel will display the prognostic results for each component with indicated degradation. The current status panel will display information about the current operation configuration and output. The "Details" tab window will display current sensor readings such as temperature, pressures and vibration levels. An image of the system will be displayed in the top right-hand corner to allow the user to navigate between different sub-components. The "Diagnostics" screen will display information about the system's sub-components and an overall health rating for the system. The "Prognostics" window will display the RUL for each sub-component for the current loading profile. This screen will also allow the user to input a future mission profile to determine the RUL under different loading conditions of the system. The user will easily be able to located schematics and detailed drawings using the "Manuals" tab and the "Procedures" tab will give

the operator quick access to operational, emergency and maintenance procedures with a list style view control. The results from the prognostics and diagnostics pages could incorporate a list of recommendation and short cuts or hot links to emergency and maintenance procedures based on the fault condition.

The Alert frame is meant to display important information to the user about critical alarms or abnormal events that are occurring across the shipboard platform. Each shipboard system module connected to this system will determine the information displayed in this panel. The alert information will be displayed in a tabular fashion along with date and time information. Each alert message will contain a short cut to allow the user to jump directly to the shipboard system in question by means of a double clicking the message.

A sample layout for a shipboard HSI used for an OSA prognostics system module is illustrated Figure 10. The figure shows the main overview screen from an HSI demonstration system developed as part of a Navy program on Prognostic Enhancements to Diagnostic Systems. A list view of shipboard equipment may also be implemented for the deck and compartment navigation frames. This would allow operators that are not familiar with the shipboard layout to located equipment based on name and not location.

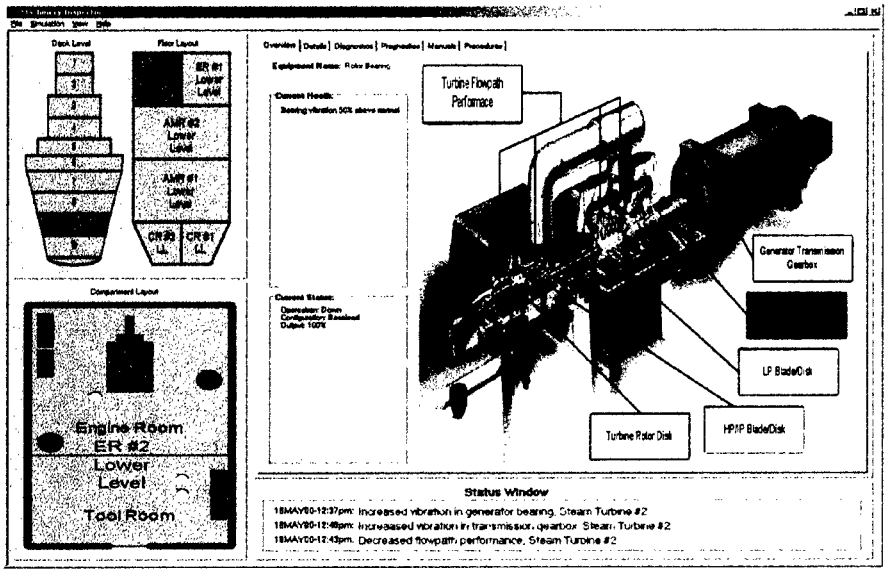


Figure 10 Prognostic Module HSI Demonstration Layout

Conclusions: A comprehensive prognostic capability for critical components and/or systems has been presented that integrates existing technologies such advanced features extraction techniques and empirical/physics-based modeling approaches. The demonstrated prognostic modules utilized a probabilistic framework for identifying confidence bounds associated with specific component time-to-failure or degradation predictions. The developed approach was also capable of integrating component reliability and inspection results with reference to operations and

maintenance. The significance of having the developed prognostic modules follow a standard OSA format was highlighted with examples of current OSA considerations identified by MIMOSA. Finally, a human system interface concept was presented for illustrating how information from a complicated health management system could be presented to an end user.

References:

1. Michael Thurston and Mitchell Lebold. "*Standards Developments for Condition Based Maintenance Systems,*" Improving Productivity Through Applications of Condition Monitoring, 55th Meeting of the Society for Machinery Failure Prevention Technology, April 2001.

2. Byington, C.S. and Kozlowski, J. D., Transitional Data for Estimation of Gearbox Remaining Useful Life, Proceedings of the 51st Meeting of the MFPT, April 1997

3. Gyde, N., "Fatigue Fractures in Babbit Lined Journal Bearings", Thesis 1969 Laboratory of Internal Combustion Engines, Technical University of Denmark, Copenhagen

4. Manual of Bearing Failures and Repair in Power Plant Rotating Equipment, GS-7352 Research Project 1648-10, Final Report, July 1991, Prepared for the Electric Power Research Institute.

5. Yahraus, W. A., "Rating Sleeve Bearing Material Fatigue Life in Terms of Peak Oil Film Pressure," 1988 Society of Automotive Engineers, Inc, No. 871685.

6. Roemer, Kacprzynski, Development of Diagnostic and Prognostic Technologies for Aerospace Health Management Applications", Proceedings of the 55th Meeting of the MFPT, April 2001.

7) Brooks, R. R., and Iyengar, S. S, Multi-Sensor Fusion, Copyright 1998 by Prentice Hall, Inc., Upper Saddle River, New Jersey 07458

PREDICTION METHODS AND DATA FUSION FOR PROGNOSTICS OF PRIMARY AND SECONDARY BATTERIES

James D. Kozlowski, Matthew J. Watson, Carl S. Byington, Amulya K. Garga, Todd A. Hay
Applied Research Laboratory
The Pennsylvania State University
State College, PA 16804-0030
814-863-3849
jdk173@psu.edu

Abstract: A method to accurately assess the state-of-charge (SOC), state-of-health (SOH), and state-of-life (SOL) of electrochemical energy sources provides significant benefit to operational systems. The model-based effort described here is focused on predictive diagnostics for primary and secondary batteries. It can also be applied to other electrochemical energy sources, such as fuel cells. This method is based on accurate modeling of the transport mechanisms within the battery and requires carefully developed electrochemical and thermal models. New features are developed from these models and are used in conjunction with several traditional measured parameters to assess the condition of the battery. Data fusion of feature vectors is used to develop inferences about the state of the system. The resulting output and any usage information available about the battery is then evaluated using hybrid automated reasoning schemes consisting of neural network and decision theoretic methods. The focus of this paper is on model identification and data fusion of the monitored and virtual sensor data. The methodology and analysis presented is applicable to mechanical systems where multiple sensor types are used for diagnostic assessment.

Key Words: Automated reasoning; condition-based maintenance; data fusion; electrochemical impedance; model-based diagnostics; predictive diagnostic techniques; state-of-charge

Introduction: Batteries are an integral part of many machines and are critical backup systems for many power and computer networks. Failure of a battery could lead to loss of operation, reduced capability, and downtime. An efficient way to monitor a battery's performance and assessment of its condition could drastically increase the reliability of these systems. The present condition of a battery is described nominally with the state-of-charge (SOC), which is defined as the ratio of the remaining capacity and the initial or rated capacity. Thus, the service history of a cell and its nominal capacity impact the assessment of SOC. Secondary batteries are observed to have a capacity that deteriorates over the service life of the cells. The term state-of-health (SOH) is used to describe the physical condition of the battery, ranging from external behavior such as loss of rated capacity, to internal behavior such as severe corrosion. The remaining life of the battery (i.e. how many cycles remain, usable charge, etc.) is termed the state-of-life (SOL), the prognostic metric. In this paper, a model-based effort is presented for predictive diagnostics of primary and secondary batteries. The flow of the model-based predictive diagnostics processing is shown in Figure 1. There are five distinct stages of the

processing: 1) measurement of signals related to diagnostics; 2) extraction of key features (such as model parameters); 3) charge, health, and life prediction; 4) decision processes that combine the predictions with knowledge and history; and 5) output of user information for display or coordination with other systems. The specific objectives of the model-based approach described here are determination of the SOC, SOH, and SOL.

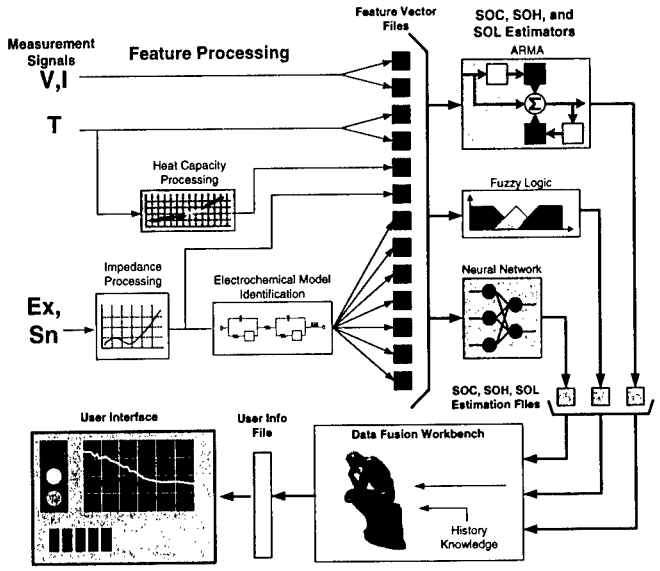


Figure 1. Flow diagram of developed predictive diagnostics processing

Model-based Approach: The general approach to model development is to formulate robustly parameterized governing equations for energy conservation and relevant electrochemical phenomena and transport processes. Lumped parameter formulation in lieu of a spatially distributed formulation offers greater applicability to the broad variety of cell chemistries and battery designs. That is, explicit geometry and configuration input are not required. The parameters and sources of the various transport, state, and conservation equations are coupled to ensure consistency with experimental observations and facilitate system classification. The model parameterization is formulated to incorporate significant aging mechanisms and pathological behavior in order to provide fault diagnostic capability. The ability to forecast future battery performance is developed by tuning system parameters through history-matching trials.

Data Fusion Techniques: A core challenge is to develop the appropriate signal processing, sensor-level data fusion, and automated reasoning to support battery diagnostics, charge control, and ultimately, prognosis of remaining cycles. Multi-sensor data fusion techniques that combine data from actual and virtual sensors provide the potential to improve detection performance and reduce the number of false alarms [1].

Automated Reasoning: The hybrid automated reasoning modules developed previously at the Pennsylvania State University Applied Research Laboratory (ARL) integrate a variety of predictive diagnostic techniques, such as neural networks, fuzzy logic, and auto-regressive moving average (ARMA) models, via decision-level data fusion[2], [3]. The outputs of these techniques are three estimates of the battery state and optimal charge control based on electrochemical and thermal data and available usage information. They are combined using hybrid automated reasoning modules, consisting of neural network and decision theoretic methods, to provide a single estimate of the battery's state. This output can be obtained as a linguistic indication or as numerical indication and is coupled with a measure of confidence. This type of tool is beneficial because it utilizes key information from multiple estimations for robustness and presents the results of the fusion assessment, rather than a mere data stream.

Measurement and Data Collection: The first step to developing model-based diagnostics is to establish the necessary and available observables (i.e., what can be measured and its sufficiency). Changes in the electrode surface, diffusion layer, and solution are not directly observable without disassembling the battery cell. Other variables, such as potential, current, and temperature, are observable and can be used to indirectly determine the performance of physical processes. This is the rationale for choosing a model-based approach. Under these constraints, the following types of measurements were selected for battery diagnostics: terminal and cell voltages, load currents, surface and internal temperatures, electrolyte pH, and electrical impedances. To ensure maximum coverage of operating modes for testing developed algorithms, test stand data were collected under the following conditions:

1. No load, fully charged
2. Once every minute while discharging
3. No load, 100% discharged
4. Once every minute while charging

An ongoing experimental test schedule is being conducted under an Office of Naval Research (ONR) sponsored battery project, where lead-acid, nickel-cadmium, lithium, and alkaline batteries are being run to failure. During a test, battery impedance data is collected along with cell and terminal voltages, load current, and temperatures at various internal and external locations on the battery. To date, over 200 data sets have been collected across the different chemistries and sizes of batteries.

Electrochemical Impedance Model Identification: Direct measurements of battery or cell condition have traditionally been very difficult for practical systems such as automotive or aviation batteries. There are, however, a variety of indirect measurement techniques that rely on the cell's response to a precise manipulation of the load [4], [5], [6]. One of the most robust and widely used methods in laboratory practice is AC Voltammetry. This technique can provide information on the electrochemical dynamics of the battery through a non-invasive interrogation of the cell. By applying a small amplitude excitation to the cell and measuring the response, the internal impedance of the cell can determine. Figure 2 represents the measured impedance of a nickel-cadmium battery that has been partially discharged.

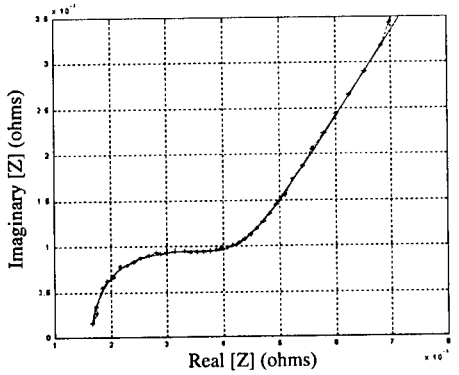


Figure 2. Impedance of 4.3 Amp-hour nickel-cadmium battery, partially discharged and fit to impedance model

Internal impedance measurements can further be used to retrieve information about the electrochemical processes that occur within the battery. This is accomplished using electrical circuit analogs such as the Randles circuit, which represents the electrode-electrolyte interface processes, and a minima search method. For model identification, a better fit of the impedance data was found using a two-electrode Randles circuit model (Figure 3).

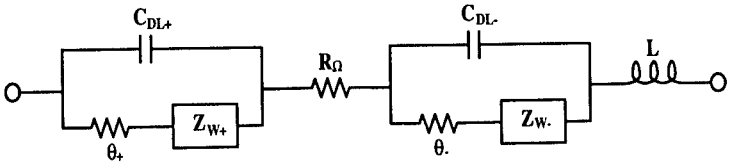


Figure 3. Two-electrode Randles circuit model with wiring inductance

The equation for this circuit is given as

$$Z_{cell}(s)=\frac{s^{1/2}\theta_{+}+\sigma_{+}\sqrt{2}}{s^{1/2}\theta_{+}C_{DL+}+sC_{DL+}\sigma_{+}\sqrt{2}+s^{1/2}}+R_{\Omega}+\frac{s^{1/2}\theta_{-}+\sigma_{-}\sqrt{2}}{s^{1/2}\theta_{-}C_{DL-}+sC_{DL-}\sigma_{-}\sqrt{2}+s^{1/2}}+sL \qquad (1)$$

In (1), s = jω (ω is frequency in rad/s), R_Ω represents the electrolyte resistance, θ represents the charge transfer resistance, C_{DL} represents the double layer capacitance, σ represents the diffusion layer coefficient, and Z_{cell} represents the Warburg impedance (Z_W) of the cell. These parameters represent the physical electrochemical processes, such as charge and mass transfer, which occur during cycling. See [4], [5], [7] for a description of these electrochemical processes.

The above parameters are extracted from the impedance measurements using a minima search method. For this approach, a simulated annealing algorithm was chosen. Unlike many local minima search methods, simulated annealing offers a global search [8]-[11]. Search regions, based on the identified parameters from previous impedance measurements, are used to minimize processing iterations. The model-identified electrolyte resistance of a nickel-cadmium battery during discharge, which was found using simulated annealing as the minima search method, is shown in Figure 4.

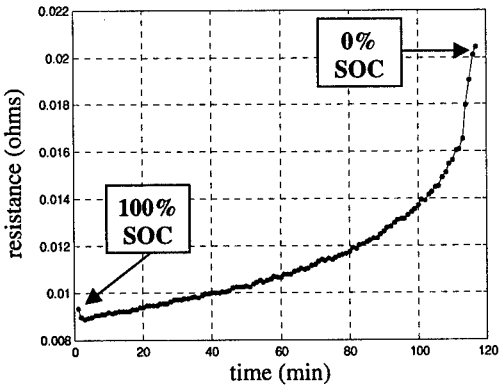


Figure 4. Model-identified electrolyte resistance of a nickel-cadmium battery using simulated annealing

State-of-Charge Prediction Models: The previous section addressed the extraction of physically meaningful parameters, such as charge transfer resistance, to more strongly connect SOC, SOH, and SOL predictions to internal battery processes. These *virtual sensor* signals (i.e., identified model parameters) also provide the decision processing with a check for bad signals. Referring to Figure 1, this section focuses on the developed SOC prediction modeling, primarily addressing the neural network and ARMA modeling and results. Work pertaining to the fuzzy logic prediction model is currently under investigation and results are being analyzed.

ARMA Modeling: Autoregressive (AR) modeling is a powerful linear modeling technique employed for predictive diagnostics [12], [3]. In order to assess battery capacity, an analytical model of battery dynamics is useful. Autoregressive moving average (ARMA) modeling is commonly used for system identification because it is linear and easy to implement. It is also a good complement to the more complex models (neural network and fuzzy logic) being used. An ARMA model was thus chosen for assessment of battery SOC and is represented by the equation:

$$y(t)=a\,X(t) + b\,X(t-1) + c_o\,y(t-1), \qquad (2)$$

where y represents SOC, X represents a vector of model inputs, and a, b, and c_o represents the model coefficients. Model coefficients are calculated during training of the model, where a least squares fit of data from a previously discharged battery is

performed [13]. The model uses instantaneous measurements, as well as past measurements of the system, to monitor changes in the system. Inputs to the model include electrochemical impedance parameters, voltage, current and temperature measurements, and past SOC predictions.

The ARMA model has been trained and tested on five different kinds of batteries with varying size, chemistry, and type: two sizes of primary poly-carbonmonofluoride $(CF)_x$ lithium (C and 2/3 A), two sizes of secondary nickel-cadmium (C and D), and one size of secondary lead-acid (12 volt).

Initial testing was performed on eight size C and nine 2/3 A $(CF)_x$ lithium batteries. Training was performed on one battery of each size and used to predict the other batteries of the same size. As shown in Table I, the model was very effective for this battery chemistry. The average prediction error for both sizes was less than 3%.

Table I. Results of ARMA model SOC predictions

Chemistry	Size	# Cells	Type	Prediction Error (%)
Lithium[1]	C	1	Primary	2.18
Lithium[1]	2/3 A	1	Primary	2.87
NiCad[2]	D	1	Secondary	3.17
NiCad[2]	C	1	Secondary	4.50
Lead-Acid	12 Volt	6	Secondary	9.13

[1] Poly-carbonmonofluoride-lithium (spiral type)

[2] Nickel-cadmium

Tests were also performed on nine size C and nine size D nickel-cadmium batteries. Similar results were obtained for these batteries as well, with average prediction errors of less than 5% (Table I). Although these batteries are secondary cells, only a few cycles from each battery were completed for analysis.

Final testing was performed on five 12-volt lead-acid starter batteries containing six cells each. Because these batteries are secondary, training was performed on an initial cycle of each battery and retrained after every additional cycle. Despite the fact that health effects make prediction more difficult, the model performed well on this chemistry. As shown in Table I, average prediction error was less than 10%.

Neural Network Modeling: An *artificial neural network* is a parallel distributed processing system inspired by biological neural networks. It consists of information processing units, called *neurons* or units, that are interconnected through *connection weights* to produce a desired output in response to its inputs. For battery SOC predictions, networks were trained to produce either a direct prediction of SOC or an estimation of initial battery capacity during the first few minutes of the run. All networks used for battery SOC estimation contained one hidden layer of neurons. The backpropagation gradient decent learning algorithm was used, which utilizes the error signal to optimize the weights and biases of both network layers.

The performance of the neural networks for direct SOC prediction was found to be quite consistent. The results for size C lithium batteries (runs 9-16) and size 2/3 A lithium batteries (runs 17-25) are given in Table II.

Table II. Errors for neural network direct SOC prediction of $(CF)_x$ lithium batteries

Network Topology [# *Hidden Neurons*]	**Size**	**Average Training Error** [*Training Set*]	**Average Testing Error**	**Maximum Testing Error** [*Run #*]
Feed Forward [*6*]	C	0.2% [*14*]	2.5 %	6.0% [*11*]
Feed Forward [*6*]	2/3 A	1.5% [*18*]	5.0 %	7.4% [*17*]
Time Delay [*7*]	C	0.3% [*14*]	3.0 %	7.1% [*11*]
Time Delay [*7*]	2/3 A	0.8% [*18*]	4.7 %	8.0% [*20*]

Networks were also trained to estimate the initial capacity of the battery during the first few minutes of the test. The SOC of the battery was then calculated directly by using the cumulative discharge current. This method can be a powerful tool for *mission planning*. Hypothetical load profiles could be used to predict whether the battery would survive or fail during a given mission, thus preventing the high cost and risk of batteries failing in the field. Results of this network on lithium batteries are given in Table III.

Table III. Error rates for SOC prediction based on initial capacity estimation with neural networks for $(CF)_x$ lithium batteries

Network Topology [#*Hidden Neurons*]	**Size**	**Average Training Error** [*Training Set*]	**Average Testing Error**	**Maximum Testing Error** [*Run #*]
Feed-forward [*5*]	C	2.4 % [*13 14 16*]	5.7 %	9.1 % [*11*]
Feed-forward [*5*]	2/3 A	3.0 % [*17 18 20*]	7.9 %	10.4 % [*23*]
Radial Basis [*6*]	C	0.6% [*13 14 16*]	4.6 %	6.8 % [*15*]
Radial Basis [*11*]	2/3 A	2.3 % [*17 18 20*]	3.4 %	4.9 % [*19*]

The SOC assessment by neural networks was very good. Although the average error is slightly higher than for the ARMA predictors, two important strengths of the neural network predictors outweigh that drawback: (i) maximum error on outliers was not significantly larger than the average error, and (ii) the network provides a conservative prediction (i.e., it does not over-predict the SOC). Both of these advantages are very important in practical systems where certification and low false alarms can impact whether a system is actually used or shelved.

SOC Modeling Remarks: Considering that very little training data are used to produce the predictions, results for both types of models are quite impressive. As more data are collected and several runs of each level of initial battery SOC become available, the

robustness of the predictors is likely to improve. The key distinction between the ARMA and neural network approaches is that the ARMA model assumes an explicit linear form of the predictor, while the neural network attempts to discover an implicit nonlinear model that captures the intricacies of the battery dynamics. If the model is of adequate degree, the ARMA model should require fewer runs than the neural network. However, the neural network can better represent nonlinearity (i.e., a variable load) and, thus, provide better generalization across the sample.

The models performed poorly on two of the tested batteries, which are examples of *outliers* that can skew most predictors. For such cases, the best predictors do not attempt to accurately predict the outliers; instead they seek rough conservative estimates that will allow the system to quickly flag the outliers. In this case, the batteries likely were *faulty* in some respect. It is the focus of SOH research to identify and assess the severity of existing or impending battery faults and this topic is briefly discussed below. The benefit of a good initial capacity estimator is a valuable capability not only during the operational scenario, but also for quality control purposes.

Fault and End-of-Life Prediction: For primary batteries, the SOC is also the SOL; once the charge is depleted the battery cannot be used again. However for secondary batteries, the SOC only represents the cycle life and not the total life of the battery because multiple discharges are possible.

State-of-Health: For secondary batteries, the life of the battery is defined by the number of usable cycles that remain until failure. For example, batteries are commonly removed from service when their discharge capacity has been reduced to 65% of the original capacity, indicating the end limit for usable cycles [14]. Other end-of-life conditions include short-circuited cells and low terminal voltage. In addition, a number of ageing mechanisms (dry-out, passivation, etc.) progress during a battery's life, resulting in its eventual failure. Each mechanism wears the battery at a different rate and simultaneous failure progression is common. Identifying which faults are occurring and to what degree will dictate the SOL prediction model that should be used. This classification of faults is an estimation of the battery's SOH. Much like the SOC approach to having three separate, parallel processing methodologies for prediction, the SOH estimation processing involves three different processing branches: statistical pattern recognition using linear discriminant functions, neural network-based pattern recognition, and fuzzy logic-based classification [15], [16]. Figure 5 demonstrates an example of battery failure identification using statistical pattern recognition. Axis labels α and β represent measured or derived parameters that are used to identify the failures.

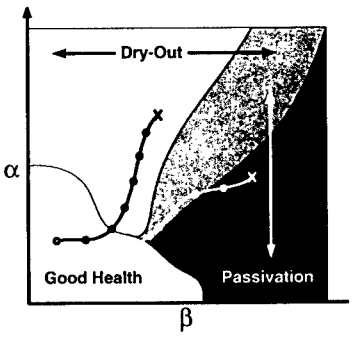


Figure 5. Failure identification using statistical pattern recognition

State-of-Life: Once faults and their severity are identified from the SOH processing, the proper SOL prediction model can be selected. Figure 6 shows a case where dry-out was identified as the dominant SOH condition in a lead-acid starter battery. As a result, a dry-out trained SOL predictor was used to predict the remaining usable cycles. Had a different dominant failure mechanism been identified from the SOH processing, a different SOL prediction model would have been used.

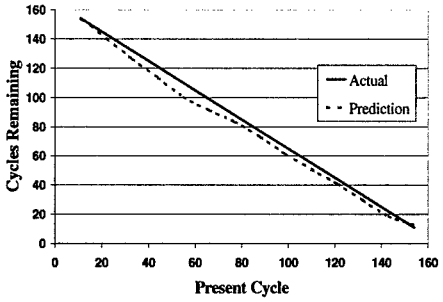


Figure 6. ARMA SOL prediction based on dry-out dominated SOH

Decision Fusion Processing: As previously mentioned, the SOC, SOH, and SOL processing makes three parallel predictions. This approach provides three assessments of the battery's condition. These three predictions are fed into a decision-processing module that determines the predictors' effectiveness relative to each other, processed sensor data, previous history, and knowledge about the battery type. The decision processing uses this information, via hybrid automated reasoning modules, to yield a combined prediction of the SOC, SOH, or SOL with a measure of confidence. Research on the decision-processing portion of the overall processing flow is currently under way. Referring to Figure 1, decision fusion represents the final stage of the processing; the output is then fed to a user interface that can display or coordinate the battery condition data.

Conclusions: Condition-based maintenance provides a means for improving the reliability of battery management in operational systems. For primary batteries, this represents using the full capacity of the battery before it is replaced. For secondary batteries, this represents cycling the battery to its true last usable cycle, rather then a conservative, statistical-based last cycle. In the case of a backup or standby battery, this represents knowledge of usage capacity prior to putting the battery online. The model-based approach described in this paper provides a framework for predicting SOC, SOH, and SOL. It has been shown that in addition to voltage, current and temperature, the internal electrical impedance of the battery ties closely to the physical processes that drive capacity and aging. A robust identification routine was developed and these identified parameters, along with measured signals, were used develop and test SOC, SOH, and SOL predictors. The developed ARMA and neural network SOC prediction models were discussed and shown to perform well across different battery chemistries and sizes. Some initial results were presented from the SOH and SOL prediction development; however, this work is still in its early stages. Finally, the framework for the decision fusion processing, which provides additional error checking and performance enhancement, was discussed. Most of the analyzed data was collected on a laboratory test stand under controlled conditions. Plans are being made to collect field data to test the developed model-based predictive diagnostics on battery systems (and other electrochemical energy sources) under real-world operating conditions.

Acknowledgement: The sponsorship of this work by the Office of Naval Research Code 331 (Dr. Phillip Abraham), under the ONR Grant N00014-98-1-0795, is gratefully acknowledged.

References:

[1] J. A. Stover, D. L. Hall, and R. E. Gibson, "A Fuzzy-Logic Architecture for Autonomous Multisensor Data Fusion," *IEEE Transactions on Industrial Electronics*, pp. 403-410, June 1996.

[2] A. K. Garga, "A Hybrid Implicit/Explicit Automated Reasoning Approach for Condition-Based Maintenance," Proceedings of the *ASNE Intelligent Ships Symposium II*, Philadelphia, 25-26 November, 1996.

[3] J.D. Erdley, Improved *Fault Detection using Multisensor Data Fusion*, M. S. Thesis, Electrical Engineering Department, The Pennsylvania State University, May 1997.

[4] D. L. Hall and A. K. Garga, "Pitfalls in Data Fusion (and How to Avoid Them)," Proceedings of *2nd International Conference on Information Fusion (FUSION 99)*, Sunnyvale, CA, 6-9 July 1999.A. J. Bard and L. R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, Wiley, New York, NY, 1980.

[5] M. Sluyters-Rehbach and J. H. Sluyther, "Sine Wave Methods in the Study of Electrode Processes," *Electroanalytic Chemistry*, vol. 4, Dekker, NY, 1970.

[6] J. P. Diard, B. LeGorrec, and C. Montella, "EIS study of electrochemical battery discharge on constant load," *Journal of Power Sources*, vol. 70, pp. 78-84, 1998.

[7] J. D. Kozlowski, T. Cawley, and C. S. Byington, "Model-Based Predictive Diagnostics for Primary and Secondary Batteries: Phase I Report," *Technical Memorandum, File No. 99-076*, Applied Research Laboratory, The Pennsylvania State University, June 17, 1999.

[8] Boukamp, B. A., "A Nonlinear Least Squares Fit Procedure for Analysis of Immittance Data of Electrochemical Systems," *Solid State Ionics*, vol. 20, pp. 31-44, 1986.

[9] MacDonald, J. R., and L. D. Potter, "A Flexible Procedure for Analyzing Impedance Spectroscopy Results: Description and Illustrations," *Solid State Ionics*, vol. 23, pp. 61-79, 1987.

[10] Pirlot, M., "General Local Search Methods," *European Journal of Operational Research*, vol. 92, pp. 493-511, 1996.

[11] Jeong, I. K. and J. J. Lee, "Adaptive Simulated Genetic Algorithm for System Identification," *Engineering Applications of Artificial Intelligence*, vol. 9, no. 5, pp. 523-532. 1996.

[12] A. K. Garga, B. T. Elverson, and D. C. Lang, "AR Modeling with Dimension Reduction for Machinery Fault Classification," Proceedings of the *MFPT Society 51st Meeting*, pp. 299-308, 14-18 April 1997.

[13] L. Ljung, *System Identification, Theory For The User*, 2nd Edition, Prentice Hall PRT, Upper Saddle River, NJ, 1999.

[14] D. Berndt, *Maintenance-Free Batteries*, 2nd Edition, John Wiley & Sons Inc., New York, NY, 1997.

[15] R. Schalkoff, *Pattern Recognition, Statistical, Structural and Neural Approaches*, John Wiley & Sons, Inc., New York, NY, 1992.

[16] D. L. McGonigal, *A Comparison of Automated Reasoning Techniques for Condition-Based Maintenance*, M. S. Thesis, Electrical Engineering Department, Pennsylvania State University, 1997.

FAILURE MODES AND ANALYSIS II

Chair: Dr. Richard D. Sisson, Jr.
Worcester Polytechnic Institute

EFFECTS OF SHOT PEENING PROCESSING ON THE FATIGUE BEHAVIOR OF THREE ALUMINUM ALLOYS AND TI-AL-4V

James Campbell
U. S. Army Research Laboratory
Weapons and Materials Directorate
AMSRL-WM-MD, Building 4600
Aberdeen Proving Ground, Maryland 21005-5069

Abstract: The fatigue strength of three shot-peened aluminum alloys (Al 7075-T651, Al 2024-T351 and Al 2014-T6) and a titanium alloy (Ti-6Al-4V) was measured to determine the differences in shot peening quality from three vendors that were given the same shot peening parameters. Shot peening produces surface roughness, a cold-worked layer and, most importantly, a residual stress layer that resists the propagation of fatigue cracks. Significant vendor-to-vendor differences in fatigue properties were found, with Vendor 1 giving the greatest fatigue lifetimes.

Key Words: Aluminum; fatigue; notched; residual stress; shot peening; titanium

Introduction: Fatigue behavior of metals is greatly influenced by the condition of the surface. Very smooth surfaces tend to show more resistance to fatigue than those that are rough. When there are compressive residual stresses at the surface, the fatigue strength is greatly improved. Over the past 50 years, shot peening has been used to improve the fatigue properties of a variety of metals, especially those that are used in aircraft applications. Shot peening is a mechanical surface treatment that introduces a roughened surface, a compressive surface layer, and increased dislocation density (cold work) near the surface. During shot peening, particles, typically steel or glass beads, are impacted onto a surface imparting significant deformation. Typically, the compressive layer improves the resistance to crack propagation, and thereby increases fatigue lifetime. In some instances where crack nucleation controls fatigue behavior, the increased surface roughness from shot peening is a detriment to fatigue strength.

The purpose of this investigation is to compare the fatigue properties of a Ti-6Al-4V alloy and three aluminum alloys that had been shot peened by three different vendors. The same shot peening specifications were given to the three vendors, one primary and two second source. Therefore, the differences in shot peening quality were assessed based on fatigue behavior.

Experimental: Three heat-treated aluminum alloys, Al 7075-T651, Al 2024-T351 and Al 2014-T6, were tested, along with a Ti-6Al-4V alloy that was in a solution treated and overaged (STOA) condition. For the Ti-6Al-4V alloy, an equiaxed grain structure was observed, as shown in Figure 1, that consisted of α and transformed β in an acicular structure. The α grain size was ~ 7.5 μm. Notched circular fatigue specimens were fabricated from the heat-treated alloys by Metal Samples of Munford, Al. They were 3" long with a gage length of 1" and a gage diameter of 0.315". The notch was centered in the gage length with a notched diameter of 0.255" and a notched radius of curvature of 0.024". The stress concentration factor k_t of the notch was calculated to be 2.4.

The machined specimens were shot peened with steel shot by three vendors (Vendor 1, Vendor 2 and Vendor 3). Specimens were shot peened over the gage and notched areas (also, the threaded portion for the Ti specimens) using specification MIL-S-13165C and the following parameters: Al (intensity- 0.006 – 0.010 N, shot size- S110, 100% coverage), and Ti (intensity- 0.005 – 0.011 N, shot size- S70 and 200% coverage).

Fatigue tests were conducted in pulsating tension using an Instron 1350 servohydraulic testing machine with an 8500 series control module. There were 11 specimens shot-peened by each vendor for the Ti alloy, while only 6 specimens from each vendor for each of the three Al alloys. The actual notched diameter for each specimen was used in determining the cross-sectional area for the Ti specimens, while the nominal diameter was used for the Al specimens. All tests were conducted at a frequency of 25 Hz and with a minimum load that was 10% of the maximum load, or R = 0.1. To determine the stresses for high-cycle fatigue and the endurance limit for the Ti specimens, many tests were arrested at 1,500,000 cycles and the fatigue specimen was retested at a higher stress level. This testing procedure is similar to the staircase test method [1] used for fatigue testing with a limited number of specimens.

Results and Discussion: S-N fatigue curves for the titanium alloy of stress vs. number of cycles-to-failure (CTF) were generated and are shown in Figures 2, 3, 4, for Vendors 1, 2 and 3, respectfully. It should be noted that the stress plotted is the applied stress σ_a not the maximum stress (k_t x σ_a) due to stress concentration at the notch. All three S-N curves are similar in shape and have a narrow stress region where there was failure by high-cycle fatigue. At this stress level, there is a large amount of scatter in the fatigue data. For example with Vendor 2, one test conducted at 72 ksi gave a fatigue life of 7.5 M cycles, while another gave 130,000 cycles at 72.5 ksi. But for each vendor, there was a different stress level where high-cycle fatigue and run-outs were observed, and this stress level was considered to be the endurance limit. For Vendor 1, the endurance limit was approximately 76 ksi, while it was ~ 74 ksi for Vendor 3 and ~ 72 ksi for Vendor 2.

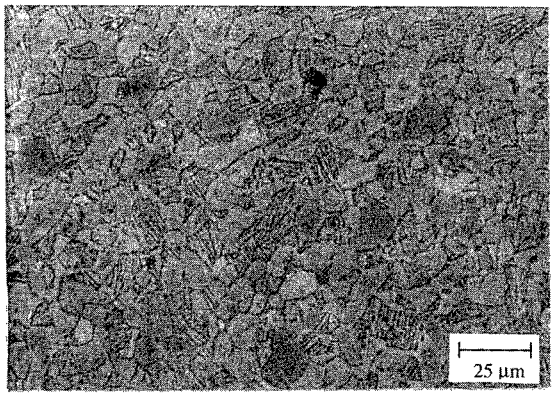


Figure 1 Microstructure of the heat-treated titanium alloy consisting of α and transformed β in an acicular structure, with an α grain size of ~ 7.5 µm. (500 X)

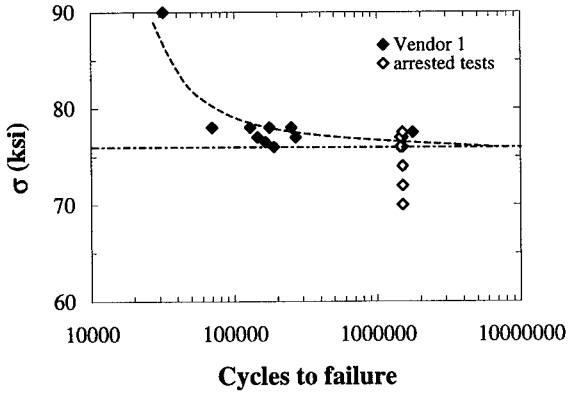


Figure 2 S-N curve for specimens shot peened by Vendor 1 with an endurance limit of ~ 76 ksi.

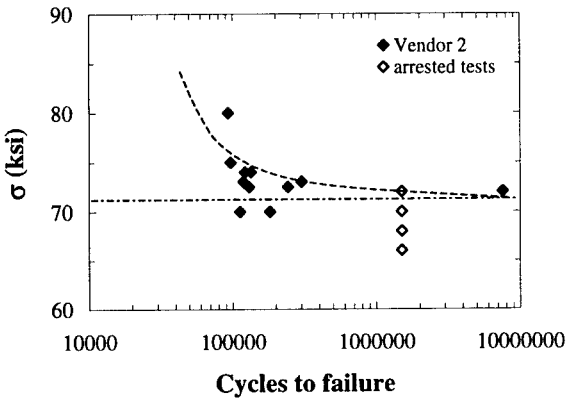


Figure 3 S-N curve for specimens shot peened by Vendor 2 with an endurance limit of ~ 72 ksi.

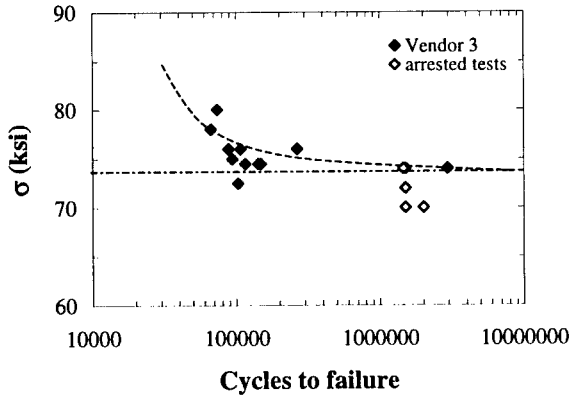


Figure 4 S-N curve for specimens shot peened by Vendor 3 with an endurance limit of ~ 74 ksi.

For the three aluminum alloys, there was less data for each S-N curve and the fatigue limit was harder to discern. Therefore, the vendor differences in fatigue behavior are easier to observe by comparing and ranking the data, as shown in Table I. Vendor 2 consistently had lowest fatigue lifetimes at several different stress levels compared to Vendors 1 and Vendor 3. This trend is clearly seen with all three alloys. Even with the data for all three alloys, the difference between Vendors 1 and 3 is rather small with Vendor 1 showing slightly longer fatigue lifetimes. With only six test specimens tested per condition, this number may be insufficient for measuring the smaller difference between Vendors 1 and 3.

Table I Vendor rank for fatigue lifetime at a given stress level for each aluminum alloy

Stress, σ (ksi)	Al 2024-T351	Al 7075-T651	Al 2014-T6
35	$V_2 \sim V_1 \sim V_3$	only V_1 tested	$V_2 \sim V_3 < V_1$
32.5	$V_2 < V_1 \sim V_3$	$V_2 < V_1 \sim V_3$	not measured
31.25	not measured	V_1 runout	not measured
30	only arrested tests	$V_2 < V_1 \sim V_3$	$V_2 < V_3 < V_1$
27, 27.5	not measured	only V_2 tested	$V_2 < V_3$

Since the specimens were fabricated, heat treated and tested in the same fashion, the difference in the fatigue behavior should be the due to differences in shot peening from each of the three vendors. Shot peening affects the surface of a metal by three mechanisms: (1) introduces surface roughness, (2) imparts cold work (higher dislocation density) and (3) forms a compressive residual stress layer. Gerdes and Luetjering [2] studied the effects of shot peening on notched Ti-6Al-4V specimens with several different microstructures. They found that the fatigue strength was determined by the compressive residual stresses retarding the crack propagation rate. Other authors [3,4] have also found that the compressive stress layer formed during peening has the most significant impact on fatigue lifetime for several other titanium alloys. Sridhar *et al.* [5] found through X-ray analysis that the compressive surface layer was approximately 0.4 – 0.7 μm deep, depending on the alloy and shot peening parameters. Similar compressive surface layer thicknesses were found in Al 2024-T3. [6] When the compressive stresses were relieved during thermal annealing [2,3], the fatigue properties were severely degraded, thereby showing the importance of the compressive layer formed during shot peening.

Surface roughness and cold work may also play a role in the observed differences in fatigue behavior. Since the surface roughness has not been measured, differences in this parameter for each vendor may be important. The nucleation of cracks is influenced by

the surface roughness and it can be an important factor when crack nucleation is the dominant mechanism controlling fatigue behavior. However, crack propagation rates through the compressive layer usually control the fatigue lifetime, since there are always local stress concentration sites that quickly nucleate cracks. [7] As for cold work, Leverant *et al.* [3] found that cold work in their studies of a Ti-6Al-4V alloy, with a similar microstructure to the current study, had very little influence on the crack growth rate and fatigue strength.

Microhardness measurements were made in an attempt to detect differences in the compressive layer and in the amount of cold work after shot peening. They were made on cross sections perpendicular to the loading direction using a Wilson Series 200 testing machine with a Vickers indenter. Measurements were taken in the center and approximately 50 μm from the surface, along with several hardness profiles. Wagner and Mueller [6] found increased dislocation densities in a 400 μm deep surface layer for Al 2024-T3. Hardness measurements for the current investigation were typically in the 305 – 325 HV range for Ti-6Al-4V and 170 – 185 HV for Al 7075-T651. In Figure 5, two depth profiles taken on a Ti specimen showed very little change in hardness from the surface to a depth of approximately 1.5 mm. For several other Ti and Al specimens, there was very little difference in microhardness measurements taken at the center of the specimen to those ~ 50 μm from the surface. In studies on fatigue behavior of shot peened titanium, Berger and Gregory [8] found that shot peening does not increase the microhardness readings substantially in the near-surface layer for a β-titanium alloy However, Rios *et al.* [7] did find significant differences in microhardness with depth below the surface for Al 2024-T351. They also developed a model that incorporates shot peening to predict fatigue lifetimes.

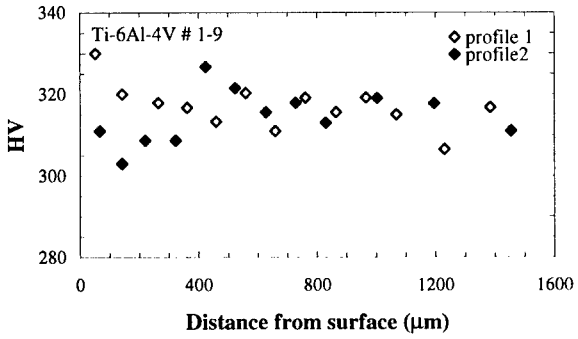


Figure 5 Vickers microhardness profiles on a Ti-6Al-4V specimen.

Conclusions: Differences in fatigue properties were observed for both Ti-6Al-4V and the three aluminum alloys that were shot peened by three different vendors. For the titanium alloy, Vendor 1 had the highest endurance limit at ~ 76 ksi, while Vendor 2 had the lowest at ~ 72 ksi and Vendor 3 was ~ 74 ksi. For the aluminum alloys, fatigue lifetimes of the vendors were similarly ranked. The differences in fatigue behavior are likely due to differences in the compressive residual stress layer formed during shot peening, which retarded the crack propagation rate.

Acknowledgement: The author would like to express appreciation to George Liu and Dr. Kirit Bhansali of the Aviation Missile Command (AMCOM) for support for this project.

References:

[1] R. C. Rice, "Fatigue Data Analysis," in ASM Metals Handbook, Mechanical Testing-Vol. 8, ASM International, Metals Park, OH, (1985) 695 – 720.

[2] C. Gerdes and G. Luterjering, "Influence of Shot Peening on Notched Fatigue Strength of Ti-6Al-4V," Proceedings of the Second International Conference on Shot Peening ICSP-2, ed. H. O. Fuchs, (1985) 175 - 180.

[3] G. R. Leverant *et al.*, "Surface Residual Stresses, Surface Topography and the Fatigue Behavior of Ti-6Al-4V," Metallurg. Trans. A, **10A** (1979) 251 – 257.

[4] B. R. Sridhar *et al.*, "Effect of Shot Peening on the Fatigue and Fracture Behavior of Two Titanium Alloys," J. Mater. Sci., **31** (1996) 5953-60.

[5] B. R. Sridhar *et al.*, "Effect of Shot Peening on the Residual Stress Distribution in Two Commercial Titanium Alloys," J. Mater. Sci., **27** (1996) 5783 - 5788.

[6] L. Wagner and C. Mueller, "Effect of Shot Peening on Fatigue Behavior in Al Alloys," Mater. Manufact. Process., **7** (1992) 423 - 440.

[7] E. R. De Los Rios *et al.*, "Modeling Fatigue Crack Growth in Shot-peened Components of Al 2024-T351," Fatigue Fract. Eng. Mater. Struct., **23** (2000) 709 - 716.

[8] M. C. Berger and J. K. Gregory, "Residual Stress Relaxation in Shot Peened Timetal 21s," Mater. Sci. Eng., **A263** (1999) 200 - 204.

Ejection-Seat-Quick-Release-Fitting
Quantitative fractography and estimation of the local toughness using the topography of the fracture surface

K.Wolf, Bundeswehr Research Institute for Materials, Explosives,
Fuels and Lubricants (WIWEB)
Landshuter Str. 70, 85435 Erding, Germany

ABSTRACT: A Quick-Release-Fitting of an Ejection Seat broke and investigations were made to estimate the fracture mechanism on the basis of fracture surface characteristics.
Fractographic based investigations normally use qualitative characteristics to purify the cause of the failure. Generally the determination of the fracture location and its origin, the kind of fracture and special features on the fracture surface are enough to describe the cause of the fracture.
The aim of this investigation is to use quantitative fractography as a tool to get more information about crack propagation mechanism and also to get an estimation concerning striations, fracture topography and stretched zone. This results in a correlation to fracture mechanic concepts.
During crack propagation striations were created on the fracture surface caused by the service induced load changes. The distance of the striations were measured to estimate crack propagation and remaining life time.
In addition a plastic stretched zone could be found on the tip of the cracks. The width of these zones gave information about local fracture toughness. The 3-dimesional zone symmetry was measured on cross sections by using stereographical methods to get more information about the crack tip and the crack propagation.
To complete the failure analysis nondestructive evaluation, metallographic examination and chemical investigations were carried out. No additional cracks could be found. Most of the failed parts showed that the microstructure, the hardness and the chemical composition of the Al-alloy were within the specification, but some of the cracked parts were manufactured with an other material as specified.

Keywords: Al-alloy, crack propagation, fatigue fracture, fracture toughness, quantitative fractography, stretch zone,

INTRODUCTION

During the use of a Quick-Release-Fitting from a parachute emergency system, a miss function came up and it was not possible for the pilot to release his seat belt. The reason for that was a broken part (bell) within the central lock (Figures 1 and 2). The bell was constructed and manufactured as thin wall, cylindric part with a centred thick part in the middle of the bell which was surrounded by a slot half the circumference. In elongation of this slot the bell broke. Many other bells were checked and eight more bells with cracks were found.

Figures 3 and 4 show the broken bell „S", the cracked bells „V1" to „V8", the new bell „N" and the entire Quick-Release-Fitting. The fracture of the bell „S" resulted in cracks along the solid central part which is surrounded by a semicircular oblong hole. The eight bells „V1" to „V8" with preinduced cracks are damaged in the same area. The bells „V1", „V7" and „V8" showed preinduced cracks on both ends of the oblong hole.

The material specified for these bells was Al-alloy AlMgSiPb; heat treatment of the bells was to follow DIN 1747, part 1; this includes an approximate minimum hardness of Brinell HB 80; The following literature was used: some information about the function of the central lock, some engineering drawings, ESIS P2-92, DVM 002-Merkblatt and some information to the material. The bells were manufactured from aluminium alloy. The constituents of the bells „S", „V1", „V3" to „V8" were within the specification requirements for Al-alloy AlMgSiPb, but the bells „V2" and „N" were fabricated from Al-alloy AlCuMgPb (Table 1).

Table 1: Chemical composition and hardness of the bells

Samples or Reference Material	Element mass ratios [%]							Hardness
	Si	Mn	Cu	Mg	Fe	Zn	Pb+[2]	HB
bell „S"	0,80	0,85	0,02	0,87	0,31	0,03	1,39	58
bell „V1"	0,75	0,40	0,06	0,64	0,29	0,02	1,57	60
bell „V2"	**0,36**	0,54	**3,70**	0,76	**0,72**	0,42	1,27	105
bell „N"	**0,44**	0,60	**3,84**	0,72	0,55	0,36	1,27	108
AlMgSiPb [3] WL 3.0615	0,6 bis 1,4	0,4 bis 1,0	≤ 0,10	0,6 bis 1,2	≤ 0,5	≤ 0,5	1,0 bis 3,0	[6]
A1CuMgPb [4] WL 3.1645	≤ 0,8	0,5 bis 1,0	3,3 bis 4,6	0,40 bis 1,8	≤ 0,8	≤ 0,8	0,8 bis 1,5	[6]

[1] remainder: Al
[2] Pb+Sn+Bi+Cd+Sb: 1,0 bis 3,0; Cd not specified
[3] according to the drawing requirements the clocks were fabricated from AlMgSiPbF28; the values of AlMgSiPb were taken from DIN 1725, part 1 vom Feb.1983
[4] the material destinated in according to the chemical analysis
[5] other elements analysed: Cr 0,10; Ni 0,20; Bi 0,20; Sn 0,20.

Remarks: The chemical analysis of the bells "V3" to "V8" are not listed in table 1 because they had the same composition like the broken bell "S" and the cracked bell "V 1 " - these bells compare favorable to the requirements as shown in the table; the elements which not met the requirements are **bolted** in the table.

Metallographic samples were taken through the cross-section of the fracture origin on „S", „N" and „V8". The cleanliness of the material was relatively good for the required alloy, despite of only few inclusions, but „N" showed many precipitates and bands (Figures 5 and 6). Microhardness measurements were performed on cross-sections of these parts. The required hardness as specified on the governing engineering drawing

for the component was 80 HB minimum. The hardness values obtained for „S" (58 HB) and „V8" (62 HB) were similar and did not confirm the specification. The hardness of „V1", „V3", „V6" and „V7" were within this range (60 HB to 65 HB) and also low, but „V4" (81 HB), „V5" (103 HB) and „V2" (105 HB) showed relatively high values.

The stress distribution around the crack initiation area of the bell was calculated with the help of Finite-Element-Method and was found in good agreement with the different stages and locations of the broken bells (Figure 7 and 8).

To find out the damage cause usually qualitative fractographic investigations are sufficient in addition to other investigations. Normally the fracture origin, the kind of fracture and special features on the fracture surface are described. The goal of this investigation was to get information about remaining life span depending on crack length and to answer questions about possible influences of the different materials. Therefore it was necessary to determine the crack propagation and local fracture toughness by means of quantitative fractography. The approach included examination of the fracture and cracks concerning crack propagation, development of local fracture toughness and fracture mechanics evaluation of the stresses required to initiate the cracks.

Results

Fracture Examination

The bells „V1" to „V8" showed precrack extensions from 0.4 mm to 1.5 mm whereby the plate thickness was 1.5 mm. The circumference precrack extension runs from 4.5 mm up to 26 mm. Only „V4" and „V8" showed precracks on both sides (up to 0.3 mm). The longest precrack was found at „V7" (26 mm). The results of the SEM-investigation showed that „S" and „V1" to „V8" cracked due to fatigue fracture (Figure 9). The fatigue crack on „S" was initiated by etch-pits and pittings underneath the 10 µm eloxal layer (Figure 10). The crack propagation took place from the outside of „S". The crack propagation showed in addition to the operational load marks signs which points to the load of the bell during „put on-lock-take off". Because of this operation the mass of the solid central part moves away from the mid part of the bell. Due to this loading the mid part is overcome by bending. The pits and pittings promoted the precracks in this area.

The fatigue fracture surface of the broken bell and the cracked bells were examined with respect to the striations at the origin area, the middle area and the crack tip area to quantify the crack propagation. Figure 11 shows the measured values (striations per mm) as a function of the crack length. These values were calculated by regression to get an average curve [1;2;3]. The integration of this curve lead to the crack propagation curve (Figure 12). The comparison of the derivated crack propagation (dl/dN) of the striation width with literature values shows that the crack propagation was uncritical at most investigated bells (Figure 13). This lead to the conclusion that for cracked bells even with relatively long cracks no critical stage can be expected during the time fixed life span,

Fracture Mechanics Evaluation

The area between the primarily crack (fatigue crack) and the final fracture (stable crack growth) is called „Stretched Zone". Ductile materials show blunting at the crack tip due

to mechanical stress, that means the beginning of the stable crack growth due to slip or shear processes. The development of the streched zone depends on the stress, the load, the velocity, the crystal structure and the texture. Quantitative statements to the local fracture toughness can be made by measuring the stretched zone width (SZW), the stretch zone height (SZH) and/or the crack tip opening displacement (CTOD).

The stretched zone measurements were done on stereo image paires by scanning electron microscopy on both corresponding fracture surfaces of the specimens „V1" and „V2" (Figure 14). Measurements along several base lines through the stretched zone achieved the streched zone width (SZW) calculated by digital image processing [4;5;6;7]. Based on the ESIS P2-92 and DVM 002-Merkblatt the mechanical properties for the local crack initiation were calculated. The crack initiation toughness defined by the J-Integral is connected with the stretched zone and can be set as

$$J_i = \frac{E \cdot SZW}{0{,}4 \cdot d^*}$$

J_i = Fracture resistance at crack initiation
E = Young`s modulus
d* = proportionality constant

By taking in the average SZW-values into this equation the bell V1 shows a J_i-value of 28 N mm/mm^2 and the bell V2 shows a J_i-value of 23 N mm/mm^2 On the other hand there exist a relation between J_i and K_{Ic}:

$$K_{Ic} = \sqrt{\frac{J_i \cdot E}{1 - \nu^2}}$$

ν = Poisson`s ratio
K_{Ic} = Fracture toughness

After calculation the value for bell V1 results in 47 MN/ $\sqrt{m^3}$ and the value for bell V2 results in 42 MN/ $\sqrt{m^3}$. Because of the little difference in local fracture toughness the results show that different materials by itself did not cause additional problems.

CONCLUSION

A bell from a Quick-Release-Fitting broke due to fatigue fracture. Other cracked bells were found and showed fatigue fracture too, but differences in crack depth. Chemical analysis results pointed out that the bells were manufactured from different aluminium alloys. Metallographic examinations also resulted in different microstructure and hardness. The crack initiation site for all failed bells was located on the oblong hole in the middle of the bells, there were etch pits and pittings underneath the oxide layer which promoted the precracks. Striations were examined on some cracked bells and the derivated crack propagation was within an uncritical stage compared with literature values. The local fracture toughness of both materials showed little difference.

To improve the whole fitting area in termes of reducing the load in critical areas and to elongate lifw span the focus should be set on the optimisation of the construction which has been proposed. Additionally clear requirements should be given for the heat treatment of these parts.

The cause of the fracture and the cracks can be lead back to load and construction. Only an optimal harmonizing of material, load, manufactoring and construction leads to an optimal part. A proposal for modifiing has be done.

ACKNOWLEDGEMENT

I would like to express my sincere appreciation to the head of WIWEB Prof. Kunz, Department Head Dr. Kohlhaas and Dr. Woidneck and Branch Head Mr. Gedon. Mr. Muellera for his professional assistance in conducting tests and SEM-work.

REFERENCES

1. J.E.Forsyth, D.A.Ryder, Some Results Derived from the Microsoft Examination of Crack Surfaces, Aircraft Engineering, April 1960
2. P.G.T.Howell, A.Boyde, Comparison of various Methods for reducing Measurements from Stereo-Pair Scanning Electron Micrographs to "Real 3-D Data", Scanning Electron Microscopy Symposium, Chicago, 1972
3. J.D.Landes, J.A.Begley, G.A.Clarke, Elastic-Plastic Fracture, Symposium ASTM, Atlanta, GA, Nov.1977
4. O.Kolednik, P.Stüwe, Abschätzung der Rißzähigkeit eines duktilen Werkstoffs aus der Gestalt der Bruchfläche, Metallkunde, Bd. 73, 1982
5. DVM-002, Ermittlung von Rißinitiierungswerten und Rißwiderstandskurven bei Anwendung des J-Integrals, Juni 1987
6. O.Kolednik, Stereogrammetrische Untersuchungen des Rißwachstums bei duktilen Materialien, Gefüge und Bruch 9, Juni 1989
7. K.-H.Schwalbe,A.Cornec,K.Baustian,M.Homayan, Intercomparison of Fracture Toughness Measurements of ductile Materials, Synthesis Report, Geesthacht, July 1990

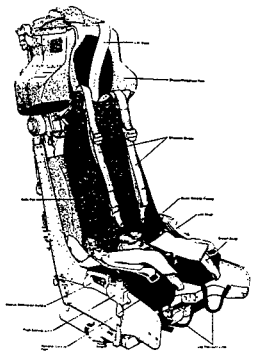


Figure 1
Ejection-Seat with the Quick-Release-Fitting

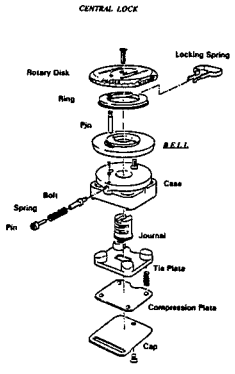


Figure 2
Schematic representation of the Quick-Release-Fitting with the bell

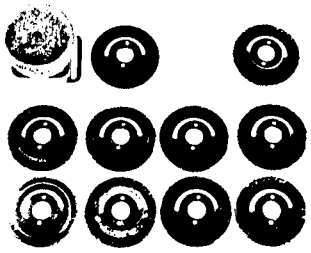


Figure 3 1:4,5
Overall view of the central lock, the unused new bell N, the damaged bell S and the cracked bells V1 through V8

Figure 4 1:1,5
Bell S with the broken central part - view direction from the outside

Figure 5 200:1
Micrograph of a cross-sectional microstructure with only few precipitations - bell S

Figure 6 200:1
Micrograph of a cross section of the new bell N with numerous precipitations and distinct grains

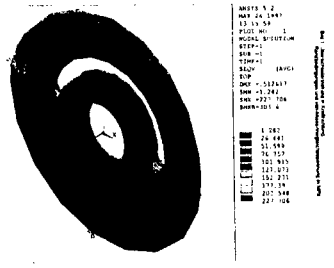


Figure 7
Overview of the von-Mises-Equivalent Stress in MPa

Figure 8
Detail of the von-Mises-Equivalent Stress in MPa

Figure 9
SEM-micrograph of the crack area with striations – bell S

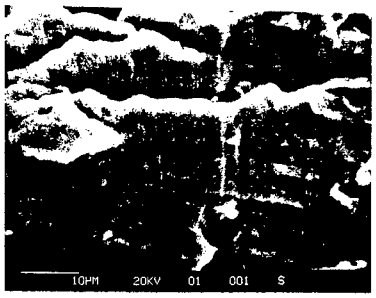


Figure 10
SEM-micrograph showing preinduced crack area with a crack origin underneath the eloxal layer – bell S

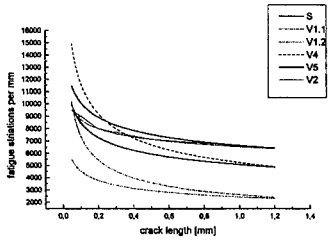


Figure 11
Results of the fracture surface measurements - „striations per millimeter as a function of the crack length"; the doted curve was determined by recression calculating.

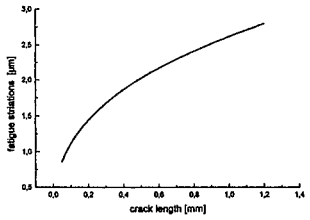


Figure 12
Crack extension curve, calculated from the average curve shown in Figure 9

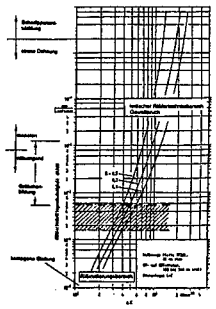


Figure 13
Comparison between the crack growth rate from the fracture surface (hatched) and the literature results for this material

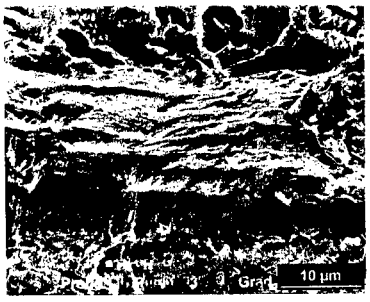


Figure 14
SEM-micrograph from the crack tip (SZW) - bell S1

A COMPARISON OF FATIGUE DESIGN METHODS

R. J. Scavuzzo

Professor Emeritus
The University of Akron
Akron, OH 44325-0301

Abstract: There are two basic fatigue-testing methods: fatigue using a maximum cyclic force, the stress-life method, and fatigue using a maximum cyclic strain, the strain-life method. Wöhler first tests to establish a S-N diagram were based on a rotating beam with a constant maximum force loading. Subsequently, the R. R. Moore test that used four-point loading with a constant load was developed and became one of the standard tests. Most machine design textbooks teach fatigue design methods based on these basic tests. Because of research and development activities in the 1960s, cyclic strain methods were developed Low-cycle fatigue analyses based on these new methods have been found to be more accurate in this low-cycle regime. Of course endurance limits of both methods do not change; differences are in predicting low-cycle fatigue life.

This paper presents a short comparison of these two fatigue design methods.

Key Words: Fatigue, Fatigue Analysis, Load Fatigue, Strain Fatigue, Low-cycle Fatigue

Introduction: In the 1950s and 1960s, cyclic thermal stress in nuclear reactors became an object of research. Very high elastic thermal stresses are often calculated in nuclear reactor components because of high temperature gradients. It was recognized that these stresses were fundamentally different from constant load stresses. A small amount of yielding decreased thermal stresses significantly. As a result, elaborate test programs were conducted to thermally cycle components to develop high thermal stresses and initiate fatigue failures. Thermal stresses are like residual stresses; a small amount of yielding relieves these stresses. As thinking matured in this area, researchers realized that the same results could be obtained by mechanical strain cycling specimens rather than trying to simulate the cyclic thermal conditions. Tests could be run in a much shorter time with much less cost and control on the temperature of the specimen and the actual cyclic strains were much more accurate. Thus, data were based on cyclic mechanical strain tests in lieu of cyclic thermal strain tests. Manson [1] and Coffin [2] contributed significantly in these areas. Design procedures for the design of nuclear pressure vessels were developed in the 1960s based on these data. The ASME Boiler and Pressure Vessel Code [3] presents these methods and has expanded the procedures to other pressure vessels besides nuclear pressure vessels. B. F. Langer [4], while at the Bettis Atomic Power Laboratory run by Westinghouse Electric Corporation, contributed significantly to these code procedures.

The endurance limit is the same developed by either test method. Differences in the two methods occur in the low-cycle regime. The cyclic strain method is much more accurate in this area. As a result, fatigue design methods in the nuclear industry as well as the aerospace field and others make use of these cyclic strain procedures. The fact that stress-life design methods, based on constant load data, are usually taught in undergraduate mechanical engineering programs adds to confusion in this area.

A number of more recent textbooks on fatigue cover both load cycling and strain cycling fatigue. Bannantine, Comer and Handrock, all former students of Professor JoDean Morrow at the University of Illinois, cover both methods very well in their text [5]. The first chapter is devoted to the "stress-life" method and the second to the "strain-life" method. Chapter 6 compares these methods. The text by Fuchs and Stephens [6], which is more scientific in approach, is an in depth presentation of many aspects of the strain cyclic method. Collins [7] also presents strain cyclic design methods and is an excellent contribution. However, the third edition of Shigley's mechanical engineering design textbook [8] and Juvinall's textbook [9] only cover the stress-life method. The most recent edition of Shigley's textbook does include some aspects of the strain-life cyclic method.

In this paper, the stress-life method is presented based on References [5,6,7]. The strain-life cyclic design method is taken from the ASME Code [3] that is based on the work B. F. Langer [4] and others. This short review paper cannot treat the subject thoroughly and the reader is referred to References [5,6,7] for additional insight. Fracture mechanics concepts are used to predict fatigue crack growth and finial fracture [5-7]. Crack initiation can be related to the cyclic Von Mises stress [9]. These topics are not considered.

The endurance limit is required in the cycles versus life graph, the S-N curve, of both methods and is reviewed first. Then, the stress-life and strain-life methods are treated.

Endurance Limit: The endurance limit, S_e, is a constant alternating stress below which failure will not occur. Steels have an endurance limit; most nonferrous alloys do not have an endurance limit. The endurance limit of steels can be approximated by the fact that a mirror polished laboratory specimen with a 0.3 inch diameter has a value of about ½ of the ultimate strength, S_u.

$$S_e = ½\ S_u \tag{1}$$

Also, since the ultimate strength is related to the Brinell Hardness Number (BHN), the endurance limit can also be approximated with this hardness measurement.

$$S_u \cong 500\ \text{BHN} \tag{2}$$

and

$$S_e \cong 250\ \text{BHN} \tag{3}$$

This relationship is depicted graphically in Fig. 1. Note that after the endurance limit

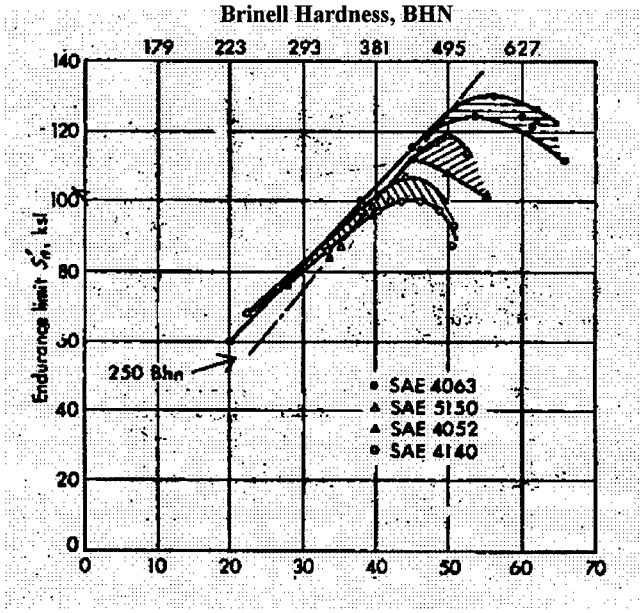


Fig. 1 Relationship between the endurance limit and hardness [9].

reaches 100 ksi the relationship between hardness and the endurance limit is lost for all alloys. The endurance limit can be increased or decreased by the following factors:

1. Surface finish
2. Size Effect
3. Temperature Effect
4. Environment
5. Surface Treatment

Surface Finish: The standard test specimen has a polished finished. Any other surface finish will decrease the endurance limit. For example, a fine ground or commercially polished surface will reduce the endurance limit from 10% to 28% and depends on hardness or ultimate strength. Below an ultimate strength of 140 ksi (280 BHN), the reduction is 10%; for higher hardnesses the decrease reaches 28% of the laboratory specimen. Other surface effects are as follows: machined surface 20% to 50%, hot-rolled from 28% to 78% and as-forged from 45% to 86%. Surface stress concentrations from scratches, pits machining marks, changes in surface strength as well as tensile residual stresses cause these reductions from these various surface finishes.

Size Effect: The diameter of the standard laboratory specimen is 0.3". As the fatigued specimen increases to about 2", the high cycle fatigue strength is deceased. There are two main reasons for this reduction: probability of a defect is increased with size and stress gradients from bending or torsion are decreased allowing more volume to be highly stressed.

Of course, tensile stresses have no gradient and a reduction of 10 % is recommended by Juvinall [9]. A recommended empirical equation [5] is as follows:

$$\text{Csize} = \begin{cases} 1.0 & \text{if } d \le 0.3\text{in.} \\ 0.869d^{-0.097} & \text{if } 0.3 \le d \le 10.0\text{in.} \end{cases} \quad (4)$$

Temperature Effects: The ASME Boiler and Pressure Vessel Code does not alter design S-N Curves below 320^0 C (600^0 F). However, other metals such as aluminum decreases significantly above 200^0 C (400^0 F) or less. The fatigue strength of Titanium decreases above room temperature. Thermoplastics are so sensitive to a temperature increase that heating associated with cyclic stresses can significantly reduce fatigue strength. Thus, the cyclic rate must be specified in data and taken into account in design. Metals are not sensitive to the cyclic rate unless rates are in the acoustic range,

Environment: A corrosive environment can reduce the endurance limit to a fraction of its value in air. Water can reduce the endurance limit of carbon and low alloy steel by more than a factor of three. For example, steel SAE 1050 steel with an ultimate strength of 120 ksi, the endurance limit is reduced to 20 ksi in water from about 60 ksi in air [5]. In adverse environments, alloy steels must be used if high endurance limits are required.

Surface Treatments: Surface treatments that increase the strength of the surface or develop compressive residual stresses on the surface or cause both effects can improve high cycle fatigue strength measurably. Helpful surface treatments are shot peening, roll hardening, nitriding or carburizing the surface can, at times, double the endurance limit over the untreated value. On the other hand, decarburization from forging, hot rolling, etc. can decrease the limit by over a factor of two. Grinding and other machining operations develop tensile residual stresses that decrease strength [5,8,9]. Electro plating of steel with hard metal such as chromium or nickel also develops tensile residual and can reduce strengths by over 50% [10].

Stress Concentrations: Stress concentration effects both methods are based on the fatigue strength reduction factor, K_f, defined as follows:

$$K_f = \frac{\textit{Unnotched Fatigue Strength}}{\textit{Notched Fatigue Strength}} \quad (5)$$

This factor is related to the theoretical stress concentration factor, K_t, and the notch sensitivity, q. The notch sensitivity is a function of material, hardness and notch radius or the volume of material affected by the stress concentration. Thus as the notch radius becomes larger, the sensitivity becomes larger. The two design methods differ in one aspect: in the stress-life q is a function or cycles [7,9] whereas in the strain-life method q does not vary with cycles and, therefore, K_f, does not vary with cycles.

$$q = \frac{K_f - 1}{K_t - 1} \quad (6)$$

The notch sensitivity is considered at times to be a material property. Thuw, given q and K_t the fatigue reduction factor can be calculated.

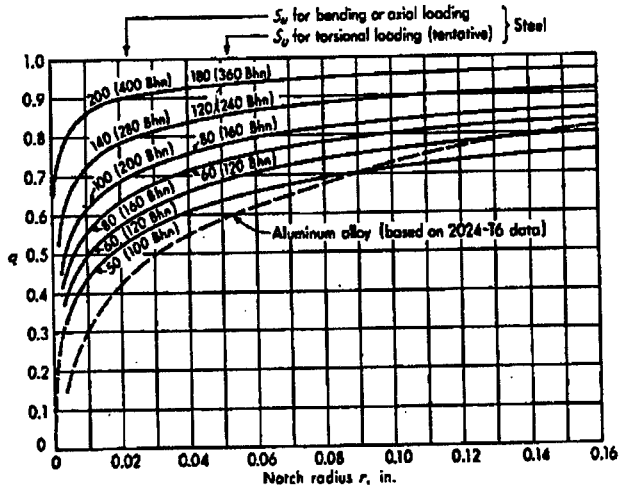


Fig. 2 Notch Sensitivity as a function of hardness, notch radius and

S-N Curve: The cycles to failure curve (S-N Curve) of the stress life method is approximated by using the fact that the cycles to failure at 1,000 cycles is about 90% of the ultimate strength. Using this point and the endurance limit, a straight line on log-log approximates the S-N curve as shown on Fig. 3 where S in the alternating stress..

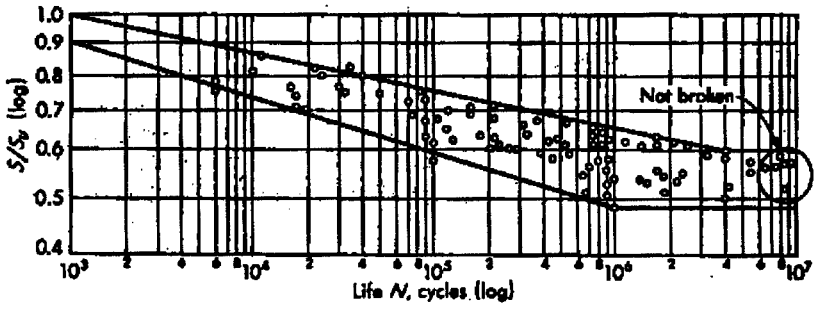


Fig. 3 Stress-life S-N Curve showing Data [8].

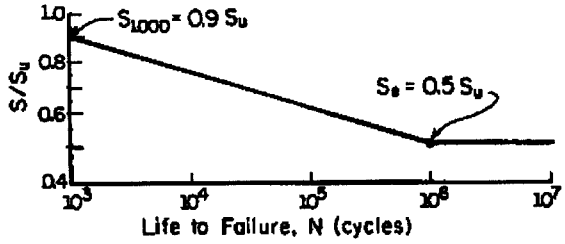
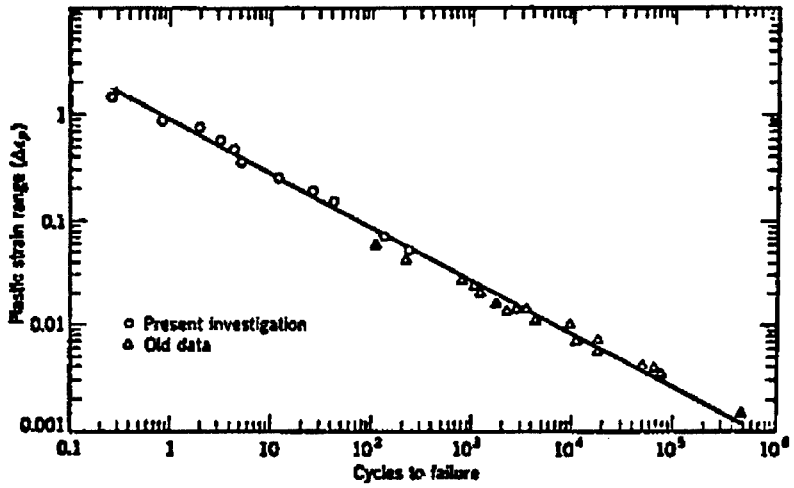


Fig. 4 Approximate S-N Curve for the Stress-Life Method [5].

The S-N curve presented in Fig. 4 is accurate in the cyclic range near the endurance limit (>10^5). In the low cycle range results are not accurate and the strain-life method must be used [5]. Mean stresses are used in the Goodman Equation assuming a nominal mean stress. The fatigue strength reduction factor is applied only to the calculated alternating stress, S_{ca},

$$\frac{S_m}{S_u}+\frac{K_f S_{ca}}{S}=\frac{1}{FS} \tag{7}$$

Strain-Life Method: As implied, this design method is based on strain versus cycles data. Fig. 5 is a plot of the plastic strain range (peak-to-peak measurements and not the alternating value which is ½ the range). The slope the curve is about –½; actual slopes will vary with material [2,6,7] but in the procedure developed by Langer a value of –½ is used.



Plastic strain range versus cycles to failure, 2S aluminum.

Fig. 5 Plastic strain range versus cycles to failure.

Langer developed the following equation with this assumption of the plastic portion of the S-N curve has a negative slope of – ½.

$$\Delta \varepsilon_r^p = \frac{C}{\sqrt{N}} \tag{8}$$

where N is the cycles to failure and C is a constant to be evaluated. By assuming that a tensile test is ¼ of a cycle, the cyclic plastic strain of Eq. (8) can be related to the true strain at fracture. The true strain at fracture can be determined from the Reduction in Area, RA in percent, from a tensile test.

$$\bar{\varepsilon}_f = \ln \frac{100}{100-RA} \tag{9}$$

Substituting into Eq. (9) yields,

$$C=\frac{1}{2}\ln\frac{100}{100-RA} \tag{10}$$

Eq. (8) becomes

$$\Delta\varepsilon_r^p=\frac{1}{2\sqrt{N}}\ln\frac{100}{100-RA} \tag{11}$$

The alternating component of plastic strain is ½ of this value. Langer assumed that the alternating component of elastic strain is the endurance limit divided by the elastic modulus, E. Thus the total alternating strain is as follows:

$$\Delta\varepsilon_a=\frac{1}{4\sqrt{N}}\ln\frac{100}{100-RA}+\frac{S_e}{E} \tag{12}$$

The approach taken in this analysis was that stresses are calculated by elastic analysis. Stresses that exceed the yield point are included and give an approximation to the strain. Thus Eq. (12) is multiplied by the elastic modulus E to obtain units of stress. Analytically developed values are compared to the design curves. Thus, if Eq. (12) is multiplied by the elastic modulus, E, Langer's equation is developed in units of stress. Present FEA codes allow the calculation of strain ranges directly and can be used in lieu of an elastically calculated stress to evaluate life.

$$S_a=\frac{E}{2\sqrt{N}}\ln\frac{100}{100-RA}+S_e \tag{13}$$

ASME Data: Fig. 6 and Fig. 7 are cyclic strain data for low carbon and low alloy steel, respectively. Langer's equation, Eq. (13), is plotted on the graphs. This procedure uses the concept of the worst-case mean stress that is also shown on these two graphs. It is assumed that the highest value of mean stress that can be obtained is the calculated alternating stress, S_{ca}, taken from the yield point, S_y.

$$K_f S_m = S_y - K_f S_{ca} \tag{14}$$

There are three cases to be considered for application to the Goodman equation:

Case 1 $K_f (S_{ca} + S_m) < S_y$ (Stresses are elastic)

$$\frac{K_f S_m}{S_u}+\frac{K_f S_{ca}}{S}=\frac{1}{FS} \tag{15}$$

Case 2 $K_f (S_{ca} + S_m) > S_y$; $K_f S_{ca} < S_y$

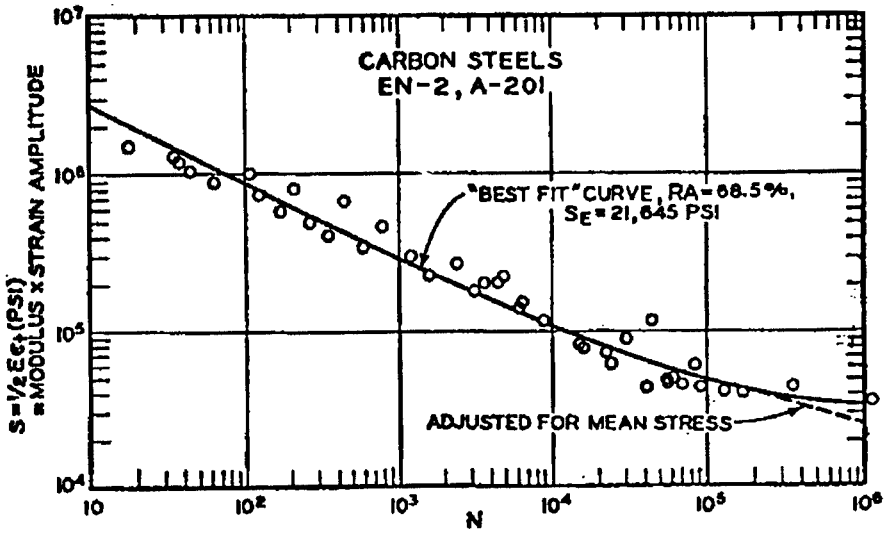


Fig. 6 Alternating Strain Data in Units of Stress for Low Carbon Steel.

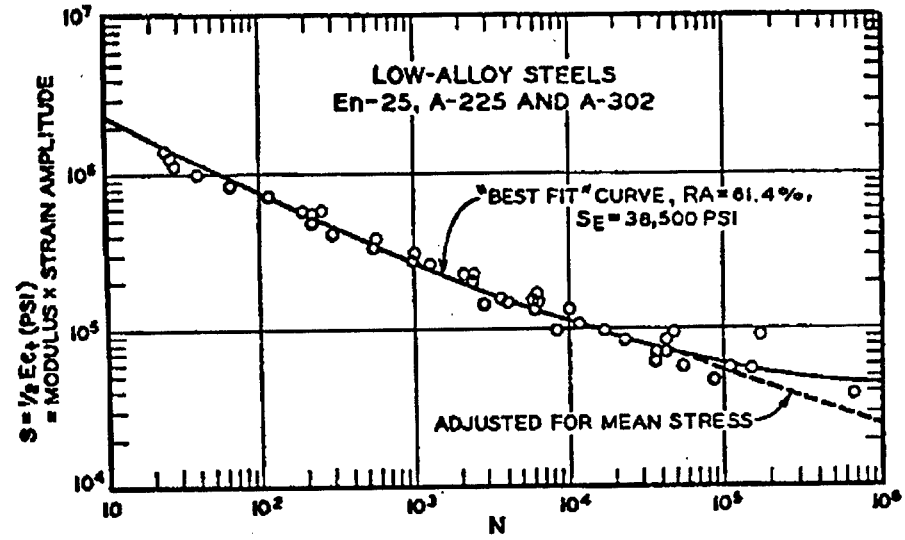


Fig. 7 Alternating Strain Data in Units of Stress for Low-Alloy Steel.

$$\frac{S_y - K_f S_{ca}}{S_u} + \frac{K_f S_{ca}}{S} = \frac{1}{FS} \tag{16}$$

Case 3 $K_f S_{ca} > S_y$ (Cyclic plasticity)

For this case the effective mean stress is zero. The Goodman equation reduces to the following:

$$\frac{K_f S_{ca}}{S} = \frac{1}{FS} \tag{17}$$

Summary: Presented is a short summary of a few differences in the stress-life method versus the strain-life method. First the data in one case is based on the cyclic loads and bending stresses based on the load calculations. In the strain method, cyclic strain data is multiplied by E to have units of stress. However, "stress" values exceed a million psi. Values are strain times the modulus. The effects of stress concentration are treated differently. In one case the usual procedure is to use nominal values of mean stress where in the strain method stress concentrations are included in both the mean and alternating components. In the stress-life method K_f is a function of cycles; in the strain-life method, K_f is constant with cycles. In this short paper many factors had to be omitted and the reader is referred to the referenced material for addition insight.

References:

[1] Manson, S. S., *Thermal Stress and Low Cycle Fatigue*, McGraw-Hill, 1966.

[2] Coffin, Jr., L. F., "A Study on the Effects of Cyclic Thermal Stresses on Ductile Metal," Trans. ASME, Vol. 74, 1954, pp. 931-950.

[3] American Society of Mechanical Engineers, "ASME Boiler and Pressure Vessel Code," ASME, 3 Park Ave., New York, NY 10016-5990.

[4] Langer, B. F., "Design of Pressure Vessels Involving Fatigue," *Pressure Vessel Engineering*, R. W. Nichols, Editor, Elsevier Publishing Co., Amsterdam 1971.

[5] Bannantine, J.A., J. J. Comer and J. L. Handrock, *Fundamentals of Metal Fatigue Analysis*, Prentice Hall, 1990.

[6] Fuchs, H. O., and R. I. Stephens, *Metal Fatigue in Engineering*, Wiley-Interscience Publications, 1980.

[7] Collins, J. A. *Failure of Materials in Machine Design*, Wiley-Interscience Publications, 1981.

[8] Shigley, J. E., *Mechanical Engineering Design*, McGraw-Hill, 3rd Ed., 1977.

[9] Juvinall, R. C., *Engineering Considerations of Stress, Strain and Strength*, McGraw-Hill, 1967.

[10] Grover, H. J., S. A. Gordon and L. R. Jackson, *Fatigue of Metals and Structures*, NAVWEPS-00-25-534, Government Printing Office, Washington, D.C., 20402, 1954.

SENSORS AND AUTOMATED REASONING

Chair: Dr. Sally Anne McInerny
University of Alabama

ROBUST LASER INTERFEROMETER (RLI) FINDINGS RELATIVE CONDITION MONITORING AND DIAGNOSTICS/PROGNOSTICS ENGINEERING MANAGEMENT

Martin F. Karchnak
Epoch Engineering, Inc.
814 West Diamond Avenue
Suite 105
Gaithersburg, MD, 20878
martyk@epochengineering.com

Andrew J. Hess
Naval Air Systems Command
Patuxent River, MD 20670-1463
hessaj@navair.navy.mil

Theodore Goodenow
2612 Nutcracker Way
Palm City, FL 34990
(561) 219-8684
goobert@gate.net

"Abstract" In recent years, a series of machinery condition and/or machinery health measurement projects have employed a Robust Laser Interferometer (RLI), a non-contact vibration measurement systen introduced for wideband vibration measurements. The specific implementation is described. Wideband (0-262.144 kHz), large dynamic range (up to 180 dB demonstrated in acceleration) measurement data is presented for differing machinery monitoring measurement projects. A discussion is presented regarding not only the availability of high frequency information, but also regarding its value for diagnostics and prognostics. Examples selected include direct spectrum/time series information, as well as envelope processed information.

System advantages are discussed in terms of dynamic effects, environmental effects, and ease of use. Specific examples and measurement history are provided to highlight individual parameters such as bandpass, sensitivity/signal-to-noise, dynamic range, amplitude frequency response, low frequency performance, high frequency performance, and user friendliness in the vibration measurement scenario.

Invited paradigm shifts are discussed, including the availability of improved "fault inception" tracking data and new measurement opportunities. Two implementation design approaches are discussed, point-and-shoot and fiber-optic routing. A summary is presented, including an assessment of growth potential.

Key Words: Diagnostics; measurements; prognostics; vibration

1.0 Introduction

1.1 Concept: In recent years, a series of machinery condition or machinery health measurement projects have employed a Robust Laser Interferometer (RLI), a non-contact vibration measurement system developed by Epoch Engineering, Inc. (EEI). The concept of laser interferometry is illustrated on the left half of Figure 1.0. As noted in this figure, a laser interferometer is conceptually similar to a Doppler radar, comparing a reflected signal from the monitored point with the transmitted signal. The difference in the signals contains the desired information.

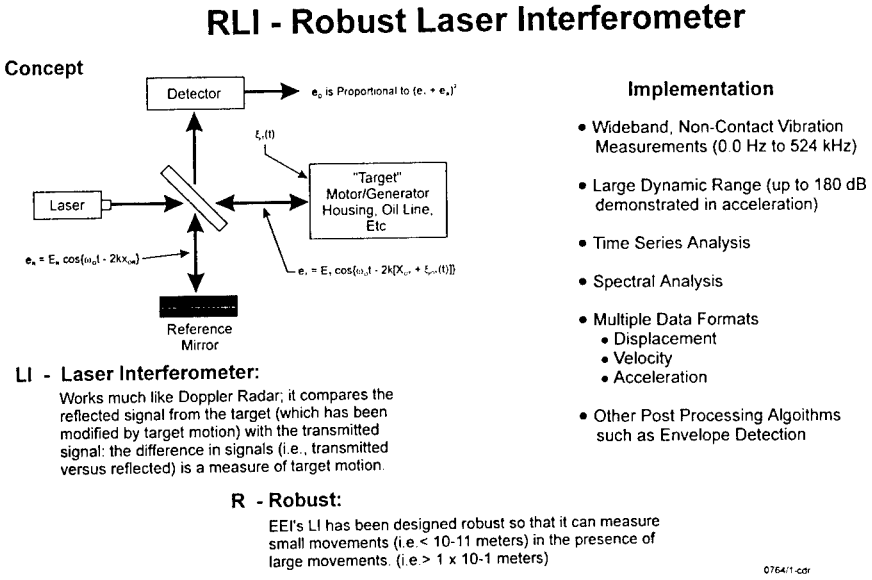


Figure 1.0 System

1.2 Implementation: EEI's laser interferometer has been designed robust so that it can measure small movements (i.e., $<10^{-11}$ meters) in the presence of large movements (i.e., $>1x10^{-1}$ meters). With the exception of the laser system, the RLI is self-contained in a rack-mounted Pentium II based PC. The master clock for data acquisition is a high resolution 10 MHz clock crystal. A Microsoft operating system is used. The laser system is an EEI-modified commercial system. The Laser Interferometer Head (LIH) for the system is connected to its pre-processing and support electronics module by a 20-foot cable, allowing great freedom in positioning the LIH. For test cell measurements, the monitor, keyboard and mouse to control the system were often located remotely, connected by up to 300 ft of CAT-5 cable. The RLI system design provides the following capabilities:

a. Wideband measurements (0.0 Hz to 524 kHz)
b. Large Dynamic Range (up to 180 dB previously demonstrated in acceleration)
c. Time Series Analyses
d. Spectral Analyses
e. Multiple Data Format - (1) Displacement, (2) Velocity, (3) Acceleration
f. Post Processing

2.0 Wideband, Large Dynamic Range Measurements

2.1 Bandwidth: RLI measurement projects have been conducted for the Nuclear Regulatory Commission (NRC), the US Army, the US Navy, and the US Air Force, inter

alia. Examples included in the remainder of this paper were selected from the projects documented in the bibliography for this paper. Figure 2.0 illustrates wideband spectrum examples from a typical RLI rotorcraft gearbox measurement (left side) and a typical turbine engine measurement (right side). The gearbox spectrum, from measurements in a Sikorsky Aircraft test cell, is from 0 Hz to 262.144 kHz. Harmonics of an input gear mesh fundamental are readily visible in the spectrum to well over 200 kHz. Strong sidebands are also evident. The turbine engine spectrum, from Joint Strike Fighter seeded fault testing at Pratt and Whitney, FL. is also shown in Figure 2.0. The measurement point is at the end of an oil line, approximately 17 inches from the turbine engine bearing #1 being monitored. The RLI broadband measurements on this case were also from 0 Hz to 262.144 kHz. Substantial character can be noted in the spectrum, particularly below 50 kHz. In recent measurements associated with rotorcraft hanger bearing health monitoring, RLI broadband data from 0 Hz to 524 kHz was utilized (see reference [1]).

2.2 Dynamic Range: In the course of taking measurements on an S-61 rotorcraft gearbox in a test cell, the RLI has demonstrated the ability to measure a vibration in excess of 30 dB relative to one g, in the presence of an observed noise floor at a lower frequency of about -150 dB relative to one g, as illustrated in the lower right hand corner of Figure 2.0. Other large dynamic range examples have been observed in RLI measurement data [such as the measurement of a 2845.8 Hz main gear mesh vibration of 46.7 g's (33.2 dB rel 1g) and the measurement of the 5.96 Hz second harmonic of the main rotor rotational vibration of 0.000166 g's (-75.6 dB rel 1g) in an observed noise floor at low frequency (about 1 Hertz) of -115 dB rel 1g (0.0000018 g) for an H-53 rotorcraft gearbox test (see reference [2]). Clearly RLI can make the wide bandwidth, large dynamic range, non-contact, vibration measurement.

3.0 Higher Frequency Information

3.1 Direct Spectrum/Time Series: The upper left portion of Figure 3.0 presents a #1 bearing velocity spectrum for a turbine engine at idle, with a seeded fault (indents on the inner race). The measurements were taken at the end of an oil line and were from 0.0 Hz to 262.144 kHz. The frequency band from about 14 kHz to 20 kHz is interesting. The bottom left portion of Figure 3.0, which presents the velocity amplitudes in this frequency band, indicates the presence of impulsive events. Examination of these impulsive events (typically 10 microseconds or less) indicates that they occur in internals of time that are readily related to the seeded fault for the bearing (see reference [3]). It should also be noted that the measurement point was on a very hot operating turbine engine. Contact sensing does not provide correct information on hot surfaces at these frequencies.

3.2 Envelope Processing: The RLI provides envelope processing (see reference [4] - IEEE paper), which is particularly robust as a result of the correct higher frequency measurements made with the RLI. The right side of Figure 3.0 provides an example of RLI envelope processing of bearing measurement data from the oil line measurement point. In reality, there are pages of vibration excitation sources and frequencies in a

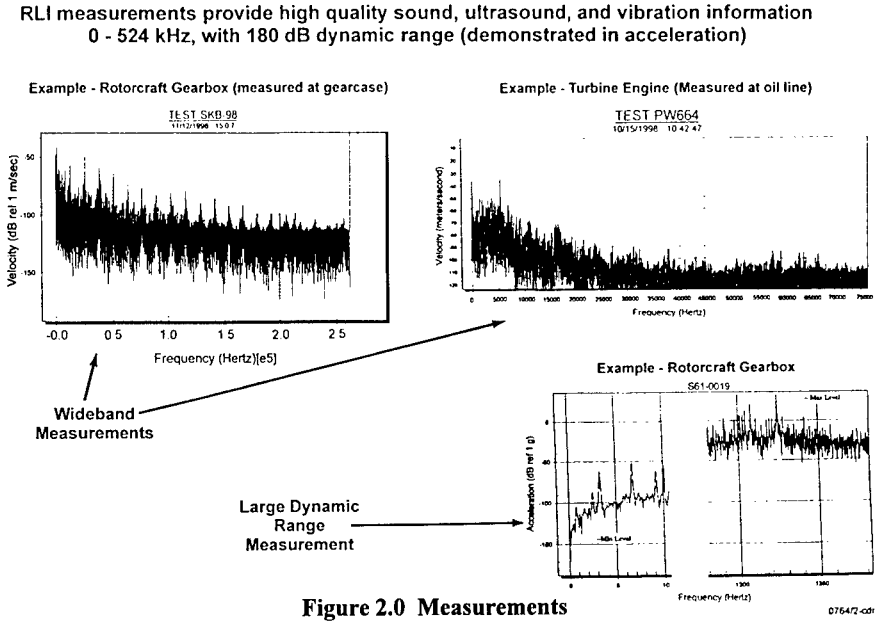


Figure 2.0 Measurements

turbine engine. One example is the first fan rotor, identified as "fs1". The top and bottom portions of the right side of Figure 3.0 illustrates the results of applying the RLI envelope processing for the case where the bearing is under stress. For the top portion, the envelope processing is applied for the frequency range of 26 kHz to 100 kHz; for the bottom portion, the envelope processing is applied for the frequency range of 100 kHz to 200 kHz. Clearly, there is information of value in the higher frequencies measured by the RLI.

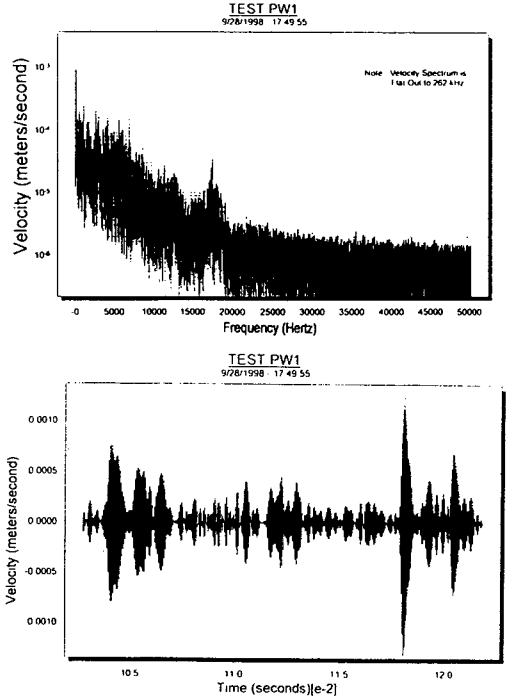
4.0 System Advantages

4.1 Overview: The left side of Figure 4.0 presents a comparison of RLI and contact sensing system considerations. The "concerns" identified as "dynamic" and "environmental" are from the reference [5] article in Sensors Magazine, March 1999. Similarly, the quantitative values for contact sensing for these two categories were also gleaned from the same reference. The right side of Figure 4.0 provides minor elaboration of a few of these comparisons to illustrate typical, practical advantages of RLI for the user ----- beyond the fact that RLI provides correct measurement of amplitude and frequency over a wide bandwidth, with a large dynamic range.

4.2 Examples: The dashed lines on the graph to the right of the arrow marked i) depict the typical frequency response (from specification sheets) for typical contact sensors.

There is information of value in the higher frequencies measured by the RLI

Quality Direct Spectrum and TimeSeries Analysis
[Example - JSF Bearing #1 Seeded Fault]



Quality Envelope Processing
[Example - Broadband Indicated Stress]

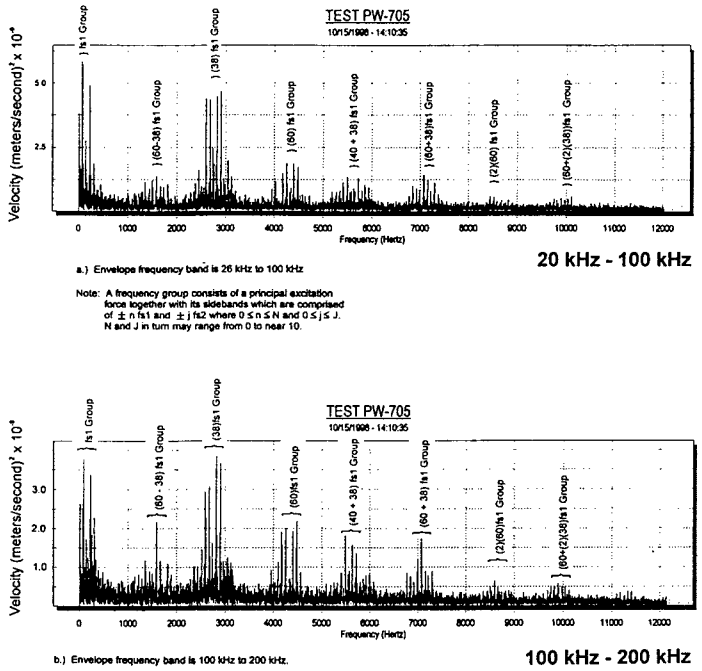


Figure 3.0 Information

076403-cdr

The solid straight line depicts the flat RLI frequency response, alleviating the user of frequency dependent uncertainties and compensation concerns.

The curved lines to the right of the arrow marked ii) depict typical monitoring problems with contact sensing as a function of frequency. Sophisticated and expensive mounting techniques can enable some marginal improvements in performance with increasing frequency, but only non-contact sensing such as that provided by the RLI can alleviate this concern totally.

The dashed lines on the graph to the right of the arrow marked iii) depicts the typical temperature response at the measurement point (from specification sheets) for typical contact sensors. The solid straight line depicts the flat RLI response, alleviating the user of temperature dependent uncertainties and comparison concerns.

For contact sensing at higher frequencies (at frequencies where accelerometer measurements are not even attempted), Acoustic Emission (AE) sensors are sometimes employed. The graph to the right of the arrow marked iv) depicts AE sensor spectrum output when subjected to an input consisting of two frequencies. The RLI, pointed at the AE sensor case, reported only the two input frequencies present in the particular laboratory comparison. Even with a single frequency input, the AE sensors routinely reported a complex spectrum at the output (see reference [6]).

For contact sensing at lower frequencies and approaching 0 Hz, accelerometer systems' performance becomes unreliable. For more complex signals, substantial artifact issues can arise (see reference [2]). The graph to the right of the arrow marked v) indicates the level of difference in reporting low frequency information observed between a quality accelerometer system (the top spectrum) and the RLI (the bottom spectrum. In a number of "simultaneous measurement" situations, RLI was able to report the measurement of the frequency and amplitude of low frequency excitations (such as the main rotor rotational frequency for a rotorcraft -- known to be present from design characteristics). Accelerometer system measurements' noise floor often "masked" such opportunities for accelerometer systems.

Non-contact sensing by the RLI has provided the opportunity to obtain quality vibration information in situations where contact sensors are not practical. The situations to the right of the arrow marked vi) identifies a few where RLI has proved itself to be practical. Clearly RLI has substantial advantages over contact sensing.

5.0 Invited Paradigm Shifts

5.1 Flexibility: The two graphs providing data marked "test 213" on the left side of Figure 5.0 provide a straightforward, simple, example of the value of RLI bandwidth. Both are from the 0-262.144 kHz wide measurement of the turbine engine bearing seeded fault discussed in Section 3.1. The upper one is the time series for the bandwidth of 9 kHz to 15 kHz; the lower one is the time series for the bandwidth of 36 kHz to 56 kHz.

Comparison

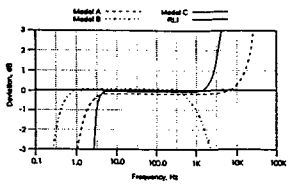
Concern	Typical Contact Transducer System	RLI
Dynamic[1]		
Sensitivity/Signal to Noise	Higher sensitivity leads to larger units	High sensitivity across bandwidth
Dynamic Range	~ 80 dB	180 dB (demonstrated in acceleration)
Amplitude Frequency Response	± 5% to approximately 20% of resonance frequency	Flat
Phase Frequency Response	Time delay[f (frequency)]	Near instantaneous; can be < 1 microsecond
Amplitude Nonlinearity	Approximately 1%; generates artifacts in data	< 0.01%
Transverse Sensitivity	1 - 5% at low frequencies	0% of axial for flat surface
Mass and Size	Attachment, resonance and mass loading issues	N/A for RLI @ measurement point
Environmental[1]		
Temperature	Non-linear as f (temperature)	Flat with temperature
Thermal Transient Sensitivity	Multiple concerns	N/A for RLI
Base Strain Sensitivity	Mounting issue	N/A for RLI
Electromagnetic Sensitivity	Serious concern; often leads to shielding	N/A for RLI @ measurement point
Acoustic Sensitivity	Problem when SPL > 100 dB	N/A for RLI @ measurement point
Radioactive Environments	Materials/life issues	N/A for RLI @ measurement point
Humidity	Leakage paths	N/A for RLI @ measurement point
Grounding/Electrical Isolation	System noise concern	N/A for RLI @ measurement point
Placebo/Noise Monitor Channel	Added complication	N/A for RLI
Other		
Bandpass	Typically 0-20 kHz; more costly units to 40-50 kHz	0-524 kHz
High Frequency Performance	Limited to approximately 40-50 kHz	Linear to 524 kHz+
Low Frequency Performance	Artifacts related to non-linearity	linear to "near dc"
Type Measurement	Contact	Non-contact
Installation Mounting	Degradation as f (frequency)	N/A for RLI
Re-calibration	Required at regular intervals	N/A for RLI
Information Robustness	Artifacts in data at all frequencies; pronounced at low and high frequencies	Extremely robust; correct measurement of both amplitude and frequency

0637-mtw6

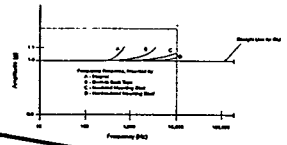
[1] Dynamic and Environmental Effects on Performance Parameters from article by Jon Wilson, *Sensor* Magazine, March 1999

The RLI non-contact measurements are preferable to contact system measurements [such as accelerometer systems and/or acoustic emission sensor system measurements.]

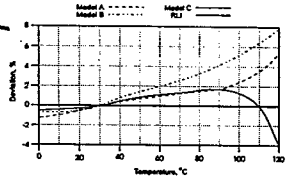
i) Typical Frequency Response



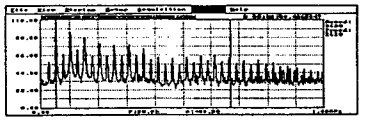
ii) Mounting Impact



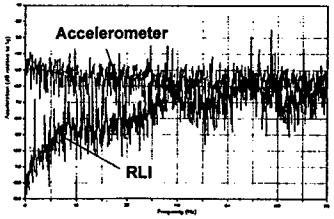
iii) Typical Temperature Response



iv) Gearbox Simulation



v) Accelerometer



vi) Hostile Surface Measurements

- "Hot" Jet Engines
- "High Voltage" or Nuclear Powered Submarine
- "Inaccessable" Hanger Bearings

Figure 4.0 System Advantages

While both contain the same "time series periodicity information", the visibility of the periodicity via the shock pulses at the higher bandwidth is clear. The flexibility of the RLI enables intelligent use of a priori design and related physics considerations to enable proper selection of where to seek information of interest.

5.2 Improved Tracking Data: The two "middle sheet" amplitude tracking graphs on Figure 5.0 represents time series for comparable bandwidth measurements of the same "fault" under two different conditions, two days apart.

The upper one indicates a "nominal "maximum velocity of 0.0002 m/sec; the lower one indicates a "nominal" maximum velocity an order of magnitude higher, namely approximately 0.002 m/sec. With its correct measurement of both amplitude and frequency, the RLI enables precise monitoring of events of interest, such as those associated with fault growth. The upper right side figure is from the Joint Strike Fighter Seeded Fault testing where a particular seeded fault decreased with bearing operating time. From all of the sensors used for the particular seeded fault test series, only the RLI provided information which indicated that the "sharply defined" defect was polished away. This was confirmed by physical inspection (see reference [1]).

5.3 New Opportunities: The bottom of Figure 5.0 provides a time zoom on typical acoustic emissions observed by RLI from the JSF seeded fault testing. [The "wideband" down the middle illustrates the approximate background noise over the measurement bandwidth]. Not only do the individual pulse like events have amplitude up to and greater than 5 times the noise baseline, but they also exhibit a great variation in the characteristics of the individual AE events. The variations extend from a variation in pulse length and shape to variation in the individual frequency content. The availability of this type of RLI data could be expected to provide various "knowledge gains"/opportunities. Clearly, RLI measurements can both enable and seduce increased reliance upon trending and health monitoring diagnostics and "prognostics".

6.0 Future

6.1 Proven Performance - Point and Shoot: The table on the left side of Figure 6.0 illustrates both the application demonstrations and functional demonstrations provided to date with the RLI non-contact "point and shoot" information. Many additional opportunities still exist for validating RLI value added, particularly in niche areas related to gears, bearings, electrical systems, structures, and other physical phenomonologies that exhibit either wideband spectrums or spectrums with large amounts of information concentrated at higher frequencies.

6.2 RLI Fiber Optic System: RLI fiberoptic system routed measurements have also been demonstrated (See reference). Comparison measurements between the baseline "point and shoot" RLI and RLI "fiberoptic routed" measurements have been made for various scenarios, including background noise comparisons, comparative measurements of small vibrations in the presence of large vibrations, comparative measurements of

The RLI measurements are of such quality and robustness that they will both enable and seduce increased reliance upon trending and health monitoring diagnostics and "prognostics"

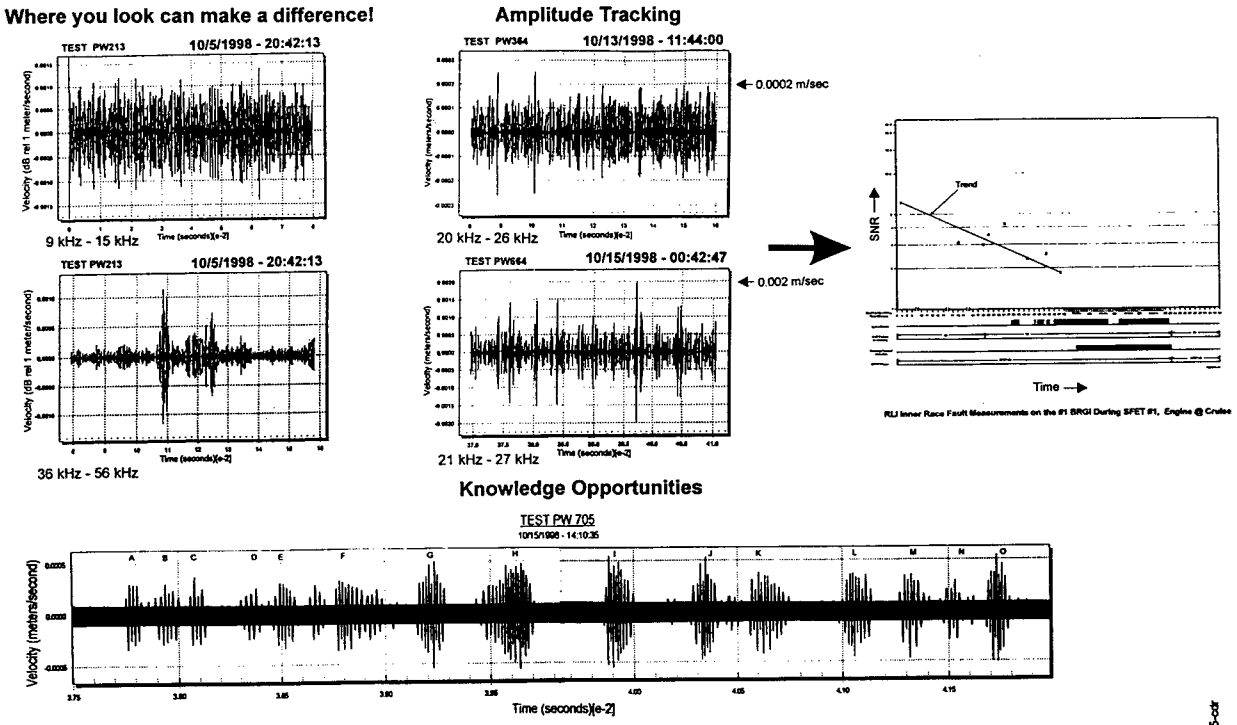


Figure 5.0 Invited Paradigm Shifts

0764/5-cdr

large amplitude complex vibrations, and high frequency measurement comparisons. The data on the right side of Figure 6.0 illustrates one set of data, that being the high frequency (260 kHz to 262 kHz) background noise for the baseline (point and shoot) RLI and a fiber-optic routed RLI measurement. [Demonstration measurements were made through 20 meters of fiber-optic cable]. Clearly RLI can spawn a new generation of sound, ultrasound and vibration measurement systems.

7.0 Summary

Figure 7.0 summarizes RLI findings. RLI also has substantial growth potential. The RLI system is based upon employment of an inexpensive laser, optics, and computer technology --- all technologies with histories of performance growth and cost reduction. [For example, RLI has been in development for about a decade. According to the reference [7] article in the March 2000 Scientific American, "... the average price per megabyte for hard-disk drives plunged from $11.54 in 1988 to $0.04 in 1998, and the estimate for last year is $0.02]." RLI provides a flexible, meaningful capability. As paradigm shifts evolve in health monitoring, diagnostics and prognostics of rotating machinery and other infrastructure components, RLI is positioned to provide multiple stand-alone and/or integrated configurations.

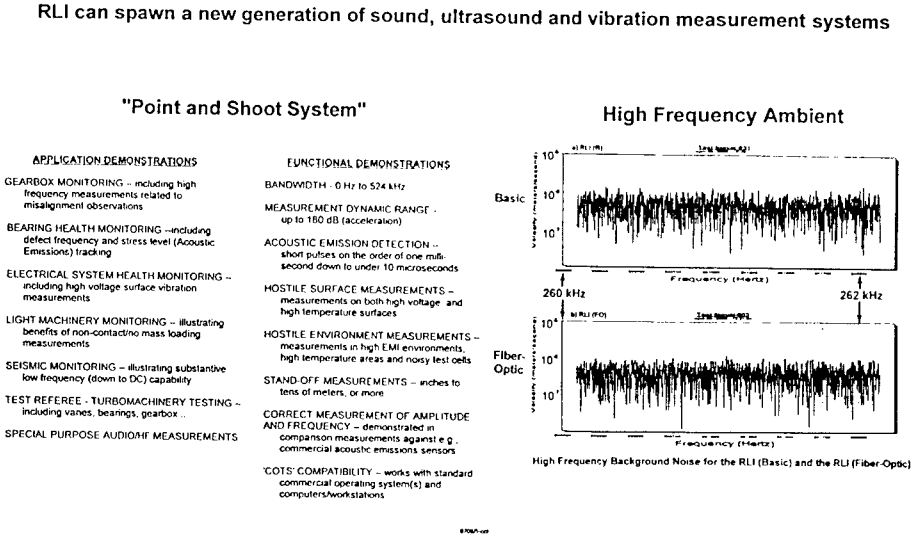


Figure 6.0 Future

Summary

Findings

- RLI measurements provide high quality vibration information 0 - 524 kHz with 180 dB dynamic range (demonstrated in acceleration).
- There is information of value in the higher frequencies measured by RLI.
- PRI measurements are preferable to contact system measurements [such as accelerometer or Acoustic Emission (AE) sensor measurements].
- RLI measurements are of such quality and robustness that they will both enable and seduce increased reliance on trending and health monitoring diagnostics and prognostics.
- RLI can spawn a new generation of sound, ultrasound and vibration measurement devices.

Growth Potential

- System based upon lasers, optics, computational power and memory - all technologies with histories of performance growth and cost reduction.
- Flexible, meaningful capability - paradigm shifts should be anticipated.
- Multiple stand-alone and/or integeated configurations feasible.

076477-ca

Figure 7.0

REFERENCES

1. *The Fiberoptic Laser Interferometer for Helicopter Gearbox Vibration Measurement;* Goodenow, Theodore C., Shipman, Robert L. and Karchnak, Martin F., 2000, Epoch Engineering, Inc. -- Prepared for Naval Air Warfare Center Aircraft Division, Patuxent River, MD 20670-5304 under contract number N00421-98-C-1255

2. *New Laser Acoustic Sensor for Gearbox Vibration Monitoring and Diagnostics,* Goodenow, Theodore C., Shipman, Robert L. and Holland, H.M., 1997, Epoch Engineering, Inc. Prepared for Naval Air Warfare Center Aircraft Division, Patuxent River, MD 20670-5304 under contract number N00421-95-C-1185

3. *Incipient Flaw Detection on an Operating Gas Turbine Engine with the Robust Laser Interferometer,* Goodenow, T.C., Shipman, R.L. and Clark, L.J., 1999, Epoch Engineering, Inc.

4. *Acoustic Emissions in Broadband Vibration as an Indicator of Bearing Stress,* Goodenow, Theodore C., Hardman, William and Karchnak, Martin, 2000, Epoch Engineering, Inc. 2000 IEEE Aerospace Conference presentation

5. *"A Practical Approach to Vibration Detection and Measurement, Part 2: Dynamic and Environmental Effects on Performance,"* Wilson, Jon, The Dynamic Consultant, Sensors, March 1999

6. *Acoustic Fault Detection for Rotorcraft Transmissions and Engines,* Goodenow, Theodore C., Shipman, Robert L., Clark, L.J., et al, 1998-1999, Volumes I & II, Epoch Engineering, Inc. Prepared for Aviation Applied Technology Directorate, U.S. Army Aviation and Missile Command, Fort Eustis, VA 23604-5577 under contract number DAAJ02-97-C-0028

7. *Avoiding A Data Crunch,* "Scientific American," May 2000

APPLICATION OF TORSIONAL VIBRATION MEASUREMENT TO SHAFT CRACK MONITORING IN POWER PLANTS

Ken Maynard, Applied Research Laboratory,
Martin Trethewey and Charles Groover, Dept. of Mechanical Engineering

The Pennsylvania State University
State College, PA 16801

Abstract: The primary goal of the this project was to demonstrate the feasibility of detecting changes in shaft natural frequencies (such as those associated with a shaft crack) on rotating machinery in electric power generation plants using non-contact, non-intrusive measurement methods. During the operation of power plant equipment, torsional natural frequencies are excited by turbulence, friction, and other random forces. This paper primarily addresses the results of field application of non-intrusive torsional vibration sensing to a hydro station and to large induced-draft (ID) fan motors. Testing reaffirmed the potential of this method for diagnostics and prognostics of shafting systems. The first few shaft natural frequencies were visible, and, for the hydro station, correlated well with finite element results (finite element results are not available for the ID fan motors). In addition, several issues related to the development of the non-intrusive transducer were revealed.

Key words: Shaft cracking; condition based maintenance; failure prediction; torsional vibration.

Background: The detection of shaft natural frequencies in the torsional domain requires that the signal resulting from excitation of the rotating elements by turbulence and other random processes is measurable. If measurable, these natural frequencies may be tracked to determine any shifting due to shaft and blade cracking or other phenomena effecting torsional natural frequencies. Difficulties associated with harvesting the potentially very small signals associated with shaft vibration in the torsional domain could render detection infeasible. Thus, transduction and data acquisition must be optimized for dynamic range and signal to noise ratio [1, 2, 3].

The advantage of using shaft torsional natural frequency tracking over shaft lateral natural frequency tracking for detecting cracks in direct-drive machine shafts is twofold:

- A shift in natural frequency for a lateral mode may be caused by anything which changes the boundary conditions between the rotating and stationary elements: seal rubs, changes in bearing film stiffness due to small temperature changes, thermal growth, misalignment, etc. So, if a shaft experiences a shift in lateral natural frequency, it would be difficult to pinpoint the cause as a cracked shaft. However,

none of these boundary conditions influence the torsional natural frequencies. So, one may say that a shift in natural frequency in a torsional mode of the shaft must involve changes in the rotating element itself, such a crack, or perhaps a coupling degradation.

- Similarly, finite element modeling of the rotor is simplified when analyzing for torsional natural frequencies: these boundary conditions, which are so difficult to characterize in rotor translational modes, are near non-existent in the torsional domain for many rotor systems. This means that characterization of the torsional rotordyanamics of a system is much more straightforward, and therefore likely to better facilitate diagnostics.

Detection of the small torsional vibration signals associated with shaft natural frequencies is complicated by transducer imperfections and by machine speed changes. The use of resampling methods has been shown to facilitate the detection of the shaft natural frequencies by: (1) correcting for torsional transduction difficulties [2] resulting from harmonic tape imperfections (printing error and overlap error); and (2) correcting errors as the machine undergoes gradual speed fluctuation [3, 4]. In addition, correction for more dramatic speed changes was addressed in [4]. These corrections made laboratory testing quite feasible.

Transducer setup and methodology: The transducer used to detect the torsional vibration of the shaft included a shaft encoded with black and white stripes, an infrared fiber optic probe, an analog incremental demodulator and an A/D converter. The implementation of the technique under laboratory conditions was previously presented in [2, 3]. Figure 1 shows a schematic of the transducer system.

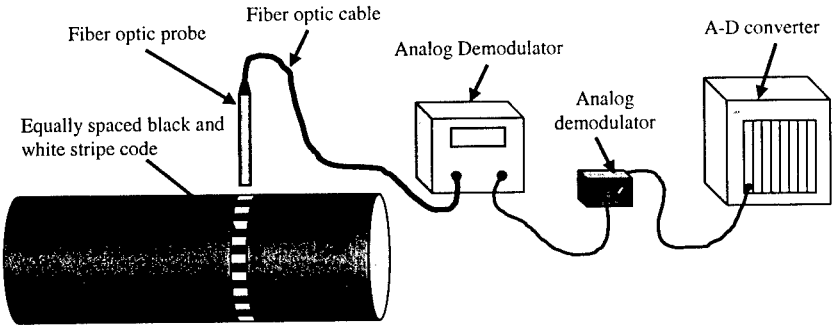


Figure 1: Schematic of transducer setup for torsional vibration measurement

Field implementation: The methodology was implemented on two power plant machines: one a hydroelectric plant turbine generator that has experienced cracking on its newly redesigned turbine rotor; the other a motor on an induced draft (ID) fan at a super-critical coal-fired plant that has experienced cracking of the web-shaft welds.

Hydro turbine: The hydro plant consists of five 3 MW electric turbine generators sets. The plant was originally built in about 1910, but it has recently been redesigned to eliminate an underwater, wooden (lignum vitae) bearing and improve efficiency. The layout of a unit is shown in Figure 2 and Figure 3.

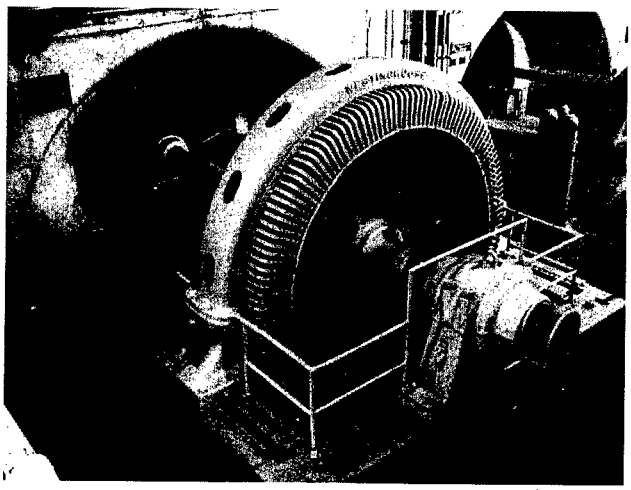


Figure 2: Hydro turbine-generator set

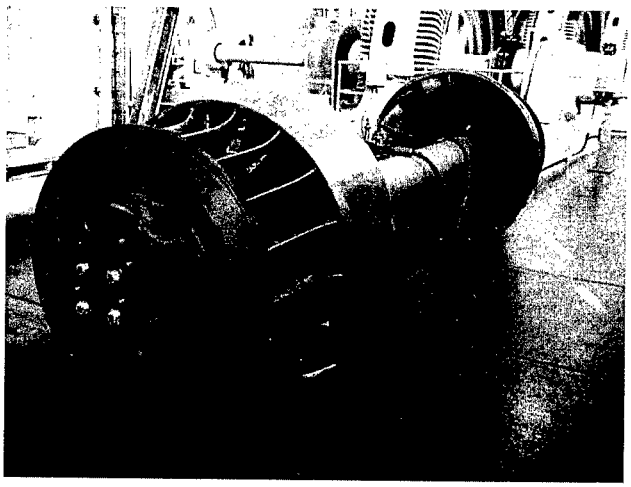


Figure 3: Disassembled Hydroturbine

However, in the last five years, three of the newly designed turbine rotors have experienced severe cracking. Instrumentation and analysis was performed on one of the units that had not experienced cracking to demonstrate the feasibility of detecting shaft

natural frequencies. Figure 4 shows the optic probe, tachometer, and encoded tape placement.

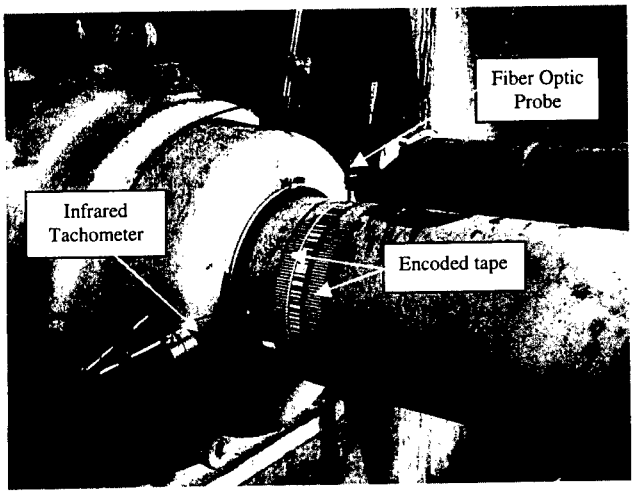


Figure 4: Optic probe, encoded tape, and tachometer placement

The data was analyzed using the double resampling technique [3,4] to eliminate the adverse effects of the presence of running speed and its harmonics on frequency identification. The results of four test runs are shown in Figure 5. Note the peaks at about 16 Hz and 41 Hz. These correspond well to the finite element model torsional frequencies of 16 Hz and 40 Hz.

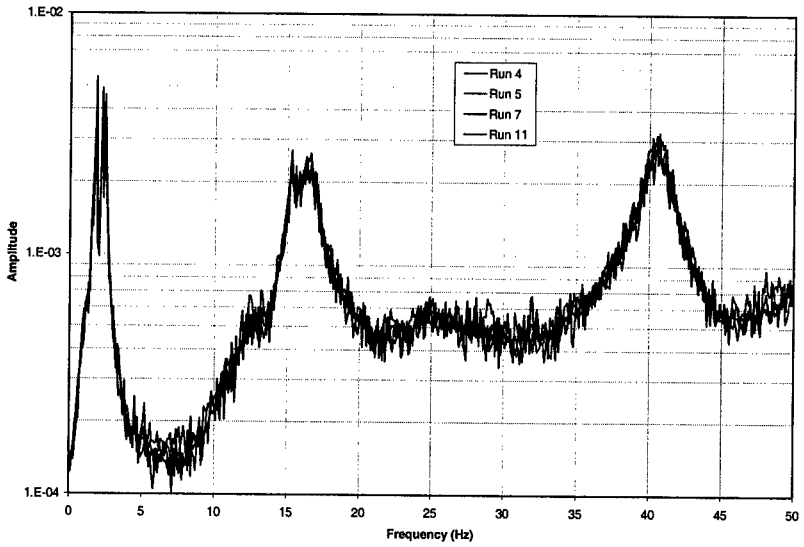


Figure 5: Torsional spectrum of hydro unit shaft motion

The frequencies below 5 Hz are somewhat enigmatic. Since the operating speed of the unit is 300 RPM, or 5 Hz, it was at first assumed that these frequencies correspond to fluid whirl, which generally occurs at speeds between 0.42 and 0.48 times operating speed [8]. However, the shaft lateral vibration data exhibited none of the signs of whirl. In addition, the three closely spaced subsynchronous peaks were stable and repeatable from run to run, as seen in Figure 6. Such stability and repeatability for three closely spaced frequencies does not correspond to the whirl phenomenon. In addition, similar spectral components have since been observed on hydro units at other sites. So, we hypothesize that these subsynchronous frequencies corresponds to the "rigid body" torsional mode on torsional springs corresponding to the bearing film stiffness in shear. Further investigation will be necessary to confirm this and to clarify the significance of these spectral components.

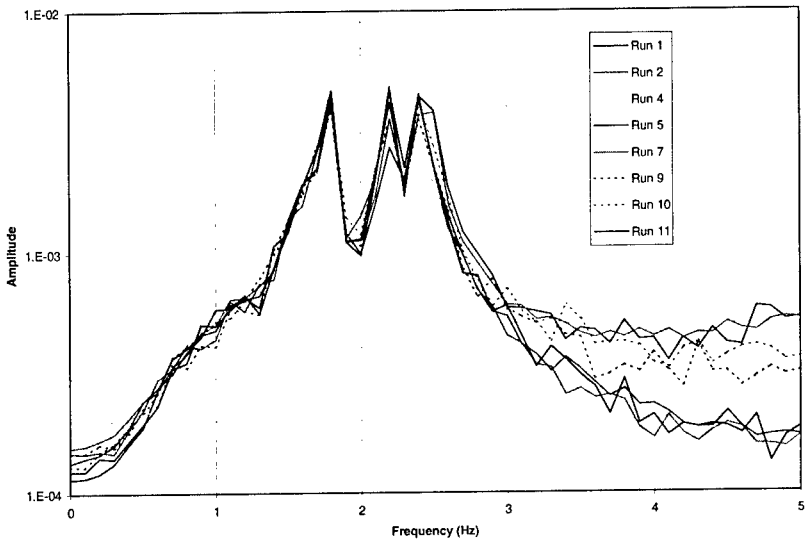


Figure 6: Subsynchronous torsional spectrum for hydro unit shaft

Several issues arose during the on-site data acquisition and analysis. Figure 7 shows some of the data of Figure 1 along with runs that had significant distortion due to tape errors. When the tape was changed, or even the axial location of the transducer was changed on the same tape, the spurious frequencies shifted. These spurious frequencies seem to be related to the encoded tape, and often interfered with the identification of shaft natural frequencies.

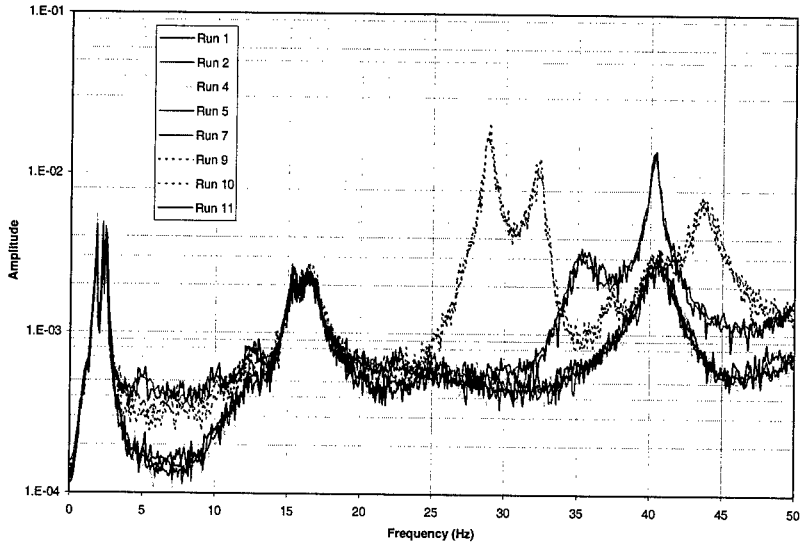


Figure 7: Torsional spectra showing encoded tape error spectral content

ID fan motors: The motors on the fossil-fired induced draft fan were constructed using rectangular cross section webs from the shaft to the rotor coil supports. The square end on the web was then welded to the circular shaft without machining to match the contours. The result has been a number of failures of the motors due to failure of the web welds. Two of these motors were instrumented to detect shaft natural frequencies and establish a baseline to track the changes that may be associated with web weld failure. Figure 8 shows the fan motor.

Figure 8: ID fan motor: (a) Motor housing; (b) scaled with minivan

Installation of the tape was more difficult on the ID fan than on the hydro unit due to the shaft size and the tight quarters. Figure 9 shows the installation of the transducer system.

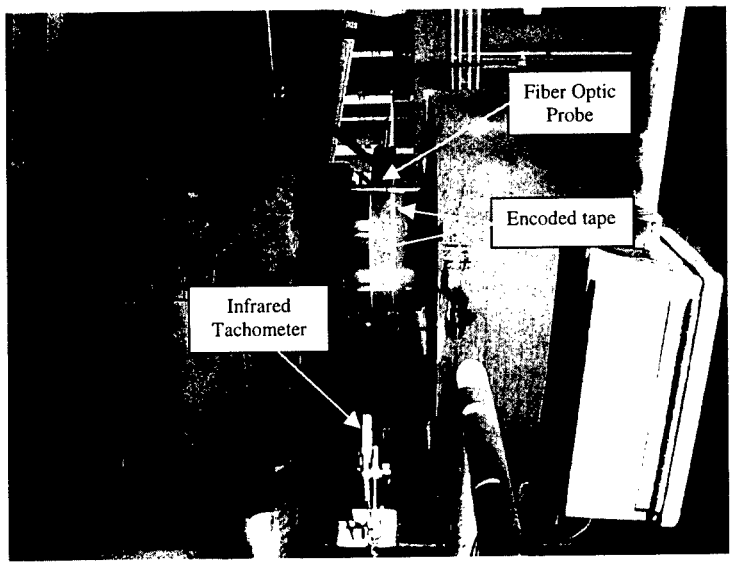


Figure 9: Tape, fiber optic probe, and tachometer installation on ID fan

In addition, the butt joint misalignment of the ends of the tape appeared to be exacerbated by thermal growth of the shaft. It was observed that a space between the ends appeared after heat up of the unit. This underlap, in some cases, caused saturation and malfunction of the analog demodulator. Figure 10 shows the results for one of the fans. The first mode appears to be about 10 Hz. Once again, it was observed that changing tapes or changing the shaft axial position of the optical probe on the encode tape changed some of the spectral content above 20 Hz. It is difficult to assess the remainder of the spectrum with high confidence due to the spectral content of the tape. However, most likely the second and third modes are at about 16 Hz and 19 Hz.

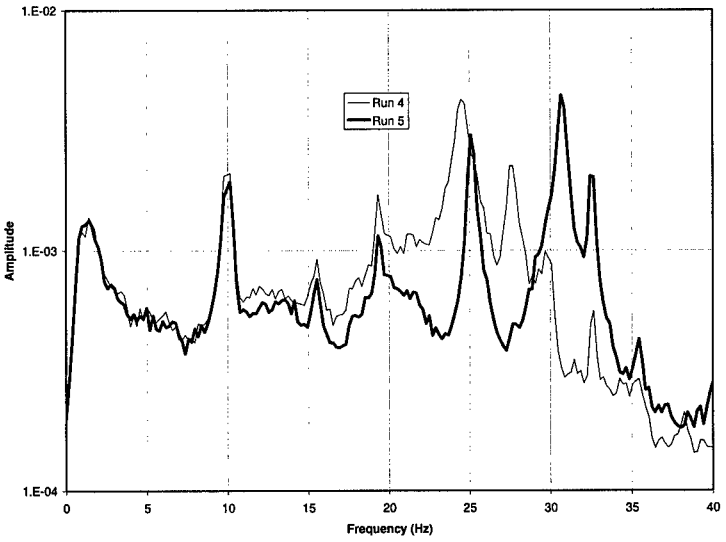


Figure 10: ID Fan motor torsional spectrum

Summary and conclusions: The techniques developed for detecting torsional natural frequencies in the laboratory were implemented on power plant machines that have experienced shaft cracking. The goals of the implementation project were to demonstrate the feasibility of field application, and to establish a baseline for each class of machine. The data acquired clearly demonstrated the feasibility of field implementation, and established baseline natural frequencies.

However, interference from tape related spectral content was experienced. This interference was not experienced in the laboratory due to shaft size, access, and environmental differences. It is believed that this spectral content is associated not with tape printing error or overlap, but was introduced by the installation.

Future work: Correction of the installation errors must be accomplished to remove ambiguity and make the technology widely accessible. This work is currently underway.

Acknowledgement: This work was supported by the Southern Company through the Cooperative Research Agreement *Torsional Vibration and Shaft Twist Measurement in Rotating Machinery* (SCS Contract Number C-98-001172). The content of the information does not necessarily reflect the position or policy of the Government, and no official endorsement should be inferred.

References:

1. Vance, John M., *Rotordynamics of Turbomachinery*, John Wiley & Sons, New York, 1988, pp. 377ff.
2. Maynard, K. P., and Trethewey, M., "On The Feasibility of Blade Crack Detection Through Torsional Vibration Measurements," Proceedings of the 53rd Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, Virginia, April 19-22, 1999, pp. 451-459.
3. Maynard, K. P.; Lebold, M.; Groover, C.; Trethewey, M., Application of Double Resampling to Shaft Torsional Vibration Measurement for the Detection of Blade Natural Frequencies, Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA, pp. 87-94.
4. Groover, Charles Leonard, "Signal Component Removal Applied to the Order Content in Rotating Machinery," Master of Science in Mechanical Engineering Thesis, Penn State University, August 2000.
5. McDonald, D, and Gribler, M., "Digital Resampling: A Viable Alternative for Order Domain Measurements of Rotating Machinery," Proceedings of the 9th Annual International Modal Analysis Conference, Part 2, April 15-18, 1991, Florence, Italy, pp. 1270-1275.
6. Potter, R., "A New Order Tracking Method for Rotating Machinery," Sound and Vibration, September 1990, pp. 30-35.
7. Hernandez, W., Paul, D., and Vosburgh, F, "On-Line Measurement and Tracking of Turbine Torsional Vibration Resonances using a New Encoder Based Rotational Vibration Method (RVM)," SAE Technical Paper 961306, Presented at the Aerospace Atlantic Conference, Dayton, OH, May 22-23, 1996.
8. Sawyer, John W., Ed., *Sawyer's Turbomachinery Maintenance Handbook*, 1st Ed., Vol. II, Turbomachinery International Publications, 1980, p. 7-34ff.

METHODS TO ESTIMATE MACHINE REMAINING USEFUL LIFE USING ARTIFICIAL NEURAL NETWORKS

Magdi A. Essawy, Ph.D.,
Engineering Studies Program, P.O. Box 8047
Georgia Southern University, Statesboro, GA 30460
Phone: 912-486-7583, e-mail: *essawy@GaSou.edu*

Abstract: In this paper, a general methodology for remaining useful life estimation based an indirect methodology is presented. Gearbox failure data, recorded using a mechanical test bed at the Applied Research Laboratory, Penn State University, is used. The machine remaining useful life estimation method used in this paper is indirect method, in the sense that it predicts first the behavior of some system parameters known to be sensitive to the machine operating status, use those predicted values in order to find the predicted machine status through the fuzzy system definitions, and then estimate the remaining useful life by measuring the time from the present time to the time where the death status was detected. Some machine parameters such as temperature, vibration spectrum and level, and acoustic emission, are used in such analysis. Machine operating regions are divided into normal operation, abnormal operation, and no operation or death. Every parameter limits is defined in each region. Prediction models are used to predict the time trajectory of the machine parameters starting from some history measurements. Those predicted trajectories could be used to determine the machine death status point in time. The remaining time to death can be estimated form such models within some appropriate certainty and error tolerance. Neural networks and fuzzy logic system modeling techniques are used for machine parameter prediction due to their known ability for non-linear system modeling, robustness, generalization, and modeling decision uncertainty.

Key Words: Decision making; diagnosis; fuzzy logic; maintenance; neural networks; prediction; prognosis; remaining useful life; vibration analysis.

Introduction: Machine remaining useful life of running machinery is very important information if known within a certain confidence level and tolerance. If machine remaining useful life is known with some certainty and within some acceptable tolerance they can be used in potential system planning [1,2]. That type of planning will lead to more efficient production, less down times, less inventory size, cost saving, and smooth system upgrade. If a machine death time is known within a certain acceptable error limits, an early planning can be made to have a replacement in time, which might lead to a big saving in cost, an appropriate selection for installation time, and avoidance to sudden machine breakdown.

However, prediction is one of the hardest problems to solve especially for non-linear and chaotic systems [3,4]. Most of the real life systems belong to non-linear and chaotic systems. Even though prediction cannot be achieved with high accuracy for such systems, but for very short time in the future, knowing something about the future is important

even if not very accurate. For example, weather forecasting can be achieved with reasonable accuracy only for the coming few days. However, it is also important to predict weather for the next weeks, months and years even with very low certainty and very high prediction error.

Literature has focused some attention in the past few years for finding techniques for estimating machine remaining useful life (RUL) [1,5]. This problem is still in need for some extra efforts in the coming years, to come up with improved models and methods. Most of the RUL estimation methods are based on direct methods that use some history of machine measurements in order to directly estimate the machine remaining useful life or time to death [5]. In this paper, machine RUL will be assumed to be the remaining time to death. And death will be defined as the time when the machine will be no longer useful, which can be due to a major defect in the machine, very low efficiency operation, the machine becoming out-dated, or machine becoming impossible to operate. The machine RUL method presented in this paper is an indirect method that is based on prediction of the future time trajectory of some machine parameters. Those parameters are correlated to the machine different operating status regions. Knowing the correlation of the deviation of some parameters from some nominal value to the machine status, and the parameter predicted deviation in a specific time would lead to good knowledge of the remaining time to reach such operating status. Neural network prediction models in conjunction with some fuzzy logic based decision-making algorithms are use to implement this indirect methodology.

Neural network parameter prediction-models are used due to their ability for non-linear system modeling, and generalization [6,7]. Fuzzy logic operating region-locators are used due to their ability to model uncertainty and continuous logic variables in real world problems [8-10].

Machine Failure Data [11]**:** The gearbox failure data used in this paper are obtained through the Applied Research Laboratory (ARL), at Penn State University. The data was recorded at the ARL using a MDTB (Mechanical Diagnostic Test Bed) that is functionally a motor-driven-train-generator test stand. The gearbox is driven at a set input speed using a 30 Hp, 1750-rpm AC (drive) motor, and the torque is applied by a 75 Hp, 1750 rpm AC (absorption) motor. The maximum speed and torque are 3500 rpm and 225 ft-lbs, respectively. The speed variation is accomplished by varying the frequency to the motor with a digital vector drive unit. A similar vector unit capable of controlling the current output of the absorption motor accomplishes the variation of the torque. The MDTB has the capability of testing single and double reduction industrial gearboxes with ratios from about 1.2:1 to 6:1. The gearboxes are nominally in the 5-20 Hp range. The system is sized to provide the maximum versatility to speed and torque settings. The motors provide about 2 to 5 times the rated torque of the selected gearboxes, and thus the system can provide good overload capability.

Ten accelerometers and an acoustic microphone are placed on the test bed. The microphone, placed in proximity to the test bed, provides a frequency range up to 22 kHz,

which is almost twice the bandwidth of human audible range. A total of 32 thermocouples are available for temperature readings on the MDTB. The highest sampling speed required was 20 kilo Samples (kS)/s for the accelerometers and 44.1 kS/s for the microphone. The thermocouples are sampled at 1 S/s.

Methodology: The machine remaining useful life (RUL) estimation methodology developed in this paper is based on a machine parameter prediction technique along with knowledge about the different operating regions of the machine. In this method, it is assumed that the operating status of a specific machine is reflected into clear changes in a set of its parameters, such as vibration, temperature, current, voltage, power, speed, etc. These define the machine state trajectory in a multidimensional space. Those parameters that are most sensitive to the machine operating status should be selected for the analysis. Mapping of the machine state trajectory to individual two-dimensional subspaces is used in order to simplify the analysis. Certainly, a prior analysis to the machine and its parameters, and their correlation to the change of operating status are needed.

Three operating status regions are assumed Normal Operation (Health), Abnormal Operation (Sickness), and Death (no operation, or non-useful operation), as shown in Figure 1. In reality there is no sharp changes between those regions and the borderlines plotted on the graph are artificial borderlines to approximate the different operating regions. A more realistic representation was developed using fuzzy logic description methods. This representation is illustrated in Figure 2. This fuzzy membership function representation allows easier handling of the terminology, continuous representation of logical functions, smooth transition of status, and easier decision-making process. This fuzzy representation for machine status is an integral part of the machine RUL estimation process developed in this paper. However, the actual measurements are used to tune those fuzzy membership functions to each type of machine separately.

When a machine is in a normal operating status, it is guaranteed that all of its parameters will be bounded in a specific region. This region will be very narrow in the first operating period, and will be in a close proximity with the rated values of the machine parameters. However, it will become wider as the machine becomes older. This transition will happen gradually. Some parameters will increase while others will decrease due to the machine degradation process. Examples for the accelerated degradation trajectories reflected into vibration information measured by accelerometers mounted on the MDTB, or what can be called a machine state transition on two-dimensional maps are shown in Figures 3 and 4. Figure 3 shows the root mean square (rms) value of the machine vibration versus time in seconds, measured during run #10 using accelerometer # 2. And Figure 4 shows the rms value of the machine vibration versus time in seconds, measured during run #10 using accelerometer # 5. These transitions are driven by the actual internal physical changes in the machine, which can take place in any similar active system, such as any rotating machinery. For the same type of machine, some units may experience an increasing trend of some of their parameters while others experience a decreasing trend, due to their unique manufacturing and operating conditions. However, this increase or decrease in itself may not correlate to the machine operating status as long as it is

bounded within certain limits. In other words the relative change in a specific parameter is more indicative of the machine operating status than its absolute value. This is why those operating regions were generated based on the deviation from the baseline value that indicates the machine condition at its birth (when it first came online). Some machine maintenance may also create a sudden shift of the machine parameter trajectory from one operating region to another, such as from the abnormal to the normal region, and needs to be taken into account during the analysis.

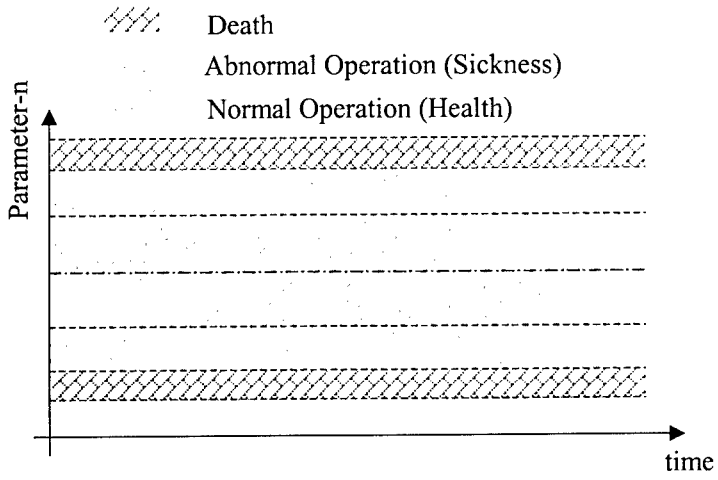


Figure 1. Graphical representation of the definitions for machine operating status regions.

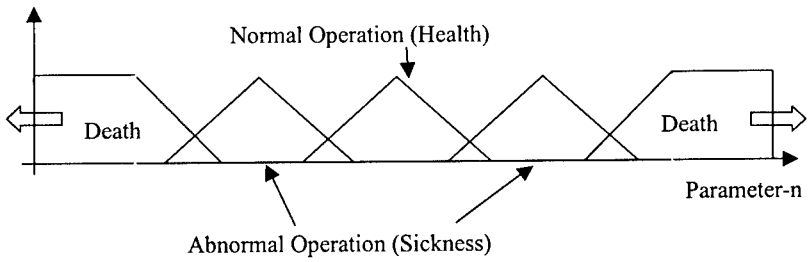


Figure 2. Fuzzy membership function definitions for machine operating status regions.

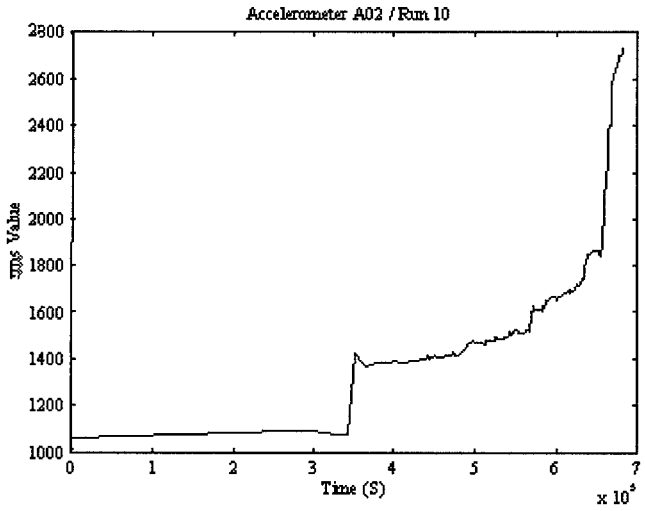


Figure 3. The rms value of the machine vibration versus time in seconds, measured during run #10 using accelerometer # 2.

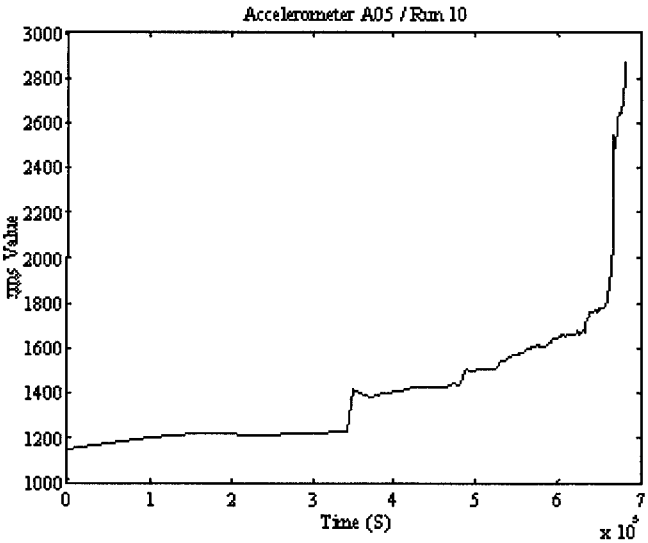


Figure 4. The rms value of the machine vibration versus time in seconds, measured during run #10 using accelerometer # 5.

When a machine parameter transition is recorded in that manner, machine monitoring and diagnosis can be performed using those actual online measurements. Machine monitoring is very useful for many operation and maintenance applications. The remaining useful life estimation is crucial to many other planning and maintenance considerations. The machine remaining useful life estimation will be based not on the measured parameter value but on the predicted parameter value from some measured history data. The state trajectory of the predicted parameter transition will indicate when the machine move from one operating region to another. Now the problem has been simplified to a straightforward prediction problem. Even though prediction of non-linear systems is very hard to achieve, it is hoped that some appropriate prediction models will be built and improved with time. Those models will use some history values in order to predict future values. Linear systems are the easiest to predict, where few history points are enough to predict long time in the future. Unfortunately, linear systems almost do not exist in practice, and prediction problem becomes one of the most challenging problems to solve. Some non-linear systems though are predictable within limits and with variable prediction errors. Chaotic systems, which are a category of nonlinear systems, are not predictable due to their sensitive dependence on initial conditions [3,4,12]. Meaning that a minute change in the initial operating point might lead to a completely different time trajectory, which makes the prediction problem for such systems almost impossible to solve, at least in the time domain.

It is assumed that the systems under discussion in this paper are non-linear and are not chaotic. This means that such systems are predictable within limits and with some prediction error, based on the nature of the system and the prediction method used. The prediction time step is also a factor in the prediction error. Prediction time step is decided based on the nature of the system and the solution method used. Generally, one time step can be predicted with very high accuracy. One time step prediction may give a prediction trajectory that is very comparable to the actual trajectory. However, the prediction extent in that case is very limited, only one time step in the future which might not be very useful in case of prediction of machine remaining useful life. In case of prediction of machine RUL iterative prediction can be used. In the iterative prediction scheme, every predicted point is added to the previous history points as if it was an actual point and used to predict the next future point. If one-step prediction is used, a very small prediction error is expected, but when iterative prediction is used the error is multiplied every time prediction is repeated, which will create a big deviation of the predicted trajectory from the actual trajectory. This deviation is expected to grow more with larger prediction time. Certain confidence level or certainty in the prediction and consequently in the RUL prediction needs to be established. For example, if this model predicts the machine RUL is time (t) then a certainty (C) for that decision needs to be provided to the user, in order for that information to be useful for practical applications. This certainty will be formulated as a function of the accurate prediction probability and the degree of fuzzy membership of the operating status on which the decision was made. The accurate prediction probability will be computed using two methods. The first method assumes a uniform probability distribution, meaning that the one step prediction is achieved with the

same probability anywhere in the operating spectrum. In this case the total iterative prediction probability of accurate prediction can be computed as:

$$P^t = (P^1)^n \tag{1}$$

Where P^t is the total accurate prediction probability and P^1 is the one step accurate prediction probability.

The second method assumes that the distribution of the prediction probability is changing over time, and the total iterative prediction probability can be computed as:

$$P^t = P^1 P^2 \Lambda\ P^n \tag{2}$$

Where $P^1, P^2, \Lambda\ P^n$ are the accurate prediction probabilities at time steps 1, 2, ... n.

The degree of fuzzy membership of any system parameter is estimated using the fuzzy membership function definitions similar to those shown in Figure 1. A simple rule base will be used to decide the operating status of the machine at any point. After plugging the different parameter values into that fuzzy system, a decision will be made about the status of the system. This decision is a fuzzy set, which results from the fuzzy system inferencing process that involves both implication of individual rules and aggregation of the collective rules. Defuzzifying this output, a crisp number reflecting its degree of membership to a specific operating region will be given. That number (Z^{death}), estimated using the fuzzy output membership functions, along with the accurate prediction probability (P^t) defined above will be used to generate a total certainty level in the decision as follows:

$$C = Z^{death} P^t \tag{3}$$

And the machine estimated RUL would be computed as:

$$RUL = n\Delta t \tag{4}$$

Where n is the number of points predicted until a death region was located, and Δt is the prediction time-step.

In addition to the certainty in that decision, an estimated error margin, or tolerance, needs to be provided, and that will be computed as an error bar around the estimated RUL as:

$$RUL = n\Delta t \pm n\Delta t\sigma_c \tag{5}$$

Where σ_c is the estimated standard deviation of the iterative prediction at the current estimation point.

A more conservative estimate can be computed as:

$$RUL = n\Delta t \pm n\Delta t(1-C) \tag{6}$$

And a less conservative or more optimistic estimate can be computed as:

$$RUL = n\Delta t \pm n\Delta t(1-C)\sigma_c \tag{7}$$

Prediction Models: There are several methods in the literature for non-linear system prediction [3,6,8,10,12]. Some of those are based on time series prediction [13]; others are based on multiple input single/multiple output non-linear system modeling [6,8]. Neural network models are the easiest and fastest to build in addition to many other advantages such as robustness, generalization, learning, and model free estimation [6,8,10]. Neural networks are capable of modeling non-linear systems [7,13]. Neural networks were adopted before for time series prediction and multiple input single/multiple-output system modeling [6,7,13]. The neural networks multiple-input single-output models will best suit the problem in hand. Since the RUL estimation deals, most of the time, with dynamic systems and components, dynamic neural network are preferred over static neural networks, for such applications, due to their ability for modeling of system time behavior. Therefore recurrent neural networks are used to predict the machine state trajectory for the problem in hand, especially that this type of neural networks is known to be capable of modeling system time behavior.

Conclusion: A general methodology was developed for estimation of machine remaining useful life using history data. This method is indirect method that starts with defining parameters sensitive to the machine operating regions and transitions, defining some fuzzy operating regions, and then building prediction models for those parameters. If machine parameters time trajectory can then be predicted with some known accuracy, and a fuzzy logic decision making system can detect the machine operating status with some certainty, then an estimate for the machine remaining useful life can be computed with a known certainty and a known tolerance. In this method, neural network models are used to predict the future trajectory of the machine parameters from some measured history data with some estimated probability of success. Neural network models are adopted due to their known ability for modeling non-linear system behavior. However the dynamic neural network models are expected to outperform the static neural models for such applications due to their ability for modeling of system time behavior. The output of those prediction models is plugged into a fuzzy logic decision-making system in order to locate the machine operating regions at any time with some certainty. These methods are tested using practical failure data for gearboxes from machine diagnostic test-bed at the Applied Research Laboratory. Some of those actual failure data have been analyzed, but the prediction models have not been yet fully developed for this type of data and methodology.

References:

1. C. M. Talbott, *"Remaining Life Test for Condition-Based Prognosis,"* Maintenance And Reliability CONference (MARCON 99), Knoxville, Tennessee, May 10-12 1999, pp. 26.01-26.08.
2. C. M. Talbott, *"Prognosis of Remaining Machine Life Based on Condition,"* the 52nd Meeting of the Society for Machinery Failure Prevention Technology (MFPT), Virginia Beach, VA, March 30-April 02, 1998, pp 293-302.
3. M. Casdagli, "Non-linear Prediction of Chaotic Time Series," Physica D 35, pp 335, 1989.
4. G. Sugihara, R. M. May, "Non-linear Forecasting as a way of Distinguishing Chaos from Measurement Error in Time Series," Nature 344, pp 734, 1990.
5. C. M. Talbott, *"Prognosis of Residual Machine Life,"* Maintenance And Reliability CONference (MARCON 98), Knoxville, Tennessee, May 12-14 1998, pp. 29.01-29.10.
6. J. C.Principe, J-M Kuo, "Dynamic Modeling of Chaotic time Series with Neural Networks," Advances in Neural Information Processing Systems 7, Ed. Tesauro, Touretzky, Leen, pp 311-318, 1995.
7. Magdi A. Essawy, and Mohammad Bodruzzaman, *"Neural Network-Based Monitoring and Control of Abnormal Chaotic Behavior in Fluidized Bed Systems,"* the Artificial Neural Networks in Engineering (ANNIE) '97, Nov. 9-12, 1997, pp 647-652.
8. B. Kosko, "Neural Networks and Fuzzy Systems, a Dynamical Systems Approach to Machine Intelligence, " Printce-Hall, Inc, 1992, pp12-36.
9. X. J. Zeng, and M. G. Sengh, "Decomposition Property of Fuzzy Systems and its Applications," IEEE Trans on Fuzzy Systems, Vol 4, No 2, May 1996, pp 149.
10. S R. Jang, "Neuro-Fuzzy Modeling and Control," Proc of the IEEE, March 95.
11. C. S. Byington, J. D. Kozlowski, *"Transitional Data for Estimation of Gearbox Remaining Useful Life,"* MDTB report, Aplllied Research Lab, Penn State University.
12. J.D. Farmer, J. Sidorowich, "Predicting Chaotic Time Series, " Phys. Rev. Lett. 59, pp 845, 1987.
13. Magdi A. Essawy, and Mohammad Bodruzzaman, *"Iterative Prediction of Chaotic Time Series Using a Recurrent Neural Network,"* the Artificial Neural Networks in Engineering (ANNIE) '96, Nov. 10-13, 1996, pp 753-762.

AUTOMATED RECOGNITION OF ADVANCED VIBRATION FEATURES FOR MACHINERY FAULT CLASSIFICATION

Katherine McClintic, Robert Campbell, Gregory Babich,
Amulya Garga, Jeffery Banks, Michael Thurston, Carl Byington

Applied Research Laboratory
The Pennsylvania State University
P.O. Box 30
State College, PA 16804-0030

Abstract: Advanced condition monitoring systems use pattern recognition and automated reasoning on features extracted from sensor data to assess the current health of a component. This paper will evaluate pattern recognition techniques for classifying the "stage of fault" using transitional failure data for commercial grade gearboxes. Features will be extracted from accelerometer data obtained on the Mechanical Diagnostic Testbed (MDTB) at Penn State Applied Research Lab. The ARL CBM Features toolbox, a MATLAB-based toolbox containing most of the traditional HUMS features and several novel features, will be used to perform feature extraction. Several classifiers and training methods will be evaluated, as well as the effect of using different dimension-reduction techniques on classification. The results obtained using the transitional failure data sets will contribute to enhanced health monitoring techniques and improved machinery health prognostic estimates.

Keywords: Classification; gearbox tooth breakage; MDTB; pattern recognition

Introduction: Penn State University Applied Research Laboratory (ARL) is contributing to the diagnostics and prognostics development for aircraft systems using statistical pattern recognition and sensor fusion. Analysis was conducted using a software package developed by ARL called the Shell Enhanced Pattern Recognition Advanced Toolbox (SEPARAT). For this analysis, features were extracted from gearbox run-to-failure accelerometer data acquired on the Mechanical Diagnostics Test Bed (MDTB) at ARL. Based upon borescope ground truth, the data was segmented into three classes: no failure, 1-2 teeth broken, and 2-8 teeth broken. Various classifiers, dimensionality reduction techniques, and training methods were evaluated for their ability to classify stage of fault.

Feature Extraction: In principle, information concerning the relative condition of the monitored machine can be extracted from the vibration signature, and inferences can be made about the health by comparing the vibration signal with previous signals to identify any anomalous conditions that may be occurring. In practice, however, such direct comparisons are not effective mainly due to the large variations between subsequent

signals. Instead, several more useful techniques have been developed over the years that involve feature extraction from the vibration signature [9]. Generally these features are more stable and well behaved than the raw signature data itself. In addition, the features constitute a reduced data set, because one feature value may represent an entire snapshot of data, thus facilitating additional analysis such as pattern recognition for diagnostics and feature tracking for prognostics. Moreover, the use of feature values instead of raw vibration data will become extremely important as wireless applications, with greater bandwidth restrictions, become more widely used.

The feature extraction method may require several steps, depending on the type of feature being calculated. Some features are calculated using the "conditioned" raw signal, while others use a time-synchronous averaged signal that has been filtered to remove the "common" spectral components. ARL developed a CBM Features Toolbox that allows these features to be calculated systematically.

Pattern Recognition Overview: The classification techniques used for this analysis are included in SEPARAT as well as neural networks, Gaussian classifiers, statistical analysis and feature reduction techniques. A discussion of some of the key pattern recognition terminology is provided below.

Feature extraction, as discussed above, is the process of reducing measured signals into feature vectors. Classifier design, also called training, is the process of determining feature space partitions so that unlabeled vectors can be given a class label. Evaluation is the process of testing the design of both the classifier and its inputs. If the evaluation is unsatisfactory, other classifier structures, features and/or attributes, must be sought; otherwise, a satisfactory evaluation indicates the selected attributes, features, and classifier can be incorporated into the application.

Optimal Classification: The goal of pattern classification is to assign a physical object or process to one of c pre-defined classes [1]. The idealized Bayes decision strategy yields a classifier that is optimal (i.e., the classification error rate is minimal). This concept is paramount to pattern classification regardless of the particular technique used. Let $\mathbf{x}$ be a random variable with d-components (features) which obeys the class conditional probability density function $p(\mathbf{x}|\omega_i)$, where ω_i represents one of c possible objects or processes that are of interest, and $P(\omega_i)$ represents the a priori probability that ω_i occurs. The state-conditional a posteriori probability can be expressed by Bayes rule [1]:

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}, \tag{1}$$

where

$$p(\mathbf{x}) = \sum_{i=1}^{c} p(\mathbf{x}|\omega_i)P(\omega_i). \tag{2}$$

The optimal decision rule with the smallest possible error is given as [1]:

Decide that $\mathbf{x}$ belongs to class ω_i iff $P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x})$ for all $j \neq i$. (3)

An equivalent decision rule is given by:

Decide that $\mathbf{x}$ belongs to class ω_i iff $g_i(\mathbf{x}) > g_j(\mathbf{x})$ for all $j \neq i$ (4)

where the discriminants, $g_i(\mathbf{x})$, are defined in the present context of the optimal classifier as

$$g_i(\mathbf{x}) = P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}. \tag{5}$$

The decision boundaries between the classes labeled ω_i and ω_j consist of the points in feature space where $g_i(\mathbf{x}) = g_j(\mathbf{x})$. Decision boundaries partition d-dimensional feature space into the decision regions that are used to classify unlabeled feature vectors.

Discriminant Functions: Because Eq. (4) compares all discriminant function outputs to find the maximum, only the relative values of the outputs are important. Therefore, equivalent changes can be made to each discriminant function without affecting the classification results. In other words, the decision boundaries are not changed. As long as each discriminant function is changed in the same way, the classification results will not be changed. Thus Eq. (5) can be simplified by removing the scaling constant $p(\mathbf{x})$ while giving the same classification results [1]:

$$g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i). \tag{6}$$

The Bayes decision rule defines the lowest possible error rate for a given problem. However, the ideal Bayes approach is not truly practical because it requires a priori knowledge of the distribution and its parameters for each class, which are rarely, if ever, known [1]. In practice, one must either assume class-conditional density models and estimate their parameters or estimate the probability densities, either explicitly or implicitly, from observed data. There are several well-known statistical techniques available. In general, they can be grouped as parametric or nonparametric. Parametric approaches assume that the functional form of the class-conditional density functions, which are described by some parameters, are known. Nonparametric approaches do not assume anything about class-conditional distributions.

Nonparametric Approaches: Some approaches, such as the minimum-distance classifier, have widely been used because they offer intuitive appeal and computational simplicity. Other nonparametric approaches, such as the minimum-squared-error algorithm, use the data to optimize a family of linear discriminant functions. When using nonparametric classifiers, class labels are traditionally assigned based on formula (4).

Minimum distance classifiers are widely referenced throughout the literature [1,2,6,7]. With this type of classifier, unknown feature vectors are assigned the class membership of the nearest sample mean. The discriminant function can be written as

$$\begin{aligned} g_i(\mathbf{x}) &= -(\mathbf{x}-\mathbf{m}_i)^t G_i^{-1}(\mathbf{x}-\mathbf{m}_i) \\ &= -\mathbf{x}^t G_i^{-1}\mathbf{x} + 2\mathbf{x}^t G_i^{-1}\mathbf{x} - \mathbf{m}_i^t G_i^{-1}\mathbf{m}_i \end{aligned} \tag{7}$$

where G_i is a positive definite symmetric weighting matrix. Often the sample covariance matrices are used for G_i; the resulting classifier is quadratic. Another variation of the minimum distance classifier involves using the same weighting matrix is used for each class (i.e., $G_i = G$), which results in a linear discriminant function. Using a common

weighting matrix in Eq. (7), multiplying by ½, and dropping the common quadratic term, the linear minimum-distance discriminant function is obtained:

$$g_i(\mathbf{x}) = \mathbf{x}^t G^{-1}\mathbf{m}_i - \frac{1}{2}\mathbf{m}_i^t G^{-1}\mathbf{m}_i . \tag{8}$$

Two commonly used weighting matrices are the identity matrix and the pooled covariance matrix, which results in the Euclidean and Fisher minimum-distance classifiers, respectively [1,6,7]. Minimum-distance classifiers are trivial to train and implement. Training only requires calculation of the sample means and weighting matrices. Classification only requires calculation of Eq. (7) or Eq. (8) for $i = 1, 2, \ldots, c$, followed by comparisons of the discriminant values (Eq. 4).

The linear classifier is particularly attractive because of its computational simplicity during classification. In some cases, linear discriminant functions arise naturally due to the distribution of the data. The minimum-distance classifier is linear when common weighting matrices are used. The following paragraphs discuss a case were the structure is assumed linear and the weights are found based on that assumption. The general form of the linear discriminant function is given by

$$g_i(\mathbf{x}) = \mathbf{x}^t \mathbf{w}_i + w_{i0}, \tag{9}$$

where $\mathbf{w}_i$ and w_{io} are the weight vector and bias term for the ith class respectively. A family of linear discriminants can be written as

$$\begin{aligned} \mathbf{g}(\mathbf{x}) &= \begin{bmatrix} g_1(\mathbf{x}) & g_2(\mathbf{x}) & \cdots & g_c(\mathbf{x}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{x}^t & 1 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ w_{10} \end{bmatrix} & \begin{bmatrix} \mathbf{w}_2 \\ w_{20} \end{bmatrix} & \cdots & \begin{bmatrix} \mathbf{w}_{c1} \\ w_{r0} \end{bmatrix} \end{bmatrix}, \\ &= \mathbf{y}W \end{aligned} \tag{10}$$

where $\mathbf{y}$ is the 1-by-(d+1) augmented feature vector and W is the (d+1)-by-c weight matrix. One method of training the discriminants is to solve the matrix equation

$$YW = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_c \end{bmatrix} W = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_c \end{bmatrix} = B, \tag{11}$$

where Y is the n_i-by-(d+1) matrix of augmented training samples for the ith class, and B is the corresponding n_i-by-c target matrix where the ith column contains all ones and the other elements are zeros [1]. Equation (11) can be solved using the pseudo-inverse of Y [1,4]:

$$W = Y^{+}B. \tag{12}$$

This approach minimizes the *trace* of squared error matrix $(YW\text{-}B)^t(YW\text{-}B)$; the resulting discriminant functions used in conjunction with (Eq. 4) are called the minimum-squared-error classifiers.

Error Rate Estimation: Choosing which classifier to use for a specific problem is often difficult. To aid in the selection of an appropriate classifier, rigorous analyses can be made to compare the performance of competing designs. In general, error estimation is

accomplished by designing a classifier on training data, labeling test data, and counting the number of errors (mislabeled samples) to estimate the error rate e. Given that $n(\omega_i)$ is the number of samples from the class ω_i incorrectly labeled by the classifier, a typical error estimate is given by

$$e = \sum_{i=1}^{c} \frac{n(\omega_i)}{n_i} P(\omega_i) \tag{13}$$

Several methods can be used to segment available data into training and test sets. When using the resubstitution procedure, the classifier is trained and tested using the same data. This results in an optimistically biased error rate [1,2,6,7]. The resubstitution error rate estimator is good for finding a lower bound on the Bayes error rate [2]. The available data can be split into two mutually exclusive sets for training and testing. This is known as the holdout procedure, which results in an unbiased estimate of the error rate [2]. However, using the holdout procedure requires the data to be segmented either manually or by some clustering algorithm [2]. In many situations, collecting data is very expensive which results in small data sets. One technique that is useful for small data sets is the leave-one-out procedure [1,2,6,7]. In this procedure, all but one of the available training samples is used to train a classifier. The classifier is then tested with the sample that was left out. This process is repeated until all of the available training samples have had their turn as a test sample. The leave-one-out error estimator is nearly unbiased but has a large variance [6,7].

Experimental Facilities: Without a recorded progression to failure, the ability to perform prognostics, the capability to predict remaining useful life, is nearly impossible. This need for high-fidelity transitional data was identified several years ago at ARL. Since then several test beds have been developed to address this shortcoming. The MDTB was created to provide a realistic test stand that effectively represents an operational environment and is able to bridge the chasm between typical university-scale test facilities and real-world applications.

The MDTB was constructed to collect calibrated, transitional data of both gear and bearing failures for commercial gearboxes and transmissions. The MDTB is instrumented with 52 sensors including 31 thermocouples, three internal temperature probes, seven single-axis accelerometers, a tri-axial accelerometer, a microphone, an acoustic emission sensor, an oil analysis sensor, a tachometer, two sets of torque and speed sensors, an infrared (IR) camera, and a borescope. Data are sampled using 16-channel, 16-bit DAQ boards. The sampling rate for the accelerometers is 20 kHz. Ten second snapshots of data are stored in a binary format to a PC. A detailed description of the PSU-ARL MDTB can be found in Reference12.

Data and Class Selection: As stated previously, the data to be used for the current effort should facilitate the development of prognostics by allowing features to be tracked during failure progression.

At the same time, if the data is to be realistically divided into classes, the system health status must be known by some "ground truth" capability. Recognizing this, researchers selected the transitional data from MDTB Run 14 for analysis. Run 14 employed a 3.33-ratio, single-reduction helical gearbox, and the test culminated with eight broken gearteeth. The ground truth for gearbox health is provided by the several borescope images that were obtained (Figure 1). Measurements were made and recorded periodically throughout the run using a variety of torque loads (an excerpt is shown in Figure 1) over the entire accelerated fault evolution. Accelerometer 3 (axial direction) was used for our analysis.

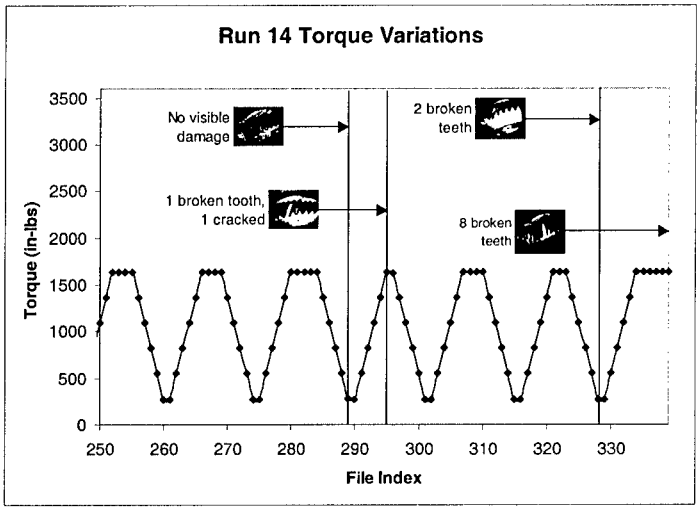


Figure 1: Run 14 Torque Variations, with Borescope Images Showing Gearbox Condition

The MDTB borescope data provides state points on the damage accumulation curve, but does not clearly identify the transition points (the ground truth occurs some time after the actual damage event). The data was divided into three classes: 1) no damage (snapshots 0-295), 2) 1-2 broken teeth (snapshots 296-328), 3) 2-8 broken teeth (snapshots 329-338). Although the faults began to occur before the borescope images were taken, basing the class boundaries solely on the borescope images is probably adequate, given the limited ground truth knowledge.

The features used in this analysis include: FM4, M8A, NA4*, INTR, INTSRC, and INTPK. FM4, M8A, and NA4* are common features that can be readily found in the literature [9,10]; the preprocessing is described in Reference 9. INTR (RMS), INTSRC (standard deviation of the rectified signal) and INTPK (spectrum peak magnitude of output shaft frequency) are features that were calculated on the ARL-developed interstitial signal [11].

Classification Results: Multiple cases were explored including: 1) using only the kurtosis to provide a baseline for the other results, 2) using six advanced features (FM4, NA4*, M8A, INTR, INTSRC, and INTPK) for the entire data set, and 3) using the six advanced features for only the high-torque data. Within each of these categories, a variety of classification and training methods were utilized. The advantage of using the advanced features should be apparent by the improvement in the error rate over the baseline.

Three classification methods are available in the SEPARAT toolbox: Parametric, Nonparametric, and Neural Network. The Nonparametric methods were used for the current investigation because there was not a good fit of the data with the built-in parametric technique (e.g., Gaussian classifier).

Baseline: Classification Using Kurtosis Alone: As can be seen in Table 1, using Kurtosis alone results in very high error rates (greater than 38%). Recall that the resubstitution training method provides a lower bound on the error and, thus, is the optimal result that can be achieved.

Table 1: Classification Errors when using Kurtosis Alone.

Classification Technique	Training/Testing Method	Error Rate
Fisher	Resubstitution	38.83%
MDE	Resubstitution	38.83%
MSE Linear	Resubstitution	66.86%
MSE Quadratic	Resubstitution	66.82%
Quadratic	Resubstitution	38.83%

Classification Using Advanced Features on the Entire Data Set: Results are provided in

Table **2** for simultaneous consideration of the six advanced features (classification using feature fusion). The results shown were obtained using resubstitution training, while the confusion matrix and error rate in Table 3 is for Leave-One-Out (LOO) training and the Minimum-Squared Error (MSE) classifier. We also transformed the six features into two space using a Fisher mapping technique, which yielded results slightly worse than in six space.

Given that resubstitution may be an optimistically biased method, an additional evaluation was performed on the most accurate classifier using LOO training to provide more realistic results. Results from this training method yielded a larger, yet more realistic error than was achieved using resubstitution (Table 3).

Table 2: Classification Errors using the Advanced Features

Classification Technique	Training/Testing Method	Error Rate
Fisher	Resubstitution	10.83%
MDE	Resubstitution	24.92%
MSE Linear	Resubstitution	16.79%
MSE Quadratic	Resubstitution	6.67%
Quadratic	Resubstitution	9.10%

Table 3: Confusion Matrix for MSE Quadratic Classifier using LOO Training Method and Advanced Features (Overall error = 18.63%)

	Class 1	Class 2	Class 3
Class 1	277	0	2
Class 2	0	28	5
Class 3	1	3	6

Feature Reduction: In many instances, multiple features will have a similar information basis and will not add significantly to the classification effectiveness. SEPARAT has the capability of ranking the features in terms of their contribution to class separability by exhaustively enumerating all possible combinations of features and noting which subsets maximize a selected criterion function (e.g., Fisher's criterion). The number of features to be included, r, is then selected by observing the plot and looking for a point where additional features do not seem to increase the criterion function. The vertical line is then dragged to a position between the r^{th} and $r^{th}+1$ points on the curve to choose the first r features within the ranking.

Figure 2 (left) is an output of SEPARAT that shows a ranking of the six features using Fisher's criterion. The figure shows that the use of more than the first three ranked features may not add significantly to the results and, thus, may not significantly impact the results. When using only the three highest ranked features (FM4, NA4*, M8A), the error increased from 6.67% (from fusion of the six features) to 13.58% (fusion of the three features). As the slope of the features in the feature ranking criterion plot approaches zero, the impact of these additional features on the results will diminish.

An additional evaluation was performed to map the six features into two space using the Fisher mapping technique. These results show that the six features can be transformed into two space with a resulting classification error of 10.83% using the LOO training method.

Figure 2 (right) shows the decision boundary results for this evaluation. This figure is useful for visualizing the distance between each classification point and the decision boundaries. In this case, significant overlap occurs between classes 2 and 3, which can be reasonably expected. Because the ground truth used to separate the classes is not

associated with a specific discrete degradation event, one should expect cross-over of classification in the neighborhood of the class boundary.

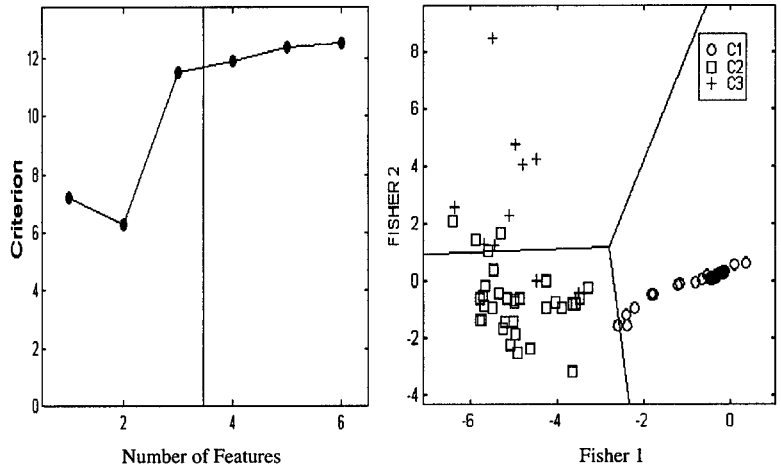


Figure 2: (Left) Feature Reduction Criterion Function. (Right) Decision Boundaries for Mapped Fisher in two-space.

Classification using Advanced Features and High Torque: The classification was repeated with a significantly truncated data set; using only the snapshots associated with high operational torque values. Various classifiers using the six advanced features were used, as well as classification using the Fisher mapping from six to two space. The results obtained using the truncated data set are provided below in Table 4.

Table 4 : Classification Error Rates for High Torque data

	Classification Technique	**Training/Testing Method**	**Error Rate**
Advanced Features	Fisher	Resubstitution	5.28%
	MDE	Resubstitution	21.39%
	MSE Linear	Resubstitution	13.06%
	MSE Quadratic	Resubstitution	11.94%
Fisher Mapping	Fisher	Resubstitution	4.17%

These results show very successful classification when only the high torque data is used for classification: less than 6% error using Fisher classifier and less than 5% error when the features are mapped into two space. These results demonstrate the advantage of identifying operational influences on the features, and then using this information to enhance the classification.

Conclusion: The results of the SEPARAT analysis show that classification performance can be improved by using some advanced diagnostic features and accounting for operational parameter (e.g., torque) changes during the failure progression. The effects of using a reduced set of features on the classification performance were evaluated using feature ranking and reduction methods as well as feature mappings from six space into a reduced feature space. These analysis results show the relative improvements that could potentially be gained by incorporating advanced features and mapping techniques into the classification scheme. The development of an optimal classification scheme would require a more critical selection of diagnostic features than those used herein.

Acknowledgements: This work was supported by the Office of Naval Research under the Accelerated Capabilities Initiative in Human Information Management through a subcontract by CHI Systems, Inc. (CHI-9803-002). The content of the information does not necessarily reflect the position or policy of the Government, and no official endorsement should be inferred.

References:

1. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, NY, 1973.
2. K. Fukunaga, *Statistical Pattern Recognition*, 2[d] ed., Academic Press, San Diego, 1990.
3. J. P. Hoffbeck and D. A. Landgrebe, "Covariance Matrix Estimation and Classification With Limited Training Data," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 18, No. 7, pp. 763-767, July 1996.
4. G. E. Golub and C. H. Van Loan, *Matrix Computations*, 2[d] ed., The Johns Hopkins University Press, Baltimore, 1993.
5. K. Jain and M. D. Ramaswami, "Classifier Design with Parzen Windows," in *Pattern Recognition and Artificial Intelligence*, pp. 211-228, E. S. Gelsema and L. N. Kanal, eds. Elsevier Science Publishers B.V. (North-Holland), 1988.
6. K. Jain, R. C. Dubes, and C. C. Chen, "Bootstrap Techniques for Error Estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-9, NO. 5, pp. 628-633, September 1987.
7. S. J. Raudys and A. K. Jain, "Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No. 3, pp. 252-264, March 1991.
8. E. Parzen, "On Estimation of a Probability Density Function and Mode," *Ann. Math. Stat.*, 33, pp. 1065-1076, September 1962.
9. McClintic, K., et al, Residual and Difference Feature Analysis with Transitional Gearbox Data, 54th Meeting of the MFPT, Virginia, May 2000.
10. Lebold, M., et al, Review of Vibration Analysis Methods for Gearbox Diagnostics and Prognostics, 54[th] Meeting of the MFPT, Virginia, May 2000.
11. Maynard, K.P., *Interstitial Processing: The Application of Noise Processing to Gear Fault Detection*, Proceedings of International Conference on Condition Monitoring, Swansea, UK, 12-15 April 1999.
12. Byington, C.S., Kozlowski, J.D., "Transitional Data for Estimation of Gearbox Remaining Useful Life", 51st Meeting of the Society for Machinery Failure Prevention Technology (MFPT), April 1997.

DEVELOPMENT OF DIAGNOSTIC AND PROGNOSTIC TECHNOLOGIES FOR AEROSPACE HEALTH MANAGEMENT APPLICATIONS

Michael J. Roemer
Gregory J. Kacprzynski
Impact Technologies, LLC
125 Tech Park Drive
Rochester, NY 14623
(716) 424-1990
mike.roemer@impact-tek.com

Emmanuel O. Nwadiogbu
Honeywell Engines and Systems
Phoenix, AZ 85254
(602) 231-7323
emmanuel.nwadiogbu@alliedsignal.com

George Bloor
The Boeing Company
P.O. Box 3707 MC 4A-47
Seattle, WA 98124-2207
(206) 655-9479
george.j.bloor@boeing.com

Abstract: Effective aerospace health management integrates component, subsystem and system level health monitoring strategies, consisting of anomaly/diagnostic/prognostic technologies, with an integrated modeling architecture that addresses failure mode mitigation and life cycle costs. Included within such health management systems will be various failure mode diagnostic and prognostic (D/P) approaches ranging from generic signal processing and experience-based algorithms to the more complex knowledge and model-based techniques. While signal processing and experienced-based approaches to D/P have proven effective in many applications, knowledge and model-based strategies can provide further improvements and are not necessarily more costly to develop or maintain. This paper will describe some generic prognostic and health management technical approaches to confidently diagnose the presence of failure modes or prognose a distribution on remaining time to failure. Specific examples of D/P strategies are presented herein that address valves, hot section lifing and performance degradation of an Auxiliary Power Unit (APU) system. In addition, a model is presented for a Power Take Off (PTO) shaft and AMAD snout bearing.

Keywords: Prognostics, Diagnostics, Aerospace

Introduction: Various health monitoring technologies have been developed for aerospace applications that aid in the detection and classification of developing system faults. However, these technologies have traditionally focussed on fault detection and isolation within an individual subsystem. Health management system developers are just beginning to address the concepts of prognostics and the integration of anomaly, diagnostic and prognostic technologies across subsystems and systems. Hence, the ability to detect and isolate impending faults or to predict the future condition of a component or subsystem based on its current diagnostic state and available operating data is currently a high priority research topic. In addition, these technologies must be capable of communicating the root cause of a problem across subsystems and propagating the up/downstream effects across the health management architecture. This paper will introduce some generic prognostic and health management (PHM) system algorithmic approaches that are demonstrated within various aircraft subsystem components with the ability to predict the time to conditional or mechanical failure (on a real-time basis). Prognostic and health management systems that can effectively implement the capabilities presented herein offer a great opportunity in terms of reducing the overall Life Cycle Costs (LCC) of operating systems as well as decreasing the operations/maintenance logistics footprint.

Generic Diagnostic and Prognostic Technologies: Prognostic and Health Management (PHM) system architectures must allow for the integration of anomaly, diagnostic, and prognostic (A/D/P) technologies from the component level all the way up through the aerospace vehicle level. In general, A/D/P technologies observe features associated with anomalous system behavior and then relates these features to useful information. Before getting into some specific examples of diagnostic and prognostic techniques applied to different aspects of an air vehicle, a brief description of some of the more common technical approaches are given. These generic descriptions will be focussed more on the prognostic algorithm side because less information is currently published in this area than on diagnostics.

Prognostics simply denotes the ability to predict a future condition. Inherently probabilistic or uncertain in nature, prognostics can be applied to system/ component failure modes governed by material condition or by functional loss. Like the diagnostic algorithms, prognostic algorithms can be generic in design but specific in terms of application. This section will briefly describes five approaches to prognostics.

Experienced-Based Prognostics:

In the case where a physical model of a subsystem or component is absent and there is an insufficient sensor network to assess condition, an experienced-based prognostic model may be the only form of prognostics that is practical. This form of prognostic model is the least complex and requires the failure history or "by-design" recommendations of the component under similar operation. Typically, failure and/or inspection data is compiled from legacy systems and a Weibull distribution or other statistical distribution is fitted to the data. An example of these types of distributions is given in Figure 1. Although simplistic, an experienced-based prognostic distribution can be used to drive interval-based maintenance practices that can then be updated on regular intervals. An example may be the maintenance scheduling for an electrical component or airframe component that has little or no sensed parameters and is not critical enough to warrant a physical model. In this case, the prognosis of when the component will fail or degrade to an unacceptable condition must be based solely on analysis of past experience or OEM recommendations. Depending on the maintenance complexity and criticality associated with the component, the prognostics system may be set up for a maintenance interval (i.e. replace every 1000+/-20 EFH) then updated as more data becomes available.

Evolutionary Prognostics:

An evolutionary prognostic approach relies on gauging the proximity and rate of change of the current component condition (i.e. features) to known performance faults. Figure 2 is an illustration of the technique. Evolutionary prognostics may be implemented on systems or subsystems that experience conditional failures such as an APU gas path degradation. Generally, evolutionary prognostics works well for system level degradation because conditional loss is typically the result of interaction of multiple components functioning improperly as a whole. This approach requires that sufficient sensor information is available to assess the current condition of the system or subsystem and relative level of uncertainty in this measurement. Furthermore, the parametric conditions that signify known performance related fault must be identifiable. While a physical model, such as a gas path analysis or control system simulation, is beneficial, it is not a requirement for this technical approach. An alternative to the physical model is built in "expert" knowledge of the fault condition and how it manifests itself in the measured and extracted features.

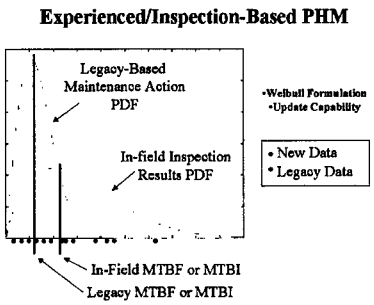


Figure 1 - Experienced-Based Approach

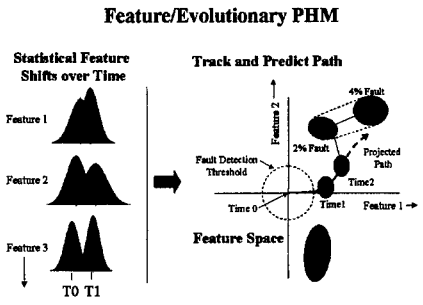


Figure 2 - Evolutionary Prognostics

Feature Progression and AI-Based Prognostics:

Utilizing known transitional or seeded fault/failure degradation paths of measured/extracted feature(s) as they progress over time is another commonly utilized prognostic approach. In this approach, neural networks or other AI techniques are trained on features that progress through a failure. In such cases, the probability of failure as defined by some measure of the "ground truth" which is required as a-priori information. The "ground truth" information that is used to train the predictive network is usually obtained from inspection data. Based on the input features and desired output prediction, the network will automatically adjusts its weights and thresholds based on the relationships it sees between the probability of failure curve and the correlated feature magnitudes. Once trained, the neural network architecture can be used to intelligently predict these same features progressions for a different test under similar operating conditions.

State Estimator Prognostics:

State estimation techniques such as Kalman filters or various other tracking filters can also be implemented as a prognostic technique. In this type of application, the minimization of error between a model and measurement is used to predict future feature behavior. Either fixed or adaptable filter gains can be utilized (Kalman is typically adapted, while Alpha-Beta-Gamma is fixed) within an n^{th}-order state variable vector. For a given measured or extracted feature f, a state vector can be constructed as shown below.

$$x = \begin{bmatrix} f & \dot{f} & \ddot{f} \end{bmatrix}^T \tag{1}$$

Then, the state transition equation is used to update these states based upon a model. A simple Newtonian model of the relationship between the feature position, velocity and acceleration can be used as an example. This simple kinematic equation can be expressed as follows:

$$f(n+1) = f(n) + \dot{f}(n)t + \frac{1}{2}\ddot{f}(n)t^2 \tag{2}$$

where f is again the feature and t is the time period between updates. There is an assumed noise level on the measurements and model related to typical signal-to-noise problems and unmodeled physics. The error covariance associated with the measurement noise vectors is typically developed based on

actual noise variances, while the process noise is assumed based on the kinematic model. In the end, the tracking filter approach is used to track and smooth the features related to predicting a failure.

Physics-Based Prognostics:

A physics-based stochastic model is a technically comprehensive modeling approach that has been traditionally used for component failure mode prognostics. It can be used to evaluate the distribution of remaining useful component life as a function of uncertainties in component strength/stress or condition for a particular fault. The results from such a model can then be used to create a neural network or probabilistic-based autonomous system for real-time failure prognostic predictions. Other information used as input to the prognostic model includes diagnostic results, current condition assessment data and operational profile predictions. This knowledge-rich information can be generated from multi-sensory data fusion combined with in-field experience and maintenance information that can be obtained from data mining processes. While the failure modes may be unique from component to component, the physics-based methodology can remain consistent across the air vehicle. An example of a physical, model-based prognostic technique is shown in Figure 3 for a rotating blade.

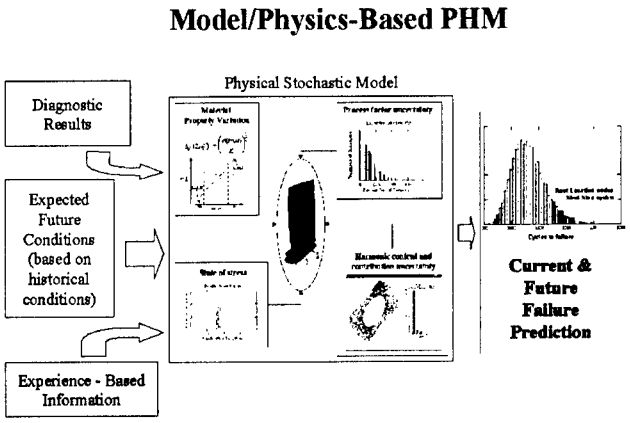


Figure 3 - Physics-Based Prognostics

Relationship to Failure Mode Model: It is important for a prognostic and health management system must to have a direct relationship to a model containing the information on how components, subsystems, and systems interact in operation. In addition, this model should contain information on how the system failure modes, sensors, and health monitoring technologies are related. This is necessary so that failure symptoms and failure propagation can be traced back to root cause failures for fault isolation purposes.

Essentially, information related to the signal and flow relationships between system components, failure modes and across system effects is linked to the sensors and A/D/P algorithms within the system architecture. This captured information is what allows A/D/P algorithms to remain as generic as possible and provides a "place holder" for the algorithms results. In a PHM system implementation, anomalous signals, indicted failure modes, diagnostic monitors or prognostic warning information is

analyzed with this information in order to isolate the root cause of a problem. In addition, the reasoners will utilize this information to prioritize maintenance/operational actions that should be taken to prevent a failure.

An example of how the HM system architecture functions within such a failure mode model representation is given in Figure 4. In this figure, an anomaly detection algorithm (A) monitors four sensors (S). If the anomaly algorithm detects an off nominal condition, then only Failure modes FM1 and FM3 are "flagged" as potential failure modes (FM2 is not a possibility because there is no connectivity within this model).

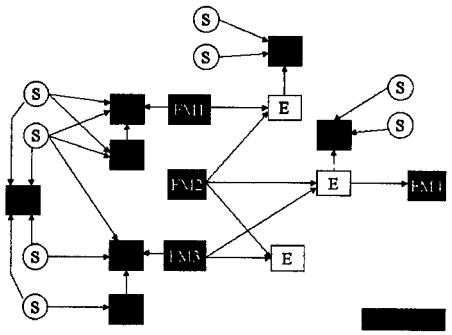


Figure 4 - Generic Representation of Failure Modes, Sensors and HM Technologies

In Figure 4, items to the left of the failure modes (FM's) are health monitoring aspects that attempt to detect the failure modes before they occur. Things to the right of the failure modes are the effects (E) of the failure mode or HM aspects that attempt the isolate which failure mode has already occurred.

Diagnostic Monitors (D) can either function as traditional BIT (Built In Tests) with 0 or 1 outputs denoting that a failure symptom or effect has been observed or they can provide "grayscale" measures of the confidence and severity of a symptom or effect. Continuing with the example, if a Diagnostic Monitor (D) were to observe a symptom of FM3, then the HM reasoner would then have some additional collaborative information to say that FM3 has the higher potential to have occurred. The Prognostic Monitor (P) on FM3 will provide the Mean Time to Failure (MTTF) with confidence bounds for that failure mode.

Let's imagine that Figure 4 represents the failure modes for a rolling element ball bearing. A physics-based prognostic model of the bearing (P) could calculate the current probability of a failure for failure mode (FM3) and project the future probability of failure based on speed and temperature (from sensed parameters) only. However, in this example, imagine that a diagnostic algorithm (D) uses data from a vibration transducer (S) to determine an unbalance or misalignment condition and uses vibration features (spike energy or kurtosis) to detect when significant spalling (FM3) of the outer race has occurred. For the majority of the life of the bearing, the diagnostic algorithms do not make any diagnostic reports and the physics-based prognostic model goes about evaluating remaining useful life. However, when the diagnostic elements diagnose higher than normal unbalance, the prognostic model would utilize this information and determine that life is being accumulated at a faster than expected rate. The HM system reasoners would then be capable of putting together these pieces of evidence to alert the maintainers to examine the bearing at an appropriate time. A prognostic module like this is presented later in this paper.

Now that the concepts for generic HM system technologies and their relationship failure mode models have been introduced, the remainder of the paper will be focussed on some specific model-based diagnostic and prognostic algorithms developed for different aerospace applications. Keep in mind that the output from these dedicated algorithms are processed within a failure mode model by the HM reasoners so that operations and maintenance decisions can be made with knowledge coming from all aspects of the system. A diagnostic algorithm for detecting unhealthy surge control valve operation and performance degradation associated with an APU is presented first. Next, prognostic algorithms for predicting when an APU will reach an EGT (exhaust gas temp.) limit or hot section remaining useful life limit is presented. The final example provides a prognostic model for a PTO (power take-off) shaft and associated bearings.

Surge Control Valve Diagnostics: Diagnostics is often defined as classification of anomalous system behavior to known fault conditions. While generic algorithms are sometimes capable of performing diagnostics, faults must be identified a-priori within an integrated modeling architecture that links anomalous conditions to particular failure modes. Often a more direct approach is to develop a specific fault classifier to diagnose a critical failure mode. Model-based diagnostic approaches are often implemented for these situations and they utilize knowledge (i.e. models) of a given system and compare expected outcomes to measured ones in order to help classify a fault condition.

It is always important to identify the sensory data that can enable a model-based diagnostic algorithm. Due to the fact that the failure mode model allows for a clear vision of inter-system relationships, the data required for diagnostics may already be present if the system is viewed on the whole. When existing or intra-system sensory data can be utilized for shared diagnostic purposes, the benefits of implementing such an approach becomes great and the expense of additional sensors is avoided. That was indeed the case for the surge control valve (SCV) diagnostic approach developed and described next.

In this case, the sensor information from the APU was used to diagnose the health of the surge control valve. Specifically, the response characteristics of the APU speed and EGT sensors after the surge control valve was commanded were used to predict the response time of the SCV and then infer the health condition. Figure 5 shows some of the data that was used to develop a Neural-Fuzzy classifier for diagnosing Surge Control and Load Control Valve sticking in a military fighter Auxiliary Power Unit (APU). To diagnose the health of the SCV or LCV (based on whether they were sticking or failed open/closed) a Neuro-Fuzzy classifier was trained on normal and faulty response characteristics of the APU responses to the valves being commanded. When the valve is sticking, the APU response characteristics are different in a predictable way (i.e. less overshoot in the EGT and speed responses).

A back-propagation neural network was trained on the overshoot levels and associated timing after either valve was commanded. A combination of laboratory results and modeling was used to develop the training data for healthy and faulty valve conditions. Based on these overshoot levels and timing, the valve response time is predicted by the neural network as shown in Figure 6. A healthy response time for the valves was in the range of 0.1 to 0.2 seconds, and anything greater than that would be suspect to sticking. Based on the valve response prediction of the neural network, a fuzzy logic reasoner translated the response time to a health measure of the valve. A value close to 1.0 was considered healthy and values lower than 0.75 would start to indicate a potential valve-sticking situation. It is important to note that the diagnostics achieved in this example was the result of a thorough understanding of the inter-component relationships captured in the modeling environment previously discussed. Also, in this case, prognostics was not feasible because of the highly

unpredictable nature of this failure mode. Hence, this module only diagnosed the health of the valve and did not attempt to predict a MTTF.

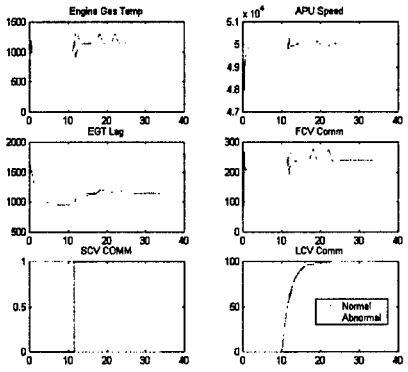


Figure 5(a) - Response in APU Speed and EGT After Surge Control Valve Command

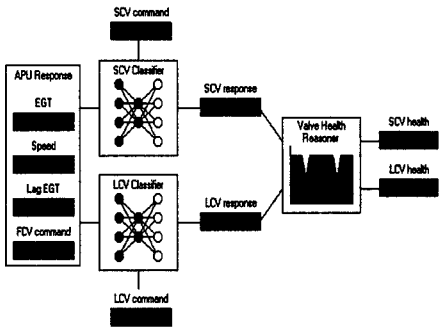


Figure 5(b) - Model-Based SCV Fault Diagnosis

APU Performance Diagnostics and Prognostics: The next example was specifically developed for monitoring the performance degradation of an APU and includes both a diagnostic and prognostic component. The combined algorithm is probabilistic in nature and utilizes statistically significant shifts in key APU performance parameters to diagnose a current level of degradation and then performs a multi-parameter, exponentially weighted projection to predict future degradation. This technique would be considered an evolutionary prognostic technique.

This feature-based diagnostic and prognostic approach relies on gauging the proximity and rate of change of the current system condition to known performance faults. This multi-parameter, evolutionary technique has already been shown to be capable of identifying degraded performance in propulsion systems (Roemer and Ghiocel, 1998).

The process involves assigning non-normal or normal Probability Density Functions (PDF's) to performance error patterns associated to known faults in N-dimensional space. Similarly, the current error exists as a PDF in the parameter space as well. The probability that the current condition (C, measured parameter shifts), may be attributed to a given fault (F, identified known fault conditions) is determined by the "overlap" (i.e. multi-dimensional integration) of their respective joint probability density functions. Figure 2 showed how this is done in 2-dimensional parameter space. If C and F can be assumed to be normally distributed (not a necessary assumption however), the probability of association (Pa) with a given fault condition F can be found using:

$$p_a = 2\Phi(-\frac{\overline{F-C}}{\sqrt{\sigma_f^2+\sigma_c^2}}) = 2\Phi(-\beta) \qquad (3)$$

where:

$\overline{F}, \overline{C}$ = the mean of the distributions F and C respectively

σ_f, σ_c = the standard deviation of the F and C distributions

The function $\Phi(\ )$ is the standard normal cumulative distribution and the β is denoted as the reliability index. The β represents the Euclidean distance between the current conditional distribution (C) and a given fault distribution (F). Hence, this approach performs diagnostics by evaluating the likelihood of the current conditions to known fault conditions and prognostics by extrapolating a fault-weighted, evolutionary path.

The evolutionary prognotics approach was applied to APU degradation data on an Auxilary Power Unit for a military fighter aircraft. APU model simulations of performance degradation were utilized along with test cell data to identify known parameter shifts to particular performance faults. Incremental effeciency degradation of the turbine and compressor sections, which simulated the effects of seal leakage or fouling, were used to build the fault paths for compressor and turbine degradation in 5-dimensional feature space. This feature space was defined by deviations from the normal parametric curves for Compressor Discharge Pressure and Temperature, Exhaust Gas Temp, Speed, and Fuel Flow which are all derivable from information on the aircraft's data bus.

The results of the evolutionary diagnostic and prognostic algorithm is shown in Figures 6a and b. At each time interval, the degree of overlap of the current error pattern is compared with compressor and turbine efficiency faults. In this case, the fault initially looks equally like a turbine or compressor efficiency fault but then continues to evolve to confidently identify a compressor fault. The blue bars in the left chart of Figure 6 represent the confidence of identifying a compressor degradation and the red bars represent the confidence associated with a turbine degradation. The 3-D plot on the right of Figure 8 represents the mean shifts of all three parameters (CDT, N1 and CDP) as a function of the APU degradation. The red line shows that the actual degradation is moving along side the model-based prediction of compressor degradation, hence giving it a higher confidence.

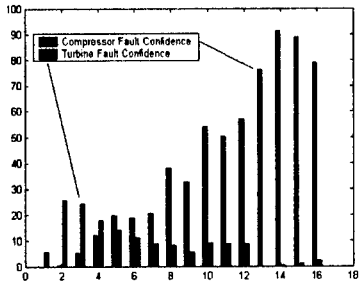


Figure 6(a) - Confidence Level in APU Compressor and Turbine Diagnosis

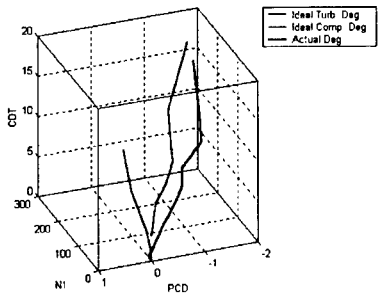


Figure 6(b) - Mean Degredation Path for Turbine and Compressor Faults

Figure 7 illustrates the concept for the prognostics algorithm in terms of what is important to predict associated with APU degradation. From an O&M point of view, any type of degradation that will lead to grounding the aircraft or putting the aircraft in danger is important to prognose. In this case, reaching an EGT limit that will prevent a main engine start (MES) will keep the aircraft on the ground and reduced air bleed from the compressor that effects the avionics cooling is also a concern. Predicting

system-wide functional failure as opposed to just isolated LRU failures raises the standard for intergrated Health Management systems. Note that the expected to time to reach these critical events is displayed along side the actual time predicted to reach the event. In this case, these are the mean number of flight hours to No Main Engine Start (MES) and the point at which Avionics would overheat due to insufficient airflow. In Figure 7, the difference between the expected and actual MTTF (no MES) is shown as 231 APU operating hours. A threshold is typically set up a priori to alert maintenance personnel that accelerated degradation is occurring in the APU which will result in an EGT limit that would prevent the main engine from starting.

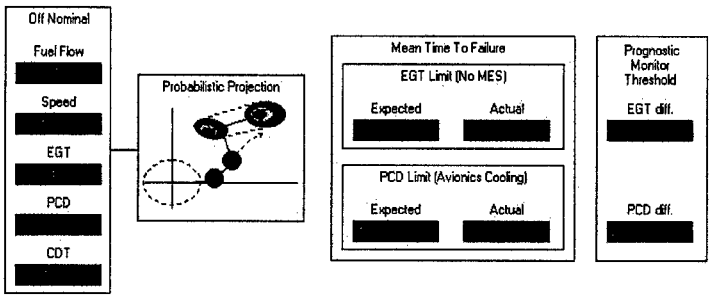


Figure 7 - Demonstration of Evolutionary Prognsotic Output

Power Take-Off Shaft Prognostics: Figure 8 shows a model-based prognsotic concept for a Power Take Off (PTO) shaft and AMAD snout bearing. In this prognostic module development, the first step is to relate processed per-rev vibration signals to different levels of PTO unbalance or misalignment. This inference (from measured vibration to unbalance/misalignment) is performed based on rigorous testing of vibrations measured on an AMAD gearbox under different levels of unbalance and misalignment on the PTO shaft. During the testing, different levels of unbalance were applied to a military fighter aircraft PTO shaft from 0.01 to 0.10 oz-in. Based on incremental levels of unbalance, the 1X vibration amplitudes were monitored and related to the level of unbalance. Some of the results from this testing are shown in Figure 9, with vibration spectrums associated with 0.01 and 0.05 oz-in unbalances shown. A simple back-propagation neural network was trained to represent this non-linear relationship between the 1X amplitudes and the unbalance level in oz-in.

Next, the unbalance prediction was applied to the rotordynamics (critical speed) model of the PTO staft including bearing stiffnesses. This model was run off-line under several different unbalance scenarios to obtain model-based estimates of the associated radial forces on the bearings under these conditions. A look-up table was then developed from which the real-time prediction of bearing radial forces are defined as a function of measured shaft unbalance (inferred from the 1X amplitudes). Based on these real-time predicted radial forces, the bearing model was utilized to assess the remaining B-10 life. A plot of the results from the bearing model are shown in Figure 10. From this figure, the B-10 bearing life can be accumulated based on the current level of unbalance force. The prognostic aspect of this model simply regresses the predicted radial bearing forces and applies them to the model to predict when the useful life of the bearing will be used up. Finally, the difference between the expected MTTF (under normal loading) was compared to the actual MTTF (predicted based on the actual operating conditions) to trigger an alarm if this difference becomes to large. In the case shown in Figure 8, the actual MTTF of the snout bearing was predicted to be 37,843 hours and the expected MTTF is 43,000

hours, resulting in a difference of over 5000 hours. This difference has triggered the PTOS prognostic monitor to report to the HM architecture discussed previously. Note, the expected and actual MTTF's are distributions as shown in the figure. Based on the severe consequences of a PTOS failure, only a small risk in the confidence interval of the distribution is acceptable and therefore the threshold of 5000 hours was choosen.

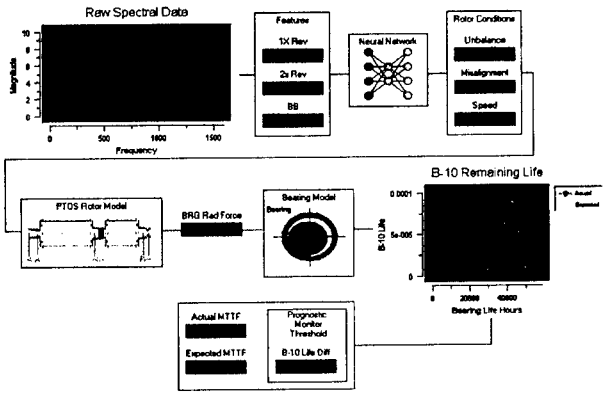


Figure 8 - Model-Based Prognostics for PTO Shaft and Bearings

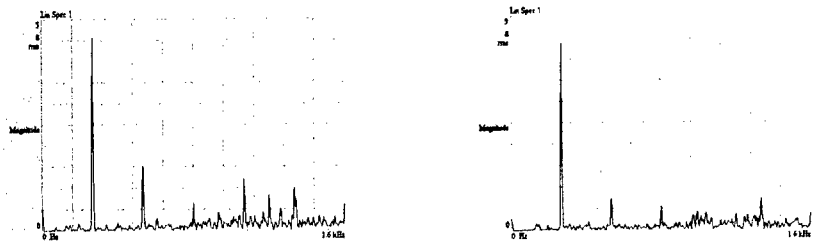


Figure 9 - Vibration Spectrums for 0.01 and 0.05 oz-in PTO Shaft Unbalance

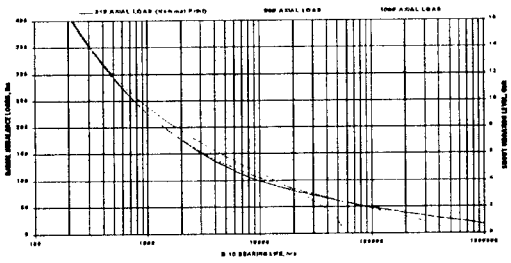


Figure 10 - B-10 Bearing Life Predictions for Various Radial Loads

APU Hot Section Prognostics: As a final example, a prognostic module was developed for the hot section blading for a military fighter APU. This hot section prognostics module is a simple extension of the OEM's life usage monitor for this APU. First, the module used the structural model to determine the creep fatigue damage done to the blading under various APU operating modes such as Main Engine Start, Landing Gear, Weapons Loading etc., as a function of speed, Turbine Inlet Temp (TIT) and Exhaust Gas Temperature (EGT). The OEM has developed event specific operating profiles that the APU is expected to see over the life of the APU. If this operating profile is adhered to, this Hot Section prognostic module will produce an actual MTTF probability distribution that is the same as the Expected MTTF; however, this is rarely the case. A particular APU may experience more Main Engine Starts (MES) than expected or have higher than expected TIT and therefore accumulate damage at a higher than expected rate. The inherent uncertainty in the remaining life is a function of many characteristics ranging from the material properties to the operating profile. Hence, the predicted probability of failure at some future time is performed by statistically sampling from these uncertainties in a probabilistic analysis.

The fundamental concept of this, and other "lifing-based" prognostic modules, is that a maintainer or auto-logistics system can simulate a future mission profile (based on past operating statistics), look at a risk-sorted list of components with potential failure modes for the mission, and be aware of all downstream effects if a failure does occur. In this manner, informed maintenance decisions may be made and aircraft readiness and availability may be assessed.

The implementation of this APU hot section is shown in Figure 11. As illustrated, the prognostic module utilizes the speed, EGT and calculated TIT as inputs to the model. As in the case of the PTO shaft model, the hot section model was run off-line under various design and off-design scenarios and a look-up table generated that related APU events at specific operating conditions to accumulated damage (in this case stress rupture due to creep). Based on this table and the monitored operating conditions, current levels of damage are continuously evaluated. The prognostic aspect utilizes the distribution of APU events specific to this APU to project future usage rates. As in the case of the other prognostic modules, this output produces two distributions on MTTF, one for the expected normal operation and one based on the actual operation. If this particular APU begins to experience high levels of hot section life, than the prognostic monitor threshold will warn the maintainers of significant differences between the actual and expected MTTF's.

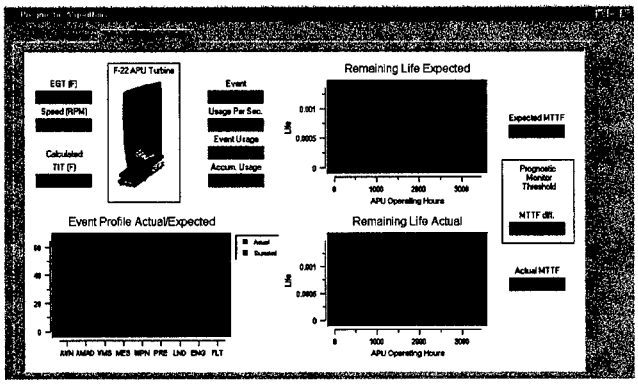


Figure 11 - APU Hot Section Prognostics Module

Conclusions:

This paper has discussed many concepts associated with next generation prognostic and health management systems for aerospace applications. First, a generic PHM architecture was presented that emphasized the integration of anomaly detection, diagnostic and prognostic reasoners through the use of an integrated model of the entire system. Next, direct links from the anomaly, diagnostic and prognostic algorithms were identified that trigger specific attributes in the integrated model for superior fault isolation and reasoning. With this type of overall algorithm integration across the system, system level reasoners can process collaborative evidence about particular failure modes more effectively so as to reduce the need for additional sensors. A few specific diagnostic and prognostic algorithmic approaches were also presented to illustrate how the results from these techniques are process by the integrated model and HM architecture.

References:

[1] Roemer, M. J. and Kacprzynski, G.J., "Advanced Diagnostics and Prognostics for Gas Turbine Engine Risk Assessment," Paper 2000-GT-30, ASME and IGTI Turbo Expo 2000, Munich, Germany, May 2000.

[2] Roemer, M. J., and Ghiocel, D. M., "A Probabilistic Approach to the Diagnosis of Gas Turbine Engine Faults" Paper 99-GT-363, ASME and IGTI Turbo Expo 1999, Indianapolis, Indiana, June 1999.

[3] Roemer, M. J., and Atkinson, B., "Real-Time Engine Health Monitoring and Diagnostics for Gas Turbine Engines," Paper 97-GT-30, ASME and IGTI Turbo Expo 1997, Orlando, Florida, June 1997.

Acknowledgments:

We would like to thank Larry Howard and Gabor Karsai of ISIS for their contributions to the failure mode model development. Also, many thanks to Robin Smith at Boeing for supporting data for validation and demonstration of some of these models.

TOTAL OWNERSHIP COST

Chair: **Dr. Kam W. Ng**
Office of Naval Research

Enhanced FMECA: Integrating Health Management Design and Traditional Failure Analysis

Greg Kacprzynski
Michael J. Roemer

Impact Technologies, LLC
125 Tech Park Drive
Rochester, NY 14623

Carl Byington
Rob Campbell
Pennsylvania State University
Applied Research Laboratory
P.O. Box 30 (North Atherton Street)
State College, PA 16804-0030

Abstract: DOD acquisition programs have recognized that operating and support costs dominate the total life cycle costs of complex military systems, and therefore should be considered up front in the design process. In order to estimate operating costs, which are predominately related to maintenance costs, a 'view' of the conceptual design must exist that can be used to evaluate the effects of system design variables upon maintenance requirements. This view is currently best embodied in the Failure Modes, Effects, and Criticality Analyses (FMECA).

Additionally, many DOD acquisition programs are interested in designing health management systems through the optimal application of system diagnostic and prognostic techniques to produce substantial safety and life cycle cost benefits. To achieve these benefits, a more systematic and accurate method to evaluate candidate health monitoring approaches during the design process must be incorporated. While the FMECA is a keystone of the maintenance planning process, it has limitations in estimating the impact of Condition-Based Maintenance (CBM) implementation on life cycle costs. CBM technology deals not just with failures, but also with monitoring the progression towards failure through detection, diagnosis, and prognosis. If we are to evaluate maintenance efforts and diagnostic/prognostic technology design choices, then the failure modes must be defined in a way that deals with incipient and evolving failures. Hence, the current paper discusses the development of a tool called FMECA++© for use by designers and end users that addresses these issues and helps to collaboratively design the optimal health management solutions for complex machinery from a cost benefit and/or availability standpoint.

We discuss the processing concept of the FMECA++© and introduce methods to optimize the expanded failure mode analysis, health management metrics, and maintainability/availability considerations. A detailed example of a health management analysis is also provided.

Key Words: FMECA, diagnostics, prognostics, health management, cost/benefit, availability

Introduction: The application of health monitoring systems serves to increase the overall reliability of a system through judicious application of intelligent condition monitoring technologies. A consistent health management philosophy integrates the results from the health monitoring system for the purposes of optimizing operations and maintenance practices through, 1.) Prediction, with confidence bounds, of the Remaining Useful Life (RUL) of critical components, and 2.) Isolating the root cause of failures after the failure effects have been observed. If RUL predictions can be made, the allocation of replacement parts or refurbishment

maintenance logistic footprints. Fault isolation is a critical component to maximizing system availability and minimizing downtime through more efficient troubleshooting efforts.

Because of its potential impact, health monitoring and management solutions should be considered during the initial design of a system. For example, implementing a health monitoring technology (defined here as the combination of sensors and algorithms) that is capable of detecting a crack in a rotating part before it gets to a critical size, may allow for a less conservative factor of safety resulting in a cheaper and lighter design that would be too risky if health monitoring was not utilized. This link between the health management system design and the overall system design is shown in Figure 1.

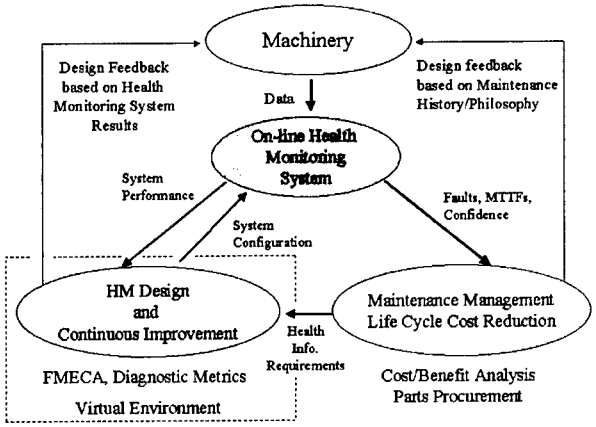


Figure 1 - Health Management with System Design

In this figure, health management system design is shown within the dotted line depicted as a "Virtual Environment". The concept illustrated allows the health management system designer to influence the "top level" system design (shown as "machinery") and assess the downstream availability and life cycle costs associated with the "whole" system including its health management. The final availability and overall life cycle cost relationships must be estimated based on the potential designs offered and an optimization performed based on the design trade-offs.

Because an initial system FMECA is performed during the design stage, it is a perfect link the critical overall system failure modes and the health management system that is designed to help mitigate these failure modes. Hence, a process will be demonstrated that links this traditional FMECA analysis with health management system design optimization based on failure mode coverage, availability, and life cycle cost analyses.

Role of FMECA in Health Management: Traditional Failure Modes, Effects and Criticality Analysis (FMECA) is typically performed in conjunction with the design process[1]. FMECA's historically contain 3 main pieces of information as described below:

[1] In this case, "Design" refers to all aspects of the system (components, control, etc.) with the exception of sensors and software used for condition monitoring.

- A list of failure modes for a particular component
- The effects of if the failure mode occurred ranging from a local level to the end effect
- The criticality of the Failure mode (I – IV), where (I) is the most critical

While this type of failure mode analysis is beneficial in getting an initial measure of system reliability and identifying candidates for redundancy, there are several areas where fundamental improvements can be made so that FMECA's can assist in health monitoring design. Four important FMECA improvements are described next.

1) Traditional FMECA does not address the precursors or symptoms to failure modes.

To move maintenance from reactive to proactive, it is important to focus on both system and component level indications that the likelihood of a failure mode has increased. Failure mode symptoms that occur prior to failure are these indications. An example of failure mode symptoms associated with a bearing would be an increase in spike energy or an increase in the oil particulate count.

2) Traditional FMECA does not address the sensors and sensor placement requirements to observe failure mode symptoms or effects.

The right data is essential to a health monitoring system. It is also important to have an optimal level of failure mode coverage so that enough collaborative information is available to detect and isolate failures. However, the authors' experiences have reinforced the fact that simply adding more sensors is impractical and ultimately reduces system reliability. By including sensors and sensor placement into the FMECA analysis, the location of a particular sensor for the optimum observational quality becomes more apparent. A simple example of this sensor placement issue might be the use of a downstream pressure sensor, necessary for a control function, which can also be used to monitor performance characteristics of upstream components. Moreover, in some cases, a simple change in the specifications of the sensor may provide monitoring capability in addition to the desired basic control function. Increasing the dynamic range or bandwidth of an accelerometer or pressure sensor are typical examples.

3) Traditional FMECA does not address health management technologies for diagnosing and prognosing faults.

The natural extension of including sensors in the FMECA is inclusion of diagnostic and prognostic technologies for observing or predicting failure modes and effects. Because several different diagnostic and/or prognostic technologies can be used for detecting a common failure mode, acquisition and implementation considerations must also be examined.

4) Traditional FMECA typically focuses on subsystems independently.

System level symptoms or system level effects are not fully realizable because subsystem interactions are typically not considered. This is a natural result of the communications barrier between the numerous teams and venders responsible for the development of a piece of complex machinery. As a result, unnecessary sensors or Health Management (HM) algorithms may be implemented or possibly overlooked entirely.

With these shortcomings in mind, a new approach has been developed as an extension to a traditional FMECA that can be used in the design of health monitoring and management systems.

Approach to Health Management Design: Figure 2 provides an overview of the approach to health management system design optimization. A basic description of each block will be given here, while details associated with each block will follow. First, a function block diagram of the system must be created that models the energy flow relationships between components. This functional block diagram provides a clear vision of how components interact with each other across subsystems. On a parallel path, a tabular FMECA is created that corresponds to a traditional FMECA except it contains failure mode symptoms, as well as sensors and diagnostic/prognostic technologies.

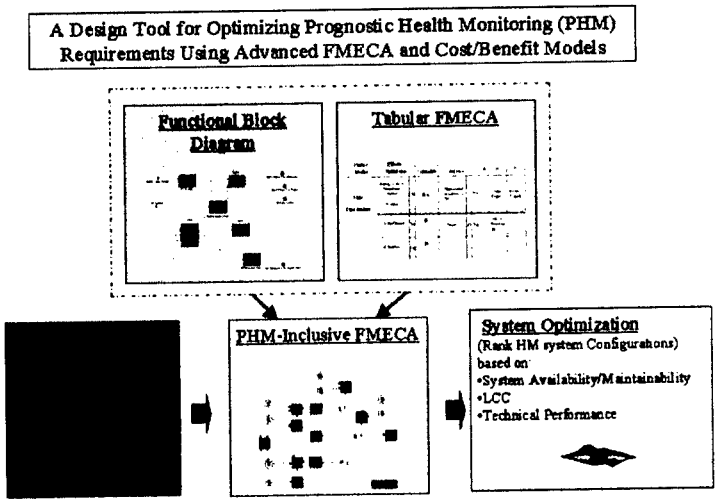


Figure 2 – Organization of the FMECA++ tool

The information from the functional block diagram and the tabular FMECA is automatically combined to create a graphical health management environment that contains all of the failure mode attributes as well as health management technologies. Once the graphical health management system has been developed, attributes are assigned to the failure modes, connections, sensors and diagnostic/prognostic technologies. The attributes are information like historical failure rates, replacement costs, false alarm rates etc., which are used to generate a fitness function for assessing the benefits of the health management system configuration. The "fitness" function criteria includes system availability, reliability, and cost. Some of these attributes must be manually determined if known, while others are related to the attributes of the diagnostic/prognostic technologies which can be determined from independent measures of performance and effectiveness tests. Finally, the health management configuration is automatically optimized from a cost/benefit and/or availability standpoint using a genetic algorithm approach. The net result is a configuration that maintains the highest system reliability to cost/benefit ratio.

Concept of the Functional Block Diagram: The Function Block Diagram (FBD) contains an integrated representation of how components, subsystems and systems interact with one another. It is not a simulation, only a hierarchical map of physical energy flows (i.e. torque transfer, current,

pressure). This energy flow map serves as the backbone for the health management design environment because it contains the failure mode symptoms and effects as well as captures their temporal paths. Figure 3 shows an example of a functional flow diagram at a "system" level. One could select any of the components to reveal specific interactions between its associated subsystem components. This FBD was created with a DARPA owned program called GME developed by ISIS Inc. at Vanderbilt University [7]. Other generic modeling software can also be used to build a FBD.

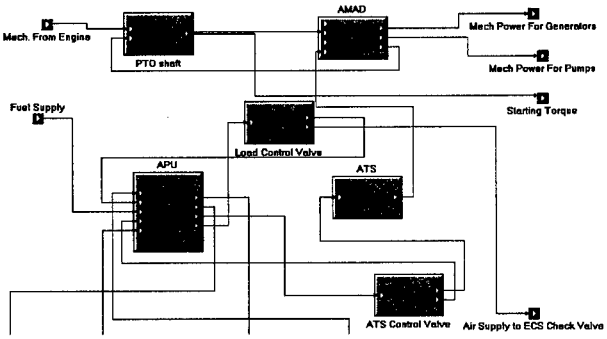


Figure 3 – Functional Block Diagram

As previously mentioned, with this approach, traditional FMECA analyses were enhanced with the addition of sensors, health monitoring technologies and failure symptoms. Figure 4 shows an example of an enhanced FMECA performed on a portion of a fuel system for a F-100 engine.

In this example, as with traditional FMECA, the failure mode is provided along with its effects (ranked from top to bottom as primary, secondary, tertiary, etc.). The Criticality or Frequency of Occurrence of the failure mode is ranked from A to E where:

A = Frequent, B = Probable, C = Occasional, D = Remote, E = Improbable

In practice, this Criticality letter would be associated with a specific probability of failure range.

The Severity of the failure mode is ranked from I-IV where:

I – Catastrophic, II – Critical, III – Marginal, IV - Negligible

The Criticality and Severity are symptoms of a failure mode used in optimizing the health management design discussed later.

In Figure 4, the first FMECA enhancement is that failure mode symptoms have been added to the "effects" column and are shaded in blue (or light gray). Failure mode symptoms are events that can be observed prior to the failure mode occurring or when the failure mode is in a very early stage of development. The effects that are shown in yellow (or dark gray) are downstream failure modes. In the case where an effect is a downstream failure mode, the failure mode of focus could be considered a failure mode precursor.

IDEAL						= Symptoms				
Module:			Engine Fuel System			= When an effect is a failure mode				
Component:			FuelManifold	Blue Sensor		= Existing Sensor				
Failure Mode	**Sev.**	**Crit.**	**Symptoms/Effects**	**Module**	**Component**	**Seasors**	**S_Module**	**S_Component**	**Diagnostics**	**Prognostics**
Faulty check valve	4	E	Increased fuel pressure in fuel supply lines	Engine Fuel System	Fuel Lines	Fuel Pressure	Engine Fuel System	Fuel Control Unit	BIT - Fuel Blockage	-
			Uneven fuel flow to the combustor	Engine	Combustor	EGT	Engine	Combustor	BIT - EGT	-
			[redacted]	Engine	Combustor	-	-	-	-	-
			Decrease in engine performance	Engine	Engine Power Section	EGT	Engine	Combustor	1) Compressor Fault Detector 2) Turbine Fault Detector	Engine Performance Prognostics
						Fuel Flowrate	Engine Fuel System	Fuel Control Unit		
						Engine Speed	Engine	Rotor		
						Inlet Temp	Engine	Compressor		
Plugged Strainer	4	E	Increased fuel pressure in fuel supply lines	Engine Fuel System	Fuel Lines	Fuel Pressure	Engine Fuel System	Fuel Control Unit	BIT - Fuel Blockage	-
			Uneven fuel flow to the combustor	Engine	Combustor	EGT	Engine	Combustor	BIT - EGT	-
			[redacted]	Engine	Combustor	-	-	-	-	-
			Decrease in engine performance	Engine	Engine Power Section	EGT	Engine	Combustor	1) Compressor Fault Detector 2) Turbine Fault Detector	Engine Performance Prognostics
						Fuel Flowrate	Engine Fuel System	Fuel Control Unit		
						Engine Speed	Engine	Rotor		
						Inlet Temp	Engine	Compressor		
Damage, Incorrect, or Missing Seals	2	D	External fuel leak	Engine Fuel System	Engine Fuel System	-	-	-	-	-
			Fire	Engine	Engine	-	-	-	-	-

Figure 4 – Tabular FMECA of a F-100 Fuel System

The "Component" column identifies the component immediately affected by the failure mode while "Module" is the subsystem in which the component resides. This functional relationship is cross-referenced with the functional block diagram. In a similar fashion, the "Sensor" column lists the sensor that can observe the symptom or effect while "S_Module" is the subsystem in which the sensor resides and "S_Component" is the component it is linked to. All sensors in this example are required for control or safety purposes. Finally, "Diagnostics" and "Prognostic" columns have been added. The "Diagnostics" column describes any discrete diagnostic (Built in Test (BIT)) or algorithms that can observe the symptom or effect. The "Prognostics" column describes any prognostic algorithms that can be used to obtain a RUL prediction on the failure mode.

Graphical Health Management Environement

The FBD and the tabular FMECA contain enough information to generate a graphical health management design and testing environment without any further human intervention. Figure 5 provides a simple representation of the graphical health management system model and will be used to illustrate the use of collaborative information to predict and isolate faults. In this figure, the "S's" represent sensors local to a component. Failure modes (FM's) are shown that originate in this component and their associated local effects. Downstream effects will propagate up to the next higher level. Diagnostic monitors and prognostic monitors are also present in this model. Consider the following example.

The diagnostic monitor (D1) could identify that the symptoms of either Failure Mode 1 (FM1) or Failure Mode 2 (FM2) have developed. If, in addition to this observation, the prognostic monitor (P) linked to "FM1" determines that "FM1" has a high probability of failure, "FM1" can be assigned more risk than "FM2". Now consider if "P" and "D1" did not exist. In this scenario, there is nothing in this health management configuration that can predict "FM1" or "FM2" before they occur. However, the effect of "FM1" is a symptom of "FM3" and, in this case, there is potential that the fault path could be prevented with "D2" before higher level effects develop. Therefore, if "FM3" is found to have occurred and "D2" did not alarm, "FM2" would be the more likely root cause (accounting for the false-negative potential of the "S4"/"D2" combination) and fault isolation potential is improved.

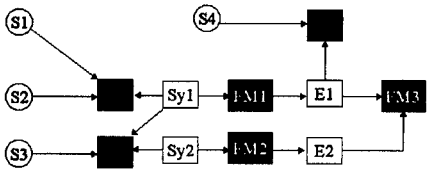


Figure 6 – Generic Graphical FMECA Representation

Health Management Attributes: To autonomously evaluate the cost/benefit of a HM system configuration, all aspects of the system must ultimately be assigned, or modify, a dollar value. Some of these "attributes" are more easily derived that others. All attributes can be grouped into "Cost Related" or "Technical Related". Cost related attributes relate to true dollar values such as hardware cost or component replacement cost while some technical related attributes are complexity factor or sensor observational quality. The FMECA++© aspects that are assigned attributes within a HM system include:

1. Failure modes (FM)
2. Sensors (S)
3. Connections (Sy/E)
4. Diagnostics (D)
5. Prognostic (P)

Each of these health management system "building blocks" that make up the Integrated HM model have "attributes" that contribute to the overall health management system configuration cost function. A description of each of these "building block" attributes is provided next.

Failure Modes - Failure Modes have been assigned a minimum of 5 attributes. These are:

1. Criticality (A-E) or Failure Rate (0-1) – (Pf)
2. Severity level (I - IV) – (S)
3. Consequential Cost of Failure Mode occurring – (CC)
4. Cost of a False Detection (CF)
5. Cost saved with Planned Maintenance (M)

The "Severity" (S) is a multiplier in the cost function that may represent the safety factor of a particular failure mode. The "Consequential Cost" (CC) is the sum of replacement, refurbishment, maintenance etc. costs for a particular failure mode as well. The downstream effects of a failure mode are naturally accounted for in the integrated FMECA++^{c} model. The "Cost of False Detection" (CF) represents the cost of an inspection maintenance event, reduced availability etc. Finally, the "Cost Saved with Planned Maintenance" (M) is the benefit realized by being able to predict when (with confidence bounds) a failure will occur.

Clearly, the failure mode attributes do not specifically address a number of maintenance related and availability issues. A number of these issues are introduced in a companion paper.

Sensors - Sensors were defined in the model as components for measuring physical quantities such as temperatures, pressures and currents. The attributes assigned to the sensors include:

1. Acquisition and Implementation Cost (AIC)
2. Criticality (A-E) or Failure Rate - (SPf)
3. Weight Cost - (W) (for aerospace applications)
4. Observational Quality (0-1) – (OQ)

The total "cost" of a particular sensor is a function of its utility in a variety of diagnostic and prognostic tools as well as its role in control system functionality.

The "Observational Quality" attribute of a particular sensor is a function of its type and placement with respect to the failure mode being observed. The identification of a parsimonious suite of sensors and their placement is a necessary step in the design of a health management system in order to optimize the detection and prognostic capability of the available sensors. A number of different approaches have been investigated by the authors [1] to help in the optimum sensor and placement in terms of health management. One method was via a system test and sensitivity study, wherein the observability of the identified failure mode symptoms at each potential sensor location was determined. Locations within the system with the largest overlapping of failure modes and the highest observability are used to select

potential locations for sensor placement. A key part of this process is a sensitivity matrix that quantifies the observability of different variables throughout the system for a set of failure modes.

Symptom and Effect Connection Attributes - Symptom and Effect connections within the graphical FMECA environment represent the causal and temporal links between failure modes and their effects. The only connection attribute is "Propagation Probability" – (Pp) which is the likelihood of an effect propagating downstream.

Diagnostic and Prognostic Attributes - Diagnostics can be either discrete or continuous. Discrete diagnostics are traditionally algorithms that produce 0 or 1 depending on if a threshold has been exceeded. Many types of Built In Tests (BITs) can be classified as Discrete Diagnostics. An example of discrete diagnostics is an Exhaust Gas Temperature (EGT) reading that has exceeded a predetermined level.

Continuous diagnostics are algorithms designed to observe transitional effects and diagnose a failure mode based on the method and rate in which the effect is changing. Continuous diagnostics are usually associated with observing the severity of failure mode symptoms. Examples of continuous diagnostics would be a spike energy monitor for identifying low levels of bearing race spalling or an A.I. classifier for diagnosing that a valve is sticking.

The attributes identified for Diagnostics have been broken up into Technical and Cost related. The Technical attributes include:

1) Detection Confidence score (0-1) – (DC) 2) % false positive score (0-1) – (FP)

The "Detection Confidence score" can be used to simultaneously account for true-negative and true-positive characteristics.

The Cost Attribute of Diagnostics include:

1. Development, Implementation and Tech. Maintenance Cost (DAIC)

Finally, Prognostic algorithms can use a combination of sensor data, a-priori knowledge of a failure mode and diagnostic information to predict the time to a failure or degraded condition with confidence bounds. Prognostic algorithms are linked directly to failure modes in the graphical FMECA model. Like diagnostic algorithms, both technical and cost related attributes have been identified for prognostic algorithms.

Technical Attribute:
1. Prognostic Accuracy (0-1) – (PA)

Prognostics do not have an attribute associated with false alarms. The "Prognostic Accuracy" accounts for the early detection quality of the technology. A physical prognostic model (i.e. based on an FE model) would ideally have a higher prognostic accuracy than an experienced-based model (i.e. Weibull distributions of historical failure rates).

The Cost Attribute for Prognostics is:

1. Development, Implementation and Tech. Maintenance Cost - (PAIC)

A valid concern is how the technical attributes of diagnostic and prognostics technologies can be determined. One method is addressed in [1] whereby algorithms are tested objectively from performance and effectiveness standpoints using transitional run to failure data. Of course in the absence of this type of information, and with a new sensor/algorithm combination, an educated guess may be the only option.

Health Management System Optimization - In order to optimize the core configuration of a health management system (i.e. what sensors and associated algorithms to implement) based on the enhanced FMECA approach previously described, a cost or fitness function that accounts for reliability, technical risk, complexity and overall life cycle costs must be developed. This total "fitness" function will then be minimized to arrive at potential HM system configurations. The plot on the top of Figure 6 shows system dependability as a function of cost in the absence of a health monitoring system. In this scenario, the redundancy and high factors of safety are essential to insure that critical failures maintain a low failure rate. [3] The lower plot illustrates the effect of implementing a HM system. *With effective (and dependable) diagnostic and prognostic capabilities, system redundancy can be reduced and the boundaries of the design envelope can be safely extended.* With health monitoring capability, the overall system dependability remains high while safety is not compromised.

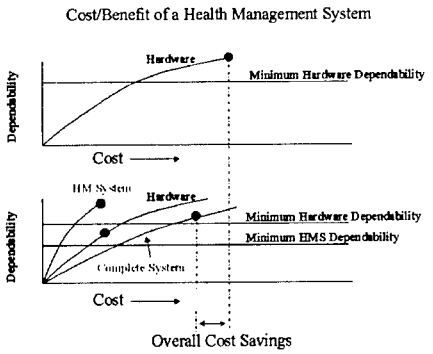


Figure 6 – Using HM to increase overall system reliability

The health management design environment contains a sufficient amount of information to generate and evaluate a fitness function for the configuration. This fitness function is of the form:

For each Failure Mode – FM(i)

*Step 1) Probability of Failure * Severity *Consequential Cost of FM(i) +(Downstream Failure Mode Consequential Costs) * Probability of Propagation*

*Step 2) *HM risk reduction attributed to FM(i)*

Step 3) + Cost associated with False Alarms on FM(i)

Step 4) + Total Cost of all HM technology

Specifically, the formulation is as follows:

Step 1 and 2 =

$$\sum_{FM_i}^{FM_N}\left\{\left[\prod_{D_{FM}} DC\cdot\frac{\sum OQ(1-SPf)}{\frac{S_D}{NsensorsD}}\cdot\prod_{P_{FM}} PA\cdot\frac{\sum OQ(1-SPf)}{\frac{S_P}{NsensorP}}\right]\cdot\left[(Pf\cdot S(CC-M)\cdot Pp)+\sum_{FM_{i+1}}^{FM_N} Rolled_Up\right]\right\}$$

Where the cost saved with planned maintenance (M) can only be realized if a prognostic algorithm is present on the failure mode.

The "Rolled Up" costs =

$$Pf\cdot S(CC)\cdot Pp\cdot\left(\prod_{D_{FM}}\left(1-\frac{\sum OQ}{\frac{S_{FM}}{NsensorsD}}\right)\cdot DC\cdot\prod_{P_{FM}}\left(1-\frac{\sum OQ}{\frac{S_{FM}}{NsensorsP}}\right)\cdot PA\right)$$

Step 3 =

$$+(1-Pf)\cdot S\cdot\left[1-\left[\prod_{S_{FM}}(1-SPf)-\prod_{D_{FM}}(1-FP)\right]\right]\cdot CI$$

Finally Step 4 =

$$+\sum_{S}(W+AIC)+\sum_{D}DAIC+\sum_{P}PAIC$$

Configuration Optimization: The HM system optimization (optimization of the previously described cost function) will operate between two boundaries; a "maximum" HM system configuration that includes the "wish list" of all potential sensors and associated algorithms that achieve complete failure mode coverage and a "minimum" configuration that is necessary for safety and control. The optimization algorithm will examine random configuration variations and calculate the "fitness" or cost for each.

A genetic algorithm optimization scheme was chosen for the HM optimization because genetic algorithms are better configured to handle optimization problems with little regard for non-linearity, dimensionality or function complexity in general. Potential cost functions generated in the FMECA++© environment can include hundreds of independent variables and thus makes it impractical to utilize traditional optimization techniques such as gradient decent or other derivative-based algorithms.

The genetic algorithm optimization scheme developed capitalizes on the benefits of both the *classic* and *elite* genetic algorithm approaches. In general, the genetic algorithm operates by evaluating the "fitness" of a "gene pool" population within a given environment. New "generations" (potential solutions) are created using a combination of "parent" genes and "mutations". Only the most "fit" genes (best solutions) are ultimately passed through the generations [5]. In terms of health management system design optimization, the "environment" is the FMECA model while the "gene pool" represents the many different health management configurations.

The HM building blocks that contribute most effectively to the minimization of the "fitness" function will be passed on to the "next generation". This process is described in the block diagram in Figure 7.

Figure 8 shows a 2-D contour of a simplified cost function associated with two variables. Normally the dimensionality would be much higher and equal to the number of possible combinations between the max. and min. configurations. This cost function was chosen to illustrate how the genetic algorithms work because it has three clear minimas, with only one as the global minima (the solution we are looking for). An initial population was generated that represents a small fraction of the possible HM configurations. Within the optimization process, aspects of this population are combined, mutated and re-evaluated.

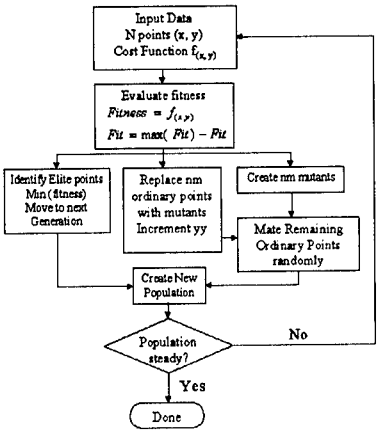


Figure 7 – Genetic Algorithm Flow Chart

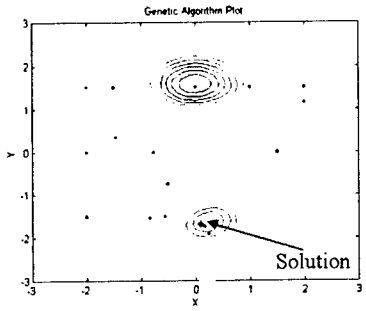


Figure 8 – Optimum Configuration

Example of HM design and optimization: Figure 9 shows the Maximum and Minimum HM configuration addressing failure modes for a bearing and bearing housing.

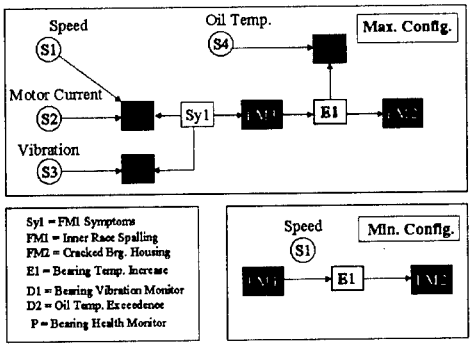


Figure 9 – Max and Min Configurations

Notice that in the Max. configuration, a diagnostic monitor is observing the vibration related symptoms of Inner Race Spalling and a prognostic monitor is predicting when the spalling will occur with a high severity using motor current, and speed. If the spalling were to occur, another diagnostic monitor (D2) will observe if oil temperature is too high and thus potentially prevent cracking of the Bearing Housing (FM2). In the Minimum configuration no health monitor capability exist and the speed sensor is present for control purposes only. If bearing spalling were to occur the risk of the housing cracking would be based entirely on the Propagation Probability between FM1 and FM2. The attributes assigned to each of the HM components in the Max. configuration are given in Figures 10 through 12.

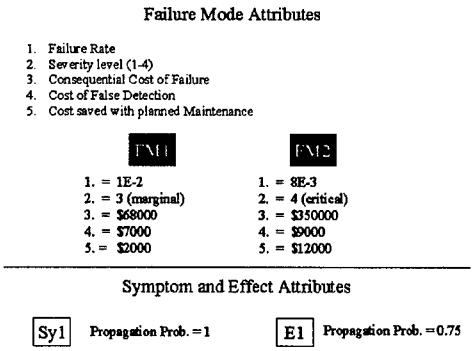


Figure 10 – Failure Mode and S/E attributes

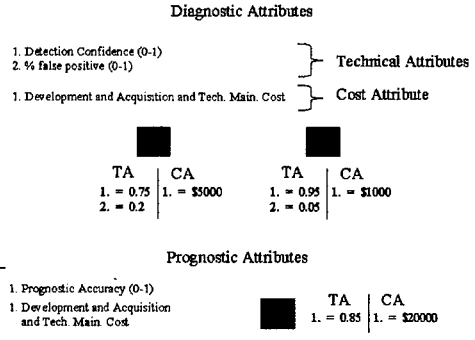


Figure 12 – D/P attributes

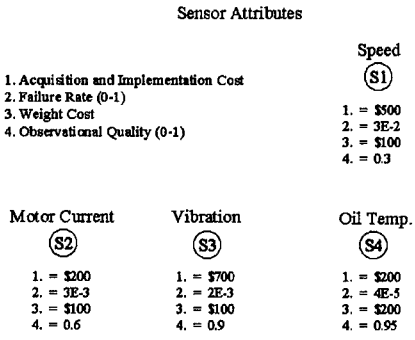
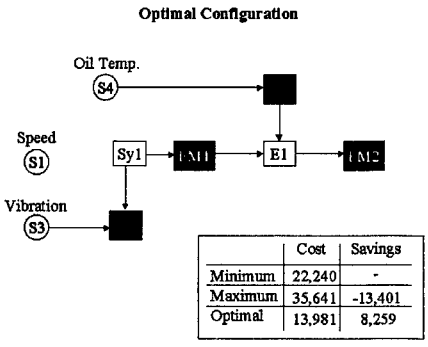


Figure 11 – Sensor Attributes



	Cost	Savings
Minimum	22,240	-
Maximum	35,641	-13,401
Optimal	13,981	8,259

Figure 13 – Optimal Configuration and Results

The results of a cost/benefit analysis of the Min and Max configurations is shown in Figure 13. In the Minimum configuration there is no benefit in terms of risk reduction from a HM system

but there is also no added cost for false alarms and HM hardware. The cost of 22,240 is the dollar value calculated for risk of both FM1 and FM2 occurring. In contrast, the maximum configuration has too much HM capability. The risk reduction of FM1 (calculated at 78%) and FM2 (10%) is not sufficient to offset the higher risk of false alarms and the significant technological development cost of prognostics in this case. The optimal configuration was found to retain both the vibration diagnostic monitor and oil temp monitor. They provided a fair amount of risk reduction (40% and 10% respectively) while maintaining good system reliability. Further optimization approaches that account for maintenance plans and system availability may be found in [11].

Conclusion: An approach has been presented that extends traditional FMECA capabilities to aid in the design of health management solutions that can for reduce total ownership costs and improve availability for complex engineered systems. This approach utilizes a graphical FMECA environment where failure modes, failure mode symptoms/effects, sensors, and diagnostic/prognostic technologies are represented. The health management system configuration can be optimized from an availability and cost/benefit standpoint with a genetic algorithm approach through analysis of the fitness attributes on HM system building blocks. The ultimate objective of this approach was to form a methodology and environment which aids condition based maintenance practices by mitigating or preventing failure modes while still keeping sensor and diagnostic/prognostic technology costs at a minimum.

[1] Orsagh R.F. and Roemer, M.J. "Development of Metrics for Mechanical Diagnostic Technique Qualification and Validation", COMADEM Conference, Houston TX, December 2000.
[2] Roemer, M. J. and Kacprzynski, G.J., "Advanced Diagnostics and Prognostics for Gas Turbine Engine Risk Assessment," Paper 2000-GT-30, ASME and IGTI Turbo Expo 2000, Munich, Germany, May 2000.
[3] Lewis, E., Introduction to Reliability Engineering, John Wiley & Sons, New York, 1987
[4] Roemer, M. J., and Atkinson, B., "Real-Time Engine Health Monitoring and Diagnostics for Gas Turbine Engines," Paper 97-GT-30, ASME and IGTI Turbo Expo 1997, Orlando, Florida, June 1997.
[5] Brooks, R. R., and Iyengar, S. S, Multi-Sensor Fusion, Copyright 1998 by Prentice Hall, Inc., Upper Saddle River, New Jersey 07458
[6] Canada, J, and Sullivan, W, Capital Investment Analysis for Engineering and Management, Copyright Prentice Hall 1996
[7] Information on the GME program can be found at www.isis.vanderbilt.edu.
[8] RAMS Software Tools from Item Software, www.itemsoft.com
[9] Dhillon, B. S., Engineering Maintainability. Houston, TX: Gulf Publishing Company, 1999.
[10] Moss, M. A., Designing for Minimal Maintenance Expense: The Practical Application of Reliability and Maintainability. New York, NY: Marcel Dekker, Inc., 1985.
[11] Yukish, Michael et. al., Issues in the Design and Optimization of Health Management Systems, 55th Meeting of the MFPT, 2001, Virginia Beach, Virginia

ISSUES IN THE DESIGN AND OPTIMIZATION OF HEALTH MANAGEMENT SYSTEMS

Michael Yukish, Carl Byington, and Robert Campbell

The Pennsylvania State University
Applied Research Laboratory
P.O. Box 30 (North Atherton Street)
State College, Pennsylvania 16804-0030

Abstract: The design of a health management system is presented as a decision problem. The decision space is affected by the choice of a particular health management system design and an employed maintenance policy. To the 1st order, the evaluation objectives consist of the conflicting goals of minimizing purchase costs, minimizing operating costs, and maximizing availability. In order to assist and even automate the decision process, data and computational tools for calculating the objectives are needed. While calculating purchase costs is straightforward, determining operating costs and availability is not. Parameters such as failure rate, criticality, component replacement cost due to unplanned and planned maintenance, and average downtime for repair are examples of data needed to determine the operating costs and availability. These types of data are not part of traditional product models. Some of the data is partially contained in the traditional FMECA, but much of it is not. This shortcoming is the motivation for tools to assist in the design of health management systems, such as the FMECA++© tool being developed by Impact Technologies and Penn State Applied Research Laboratory.

Key Words: Availability; CBM; evaluation metrics; FMECA; health management; operating cost; optimization; purchase cost.

Introduction: It is well known that health management systems can increase the overall reliability of the underlying system by providing early fault detection and diagnostic localization. In the ultimate case, a CBM system would enable one to predict the remaining useful life of critical components, and to isolate the root cause of failures after the failure symptoms have been observed. If predictions can be made, replacement part orders and repair actions can be optimally scheduled to reduce the overall operational and maintenance-related costs, while minimizing downtime and therefore maximizing system availability. These improvements in operating costs and availability are of course offset by the increase in cost of acquiring and maintaining the health management system.

Thus, the choice of what health management system to use can be abstractly considered as a decision problem, where the decision maker chooses a health management system and an accompanying maintenance policy to satisfy the conflicting goals of minimizing purchase costs, minimizing operating costs, and maximizing availability. Cast in this

fashion, the tools and techniques of multi-objective optimization and multidisciplinary design optimization can be used to find a "best" design [1].

In order to assist and even automate the decision process, data and computational tools for calculating the objectives are needed. While calculating purchase costs is straightforward, determining operating costs and availability is not. Parameters such as failure rate, criticality, component replacement cost due to unplanned and planned maintenance, and average downtime for repair are examples of data needed to determine the operating costs and availability. Again, such data is partially contained in the traditional FMECA, but much of it is not. Augmented models such as used in the FMECA++© intrinsically capture the downstream effects of the failure modes, including secondary effects as embodied in a hierarchical model. FMECA++© is envisioned to be a (graphical & tabular) representation of functional failure modes with hierarchically linked effects and symptoms to provide a blueprint for the design of a health management system. It extends a typical FMECA with information on precursor symptoms, sensor observables, diagnostic/prognostic processes and their associated metrics. The data embodied in the FMECA++© can be combined with its associated methods and tools for calculating operating costs and availability, and the problem can be cast as a multi-objective optimization problem and solved using well-known methods.

The remainder of this paper first establishes the basic problem statement for casting the choice of a health management system as a decision problem, choosing over multiple objectives. Next various methods for optimizing with multiple objectives are presented. The determination of an availability metric receives extra attention, as its calculation is less straightforward in comparison to the other objectives. Finally, the requirements imposed on a design environment in order to implement the problem structure developed in this paper are presented.

Statement of the Decision Problem: In choosing a health management system, the decision maker starts (by assumption) with a system to be monitored (S), and has the conflicting objectives of minimizing purchase cost (PC) and operating cost (OC) while maximizing availability (A). The "decision space" is the choice of health monitoring suite to employ (HM), and the choice of accompanying maintenance policy (MP). The prime dependencies of the objectives with regard to the decisions are as follows:

$$\begin{aligned} \text{Purchase Cost} &= \text{PC}(\text{S}, \text{HM}) \\ \text{Operating Cost} &= \text{OC}(\text{S}, \text{MP}, \text{HM}) \\ \text{Availability} &= \text{A}(\text{S}, \text{MP}, \text{HM}) \end{aligned} \tag{1}$$

Note that in addition to the dependence on the system to be monitored, purchase cost is shown as a function of the choice of health management system, and operating cost is shown as a function of the maintenance policy and the heath management system. Particularly in the case of operating costs, this is done to make explicit the dependency of the operating cost on both the maintenance policy and the health management system. The health management system will directly affect the operating costs to a small degree through its own life cycle costs. It will also affect OC with the ability to impact the required amount of maintenance and provide potentially large mishap cost avoidances.

Multi-Objective Optimization: The health management choice presents a multi-objective decision problem, where the objectives are conflicting. In this instance, a significant tradeoff is in the up-front purchase cost of a health management system versus the downstream savings in operating costs. An additional tradeoff is, given a health management system, choosing a maintenance policy that will minimize the operating costs versus choosing one to maximize the availability. Many methods are available for finding "best" solutions for such problems [2]. Each method attempts to capture, in some rational manner, the decision maker's preference. The methods discussed briefly here are weighted sums, minimax, goal programming, and *design by shopping*, explained below.

The most basic methods are weighted sums methods, where a scalar measure of worth is calculated by multiplying each of PC, OC and A by a weighting factor. Note that the availability term is subtracted from the total, to account for its maximization vice the other terms' minimization:

$$\begin{aligned} &\min_{\text{HM,MP}} \quad w_1\text{PC} + w_2\text{OC} - w_3\text{A} \\ &\text{where} \sum_{i=1}^{3} w_i = 1 \end{aligned} \tag{2}$$

The weighting parameters are an attempt to capture the preference of the decision maker as to the relative importance of the terms. These can be generalized to quadratic and higher systems with weighted sums:

$$\begin{aligned} &\min_{\text{HM,MP}} \quad w_1\text{PC}^k + w_2\text{OC}^k - w_3\text{A}^k \\ &\text{where} \sum_{i=1}^{3} w_i = 1, \quad k \in \{1,2,3,\ldots\} \end{aligned} \tag{3}$$

Another method is to apply the minimax criteria as follows:

$$\begin{aligned} &\min_{\text{HM,MP}} \quad \max_{w_i} \left[w_1\text{PC} + w_2\text{OC} - w_3\text{A} \right] \\ &\text{where} \sum_{i=1}^{3} w_i = 1 \end{aligned} \tag{4}$$

The minimax criteria can be interpreted as "chose HM and MP so as to minimize the worst possible choice of weights *w*." Use of the minimax criteria can be construed as an attempt to guard against incorrectly capturing a user's preference, expressed in *w*. However, designs chosen by the minimax criteria are usually considered as too conservative.

Another method is known as pre-emptive goal programming. With this method, the objectives are first ordered from most to least important. Then the optimization problem is solved for the most important objective first, and only solving for the next objective if the answer to the first problem is non-unique. So for example if the objectives are ordered {PC, OC, A}, then the problem solved first is

$$PC^* = \min_{HM,MP} \; PC(HM,MP) \tag{5}$$

If the solution for HM is not unique, then the next problem solved is

$$\begin{aligned} OC^* &= \min_{HM,MP} \; OC(HM,MP) \\ \text{s.t. } & PC(HM,MP) = PC^* \end{aligned} \tag{6}$$

and so on until a unique solution is reached.

The final method presented, known as design by shopping, does not establish a global objective at all [3]. Rather, the *pareto frontier* of the feasible results of PC, OC, and A is presented to the user, and the user decides. The pareto frontier is the set of all designs that are *non-dominated* by other designs. Assume that a choice of HM and MP result in a set of values {PC, OC, A} that define a point in the performance space. This point is non-dominated if an improvement in any one of the objectives can only be achieved by a decrease in one or more of the others. A dominated design point is one where a feasible design exists that is at least as good as the first point in all objectives, and better in at least one. The diagram below, Figure 1, shows the pareto frontier in bold for the two-dimensional case, holding purchase cost constant. Designs that are along the lower left boundary are non-dominated, in that an improvement in one aspect is accompanied by a decrement in another. Interior points are dominated, and it is reasonable to expect that a decision maker would never choose one.

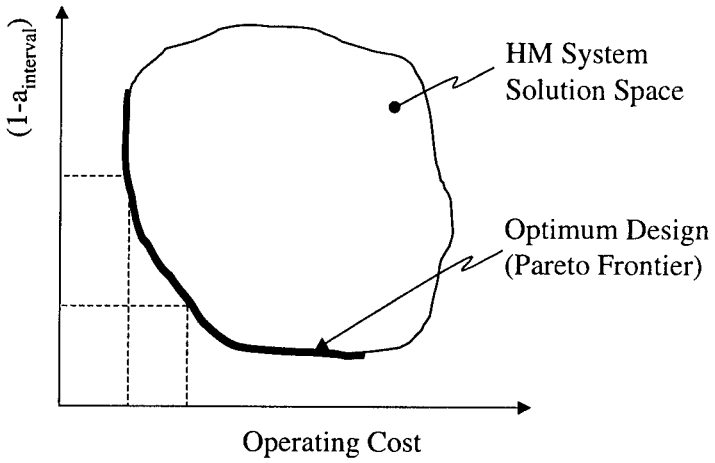


Figure 1: Pareto Frontier

Many design optimization experts would now consider it a mistake to come up with some single scalar objective that is a blend of PC, OC and A and that attempts to capture the preference of the decision maker. In their opinion it is better to let the decision maker

"shop" for the right mix by presenting designs from the pareto set, rather than trying to capture the decision maker's preference, which experience has proven difficult.

Determining Purchasing and Operating Cost: Purchasing cost can be determined using parametric cost estimating applications such as PRICE-E ™. Such tools have been widely employed in the cost estimating of conceptual through detailed design [4]. The primary inputs are weight, volume and complexity of the subsystems, the hierarchical structure of the system, and the complexity of the assemblies. If the baseline costs are known due to activity-based costing, then these numbers can be used. Additional data needed to estimate purchasing costs relates to the actual purchase, e.g., the dates for initiating purchases, buy rates, total amount purchased, and so on. The same tools can also be used to develop approximations to operating costs, based on the design data listed.

Determining availability: It is probably too hard to calculate availability in closed form for a system that forms a complex reliability block diagram. Upper and lower bounds might be calculated by making simplifying assumptions, such as choosing failure rate statistics from a restrictive set of families, or assuming the reliability block diagram is all series or parallel. But if models of the system are available, tools exist to simulate the system and determine an availability metric. A choice is to use a Monte Carlo simulation, with values for isolation, repair, and admin/logistics times for various components along with failure statistics, and simulate the system over an interval to get an estimate of availability.

Availability metric: Presented in this section is one approach to determining a measure of availability. It is important to bear in mind that, if availability is to be an objective, its computation must be such that the impact of the addition of HM is clear.

For a system that operates over some fixed interval, the availability can be determined by the equation

$$a_{\text{interval}} = \frac{T_{\text{operate}}}{T_{\text{iso}} + T_{\text{adl}} + T_{\text{repair}} + T_{\text{PM}} + T_{\text{operate}}} \tag{7}$$

where over the interval, T_{iso} is the time spent isolating faults, T_{adl} is the admin and logistics time associated with repair events, T_{repair} is the time to disassemble, repair or replace, and reassemble during repair events, T_{PM} is the time spent doing preventive maintenance, and T_{operate} is the time operating [5]. All of the T's other than T_{operate} must have the caveat that they only count if they occurred when the system was supposed to be available. So the additional constraint can be imposed

$$T_{\text{required}} = T_{\text{iso}} + T_{\text{adl}} + T_{\text{repair}} + T_{\text{PM}} + T_{\text{operate}} \tag{8}$$

where T_{required} is the time that the system is required to be operating, which may be only eight hours per day, for example. In this case, all maintenance may be performed while the system is not required to be available, ensuring 100% availability.

To gain a feel for how the addition of health management can affect the various parameters, we consider below the results of adding a partially effective and a *perfect* health management suite compared to no health monitoring [Table 1]. Reference [10] discusses previously developed metrics to associate with a HM system design, and reference [9] proposes how this could be applied to produce an operational impact on A and OC. These metrics represent a way to propagate the effectiveness of specific HM sensors and algorithms and map them to an availability effect as is shown in Table 1. To further simplify the problem of comparison, assume the available maintenance actions are restricted to preventive maintenance (PM) and replacement (REP). Further assume that a REP is either planned or unplanned.

Table 1: Comparison with degrees of Health Management

	No HM	Partially Effective HM	Perfect HM
Unplanned replacement	$T_{\text{repair}}, T_{\text{iso}}, T_{\text{adl}}$	$T_{\text{repair}}, T_{\text{iso}}, T_{\text{adl}}$. Unplanned replacements reduced corresponding to HM fault detection metrics	Unplanned replacements are totally eliminated
Planned replacement	$T_{\text{repair}}, T_{\text{iso}}$	$T_{\text{repair}}, T_{\text{iso}}$. Isolation time is reduced appropriately by diagnostic accuracy metric	T_{repair} only. Isolation time is eliminated
Preventive maintenance	T_{PM}	T_{PM} will be reduced, based upon the composite effectiveness of the HM system to predict the overall failure modes	T_{PM} will be reduced to provide predictive maintenance on all critical systems

Maintenance Policy: Complicating the decision problem greatly is the fact that the choice of maintenance policy has a critical impact on the value of the T variables. They can all be written as $T = T(\text{MP})$. For example, a HM system coupled with a maintenance policy that reads "Replace only on failure" will show no benefits of a health management system. In general, the choice of maintenance policy (MP) will have an impact on availability equal or greater in scope to the choice of HM system. Restating the equation for determining availability, (1)

$$\text{A} = \text{A}(\text{S}, \text{HM}, \text{MP}) \tag{9}$$

If we are trying to find the HM system that gives us the best availability, we must solve the optimization problem for maximum availability while including MP as a decision variable.

$$A^* = \max_{HM,\,MP} \; A(S, HM, MP) \tag{10}$$

Casting as a search for the best HM and moving the optimization with regard to MP to an inner loop, the problem gains a bi-level optimization structure.

$$HM^* = \arg\max_{HM} \left[\max_{MP} \; A(S, HM, MP) \right] \tag{11}$$

Notional Infrastructure for Supporting the Decision Process: Given the statement of the decision problem above, the suite of computational tools and applications must next be developed. One such structure for supporting the decision process is shown below, in Figure 2.

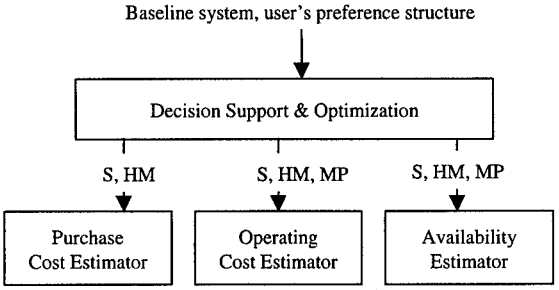


Figure 2: Decision Support Structure

At the top, the baseline system that is to be considered along with possibly some preference structure is entered. At the bottom are three separate applications, each of which analyzes a design concept to determine its value with respect to one of the three objectives. In the middle is the optimization engine, which in effect automates the search through the design space in order to find the "best" designs.

It is important to note that each of the three estimators require data about the system, both the baseline system and the chosen health management system, to be passed down, but that the constitution of the data differs from one estimator to another. The purchase cost estimator needs sizes, weights, complexities, and other manufacturing cost-related data of the system and its components. The availability estimator needs data about the components of the system, such as failure statistics as a function of loading, and about the constitution of the coupling of the components in the system, such as captured in a reliability block diagram. Therefore, before any optimization can occur, the product must be modeled in a fashion that can serve as input to the estimators.

Optimization Methods: Once having posed the decision problem and developed the appropriate data models to drive the estimators, the implementation of an optimization algorithm can be considered. The choice of an optimization algorithm is constrained by a number of aspects of the problem. First, the estimator inputs will likely contain a mix of continuous, countable, and enumerated variables. This implies a smooth optimization algorithm will not suffice for the overall problem, but may be applicable for sub-problems. However, if the problem is cast such that the maintenance policy is solved for in an inner loop and the health management choice is solved for in an outer loop, this presents a bi-level optimization problem. Bi-level optimization problems are notoriously difficult for gradient-based optimizers to work with, [6, 7].

Alternatives to the gradient-based optimization algorithms are the non-gradient methods such as simulated annealing and genetic algorithms. Genetic algorithms have the added benefit that they are conducive to exploring the pareto set of a design space, [8]. At each iteration, a new set of proposed designs are created, and the non-dominated designs are culled from the offspring. Eventually the genetic algorithm will develop a set of design points that are reasonably expected to be along the pareto set. A drawback to all of the non-gradient based methods is that they have no obvious stopping criteria, as does exist in the gradient-based methods.

Future Work: Because of its potential impact, health management solutions should be considered during the initial design of a system. However, current practice in system design does not adequately support the consideration of such solutions. It would appear that, because an initial system FMECA is performed during the design stage, it is a perfect link to the critical overall system failure modes that a health management system is designed to help mitigate. In fact, a process has been demonstrated that links this traditional FMECA analysis with health management system design optimization based on failure mode coverage and availability and life cycle cost analyses, [9]. But in order to be able to truly evaluate the relative merits of different health management system options, the systems must be modeled in a more extensive manner. New tools such as the FMECA++© are now being developed to address this shortfall, [9]. The methods presented herein can be implemented in such tools for use in the optimization of the system and the HM, thus providing the maximum benefit of HM through its impact on the system design.

Acknowledgment: This work was supported by a Navy Small Business Innovative Research Program through a subcontract with Impact Technologies (Contract No. S00139N). The content of the information does not necessarily reflect the position or policy of the Government, and no official endorsement should be inferred.

1. Sobieszczanski-Sobieski, J. *Multidisciplinary Design Optimization: An Emerging New Engineering Discipline.* in *The World Congress on Optimal Design of Structural Systems.* 1993. Rio de Janeiro, Brazil: NASA Technical Memorandum 107761.

2. Rosenthal, R., *Principles of Multiobjective Optimization.* Decision Sciences, 1985. **16**: p. 133-152.
3. Balling, R. *Design by Shopping: A New Paradigm?* in *Proceedings of the Third World Congress of Structural and Multidisciplinary Optimization (WCSMO-3).* 1999. Buffalo, NY: University at Buffalo.
4. Yukish, M., L. Bennett, and T. Simpson. *Requirements on MDO imposed by the Undersea Vehicle Conceptual Design Problem.* in *8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization.* 2000. Long Beach, CA: AIAA.
5. Dhillon, B.S., *Engineering Maintainability.* 1999, Houston, TX: Gulf Publishing Company.
6. Vicente, L. and P. Calamai, *Bilevel and multilevel programming: a bibliography review.* 2000.
7. Alexandrov, N.M. and R. Lewis, *Analytical and computational aspects of collaborative optimization.* 2000, NASA: Hampton, VA. p. 26.
8. Murata, T. and H. Ishibuchi. *MOGA: multi-objective genetic algorithms.* in *IEEE Conference on Evolutionary Computation.* 1995: IEEE.
9. Kacprzynski, G., et al. *Enhanced FMECA: Integrating Health Management Design and Traditional Failure Analysis.* in *55th Meeting of the MFPT.* 2001. Virginia Beach, Virginia.
10. Orsagh, R., et al., *Development of Metrics for Mechanical Diagnostic Technique Qualification and Validation,* Proceedings of the COMADEM, Houston TX, December 2000.

COST BENEFIT ANALYSIS MODELS FOR EVALUATION OF VMEP/HUMS PROJECT

Victor Giurgiutiu, Georgiana Craciun, and Andrew Rekers
University of South Carolina, Columbia, SC 29208
Phone: 803-777-0660, FAX: 803-777-0106, E-mail: victorg@sc.edu,
cracigeop4@spanky.badm.sc.edu, and rekers@engr.sc.edu

Abstract: Models for cost benefit analysis (CBA) of vibration monitoring (VM) and Health Usage Monitoring Systems (HUMS) are discussed. A brief overview of CBA methods with examples of its application to large government funded projects is given. The objectives and projected benefits of the South Carolina Army National Guard (SCARNG) Army Aviation Support Facility (AASF) Vibration Management Enhancement Program (VMEP) are briefly reviewed. The cost components associated with this activity at the SCARNG/AASF operational unit are identified and discussed. A list of most costly maintenance parts and operations is given. Possible cost savings and cost differing components are analyzed from CBA perspective. As the implementation of the VMEP project has just started, the last part of the paper presents the projected CBA evaluation results.

Key Words: Cost benefit analysis; Health usage monitoring system; Vibration management enhancement program; VMEP; HUMS; O&S; RT&B; SCARNG.

INTRODUCTION

Vibration Monitoring and HUMS activities are essential for reducing the operational and support (O&S) cost of Military and Civilian helicopters. In recent years, a significant number of VM/HUMS activities have proliferated in order to increase the safety, reduce maintenance cost, and eventually extend the life of existing helicopter fleets. In order for such VM/HUMS activities to be proven cost effective, a CBA must be preformed.

The cost effectiveness of the VMEP/HUMS system usage on the SCARNG AH-64A Apache and UH-60 Blackhawk helicopters is determined using a modification of the RITA-HUMS CBA software. Data to be collected will include aircraft details and operating, maintenance, VMEP/HUMS equipment, and VMEP/HUMS installation cost. The output from the CBAM software includes benefit and cost tables showing impact of in-flight and mission abort for aircraft with and without VMEP/HUMS, maintenance and availability, and an estimation of aircraft maintenance cost. Based on this data, projections are made for future O&S cost. This data is then analyzed to provide a break-even value for the VMEP/HUMS system on the SCARNG helicopter and determine the payback time horizon.

REVIEW OF COST-BENEFIT ANALYSIS METHOD

Currently, cost-benefit analysis (CBA) is largely used by government agencies. This is mainly due to the strong legislative actions taken by the Reagan and Clinton Administrations that issued Executive Orders endorsing the use of CBA. Executive Order 12886 on Regulatory Planning and Review, signed by President Clinton on September 30, 1993 requires agencies to perform cost-benefit analysis of proposed and final regulations. It revoked and replaced two executive orders issued under Reagan Administration: Executive Order 12911 requiring Regulatory Impact Assessment and Executive Order 12498 establishing the regulatory planning process. Moreover, the use of CBA by government agencies was enforced by Congress who enacted numerous statutes requiring agencies to perform CBA analyses.

When used by governmental agencies, CBA attempts to measure, over a relevant time period, the change in societal well-being resulting from the implementation of a governmental project or the imposition of governmental regulations. It can provide information to decision makers on the merits of the current project or regulation as well as offer a framework for comparing a variety of project or regulatory alternatives. Agencies' project or regulation evaluations are subject to the review of the Office of Management and Budget (OMB). In 1992 OMB issued the Circular No. A-94, which recommends the use of CBA in formal economic analyses of government programs or projects and provides general guidance for conducting CBA. Its goal is to "promote efficient resource allocation through well-informed decision-making by the Federal government".

CBA aims to present categories of costs and benefits in terms of dollars (so that the cost-benefit comparison can be performed with a common unit of measurement); therefore, agencies have to define and monetize all categories of costs and benefits determined by the project implementation. Sometimes practical problems appear such as obtaining data, evaluating benefits and costs, etc. Monetization of some benefits categories may be controversial because indirect methods are often employed to estimate a value for goods that are not generally traded in the marketplace (e.g. estimate the monetary value of a reduction in risk of premature mortality). In this sense OMB stipulated, "Analyses should include comprehensive estimates of the expected benefits and costs to society based on established definitions and practices for program and policy evaluation. Social net benefits, and not the benefits and costs to the Federal Government, should be the basis for evaluating government programs or policies that have effects on private citizens or other levels of government. [...] Both intangible and tangible benefits and costs should be recognized. [...] Costs should reflect the opportunity cost of any resources used, measured by the return to those resources in their most productive application elsewhere"(OMB – A-94).

Despite its recognized merit in providing important information and transparency in the governmental decision-making process, CBA was often criticized, especially by American academics who claim that CBA is an analytical technique that deals only with economic efficiency without considering who receives the benefits and who bears the costs. They also claim that CBA sometimes produces morally unjustified outcomes or it is not correctly used. Yet, it is important to highlight that CBA is a decision procedure or a method for achieving desirable results, and "some decision procedures are more

accurate or less costly than others". As long as it is used in the right way, meaning that under certain conditions agencies may need to modify the traditional approach of CBA, this decision procedure is justified if it is less costly than other procedures (e.g. risk-risk analysis, feasibility based assessment, etc.).

In order to place CBA in context, a good example is the Environmental Protection Agency (EPA) monitoring of drinking water contamination with lead. By law, EPA has to regulate the water cleaning against lead contamination. Therefore, EPA used CBA to evaluate three rules it has previously issued as to lead contamination of water. On the cost side EPA took into consideration the cost of treating contaminated water that enters the distribution system; the cost of maintaining water quality (pH level, temperature, etc.); the cost of replacing lead pipes; the cost of warning the public of high lead levels and informing it of precautions; and the cost of monitoring water quality. These costs were put in balance with the health benefits accrued from avoiding hospitalization and medical treatment of contaminated persons and compensatory costs for lost productivity. After aggregating all these costs and benefits, EPA concluded that the total health benefits from corrosion control alone would be $63.8 billion over a twenty-year period, which vastly exceeded estimated costs of $4.2 billion (Adler and Posner, 1999). Thus, with a large amount of data the CBA analysis was very transparent and convincing so it justified the adopted rule. Yet, without justification, EPA did not include in its final CBA the benefits from reducing lead damage to plumbing components, even if these benefits had been evaluated.

Some remarks have to be made. First, budgetary and time constraints may impede EPA, as well as other governmental agencies, from collecting all the necessary data. Second, when all data are available and easy to collect, agencies should try to monetize all costs and benefits and include them in their final CBA. This helps agencies to clearly present the effects of governmental projects and alert affected groups. Third, CBA is an important way for governmental agencies to defend their projects against critics coming from other agencies, as well as against legal and political challenges from affected groups. Finally, given its relative cheapness and transparency, CBA is considered the best procedure for agencies to use in evaluating their projects.

The use of CBA is not limited to governmental agencies. The U.S. Army also employs this technique in estimating whether its projects achieve an improvement in the allocation of resources.
CBA can provide valuable perspectives on the best ways to manage projects concerning the army infrastructure, labor force, capital stock etc. This approach is consistent with the Department of Defense and Army guidance and with the Army Regulation 11-18 establishing responsibilities and policy for the Army's Cost and Economic Analysis Program.

For the design and manufacturing of the helicopter AH-64D Apache Longbow, Boeing Helicopter of Mesa, Arizona put up a multidisciplinary team focused on meeting the Army's cost and performance requirements. This Integrated Product Development (IPD) team incorporated a manufacturing engineer, a design engineer, a tool engineer, and a stress engineer, and later on a material process engineer, purchasing personnel, and an industrial engineer who was called in to perform a CBA. During the project development, the team used the costing software Design for Manufacture and Assembly (DFMA) that

provided "a means of before-and-after comparison – not only against the previous models [six Apache Prototypes] but for individual redesign ideas that are part of the iterative process"(Parker, 1997). Thus through continuous CBA the best alternative was chosen and the new Apache Longbow innovative production strategies not only proved better performance and quality, but also brought savings of $1.3 billion over the life of the program.

MILITARY OPERATION AND SUPPORT COSTS

The Operation and Support (O&S) costs of US Army aviation are considerable. According to Defense Budget documents (DBD, 2000), the US Army spent $1,384M on aircraft in 1999, of which $930M were spent on modifications ($666M on AH-64A and D models), $36M on spare and repair parts, and $104M on support equipment and facilities. The Army flies around 180h/helicopter/year at a cost between $1,483/h for UH-60L Blackhawk to ~$5,000/h for AH-64D Longbow Apache. Of these flight hours, approximately 8% are used in maintenance test flights, including 5% for Rotor Track and Balance (RT&B), and 3% for others.

Table I ARIP AH-64 cost reduction high demand items for 1995-1996. (ILTI, 1996)

	NSN	NAME	CNT	QTY	PRICE	TOTAL COST	%COST	CUM%
1	1615-01-332-0702	BLADE, MAIN	301	468	$99,797	$46,704,996	43.81%	43.81%
2	1615-01-154-7076	STRAP ASSEMBLY	795	2,086	$6,670	$13,913,620	13.05%	56.86%
3	1615-01-312-2387	BLADE, TAIL	221	386	$18,467	$7,128,262	6.69%	63.55%
4	2840-01-345-2584	ROTOR COMPRES.	277	325	$18,933	$6,153,225	5.77%	69.32%
5	3040-01-352-1531	CONNECTING LINK	534	1,171	$4,703	$5,507,213	5.17%	74.49%
6	6115-01-224-9230	GENERATOR-ALT	274	329	$14,521	$4,777,409	4.48%	78.97%
7	1650-01-263-7856	CYLINDER ASSEMBLY	232	325	$9,355	$3,040,375	2.85%	81.82%
8	2835-01-164-5786	CLUTCH, ASSEMBLY	211	237	$12,467	$2,954,679	2.77%	84.59%
9	1615-01-235-5345	HOUSING ASSEMBLY	202	366	$6,674	$2,442,684	2.29%	86.88%
10	3010-01-364-2470	CLUTCH ASSEMBLY	247	496	$4,033	$2,000,368	1.88%	88.76%
11	4320-01-158-0893	AXIAL PUMP	227	313	$5,693	$1,781,909	1.67%	90.43%
		OTHERS				$10,199,239	9.57%	100.00%
					Total cost	$106,603,978		

A statistical study of AH-64 Apache premature failures performed by Innovative Logistics Techniques, Inc. (ILTI, 1996) indicated that 81% of parts removal occurred on some 40 items. Industry and government cooperation in addressing O&S costs improvements with emphasis on readiness drivers, high removal rates, and labor-intensive items, is required. Analysis of 2-year data from the Apache Readiness Improvement Program (ARIP) tracking AH-64 high demand items revealed that out of the ~$106M costs, 90% were expended on 11 high cost/demand items (Table I).

CBA OF HEALTH AND USAGE MONITORING SYSTEMS (HUMS)

The use of Health and Usage Monitoring Systems (HUMS) can significantly reduce the cost of helicopter O&S activities. Detection of incipient failure in critical components can prevent costly aircraft accidents. Whereas life extension and condition based maintenance is expected to significantly reduce the unwarranted replacement of 'healthy" parts. A major effect will come from the reduction of general vibration levels through improved Rotor Track and Balance (RT&B) procedures. Data recorded during the 1994-1996 period at the South Carolina Army National Guard – Army Aviation Support Facility (SCARNG-AASF) indicates that vibration reduction obtained through the implementation of RT&B function of the Aviation Vibrations Analyzer (AVA) could have saved $2M on TADS/PNVS visionics systems alone. At present, SCARNG-AASF is replacing AVA with the more advanced Vibrations Management Enhancement Program (VMEP) technology (Giurgiutiu *et al.*, 2000). Preliminary tests with the VMEP neural networks (NN) algorithms for improved RT&B have already shown 40% reduction in the number of maintenance test flights, and lower vibration levels when compared with existing AVA algorithms. As predicted, considerable O&S cost reduction is expected.

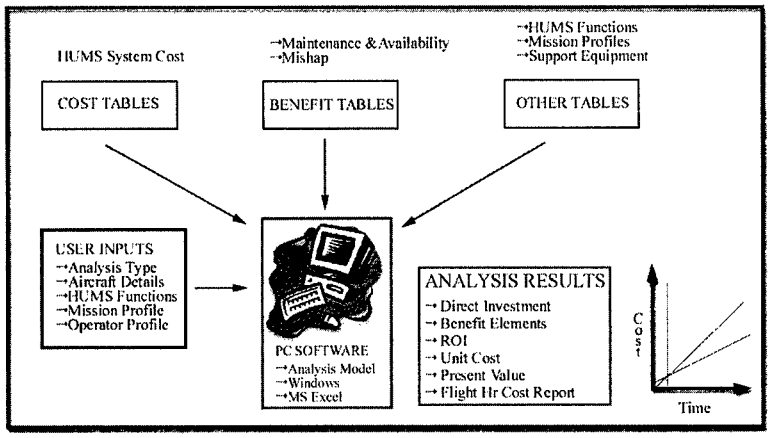


Figure 1 Overall architecture of the RITA HUMS Cost Benefit Analysis Model (RITA, 1998).

A **major obstacle** that prevents the wide spread dissemination of HUMS systems is the lack of irrefutable hard evidence that their implementation will actually reduce the helicopter O&S life-cycle costs and actually save money. Crews (1991) indicated, "many in the helicopter community have long felt that there is a direct relationship between helicopter reliability and maintainability and the level of vibrations allowed on the helicopters. This is a difficult thesis to prove for a number of reasons and skeptics have argued for hard proof that this is indeed true before they would allow significant dollars to be spent on efforts to reduce helicopter vibration." The transition from scheduled overhauls (where early removal of 'perfectly good' parts is practiced) to condition-based maintenance with 'just-in-time' replacements is expected to save considerable O&S costs. To verify this, good statistical models, carefully conducted experiments, and

statistically significant data collected on a sufficiently large sample of service helicopters are needed. Cost-benefit analysis (Nas, 1996) has been used in the past for effectiveness evaluation of space technology (Hein, 1976) and national aviation system (Noah, 1977). Booz-Allen & Hamilton, Inc. developed during 1995-1999 the RITA HUMS Cost and Benefits Analytical Model (CBAM) software (RITA, 1999) under joint funding from the US rotorcraft industry and government (Figure 1). In 1999, USC obtained access to an evaluation copy of the CBAM software for use on the VMEP project. Our analysis is planned as follows:

Thrust 1 **Cost-benefits analysis model for O&S costs at SCARNG-AASF operational unit level.** A model to track O&S costs at operational-unit level is being established for the organizational structure of SCARNG-AASF and its operational costs environment. The model will use the RITA-HUMS CBAM software framework customized to the SCARNG-AASF organizational structure. The model will track the costs associated with the acquisition, installation, and support of the VMEP-HUMS, and associated impact on the unit-level operations. The model will identify the benefits resulting from VMEP-HUMS usage in terms of increased availability, reduced mission aborts and flight mishaps, and reduction of maintenance flights for validation of replaced equipment and RT&B vibration reduction. The result of the cost-benefit analysis will be presented in terms of return on investment (ROI), present value analysis, flight hour cost reports, and a graphic representation of payback point.

Thrust 2 **Conduct statistically designed experiments at SCARNG-AASF in connection with the VMEP program.** Statistical groups of 'control' and 'exposure' helicopters will be established from the SCARNG-AASF AH-64 and UH-60 fleet. The exposure helicopters will be fitted with VMEP-HUMS equipment and will follow the VMEP-HUMS O&S procedures, while the control helicopters will be equipped, operated, and maintained in strict accordance with established Army procedures. Full data monitoring and recording will be performed on both control and exposure helicopters. The O&S data tracked during this experiment will be collected through Unit Level Logistics Systems-Aircraft (ULLS-A) and electronically transferred to the USC data repository for processing, analysis, and interpretation.

Thrust 3 **Process statistical data with data-mining algorithms to establish correlations and O&S costs trends**. Data mining is an artificial-intelligence methodology based on data analysis tools that discover data patterns and relationships suitable for prediction and extrapolation. Using data collected during the initial Thrust 2 experiments, data-mining algorithms will establish O&S cost reduction predictions that will be tested on additional data collected during follow-up Thrust 2 experiments. This iterative approach will assure model robustness and stability. Activity based costing (McDonald et al., 1998) will be used to properly track some non-technical related costs.

Thrust 4 **Verify the level of O&S cost reduction achieved through HUMS vibration management and develop cost-reduction predictions.** A number of overall cost and reliability outcomes will result from the collected data, such as: (a) time between failure (TBF) and time between maintenance action (TBMA) on critical and/or high cost components; (b) inventory costs; (c) flight time allocated to maintenance actions; (d) downtime. Besides these general trends, systematic valuation of the dollar costs and benefits associated with the VMEP program implementation will be applied. These data

will be processed with the CBAM software to reveal ROI, net present value (NPV), payback period, and future O&S cost savings.

EXPECTED BENEFITS OF USING HEALTH AND USAGE MONITORING SYSTEMS

This study will demonstrate that the use of HUMS systems may produce sizable O&S cost savings and improve affordability of Army helicopters. The benefits of the HUMS system, experienced initially by the unit-level maintenance managers, will eventually propagate through the Army Logistics Network. Process improvements at unit level, higher availability, reduced direct operating costs, and avoidance of expensive maintenance flights are the principal significant benefits expected from this study.

RULES OF A GOOD CBA MODEL

There are rules that must be followed in order to have a good CBA model. First the estimates of expected cost and benefits must be provided and clearly defined. Both the intangible and tangible benefits and costs should be included in the analysis. Cost should be defined in terms of opportunity cost and incremental costs. All cost should be inflated or deflated over the life of the analysis. The model needs to provide for the review and modification of all algorithms used in the model. The CBA model needs to calculate recurring and nonrecurring cost as well as be easy to use with good documentation of equations and help menus. These rules will provide for an efficient and well-documented CBA.

RITA-HUMS CBAM SOFTWARE OVERVIEW

There are two main components of the RITA HUMS CBAM Model. The first is the operator's module. The purpose of this module is to identify HUMS functions that provide the most value added and the maximum benefit to the operator. This includes reduction in cost per hour, the payback period, and the total dollar savings over the rotorcrafts' expected life. The second is the manufacture's module. This module gives the ability to create databases from scratch, generic databases, or an existing aircraft database. The analytical side of manufacturer's module includes the development of these aircraft databases to support the evaluation of prospective HUMS implementations. There are two distinct components in this module, the aircraft database and the analysis.

The inputs into the CBAM software include the type of analysis, aircraft details, and a definition of anticipated aircraft usage. The type of analysis includes whether the aircraft is military or commercial. The AH-64 and the UH-60 helicopters fall under the military analysis. The market value, insurance, and operating cost are entered under the input of aircraft details. The definition of the anticipated aircraft usage is inputted at this point in the software.

The outputs of the software include cost, benefit, and other tables as well as a payback graph. The cost tables include the HUMS equipment, installation, and operating cost. The benefit tables include an estimation of aircraft maintenance cost, impact of in-flight and mission abort, maintenance and availability, and mishaps for the aircraft. The individual HUMS functions, mission profiles, and support equipment are listed in other tables in the

output. The analysis results of the RITA HUMS CBAM software include the direct investment, benefit elements, return on investment, unit cost, present value analysis, flight hour cost, and a graphical representation of the payback point.

PROS AND CONS OF THE RITA-HUMS CBAM SOFTWARE

The software is user friendly when it comes to inputting the data, but several problems still exist. There is a problem with acronyms within the software. There is no explanation for the meaning of many of the acronyms involved in the software. The wizard will ask for an entry, but it will not give the definition of the acronym. There is also some confusion on why some of the data is needed in the final CBA. Because the algorithms are not clearly defined or stated, it is difficult to determine which data entries effect what in the final CBA.

One problem with the CBAM software is that it does not have the customization potential that other models have. If a cost concern is thought to affect the final outcome of the model, it is impossible to add this concern to the model. It is also difficult to see how certain inputs affect the cost benefit of the HUMS system. Certain inputs were changed dramatically with unpredictable results in the final analysis.

The help menus are useful in navigating through the input wizard, but offer very little help in the direction taken in the output. There are many pros for this model, but because it does not fit exactly the situation at SCARNG-AASF, it is difficult to get an accurate depiction of the cost savings by using this software.

CBA FOR EVALUATING VMEP

At the University of South Carolina, a research CBA module, using Excel software, has been developed. During our research, it was found that CBA is a useful way of organizing a comparison of different alternatives of a project. It can help the decision maker better understand the implications of a decision. Yet, not all impacts of a decision can be quantified or expressed in dollar terms. (e.g., intangible benefits such as aircraft availability, safety, and moral). Therefore, care should be taken to ensure that quantitative factors do not dominate important qualitative factors in decision-making.

In performing CBA for evaluating VMEP activities, we start from the baseline process and compare it to the VMEP alternative. The cost of the current process at SCARNG-AASF provides the baseline for the CBA. In our case, benefits take the form of savings and non-tangible benefits. Therefore, we first analyze the savings of the VMEP alternative by comparing the costs in the two cases. Then we discuss the non-tangible benefits of the VMEP alternative and their implications. For comparing the VMEP alternative with the baseline, we define common cost elements (Table II).

Table II Cost Elements (costs per aircraft)

	First Year		Second Year		Third Year		Fourth Year		Fifth Year		Sixth Year	
Cost Variables - (per a/c)	**VMEP**	**Baseline**	**VMEP**	**Baseline**	**VMEP**	**Baseline**	**VMEP**	**Baseline**	**VMEP**	**Baseline**	**VMEP**	**Baseline**
RT&B OCCURRENCE RATE - (a/c per year)	12	24	12	24	12	24	10	24	8	24	8	24
MAINTENANCE FLIGHT HOURS - (per a/c RT&B)	6	6	6	6	3	6	3	6	3	6	3	6
VMEP INVESTMENT - ($10K+12K/ a/c)	$22K	-	-	-	-	-	-	-	-	-	-	-
FLIGHT HOUR COST - (not including fuel)	$1.5K	$1.5K	$1.5K	$1.5K	$1.5K	$1.5K	$1.5K	$1.5K	$1.5K	$1.5K	$1.5K	$1.5K
MAINTENANCE OF HUMS ON-BOARD EQUIPMENT	$2K	-	$2K	-	$2K	-	$2K	-	$2K	-	$2K	-
PARTS - (high cost items)	$300K	$300K	$300K	$300K	$290K	$300K	$275K	$300K	$250K	$300K	$225K	$300K
MAINTENANCE FLIGHT HOURS COST - (per a/c RT&B)	$9K	$9K	$9K	$9K	$5K	$9K	$5K	$9K	$5K	$9K	$5K	$9K
OPERATIONAL FLIGHT HOURS COST	$229K	$208K	$229K	$208K	$229K	$208K	$229K	$208K	$229K	$208K	$229K	$208K

Disclaimer: To avoid un-appropriate disclosure of information, the actual dollar values have been modified, while maintaining a realistic order of magnitude.

Table III VMEP Costs for SCARNG-AASF Fleet

Year	VMEP Investment	Operational Flight Hours Costs	Maintenance Flight Hours Costs	Parts	VMEP Maintenance	Annual costs	Discount factor	VMEP Discounted Cost Flows
	I	OC	MC	P	M	AC	DF	AC/DF
1	$396K	$4.1M	$162K	$5.4M	$36K	$10.1M	1.0000	$10.1M
2	-	$4.1M	$162K	$5.4M	$36K	$9.7M	1.0300	$9.4M
3	-	$4.1M	$81K	$5.2M	$36K	$9.5M	1.0609	$8.9M
4	-	$4.1M	$81K	$5.0M	$36K	$9.2M	1.0927	$8.4M
5	-	$4.1M	$81K	$4.5M	$36K	$8.7M	1.1255	$7.8M
6	-	$4.1M	$81K	$4.1M	$36K	$8.3M	1.1593	$7.2M
Total	$396K	$24.8M	$648K	$29.5M	$216K	$55.5M		$51.8M

Table IV Baseline Costs for SCARNG-AASF Fleet

Year	VMEP Investment	Operational Flight Hours Costs	Maintenance Flight Hours Costs	Parts	VMEP Maintenance	Annual costs	Discount factor	Baseline Discounted Cost Flows
	I	OC	MC	P	M	AC	DF	AC/DF
1	-	$3.8M	$162K	$5.4M	-	$9.3M	1.0000	$9.3M
2	-	$3.8M	$162K	$5.4M	-	$9.3M	1.0300	$9.0M
3	-	$3.8M	$162K	$5.4M	-	$9.3M	1.0609	$8.8M
4	-	$3.8M	$162K	$5.4M	-	$9.3M	1.0927	$8.5M
5	-	$3.8M	$162K	$5.4M	-	$9.3M	1.1255	$8.3M
6	-	$3.8M	$162K	$5.4M	-	$9.3M	1.1593	$8.0M
Total	-	$22.5M	$972K	$32.4M	-	$55.9M		$52.0M

Yet, we cannot properly compare the two competing alternatives if we do not convert them to a common unit of measurement. Therefore, we discount future dollar values to a present value (also referred to as the discounted value). Present values are cash flows that occur now or in the immediate future and may include start-up expenses (VMEP investment) as well as any other expenses or incomes that occur at or close to the beginning of the project. Future values are the cash flows that occur sometime in the future. By converting all the future values to present values, we perform a present value analysis that will tell us what our project is worth in equivalent dollars right now. The formula we use is: $PV=FV/(1+I)^{(n-1)}$, where PV=Present Value, FV=Future Value, I=Interest Rate, and n=number of years. Tables III and IV show the annual discounted costs for both alternatives. For exemplification, we have chosen a six-year period of analysis. Cost data have been collected for estimating the costs and savings of each of the two project alternatives (baseline and VMEP) for each year of analysis. The annual costs are discounted to reflect the dollar depreciation, based on an interest rate of 3%. This interest rate was chosen based on public information about the present trends in the US economy. Figure 2 shows that in the first three years the VMEP project is more costly than the baseline, due to an initial investment of $22k/aircraft and additional VMEP

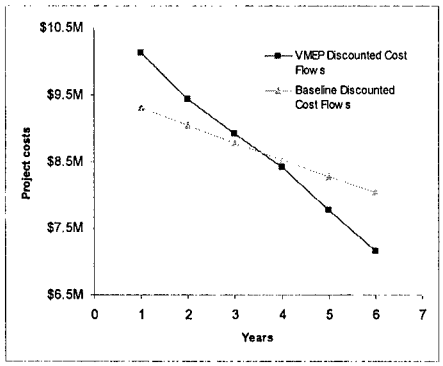


Figure 2 Annual discounted costs for VMEP and baseline alternatives

maintenance costs. However, after three and a half years the VMEP project attains the break-even point. From that point on the VMEP project costs fall sharply below the baseline costs and, consequently, savings are increasing. Furthermore, at the end of the 6-year period of analysis, the cumulative discounted cost flow for the VMEP alternative falls bellow the same cost for the baseline. Thus, a positive present value of $149,000 shows that the VMEP alternative is favorable.

The benefit variables in our analysis cannot be linked directly to a monetary value like the cost variables. They do ultimately affect the overall monetary value of the VMEP project, but cannot be linked to a dollar figure in the same way the cost variables are linked. Instead, the availability and safety variables are set up using a percentile comparison. A numeric tally is used to compare the premature parts failure, mission aborts, and the unscheduled maintenance occurrence. Moral could not be quantified in a normal scale. It is quantified by an increase in the specified year. The benefits of the VMEP project need to be looked at as non-tangible benefits and not necessarily a monetary gain. Therefore, CBA is important when VMEP and HUMS activities are essential for reducing O&S cost of Military and Civilian helicopters.

REFERENCES

Adler, Matthew; Posner, Eric. (1999) "Rethinking cost-benefit analysis", University of Chicago Law School, John M. Olin Law & Economics Working Paper No. 72, May27, 1999

Army Regulation 11-18. (1995) The Cost and Economic Analysis Program. 31 Jan. 1995

Crews, S. T. (1991) "Helicopter Vibration and its Effect on Operating Costs and Maintenance Requirements", AVCOM, Huntsville, AL, 1991.

Dasgupta, Ajit K.; Pearce, D.W. (1972) Cost-benefit analysis: theory and practice. London, MacMillan, 1972, pp 12-13.

DBD (2000) "FY 2000 Defense Budget Documents", Office of the Assistant Secretary of the Army for Financial Management and Comptroller, http://www.fas.org/man/docs/fy00/topics.htm

Exec. Order No. 12886, 3 C.F.R. 638, 639 (1993).

Giurgiutiu, V.; Grant, L. Grabill, P.; Wroblewski, D. (2000) "Helicopter Health Monitoring and Failure Prevention through Vibration Management Enhancement Program", Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, May 1-4, 2000, Virginia Beach, VA pp. 593-602.

Hein, G. F. (1976) "Cost Benefit Analysis of Space Technology", NASA, 1976

ILTI (1996) "Premature Failures", PAT Meeting, 7-9 May 1996, Innovative Logistics Technologies, Inc.

McDonald, D.; Cushman, B.; McNabb, S. (1998) "Activity Based Costing", Booz-Allen & Hamilton, Inc., DAL Conference, September 30, 1998, Seattle, WA.

Milford, S.; McNabb, S.; Sanders, T.; Phillips, K. (1998) "DAL-P Cost/Benefit Analysis", Booz-Allen & Hamilton, Inc., DAL Conference, September 30, 1998, Seattle, WA.

Nas, Tevfik (1996)"Cost Benefit Analysis: Theory and Practice", Sage Pub., 1996.

Noah, J. W. (1977) "Cost Benefit Analysis of the National Aviation System", FAA, 1977.

Office of Management and Budget. (1992) Circular A-94 Revised, Oct. 29, 1992

Parker, M. (1997) Mission Possible. Improving design and reducing costs at Boeing Helicopter. IIE Solutions, vol. 29, Dec. 1997, pp. 20-24

Pearce, D.W.; Nash, C.A. (1981) The social appraisal of projects: a text on cost-benefit analysis. New York: John Wiley & Sons, 1981, pp 3-4.

Porter, Theodore M. (1995) Trust in numbers: the pursuit of objectivity in science and public life. Princeton, N.J.: Princeton University Press, 1995, pp 148-89.

RITA (1999) "RITA-HUMS Cost Benefits Analytical Model (CBAM) software", Rotorcraft Industry Technology Association, Inc., 1999.

Unfunded Mandates Reform Act (P.L. 104-4) (1995)

FEATURE EXTRACTION

Chair: Mr. Eddie C. Crow
Pennsylvania State Univ/ARL

EXTRACTION OF BEARING FAULT TRANSIENTS FROM A STRONG CONTINUOUS SIGNAL VIA DWPA MULTIPLE BAND-PASS FILTERING

Dr. Joshua Altmann

STI Technologies
1800 Brighton-Henrietta TL Rd.
Rochester, New York, USA
Ph: (716) 424-2010
E-mail: jaltmann@sti-tech.com

Prof. Joseph Mathew

School of Mechanical, Manufacturing and Medical Engineering
Queensland University of Technology,
Brisbane, Australia

Abstract: This paper presents a new method to enhance the detection and diagnosis of rolling element-bearing faults based on discrete wavelet packet analysis (DWPA). The extraction of attenuated resonant vibrations due to impacts from localized faults in rolling element bearings is normally achieved by high-pass or band-pass filtering of the vibration signal. The main problem with this approach is the difficulty in choosing an appropriate filter range of interest. This is a serious obstacle when the bearing fault transients are buried in high levels of noise or contaminating signals. An alternative that enables the automation of the selection process and the inclusion of multiple frequency bands of interest is presented. A superior signal to noise ratio is achieved in comparison to either high-pass or band-pass filtering of the signal, as the DWPA feature extraction facilitates the equivalent of automatically selecting an optimal multiple band-pass filter.

Key Words: Adaptive Network-based Fuzzy Inference Systems, Condition Monitoring, Demodulation, Wavelet Analysis, Vibration Analysis

Introduction: The method for the extraction of high frequency transients due to bearing impact resonance is achieved at an optimal time-frequency resolution via best basis discrete wavelet packet analysis (DWPA) representation, using the Daubechies-20 wavelet [1]. Selection of the frequency band or bands of interest is achieved by analyzing the characteristics of the wavelet packets. The selection process is automated through the use of an adaptive network-based fuzzy inference system, thus removing the need for the analyst to manually identify the bands of interest. A superior signal to noise ratio is achieved in comparison to either high-pass or band-pass filtering of the signal, as the DWPA feature extraction facilitates the equivalent of automatically selecting an

optimal multiple band-pass filter. Further enhancement of the signal-to-noise ratio is achieved through hard-thresholding of the wavelet coefficients prior to reconstruction of the final signal. The main limitation of this technique is the increased computational time required over standard filtering approaches, restricting its suitability to signal conditions that are not favorable for standard demodulation due to high levels of contamination from external sources. This constraint should be able to be overcome with the implementation of parallel wavelet based digital signal processors.

This paper briefly introduces the extraction technique implemented, and then examines the performance of the DWPA multiple band-pass filtering for the extraction of a low speed rolling element bearing transient from a strong continuous signal.

Adapative Network Based Fuzzy Inference System (ANFIS): The adaptive network based fuzzy inference system, which is utilized in this investigation for wavelet packet feature extraction, is a transformational model of integration where the final fuzzy inference system is optimized via artificial neural network training. This method of integration was selected due to its ability to incorporate expert knowledge and maintain system transparency, while allowing tuning of the fuzzy inference system via neural training to ensure a satisfactory performance. The validity of the expert knowledge and the suitability of the input data chosen can then be verified by examining the structure and the performance of the final fuzzy inference system. This section describes the design and operation of an adaptive network based fuzzy inference system.

Jyh-Shing Roger Jang introduced the adaptive network based fuzzy inference system (ANFIS) in 1993 [2]. The initial membership functions and rules for the fuzzy inference system can be designed by employing human expertise about the target system to be modelled. ANFIS can then refine the fuzzy if-then rules and membership functions to describe the input-output behavior of a complex system. Jang showed that even if human expertise is not available it is possible to set up intuitively reasonable membership functions and then employ the neural training process to generate a set of fuzzy if-then rules that approximate a desired data set.

Sugeno type fuzzy inferences systems have been used in most adaptive techniques for constructing fuzzy models, due to their more compact and computationally efficient representation of data than the Mamdani or Tsukamato fuzzy systems. A typical fuzzy rule in a zero-order Sugeno fuzzy system has the form:

$$\textit{If } x \textit{ is } A \textit{ and } y \textit{ is } B \textit{ then } z = c$$

Where A and B are fuzzy sets in the antecedent, and z is a crisply defined function in the consequent. It is frequently the case that the singleton spike of the crisply defined consequent is completely sufficient to cater for a given problem's needs. If required the more general first-order Sugeno can be employed by setting the consequent to $z = px + qy + c$. Higher order Sugeno systems add an unwarranted level of complexity, for minimal remuneration. A zero-order Sugeno fuzzy inference system is used in this

investigation. The equivalent ANFIS architecture for a Sugeno fuzzy inference system is illustrated in figure 1.

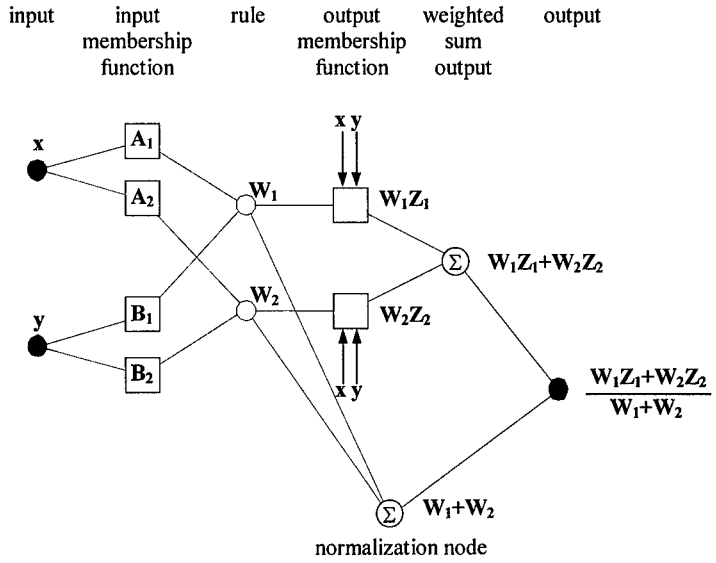


Figure 1: ANFIS zero-order Sugeno fuzzy model

Although any feed forward network can be used in an adaptive network based fuzzy inference system, Jang and Sun implemented a hybrid learning algorithm that converges much faster than training reliant solely on a gradient decent method [3]. During the forward pass, the node outputs advance until the output membership function layer, where the consequent parameters are identified by the least-squares method. The backward pass uses a back propagation gradient decent method to upgrade the premise parameters, based on the error signals that propagate backward. Under the condition that the premise parameters are fixed, the consequent parameters determined are optimal. This reduces the dimension of the search space for the gradient decent algorithm, thus ensuing faster convergence. This hybrid learning system is used in the training of the fuzzy inference systems used for bearing fault feature extraction.

Table 1: LSE-Back propagation Hybrid Learning System

	Forward Pass	Backward Pass
Premise Parameters	Fixed	Gradient Descent
Consequent Parameters	Least-Squares Estimate	Fixed
Signals	Node Outputs	Error Signals

Implementation of ANFIS for Automatic Feature Extraction: Before training sets are fed into the neuro-fuzzy network, suitable input parameters to train the network must be selected. It is necessary for the parameters chosen to enable the neuro-fuzzy network to make an intelligent extraction of the wavelet packets containing bearing fault related information. The chosen parameters must not only provide a robust foundation for the identification of wavelet packets of interest, but they must also be limited in number so as to avoid the *curse of dimensionality*. This refers to the explosion in the number of rules that occurs when the number of inputs is moderately large. The input parameters that were chosen for this process were kurtosis and the spectrum peak ratio (SPR).

Kurtosis is an effective measure of the spikiness of a signal. A high kurtosis level indicates the wavelet packet is impulsive in nature, as would be expected from a wavelet packet that contains bearing fault related features. Kurtosis is defined as,

$$Kurtosis = \frac{1}{NS_y^{\,4}} \sum_{i=1}^{N} \left(y(i) - \mu_y \right)^4 \tag{1}$$

Kurtosis was chosen over other measures of spikiness (crest factor, impulse factor and shape factor) due to its statistically robust nature. A ceiling on the kurtosis values was set at 100, as all values above this are considered exceedingly spiky.

The spectrum peak ratio was defined by Shiroishi [4] as the sum of the peak values of the defect frequency and its harmonics, divided by the average of the spectrum. Shiroishi used the spectrum peak ratio as a trending parameter to indicate the presence of localized bearing defects, which was found to be more robust than considering just the defect frequency.

$$SPR = \frac{N \times \sum_{h=1}^{n} P_h}{\sum_{k=1}^{N} A_i} \tag{2}$$

P_h is the amplitude of the peak located at the defect frequency harmonic, Ai is the amplitude at any frequency, and N is the number of points in the spectrum. In order to differentiate between wavelet packets belonging to different classes of bearing faults, three auto-regressive based peak ratios are employed, spectrum peak ratio inner (SPRI), spectrum peak ratio outer (SPRO) and spectrum peak ratio rolling element (SPRR). Calculation of the spectrum peak ratios was based on Yule-Walker auto-regressive spectral estimates of the signal using a model order of 125, equivalent to one shaft revolution. Removal of outliers in the values of P_h was used to further improve the robustness of this measure.

A total of 2048 wavelet packets were available out of the vibration data collected from a low speed rolling-element bearing test rig. These wavelet packets were individually assessed as to whether they contained bearing related fault features by visual examination of their time series and auto-regressive spectrum. They were then categorized for each fault class as containing fault related features (1), probably containing fault related features (0.66), probably not containing fault related features (0.33), or not containing fault related features (0). The wavelet packet data set included 402 containing inner race

fault defect information, 138 containing rolling element fault information and 64 containing outer race fault information. An additional 762 wavelet packets were created using mathematical models of bearings containing localized faults [5]. The additional wavelet packet data set included 42 containing inner race fault defect information, 83 containing rolling element fault information and 98 containing outer race fault information. The wavelet packets were split into three data sets, a training data set of 1000 wavelet packets, a checking data set of 1000 wavelet packets and a testing data set of 810 wavelet packets.

Given the training and checking input/output data sets, the membership function parameters were adjusted using a back propagation algorithm in combination with a least squares method. The checking data was used to cross-validate and test the generalization capability of the fuzzy inference system. This was achieved by testing how well the checking data fits the fuzzy inference system at each epoch of training, and the final membership functions were associated with the training epoch that has a minimum checking error. This was an important task, as it ensured that the tendency for the fuzzy inference system to over fit the training data, especially for a large number of epochs, was avoided.

Example of Bearing Fault Transient Extraction: The example presented in this paper illustrates the performance of DWPA multiple band-pass filtering for the extraction of bearing fault-related components from a signal principally composed of a continuous sinusoidal signal and its odd harmonics. The example is based on the low speed rolling-element bearing test rig and involves a rolling-element fault of width 0.38mm, an operating speed of 60 rpm and a radial loading of 15 kN.

As depicted in Figure 2, an external signal constructed from a sinusoid (50Hz) amplitude modulated by a low frequency carrier wave at 3Hz clearly dominates the signal. Any transient vibrations due to the bearing fault are well buried by the external signal, with no evidence of a bearing fault apparent when examining the time and frequency (linear and dB) domains.

Digital demodulation is employed in figure 3 in an attempt to isolate bearing fault transients from other signal components present. Figures 3(a+c) illustrate the signal after high-pass filtering at 500Hz and the corresponding enveloped auto-regressive spectrum. The choice of 500Hz for the high-pass filter is based on the theoretical outer race resonance frequency of 696Hz. As the example in question is operating at a low speed (60 RPM) auto-regressive spectral analysis of the enveloped signal is employed. This is due to the relatively short data length (500 points) available after enveloping of the high-pass filtered signal. Although the presence of a rolling-element bearing fault can be ascertained from the demodulated signal, a significant sinusoidal component remains. This is due to the presence of harmonics of the sinusoidal signal above 500Hz, as shown in figure 2(c).

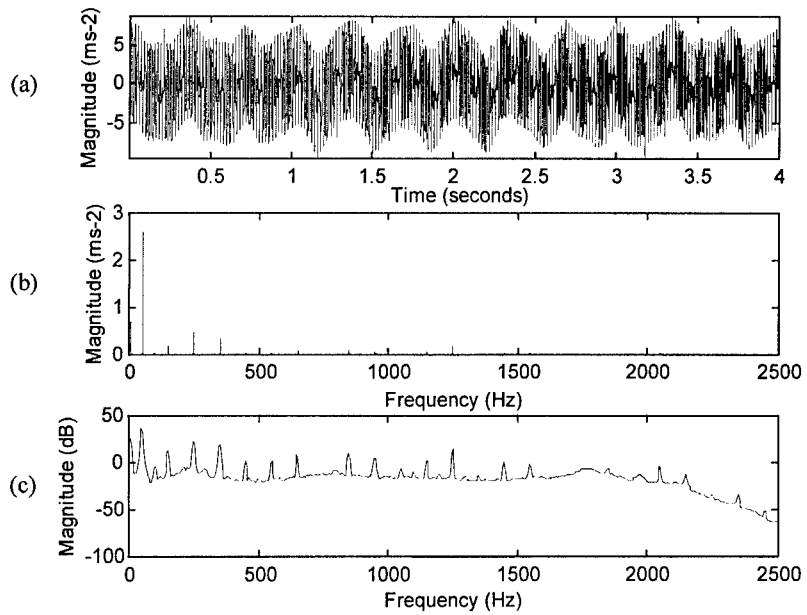


Figure 2: Time domain signal: (a) Continuous sinusoidal signal (50Hz) amplitude modulated by a low frequency carrier wave (3Hz), and a low amplitude rolling-element fault (fault width 0.38mm), (b) Linear FFT, (c) Power Spectral Density.

A band-pass filter was constructed in order to curtail the contamination from the harmonics of the 50 Hz sinusoid. Manual optimization of the filter led to a band-pass filter range of 1500Hz>BPF>2500Hz. As is evident from a comparison of figures 3(a) and 3(b) the band-pass filtered signal results in a much cleaner signal. Unfortunately there is a trade-off, with regions of bearing resonance below 1500Hz being discarded.

DWPA multiple band-pass filtering offers an alternative that surmounts the problem of extracting regions of bearing resonance that are intertwined with continuous signals. Figure 4 illustrates how this method facilitates the extraction of bearing fault-related components from a signal while rejecting the unwanted harmonics. The wavelet packets identified by the adaptive network-based fuzzy inference system as containing bearing fault-related features are indicated. To visualize the rejection of wavelet packets containing unwanted continuous signal components, the power spectral density is plotted along the vertical axis of the DWPA representation. Wavelet packets that contained the harmonic peaks present in the power spectral density plot were rejected by the adaptive network-based fuzzy inference system as containing excessive levels of signal contamination. This clearly demonstrates the ability of DWPA multiple band-pass filtering to extract only the wavelet packets composed predominately of bearing fault-related vibrations.

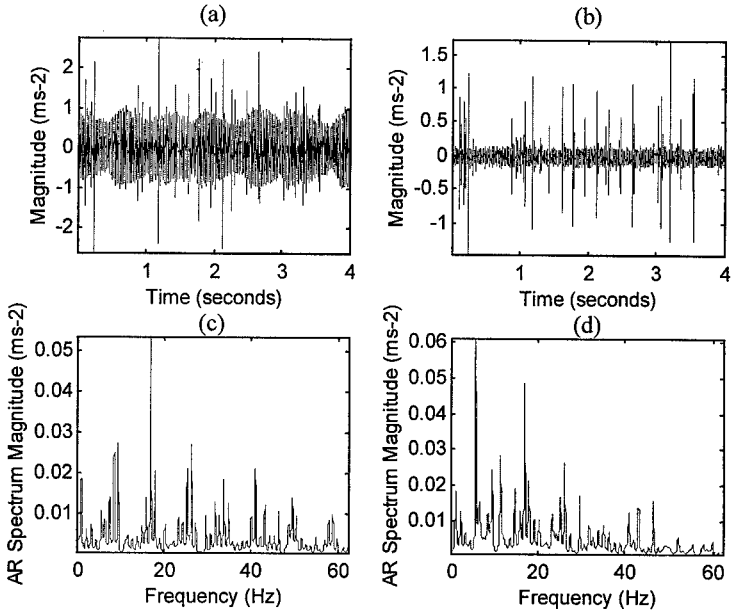


Figure 3: (a) High-pass filtered signal (>500Hz), (b) Best possible band-pass filtered signal that could be achieved (1500Hz>BPF>2500Hz),
(c+d) The corresponding enveloped AR-spectrum (Yule-Walker, model order 125)

Figure 5(a) depicts the reconstructed signal from the extracted wavelet packets. The reconstructed signal has a marginally lower level of sinusoidal contamination than the best possible band-pass filter, and the bearing fault-related transients are also stronger. This is reflected in the enveloped AR spectra shown in figure 5(c), where the peak at the rolling-element fundamental fault frequency is more than twice the magnitude of the corresponding peaks for the high and band-pass filter based spectra (compare figures 3(c+d) with figure 5(c)). Hard threshold de-noising (figures 5(b+d)) almost eliminated the remaining polluting sources of vibration. This further enhances the ability of DWPA multiple band-pass enveloped spectra to accurately diagnose the location and magnitude of bearing defects.

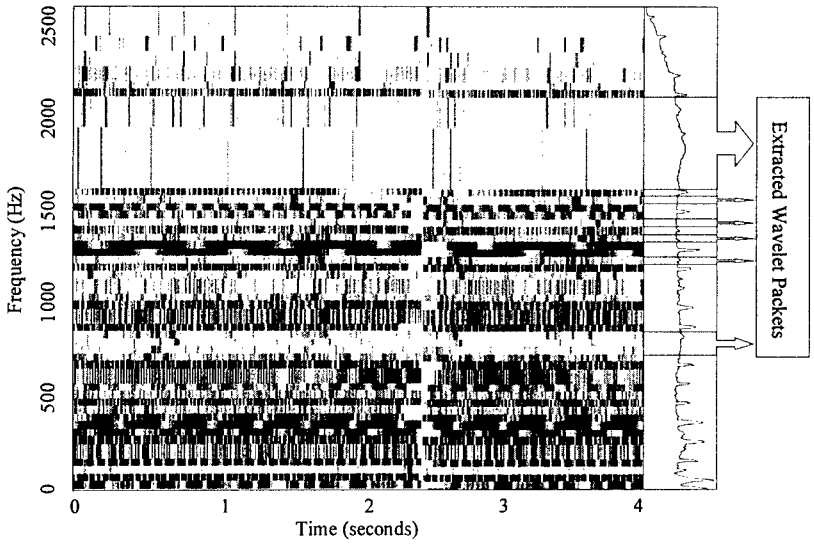


Figure 4: DWPA representation of the vibration signal and the wavelet packets selected by ANFIS as containing bearing fault-related components. The extracted wavelet packets are [(3,7)(4,10)(6,17)(6,26)(6,27)(6,30)(6,49)(6,50)(6,53)].

Concluding Remarks: In conclusion, DWPA multiple band-pass filtering performs admirably at the task of extracting bearing fault-related transients from a signal composed predominately of continuous sinusoidal components. This holds true even when unwanted contaminants from continuous components are contained within the bearing's regions of resonance. The adaptive scheme of the DWPA enables the contaminants to be isolated in wavelet packets of high frequency resolution and then discarded by the adaptive network-based fuzzy inference system. This results in the extracted bearing fault signal being relatively untainted by contamination compared to conventional filtering approaches. Used in conjunction with auto-regressive spectral analysis, the technique should provide vastly enhanced diagnostic capabilities compared to standard demodulation for low speed rolling element bearings.

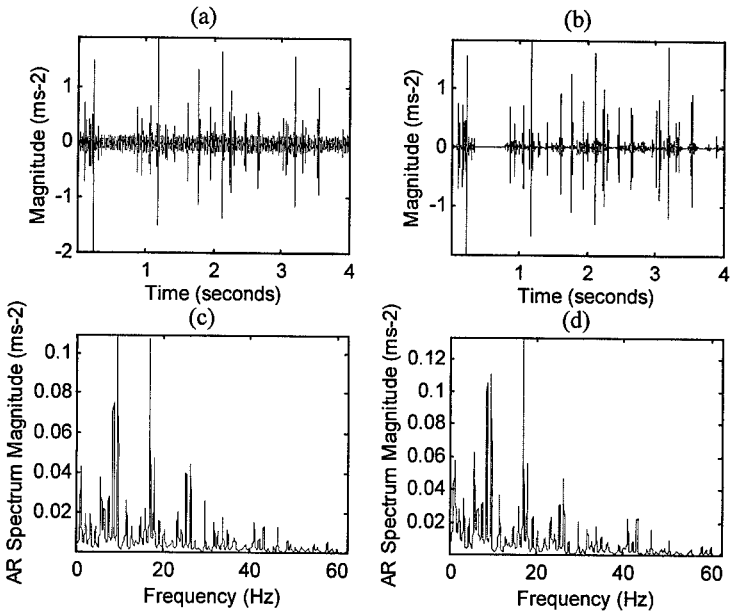


Figure 5: (a) Reconstruction of extracted wavelet packets,
(b) Hard-threshold de-noised reconstruction of wavelet packets,
(c+d) The corresponding enveloped AR-spectrum (Yule-Walker, model order 125)

References

[1] Altmann, J; Mathew, J, "Automated DWPA Feature Extraction of Fault Information from Low Speed Rolling Element Bearings", A-PVC Proceedings, December 1999.

[2] Jang, J.-S. R., "ANFIS: Adaptive Network based Fuzzy Inference Systems", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No. 3, pp. 665-685, May 1993.

[3] Jang, J.-S. R.; Sun, C.-T., "Neuro-Fuzzy Modelling and Control", Proceedings of the IEEE, March 1995

[4] Shiroishi, J.; Li, Y.; Liang, S.; Kurfess, T.; Danyluk, S., "Bearing Condition Monitoring via Vibration and Acoustic Emission Measurements", Mechanical Systems and Signal Processing, Vol. 11, No. 5, pp. 693-705, September 1997.

[5] Altmann, J.; Mathew, J., "Analytical Modelling of Vibrations due to Localised Defects in Rolling Element Bearings", COMADEM Proceedings, pp. 31-40, December 1998.

MINIMIZING LOAD EFFECTS ON NA4 GEAR VIBRATION DIAGNOSTIC PARAMETER

Paula J. Dempsey and James J. Zakrajsek
National Aeronautic and Space Administration
Glenn Research Center
Cleveland, Ohio 44135

Abstract: NA4 is a vibration diagnostic parameter, developed by researchers at NASA Glenn Research Center, for health monitoring of gears in helicopter transmissions. NA4 reacts to the onset of gear pitting damage and continues to react to the damage as it spreads. This research also indicates NA4 reacts similarly to load variations. The sensitivity of NA4 to load changes will substantially affect its performance on a helicopter gearbox that experiences continuously changing load throughout its flight regimes. The parameter NA4 has been used to monitor gear fatigue tests at constant load. At constant load, NA4 effectively detects the onset of pitting damage and tracks damage severity. Previous research also shows that NA4 reacts to changes in load applied to the gears in the same way it reacts to the onset of pitting damage. The method used to calculate NA4 was modified to minimize these load effects. The modified NA4 parameter was applied to four sets of experimental data. Results indicate the modified NA4 is no longer sensitive to load changes, but remains sensitive to pitting damage.

Key Words: Damage assessment; Damage detection; Gears; Health monitoring; Oil debris monitor; Pitting fatigue; Transmissions; Vibration

Introduction: Although various techniques exist for diagnosing damage in helicopter transmissions, the method most widely used involves monitoring vibration. Numerous algorithms have been developed for the processing of vibration data collected from gearbox accelerometers to detect when gear damage has occurred. One of these algorithms, NA4, was developed to detect the onset of gear damage and to continue to react to the damage as it spreads [1]. NA4 is a dimensionless parameter with a nominal magnitude of approximately 3. When pitting damage occurs, the magnitude of NA4 shows a significant increase above 3. Unfortunately, NA4 responds similarly to load changes. The sensitivity of NA4 to even minor changes in load has been documented in several research papers [2,3]. The magnitude of NA4 reacts to changes in load since the load change affects the running average in the denominator of this algorithm. When using this algorithm to detect gear pitting damage on helicopter gearboxes in different flight regimes, the load effect on this algorithm must be minimized. The goal of this research was to minimize the effect of load on vibration diagnostic parameter NA4 while maintaining its sensitivity to pitting damage.

Apparatus and Test Procedure: Experimental data was recorded from tests performed in the Spur Gear Fatigue Test Rig at NASA Glenn Research Center [4]. Figure 1 shows the test

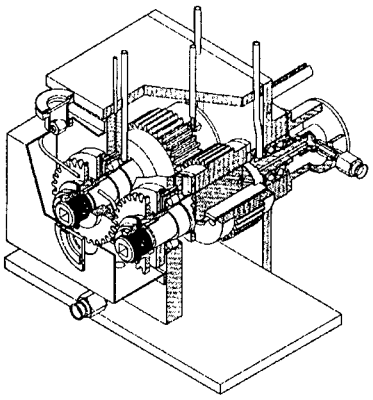


Figure 1.—Spur gear fatigue test rig.

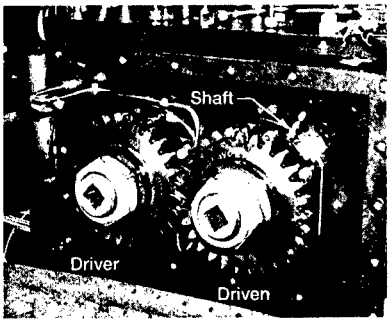


Figure 2.—Accelerometer location on spur gear fatigue test rig.

apparatus. Operating on a four-square principle, the shafts are coupled together with torque applied by a hydraulic loading mechanism that twists two shafts with respect to one another. The power required to drive the system is that to overcome friction losses in the system [5]. The test gears are standard spur gears having 28 teeth, a 3.50-in.-pitch diameter, and a 0.25-in.-face width.

Data was collected using vibration, speed, and pressure sensors installed on the test rig. Vibration was measured on the housing near a shaft support bearing using a miniature, lightweight, piezoelectric accelerometer. The location of this sensor is shown in Fig. 2. This location was chosen based on an analysis of optimum accelerometer location for this test rig [6]. Gear rotation and speed was measured by an optical sensor that creates a pulse signal for each revolution of the gear. Hydraulic pressure to the loading device was measured using a capacitance pressure transducer. Shaft torque is proportional to the pressure. The measured pressure will be referred to as load pressure in this report.

Data was also collected from an oil debris monitor (ODM). The ODM is installed on the rig to give another indication when pitting damage occurs [7]. Oil debris data was collected using a commercially available oil debris sensor that measures the change in a magnetic field caused by passage of a metal particle. The amplitude of the sensor output signal is proportional to the particle mass. The sensor measures the number of particles, determines their approximate size (125 to 1000 microns), and calculates an accumulated mass [8]. The ODM was used to automatically shut down the rig when the accumulated mass measured by the monitor exceeded a preset limit.

Speed, pressure, ODM, and raw vibration data were collected and processed in real time using the program ALBERT, Ames-Lewis Basic Experimentation in Real Time, codeveloped by NASA Glenn and NASA Ames. Pressure data was recorded once per minute. Vibration and speed data was sampled at 200 kHz for a 1-sec duration every minute. Vibration algorithm NA4 was calculated from this data and recorded every minute. Vibration algorithm FM4 was also calculated from this data. FM4 is a widely recognized vibration algorithm developed to detect changes to the vibration pattern resulting from damage to a limited number of teeth. FM4 is a nondimensional number independent of load and speed [7, 9,10].

Gears are run until initial pitting occurs on two or more teeth. Pitting is a fatigue failure of the gear material on or near the surface induced by repeated contacts. Pitting is documented by a video inspection system installed on the rig capable of following the progression of gear pitting while avoiding the need to remove the gearbox cover. The gears were inspected periodically based on a limit set on the ODM. For the purpose of this paper, different levels of pitting must be defined. Due to the limited resolution of the video camera, only wear and two levels of pitting could be monitored; initial and destructive. Initial pitting could not be verified until inspection at completion of the experiment. For the purpose of identifying the damaged gear, the gears are referred to as "driver" and "driven" as shown in Fig. 2.

Vibration Diagnostic Parameter NA4: The method used to calculate NA4 is published in several research papers and will be discussed in the following paragraphs [2,11]. The first step in calculating NA4 is to calculate the time-synchronous average of the raw vibration data. Signal time-synchronous averaging is used to extract waveforms synchronous with gear rotation from the total vibration signal. Vibration data is sampled at 200 kHz for a 1-sec duration and is then averaged synchronous to gear rotation. The desired signal which is synchronous with the gear rotation will intensify relative to the nonsynchronous signals. This time synchronous average signal is used to calculate NA4.

Several statistical and filtering operations are used to calculate NA4. First, the regular gear-meshing components are filtered from the signal resulting in a residual signal. The regular gear-meshing components are the shaft and gear-meshing frequencies and their harmonics. Variance and kurtosis are then calculated from the residual signal. The numerator, kurtosis, the fourth moment of a probability density function, is used to indicate when the distribution is more peaked than a normal distribution.

The denominator is the square of the average variance, the mean value of the variance of all previous readings in the run ensemble [11]. The NA4 is calculated as follows:

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

where

r = residual signal = shaft and meshing frequencies and their harmonics removed from Fast Fourier Transform (FFT) of time-synchronous-averaged signal

$\bar{r}$ = mean value of residual signal

N = total number of interpolated data points per reading

i = interpolated data point number per reading

M = current reading number

j = reading number

A change to the calculation of NA4 is required to minimize the effect of a fluctuating load on NA4. This change, NA4 reset, is made when the load increases or decreases by a given

percentage. For this application, a 10 percent load change was used. For NA4 reset, when the load changes by 10 percent, the denominator resets to the square of the variance of the same reading, and a new average variance is calculated starting with the reading measured when the load changed. Each time the load changes by 10 percent, the first reading in the average variance resets to the first reading when the load changed. This first reading is calculated as follows:

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

where

r = residual signal = shaft and meshing frequencies and their harmonics removed from FFT of time-synchronous-averaged signal

$\bar{r}$ = mean value of residual signal

N = total number of interpolated data points per reading

i = interpolated data point number per reading

This denominator for the readings that follow is calculated as the square of the average variance, the mean value of the variance of all previous readings starting with the first reading when the load changed. Each time the load changes ±10 percent, the denominator is reset by using Eq. (2) for the initial reading.

In addition to load changes, NA4 was also sensitive to restarts after the test rig was shut down. The shutdowns are logged automatically in the data acquisition system during each experiment. This information was used to calculate NA4 reset when the rig was restarted after a shutdown.

Discussion of Results: The analysis discussed in this section is based on data collected during four experiments, three of which pitting damage occurred. The first experiment was to verify the effect of load on the NA4 parameter. The load was increased and decreased with NA4 calculated from the vibration data. The gear set had no evidence of pitting before or after the test. A plot of load pressure, NA4, and NA4 reset for the first experiment is shown in Fig. 3. Data was collected every minute; therefore, the reading number is equivalent to minutes. Since the shaft speed is 10 000 revolutions per minute, the reading number can also be interpreted as mesh cycles equal to the reading number times 10^4.

As discussed previously, NA4 reset is the same as NA4 except the average variance in the denominator is reset each time the load fluctuated by 10 percent. From this plot, the sensitivity of NA4 to changes in load can be easily observed. NA4 appeared to track load pressure. The plot of NA4 reset shows that applying this technique minimizes the sensitivity of NA4 to load.

Although the sensitivity of NA4 to load changes can be corrected by resetting the denominator, one must verify that applying this technique does not significantly decrease the

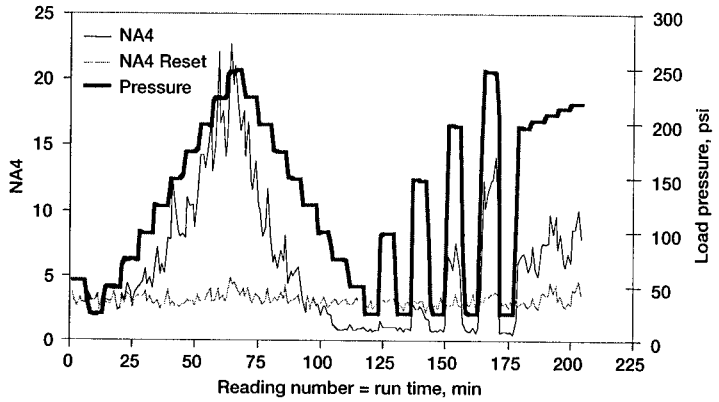


Figure 3.—Data from experiment 1 illustrating load effects.

TABLE I.—DAMAGE DESCRIPTION FOR EXPERIMENT 2

Reading number run time (min)	Damage description	Teeth damaged on driver gear	Teeth damaged on driven gear
60	Run-in wear	All	All
120	Run-in wear	All	All
1581	Run-in wear	All	All
10622	Run-in wear	All	All
14369	Wear Destructive pitting	All 6	All 6
14430	Wear Destructive pitting	All 6	All 6
14512	Wear Destructive pitting	All 6, 7	All 6, 7
14688	Wear Destructive pitting	All 6, 7	All 6, 7
14846	Wear Destructive pitting	All 6, 7	All 6, 7
15136	Wear Initial pitting Destructive pitting	All teeth All teeth 6, 7, 8	All teeth 6, 7, 8

sensitivity of NA4 to pitting damage. Data from three experiments when pitting damage occurred and the load fluctuated was used to verify resetting the denominator of NA4 did not decrease its sensitivity to pitting damage. Descriptions of the pitting damage that occurred during these three experiments are listed in Tables I to III. Photos of damage progression on a selected tooth during each experiment are shown in Figs. 4 to 6. The test gears are run offset to provide a narrow effective face width to maximize gear contact stress. Damage levels are described as follows:

(1) Wear—Layers of metal uniformly removed from the surface
(2) Initial Pitting—Pits of the initial type are less than 1/64 in. in diameter and cover less than 25 percent of the tooth contact area
(3) Destructive Pitting—Destructive pitting is more severe with pits greater than 1/64 in. in diameter and cover greater than 25 percent of the tooth contact area

TABLE II.—DAMAGE DESCRIPTION FOR EXPERIMENT 3

Reading number run time (min)	Damage description	Teeth damaged on driver gear	Teeth damaged on driven gear
0			
1573	Run-in wear	All	All
2199	Wear	All	All
	Destructive pitting		11
2296	Wear	All	All
	Destructive pitting		10, 11
2444	Wear	All	All
	Initial pitting	All	10, 11, 14
	Destructive pitting	10, 11	10, 11, 14

TABLE III.—DAMAGE DESCRIPTION FOR EXPERIMENT 4

Reading number run time (min)	Damage description	Teeth damaged on driver gear	Teeth damaged on driven gear
0			
58	Run-in wear	All	All
2669	Wear	All	All
	Destructive pitting	1, 28	1, 28
2857	Wear	All	All
	Destructive pitting	1, 6, 28	1, 6, 28
3029	Wear	All	All
	Initial pitting	All	1, 6, 28
	Destructive pitting	1, 6, 28	1, 6, 28

Initial pitting on specific teeth will only be discussed in reference to test completion. Although initial pitting most likely occurred prior to test completion, a detailed analysis of the inspection images is required to verify when it occurred and is outside the scope of this paper.

Plots of the data measured during these three experiments are shown in Figs. 7 to 13. Two different plots are shown for each experiment. The first plot is of load pressure, NA4, and NA4 reset for each experiment. The diamonds indicate when the rig was restarted after a shutdown. The second is a plot of FM4, NA4 reset, and the accumulated mass from the ODM. The triangles on the x-axis indicate the reading number that the rig was shut down for inspection. These reading numbers are listed in Tables I to III. Each experiment will be discussed in turn.

Experiment 2 is plotted in Figs. 7 to 9. Figure 7 shows the effect of the rig restarts after shutdowns on NA4 by the NA4 magnitude spikes that occur after shutdowns. Figures 8 and 9 indicate damage occurred just prior to inspection at reading 14369. Inspection at reading 14369 indicated destructive pitting first occurred on driver and driven tooth 6. The progression of damage is detailed in Table I and Fig. 4. Both NA4 and FM4 indicate an increase in magnitude when it appears destructive pitting occurred. The NA4 reset, like FM4, is less sensitive to damage as it progresses to a number of teeth and becomes more severe.

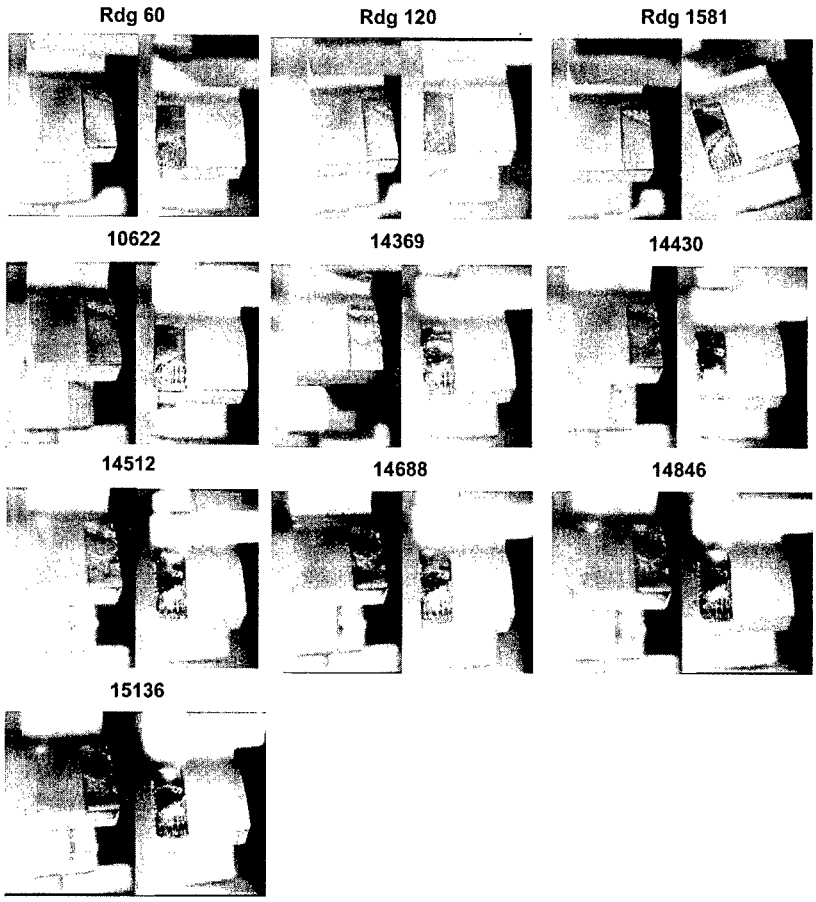


Figure 4.—Damage progression of driver/driven tooth 6 for experiment 2.

Experiment 3 is plotted in Figs. 10 to 11. Damage progression is shown in Table II and Fig. 5. Destructive pitting occurred on driven tooth 11 prior to inspection at reading 2199. From Fig. 11, FM4 and NA4 both indicate an increase in magnitude at approximately reading 1700. As seen previously, both become less sensitive to damage as it progresses.

Experiment 4 is plotted in Figs. 12 to 13. Damage progression is shown in Table III and Fig. 6. Destructive pitting occurred on driver and driven teeth 1 and 28 prior to inspection at reading 2669. From Fig. 13, FM4 and NA4 both indicate an increase in magnitude prior to inspection at reading 2669 and become less sensitive to damage as it progresses.

As seen in Figs. 7 to 13, NA4 does react to pitting damage. However, some of the response magnitude is lost with the reset operation. The NA4 reset does increase the stability of the

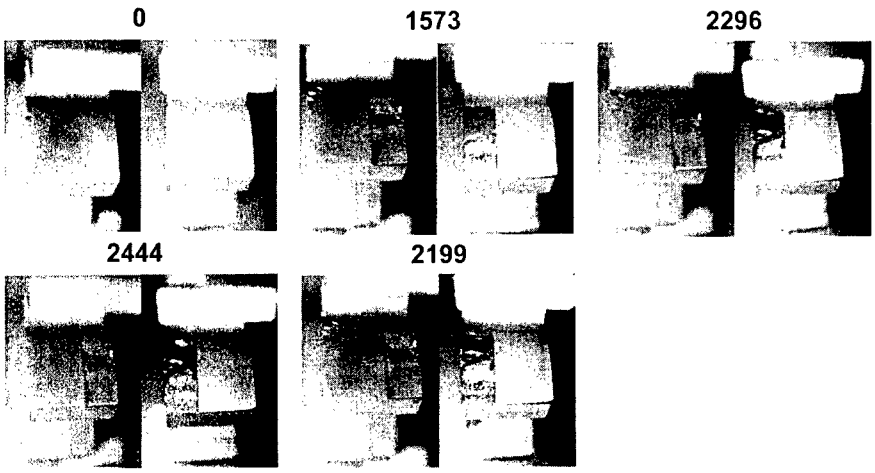


Figure 5.—Damage progression of driver/driven tooth 11 for experiment 3.

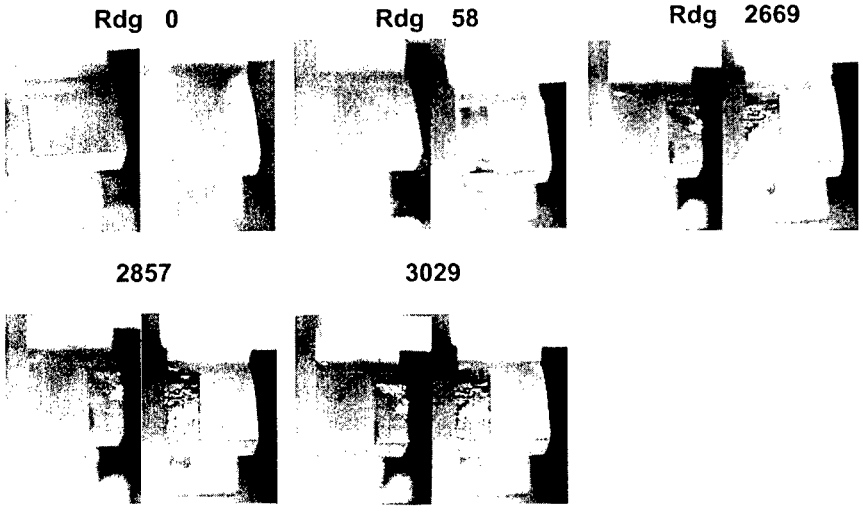


Figure 6.—Damage progression of driver/driven tooth 28 for experiment 4.

NA4 parameter enabling it to have a more consistent threshold limit. This is a key critical factor in reducing false alarm rates.

Conclusions: Operational effects, such as load and speed fluctuations, can adversely impact vibration diagnostic parameters and result in an unacceptable level of false alarms. To minimize this, current practice is to reduce the sensitivity of the vibration-based-diagnostic

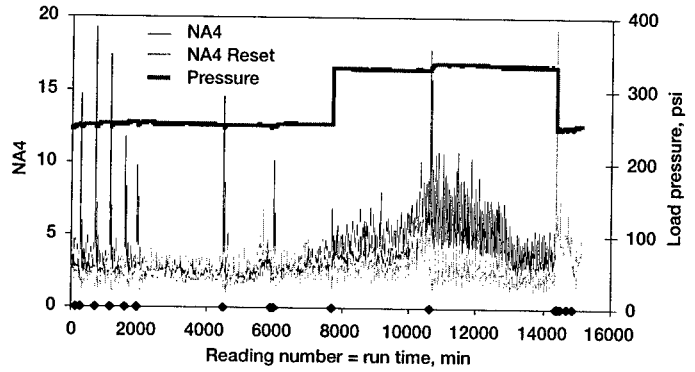


Figure 7.—Data from experiment 2 illustrating load and shutdown effects.

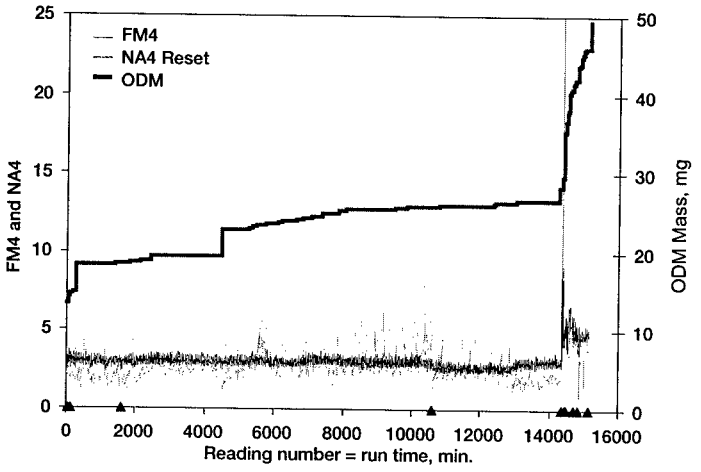


Figure 8.—Vibration, ODM, and damage data from experiment 2.

techniques. However, this also results in a decreased sensitivity of these techniques to actual damage.

The goal of this research was to minimize the effect of load on the vibration-diagnostic-parameter NA4 while maintaining its sensitivity to pitting damage. Results indicate the NA4 reset is no longer sensitive to load changes but is still sensitive to pitting damage. Both NA4 reset and FM4 indicate when destructive pitting occurs on one gear tooth. The NA4 reset, like the FM4, is less sensitive to damage as it progresses to a number of teeth and increases in severity. The magnitude of NA4 reset is less than NA4 when pitting damage occurs requiring a smaller threshold limit to indicate pitting damage. However, the magnitude of NA4 reset is significantly larger than FM4 when pitting damage begins to occur. It should be noted that successful implementation of NA4 reset requires a signal that can be

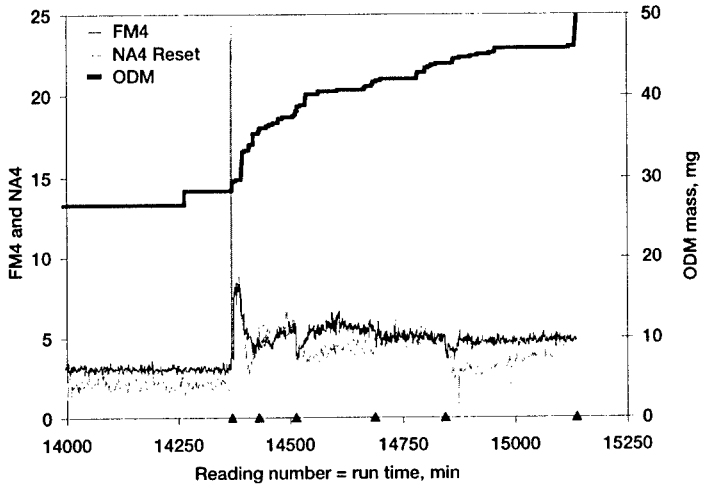


Figure 9.—Vibration, ODM, and damage data from experiment 2.

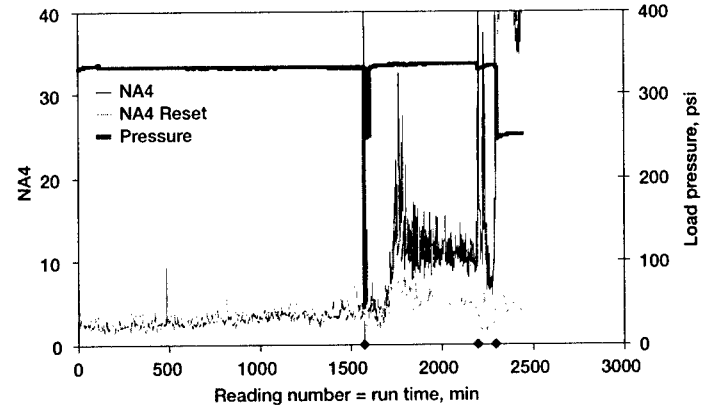


Figure 10.—Data from experiment 3 illustrating shutdown and load effects.

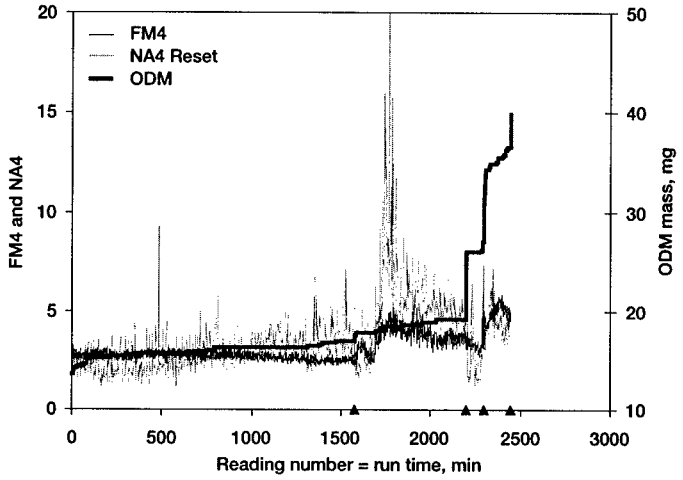


Figure 11.—Vibration, ODM, and damage data from experiment 3.

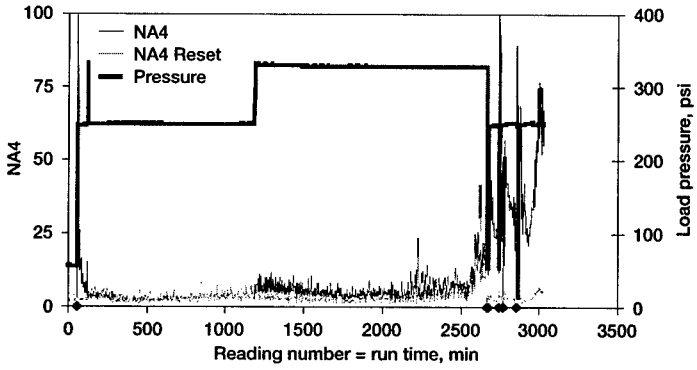


Figure 12.—Data from experiment 4 illustrating shutdown and load effects.

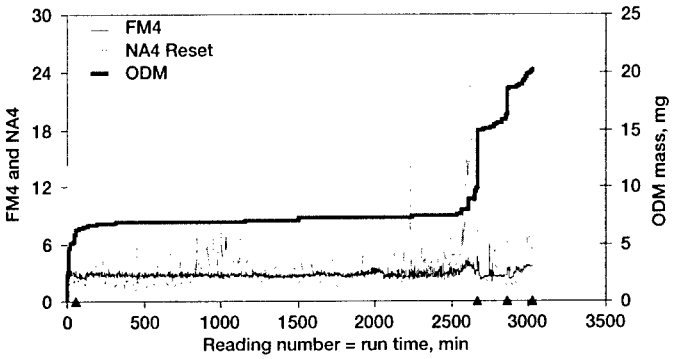


Figure 13.—Vibration, ODM, and damage data from experiment 4.

directly correlated to torque load. Additional research is required to define alert and fault threshold limits for vibration algorithm NA4 reset.

References

1. Zakrajsek, J.J.; Townsend, D.P.; and Decker, H.J.: An Analysis of Gear Fault Detection Methods as Applied to Pitting Fatigue Failure Data. NASA TM–105950, 1993.
2. Zakrajsek, J.J.; Handschuh, R.F.; and Decker, H.J.: Application of Fault Detection Techniques to Spiral Bevel Gear Fatigue Data. NASA TM–106467, 1994.
3. Zakrajsek, J.J.; Decker, H.J.; Handschuh, R.F.; Lewicki, D.G.; and Decker, H.J.: Detecting Gear Tooth Fracture in a High Contact Ratio Face Gear Mesh. NASA TM–106822, 1995.
4. Lewicki, D.G., and Coy, J.J.: Helicopter Transmission Testing at NASA Lewis Research Center. NASA TM–89912, 1987.
5. Lynwander, P.: Gear Drive Systems Design and Application. Marcel Dekker, Inc., New York, NY, 1983.
6. Zakrajsek, J.J.; Townsend, D.P., Oswald, F.B., and Decker, H.J.: Analysis and Modification of a Single-Mesh Gear Fatigue Rig for Use in Diagnostic Studies. NASA TM–105416, 1992.
7. Dempsey, P.J.: A Comparison of Vibration and Oil Debris Gear Damage Detection Methods Applied to Pitting Damage. NASA TM–210371, 2000.
8. Howe, B.; and Muir, D.: In-Line Oil Debris Monitor (ODM) for Helicopter Gearbox Condition Assessment, January 1, 1998.
9. 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 Southhampton, 1977.
10. Rock, D., Malkoff, D.; and Stewart, R.: AI and Aircraft. AI Expert, Feb. 1993, pp. 28–35.
11. Zakrajsek, J.J.; Decker, H.J.; and Handschuh, R.F.: An Enhancement to the NA4 Gear Vibration Diagnostic Parameter. NASA TM–106553, 1994.

THE USE OF HISTOGRAMS FOR DETECTION OF ELECTRICAL INSULATION BREAKDOWN

Jun Wang and Sally M°Inerny

Dept. of Aerospace Engineering and Mechanics
The University of Alabama
P.O. Box 870280
Tuscaloosa, AL 35487

Abstract: Seeded fault experiments were performed on a 50 Hp, 3 phase, 1750 RPM induction motor operated with a variable speed (pulse width modulation, or PWM, type) controller. A hole was cut in the stator housing, allowing access with an industrial heat gun. A single coil of one phase, at the point where it enters the stator laminations, was heated to a series of raised local temperatures. Current and voltage data were simultaneously measured on the target phase and recorded at a sample rate of 2.5 MHz each. Various feature extraction methods were applied to the voltage, current and instantaneous power (voltage times current) time series data. Single point statistics (rms, skewness, and kurtosis), histograms, and Fourier spectra are presented here.

Key Words: Electrical insulation; condition monitoring; variable speed motors.

Introduction: With the development of advanced power electronics and the microprocessor, induction motors for variable speed operation are predominantly fed from pulse width modulation (PWM) inverter drives. Inverter duty induction motors are designed to withstand the rigors of adjustable speed drive (ASD) operation. However, there are significant differences between applications of motors operated on sine-wave power and motors operated on adjustable frequency controls. New PWM drives employing insulated gate bipolar transistors (IGBT) technology offer switching frequencies as high as 20 kHz [1-2], generate steep-front pulses with slopes as high as $6000v/\mu s$, and can induce significant over-voltages [3-4]. Higher switching frequencies result in motor working currents that are more nearly sinusoidal, reduce the audible noise, and are said to offer improved motor efficiencies. However, higher switching frequencies also result in shorter pulse rise times and larger peak over-voltages, subjecting the motor to more severe insulation stresses.

The over voltage induced in PWM operations is unevenly distributed along the coils in a phase. Dr. Bonnett reported that as much as 85% of the peak over-voltage can be dropped across the first turn of the first coil of a phase [5]. This uneven over-voltage distribution causes a significant overstress across the end turns and may result in turn-to-turn insulation failure. The first coil voltage represents the highest possible inter-turn

voltage [4-8] and the shorter pulse the rise time, the greater the voltage dropped across the first coil [9]. For this reason, the focus of our experimental tests was on the first turn of one coil in one phase.

This paper reports preliminary results of research on the development of methods for the early detection of insulation degradation in low voltage (under 600 V) PWM controlled 3-phase induction motors. Voltage and current data were acquired on one phase of a 50 Hp induction motor operating under healthy (baseline) and faulted conditions. The tests were performed at the Motor Test Facility (MTF) at Oak Ridge National laboratories. The results of statistical and spectral analyses of the data are discussed in this paper

Experiments: The availability of 50 Hp drives and the interest of one of the sponsors, ASHRAE, in larger motors led to the selection of a 50 Hp motor as the test item. A new 50 Hp, 440 V, three phase inverter duty motor was donated by Reliance Electric, Inc. Three methods of achieving insulation degradation were considered. Two methods, running the motor at full load with an imbalance and mechanically piercing the insulation in one of the windings, were discussed first. It is difficult to control failures induced by the first of these methods and it could be argued that the second would not be a good representation of normally occurring insulation failure. John Kueck of the MTF suggested removing a section of the motor housing and locally heating a coil. It was decided that this approach would provide the best control and of the three options be the most representative of insulation failures in PWM controlled induction motors.

The end bell of the donated motor was removed and one of the first windings of a phase was identified. Then, a section of the motor housing between the stator laminations and the end bell was cut out in order to expose this coil where it enters the stator, permitting access with an industrial dryer / heat gun. Two type K thermocouples were mounted on the outer edge of the stator laminations roughly half way between the teeth and the motor housing. Only one of the thermocouples worked throughout the tests. Note that the measured temperature provides only limited information on winding temperatures, as it was a local surface temperature. A more complete description of the test plan and instrumentation can be found in Ref. [10].

Test Series One: In the first series of tests, the motor was run until steady state conditions were reached and then the exposed motor coils were heated at the point where they entered the stator coil. Measurements were made using three different PWM switching frequencies and, for each switching frequency, at three different speed and load conditions in the order indicated in Table I.

For each test cycle indicated in Table I, the exposed coil was local heated until the temperature reading stabilized, then data was recorded. The heater was then moved closer to the coil and this process repeated at higher temperatures. Thermocouple readings for the recorded data ranged from 70°C to 252°C with local heating. Preliminary analyses of the data recorded during this set of tests did not show any significant effect of the local heating on the data. A second set of tests performed a month later focused, therefore, on inducing motor failure.

All of the baseline data sets discussed in this paper were recorded during this first series of tests. The baseline test conditions were always: 10 kHz switching frequency, 1750 RPM and 50 Hp. ***Baseline data*** set 1 was recorded before Test Cycle 1; baseline data sets 2 and 3 were recorded between Test Cycles 3 and 4; baseline data set 4 was recorded between Test Cycles 6 and 7; and baseline data set five was recorded after Test Cycle 9.

Table I. Test Series One Operating Conditions.

Test Cycle	Switching Frequency	Motor Speed	HP
1	10k Hz	1750 RPM	50 Hp
2		438 RPM	2.1 Hp
3		875 RPM	7.5 Hp
4	8k Hz	1750 RPM	50 Hp
5		438 RPM	2.1 Hp
6		875 RPM	7.5 Hp
7	2k Hz	1750 RPM	50 Hp
8		438 RPM	2.1 Hp
9		875 RPM	7.5 Hp

Test Series Two: All of these tests were performed with a 10 kHz switching frequency at full speed (1750 RPM) and full load (50 Hp). A more powerful heat gun was purchased and local temperatures were raised much higher than in the previous tests. The following is a brief description of the chronology of the test cycles in this series.

Day One. Two cycles of local heating were performed and then the motor fan intake was partially blocked. With the fan intake partially blocked and no local heating, the thermocouple reading rose to 115°C.

The fan intake blockage was removed and the motor run overnight at full speed and load.

Day Two. The motor fan intake was completely blocked and the motor temperature rose to 130° C. The exposed coil was then locally heated to 360°C while data was recorded.

The locally heated area was allowed to cool, while the global motor temperature continued to rise before the next test cycle. At the start of the next test cycle, the initial temperature reading was 145° C. With local heating in place, the temperature reading reached 383° C. It was ***at the end of this test cycle that Data Set A*** was recorded.

The locally heated area was allowed to cool, while the global motor temperature continued to rise before the next test cycle. At the start of the next test cycle, the initial temperature reading was 158° C. With local heating in place, the temperature reading reached 391° C. It was ***at the end of this test cycle that Data Set B*** was recorded.

The locally heated area was allowed to cool, while the global motor temperature continued to rise before the next test cycle. At the start of the next test cycle, the initial temperature reading was 164° C. With local heating in place, the temperature reading reached 404° C. It was ***at the end of this test cycle that Data Set C*** was recorded.

The fan intake blockage was removed and the motor run overnight at full speed and load.

Day Three. The motor fan intake was fully blocked and thermal insulation was placed between the fins of the motor housing. See Figure 1. Just before local heating of the exposed coils commenced, the thermocouple reading was 152° C. Local heating of the exposed coils was begun, when the thermocouple reading reached 380° C, but before data was recorded ***the motor shut down***.

The motor was restarted and data set D acquired at a thermocouple reading of 158° C with no local heating. Local heating was applied and ***data set E recorded*** at a thermocouple reading of 360° C just as the motor failed for the second and final time.

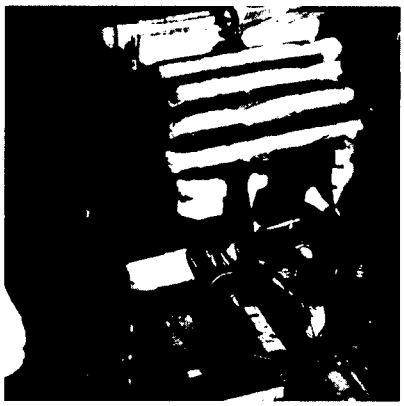


Figure 1. The Motor On the Final Day of Testing. The heat gun is shown directing heat at the exposed coils where they enter the motor stator.

Statistical and Spectral Analyses: A histogram of a given data set can be described as the "frequency of occurrence versus range of values" or "relative frequency of occurrence versus range of values." In either case, the histogram provides information on the probability that the value of a single data point in a set will lie within a specific range of values. Take, for example, a set of 1000 values randomly distributed in the range of -3.5 to 3.5 (normally distributed random numbers generated using the "randn" command in the program MATLAB). When the range is divided into 50 intervals (a.k.a. "bins") and the percentage of the total number of data points whose values lie in each of these ranges is indicated by the height of a bar, one obtains the histogram shown in Figure 2. This relative frequency histogram provides a graphical description of the statistical distribution of the data.

The mean, standard deviation, skewness coefficient and kurtosis coefficient provide simpler single number metrics that can be used to characterize a data set. For a set of N discretely sampled values of $x(t_i)$, these are defined as:

Mean (or Average Value)
$$\mu = \frac{1}{N}\sum_{i=1}^{N} x(t_i) \quad (1)$$

Standard Deviation
$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x(t_i)-\mu)^2} \quad (2)$$

Skewness Coefficient
$$\gamma_3 = \frac{1}{N}\sum_{i=1}^{N}\frac{(x(t_i)-\mu)^3}{\sigma^3} \quad (3)$$

Kurtosis Coefficient
$$\gamma_4 = \frac{1}{N}\sum_{i=1}^{N}\frac{(x(t_i)-\mu)^4}{\sigma^4} \quad (4)$$

For zero mean signals, the standard deviation is the same as the root mean square (or rms) value, x_{rms}. The skewness and kurtosis coefficients are then the third and fourth order moments normalized by the third and fourth power of the rms value, respectively. In this application, the current and voltage are zero mean signals and we will refer to their rms values rather than their standard deviations. In the case of the electrical power (the product of the current and voltage), the mean is the average value and the standard deviation does not have a familiar physical interpretation.

In addition to statistical analyses, Fourier spectra of the voltage, current and power were calculated. The spectra presented in this paper have a resolution of just under 10 Hz and were calculated using a Hanning window and 30 spectral averages with 50% overlap processing.

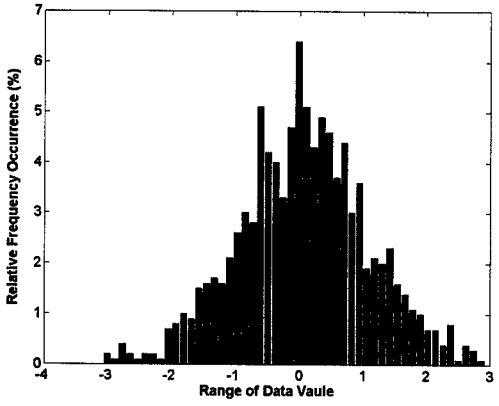


Figure 2. A 50 Bin Histogram of 1000 Numbers Whose Values are Randomly Distributed between –3.5 and 3.5. The numbers were generated in MATLAB using the "randn" command.

Results: Initial comparisons of histograms of the voltage, current and power (voltage times current) time series data indicated dramatic differences between one of the baseline and one of the fault data sets. We later discovered, however, that the differences between different baseline sets were as dramatic as those seen between the baseline and fault data sets. See Figures 3 and 4. It appears that the variability of the histograms even under nominally identical conditions is so large as to mask any variation due to motor health.

A comparison of the single number statistical metrics (mean, rms, maximum value, minimum value, skewness and kurtosis coefficients) between the baseline and fault data sets is provided in Table II. There is no discernible difference in the voltage statistics. As might be expected, the rms current shows a consistent increase. The maximum and the magnitude of the minimum instantaneous current also increase as the motor health degrades. Consistent with the increase in current, the average or mean electrical power also increases.

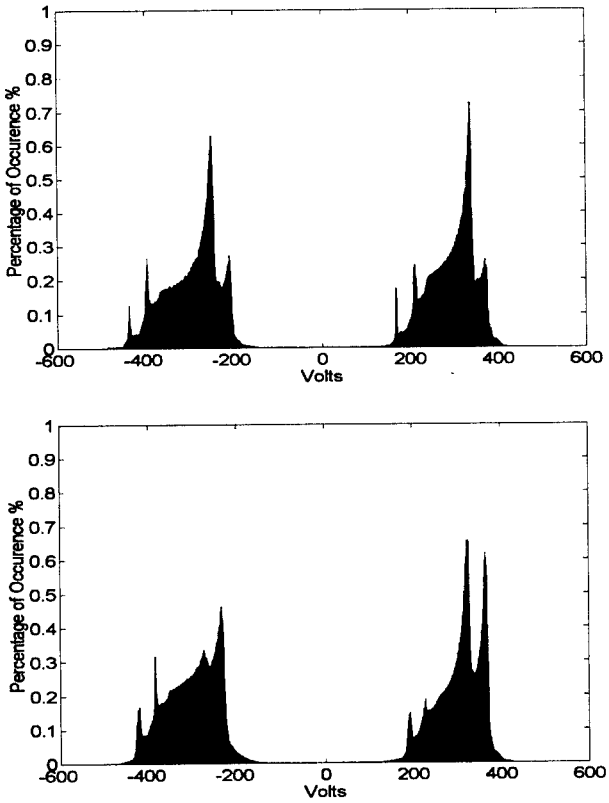


Figure 3. Histograms of Two Baseline Voltage Data Sets: 3 (top) and 4 (bottom).

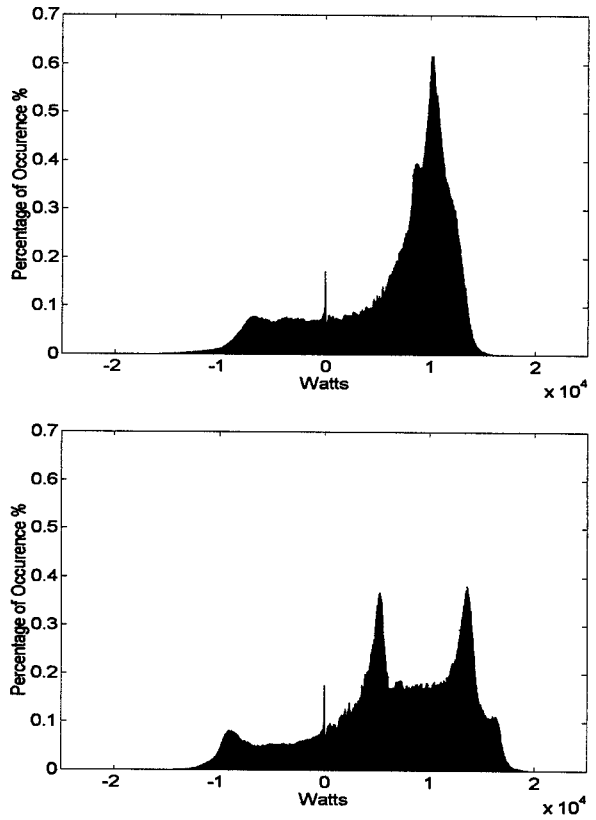


Figure 4. Histograms of Two Baseline Power Data Sets: 4 (top) and 1 (bottom). The peak at 0 Watts was traced to the use of DC coupling of the current signal.

The initial premise of these experiments was that small-scale, local insulation degradation would first be reflected in changes in the high frequency or small-scale data. For this reason, our initial spectral analyses have focused on the full 0-1.25 MHz frequency range, as opposed to the 0-3000 Hz range typically examined in electrical power / motor studies. (Bear in mind that the power consumed by an induction motor is dominated by the motor control frequency, which is 60 Hz here.) Direct comparisons of the voltage spectra of the baseline and fault data sets revealed no discernible differences. Consistent increases in spectral levels were detectable in the current (response) spectra and, consequently, the spectra of the electrical power. Spectral differences occurred primarily in two frequency ranges. For this reason, we adopted an approach that has been used to monitor tail drive shaft bearings in rotor-craft. In this approach, the rms energy in key frequency bands is calculated as a characteristic metric. The frequency ranges we chose were determined by inspection of current spectra, as indicated in Figure 5. The H1 range encompasses a peak

Table II. Single Number Statistical Metrics of Baseline and Fault Data Sets (A-E).

	Baseline Sets	A	B	C	D	E
Starting Motor Temperature	70-81° C	145° C	158° C	161° C	158° C	158° C
With Local Heating	N.A.	383° C	391° C	404° C	N.A.	360° C
V_rms Volts	296.9 to 310.1	297.4	302.0	304.4	305.5	309.7
V_max Volts	463.4 to 478.8	479.5	471.7	473.7	467.4	484.3
V_min Volts	-510.2 to -551.6	-564.3	-496.0	-512.5	-492.3	-529.3
V_kurt coeff	1.123 to 1.146	1.136	1.127	1.141	1.128	1.128
V_skew coeff	-0.0132 to -0.0194	-0.0155	-0.0144	-0.0186	-0.0155	-0.0190
I_rms Amps	**30.4 to 31.2**	**31.7**	**31.6**	**31.3**	**33.2**	**31.5**
I_max Amps	**47.3 to 49.7**	**50.3**	**50.0**	**48.8**	**52.3**	**53.0**
I_min Amps	**-47.0 to -48.8**	**-50.4**	**-49.7**	**-49.2**	**-52.2**	**-53.2**
I_kurt. coeff	1.499 to 1.511	1.512	1.509	1.501	1.501	1.508
I_skew. coeff	0.0053 to 0.0068	0.0054	0.0046	0.0063	0.0051	0.0054
P_max Watts	18640 to 22340	19740	17270	20530	18540	20870
P_min Watts	-16830 to -19730	-23900	-20250	-17430	-20610	-17440
P_mean Watts	**6186 to 6580**	**6674**	**6634**	**6622**	**7032**	**7133**
P_kurt coeff	2.752 to 3.249	3.730	3.702	2.958	3.652	3.190
P_skew coeff	-0.680 to -1.085	-1.115	-1.294	-0.731	-1.290	-1.016

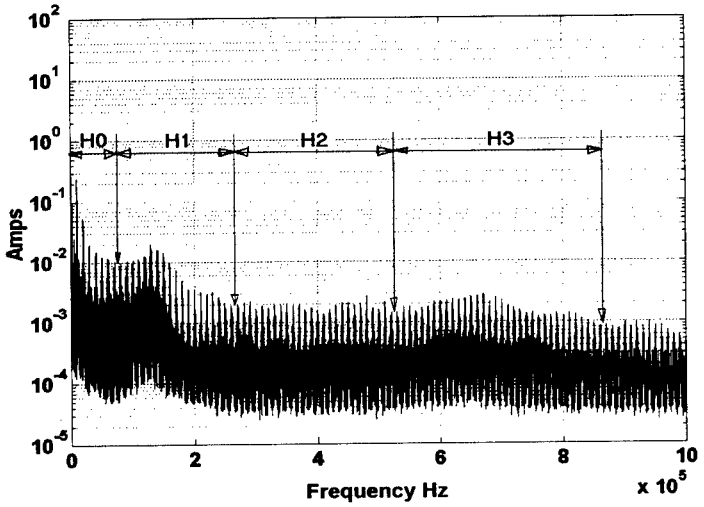


Figure 5. Spectrum of The Current from Baseline Data Set 4 Showing the Frequency Bands Selected for RMS Band Level Calculations.

in spectral levels thought to be a motor resonance effect. Note that the H0 band includes the very low frequency range (0-100 Hz) that dominates the current and electrical power, as can be seen by comparing the results for overall rms level and that in the H0 band in Table III.

The results in Table III confirm what was seen in initial comparisons of the current (and, consequently, electrical power) spectra of the baseline and fault data. Focusing on the high frequency bands not controlled by motor power consumption, the spectral levels in the H1 band (which encompasses what appears to be a motor resonance) are relatively large and show a significant increase with deteriorating motor health. At the low frequency end of the spectrum, the increase in overall (or H0) levels is consistent with the increase in rms current and average power in Table II.

Table III: RMS Current (in Amps) in the Frequency Bands Indicated in Figure 5.

		Baseline Data Sets					Fault Sets				
Frequency Band		**1**	**2**	**3**	**4**	**5**	**A**	**B**	**C**	**D**	**E**
H3	525k – 865 kHz	0.051	0.046	0.045	0.064	0.066	0.088	0.086	0.086	0.091	0.092
H2	265k – 525 kHz	0.049	0.052	0.052	0.052	0.054	0.068	0.068	0.067	0.067	0.067
H1	75k – 265 kHz	0.107	0.080	0.080	0.141	0.146	0.187	0.183	0.182	0.187	0.191
H0	0 - 75 kHz	26.34	27.03	26.99	27.00	26.27	27.48	27.35	27.15	28.72	29.01
Over -all	0 - 1.25 MHz	26.34	27.03	26.99	27.00	26.27	27.48	27.35	27.15	28.72	29.01

Conclusion: The initial intent of the seeded fault tests was to induce localized insulation failure in the first turn of one phase of a three-phase induction motor. The first series of tests raised the local motor temperature in the vicinity of the coil to over 250° C. Preliminary analyses failed to show any significant change in the characteristics of the measured current or voltage. The baseline data sets discussed in this paper were all recorded during this first series of relatively benign seeded fault tests.

The intent of the second series of tests was to induce motor failure by raising both the temperature of the entire motor (blocking the motor fan inlet and placing insulation between the fins of the motor housing) and by local heating of the exposed coils. Five of the data sets recorded during this series of tests were analyzed in this paper. Precise knowledge of the relative motor health at the time that each of these fault data sets was recorded is lacking. It is certain, however, that motor health steadily deteriorated as the tests progressed and the motor was subjected to additional thermal cycles and local heating of the exposed coil.

Statistical and spectral analyses of the recorded baseline and fault data sets were presented. Histograms of the voltage, current and electrical power time series data varied as greatly between baseline data sets as between baseline and fault data sets. No conclusions on motor health could be drawn from visual inspections of the histograms. Single number statistical metrics were also examined. Only the maximum, minimum, and rms values of the current and the mean value of the electrical power showed consistent trends with deteriorating motor health. Spectral analysis confirmed the sensitivity of the current to motor health both in the low frequency band controlled by the motor power consumption (60 Hz, here) as well as in a high frequency band that encompassed a peak in the spectrum attributed to motor resonance.

References

1. Stone, G., Campbell, S., and Tetreault, S., "Inverter-Fed Drives: Which Motor Stators Are at Risk?", *IEEE Industry Applications Magazine*, Sep/Oct 2000, pp. 17-22.
2. Lowery, T.F., "Design considerations for Motors And Variable Speed Drives", *ASHRAE Journal*, Feb. 1999, pp. 28-32.
3. Oliver, J.A., Stone, G.C., "Implications for the Application of Adjustable Speed Drive Electronics to Motor Stator Winding Insulation", *IEEE Electrical Insulation Magazine*, July/Aug. 1995-Vol. 11, No.4, p32-36.
4. Melhorn, C.J., Tang, L., "Transient Effects of PWM Drives on Induction Motors", IEEE Transactions on Industry Applications, vol. 33, No. 4, Jul/Aug. 1997, p1065-1072.
5. Bonnett, A.H., "Analysis of the Impact of Pulse-Width Modulated Inverter voltage Waveforms on AC Induction Motors", IEEE Trans. on Industry Appl. Vol. 32, No.2, Mar/Apr 1996, p 386-392.
6. Sidney B., Jason S., "Will Your Motor Insulation Survive a New Adjustable-Frequency Drive?", IEEE Transactions On Industry Applications, vol. 33. No. 5, Sep/Oct. 1997, p1307-1311.
7. Mbaye, A., Bellomo, J.P., Lebey, T., Oraison, J.M., Peltier, F., "Electrical Stresses Applied to Stator Insulation in Low-Voltage Induction Motors Fed by PWM Drives", IEE Proc. Electr. Power Appl. Vol. 144, No.3, May 1997, p191-198.
8. Oliver, J.A., Stone, G.C., "Implications for the Application of Adjustable Speed Drive Electronics to Motor Stator Winding Insulation", *IEEE Electrical Insulation Magazine*, July/Aug. 1995-Vol. 11, No.4, p32-36.
9. Stone, G., Kapler, J., "The Impact of Adjustable Speed Drive(ASD) Voltage Surges on Motor Stator Windings." *Proceedings of the 1998 IEEE/PCA Cement Industry Technical Conference* May 17-21 1998 Rapid City, SD, p133-138.
10. Wang, J., M^{c}Inerny, and Haskew, T., "Insulation Fault Detection in a PWM Controlled Induction Motor – Experimental Design and Preliminary Results," *Proceedings of the International Conference on Harmonics and Power Quality*, 1-4 October 2000, Orlando, Fl.

DETECTION AND SEVERITY ASSESSMENT OF FAULTS IN GEAR BOXES FROM STRESS WAVE CAPTURE AND ANALYSIS

James C. Robinson
Computational Systems, Inc.

Abstract: Many faults in gearboxes are accompanied by the emission of stress waves that disperse away from the initiation site at the speed of sound in metal. The wave propagation introduce a propagating ripple on the surface which will introduce a response, output, in a sensor sensing absolute motion such as an accelerometer. For an accelerometer at a fixed location, the wave propagation will be a reasonable short-term transient event lasting on the order of fractional to several milliseconds. The duration of the event will be dependent on (1) type of event e.g., stress waves from impacting will last longer than stress waves accompanying the release of residual stress buildup through fatigue cracking, (2) relative location of the sensor (accelerometer) to the initiation site, and (3) severity of the fault responsible for the stress wave emission.

For a healthy smooth running machine (gearbox), there generally will be no stress waves present. Therefore their presence is indicative of a defect which generates stress waves. Some common defects which generate stress waves are pitting in the races causing the rollers to impact, fatigue cracking in bearing raceways or gear teeth (generally at root), scuffing or scoring on gear teeth, cracked gear teeth, and others. The challenge becomes one of detecting and quantifying relative to energy and repetition rate (or lack thereof) the stress wave activity. This leads to the identification of certain faults and, with experience, their severity.

The methodology employed by CSI per the capture and analysis of stress waves are to collect a block of data consisting of peak values (in g's) which occur within discrete sequential equal time intervals determined by the resolution sufficient to identify faults. The number of time intervals over which peak values are collected are consistent with that needed to invoke spectral analysis for the desired resolution and spectral band width.

The magnitude of the stress wave packets is identified in the discrete time data block containing peak values. The presence of periodicity is identified in the spectral (frequency) domain. An alternative to spectral analysis for the identification of periodicity is auto-correlation analysis.

To illustrate this peak value (PeakVue™) analysis for fault detection and severity assessment in gearboxes, several case studies are presented. The specific faults demonstrated are bearing faults, cracked gear teeth, unstable driver speed, and torsional vibration. It will be demonstrated that the PeakVue methodology is a very beneficial tool for monitoring gearboxes.

Introduction: Many mechanical faults within industrial rotating machinery manifest themselves through modal excitation (vibration) and stress wave initiation. Modal excitation can be detected using sensors which detect absolute or relative motion. A common sensor employed for detection of absolute motion is the accelerometer. The

The method employed by CSI for stress wave analysis avoids the use of a low pass filter completely. This is accomplished by separating, as much as possibly, the stress wave activity (the short term transient activity) from the continuous activity by routing the signal from the sensor through a high pass filter set consistent with possible fault frequencies within the machine. The resultant time signal is converted into a digital signal at constant time increments for further analysis. The digital value recorded over each time increment is not the signal value at a specific time, but instead is the absolute maximum (peak) value observed over each discrete time interval. The resultant digital representations are peak values, which occurred over each time increment.

The analysis of the peak value (PeakVue™) waveform is basically (1) the identification of any periodic activity occurring at rates consistent with possible fault frequencies and (2) severity of assessment based on the level (peak value) of the stress wave activity. The presence of periodic activity is identified through spectral analysis of the digital block of data consisting of the sequential peak values. Severity level is extracted from observed peak values compared with similar faults and/or trending of the peak values.

3. Case Studies

3.1 Introduction: The case studies chosen for presentation will demonstrate an outer race and an inner race defect in separate pinion stand gear boxes. Sufficient data was available to demonstrate the importance of trending for the inner race defect case. The third case demonstrates severe cracked teeth in a Precision Tension Bridle gearbox. The fourth case is from an extruder gearbox being driven by a DC motor with speed variation. The fifth case was selected to demonstrate a torsional resonance problem present in a large crusher gearbox.

3.2 Outer Race Defect in Pinion Stand Gearbox: This pinion stand gearbox was included in the scheduled monthly condition monitoring program employing vibration analysis. The traditional vibration monitoring showed no indication of a bearing fault. In July 1997, the PeakVue methodology was introduced into the monitoring program.

It was obvious from the PeakVue data there was an outer race defect on the inlet shaft. The peak g readings were 18 g's (the normal vibration readings were showing 1.5 g's with no indication of a problem). The peak g readings in PeakVue continued to trend up (got to 38 g's in mid Sept. 1997) and then started a downward trend (14 g's in early October). The bearing was then replaced the peak g-levels on the replacement bearing was less than 1-g.

The normal vibration spectra and acceleration time waveform for data acquired on September 15 1997 are presented in Figure 1. There is some indication of a possible outer race problem but not conclusive.

The data from the PeakVue methodology acquired on the same date are presented in a Figure 2. The peak absolute g-levels are up to 38 g's with a recurring rate consistent with the outer race defect frequency. The bearing was replaced in early October 1997. A picture of the defective bearing is presented in Figure 3. The defects in the outer race are apparent.

analysis of the modal motion relative to the machine health is referred to as vibration analysis. The methodology employed in vibration analysis consists of:

1. Capture (digitally) of a time waveform from a sensor for a specified time period. The signal is first passed through a high order low pass filter prior to digitalization. The purpose of the low pass filter is to remove all frequency content which exceeds the Ngquist frequency (one half the sampling rate).
2. Transform the modified discrete time waveform into the frequency domain employing FFT methodologies.
3. Look for excessive activity compared to other similar machinery or previous history at discrete know fault frequencies.

The implicit assumption in vibration analysis is the signal being analyzed is stationary equilibrium. The spectral values are average values, which are appropriate for stationary (continuous) conditions.

Stress waves in metal accompany actions such as impacting, fatigue cracking, scuffing (scoring) abrasive wear, etc. Stress wave emissions are short term, lasting several microseconds to a few milliseconds, transient events which propagate away from the initiation site as bending (s) and longitudinal (p) waves at the speed of sound in metal. The s waves introduce a ripple on the surface which will excite an absolute motion sensor such as an accelerometer. The detection and classification of these stress wave packets provide an important diagnostic tool for (a) detecting certain classes of problems and (b) severity assessment.

In the next section, a brief discussion of the methodology employed by CSI for stress wave capture and analysis will be presented. The next section will present several case studies showing the detection and severity assessment for faults commonly experienced in gearboxes. The last section will be a summary and conclusion section.

2. Capture and Analysis Methodology for Stress Wave Activity: Stress waves are generated when impacting, scuffing (scoring), fatigue cracking, abrasive wear, etc. are present The duration of an individual event will range form fractional to several milliseconds. The rate at which individual events occur within rotating machinery generally are periodic consistent with the fault, e.g., a pitted area in the outer race will cause impacting at the outer race fault frequency.

The sensor generally employed for the detection of stress waves is the accelerometer. Since the signal is a short-term event relative to the repetition rate, the methodology employed for the detection of the events preferable should avoid any averaging. This is because there can be a large variation in repetition rate and hence the duty cycle will introduce large variations in average values independent of the severity level of the fault. Averaging negates the ability to perform severity level analysis based on trending and relative comparisons.

When executing vibration spectral analysis, the general procedure is to route the sensor analog signal through signal conditioning, which includes a low pass filter (anti-aliasing) immediately prior to conversion into a digital signal with discrete values at a constant sampling rate. The low pass filter is an averaging process; hence any short term events are averaged over the averaging time associated with the anti-analyzing filter.

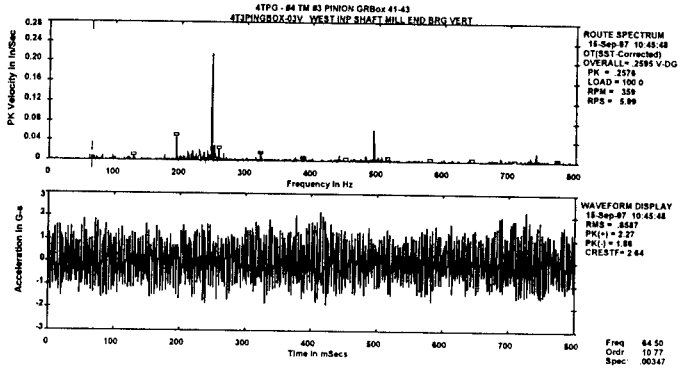


Figure 1. Vibration velocity spectra and acceleration time waveform on pinion stand gearbox on September 15, 1997.

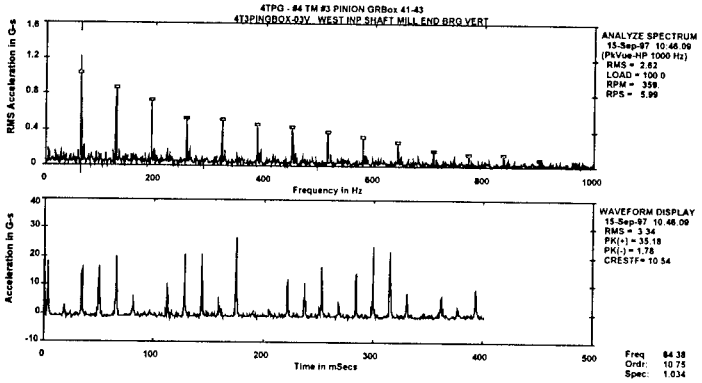


Figure 2. PeakVue spectra and time waveform on Pinion stand gearbox (same as Fig. 1.)

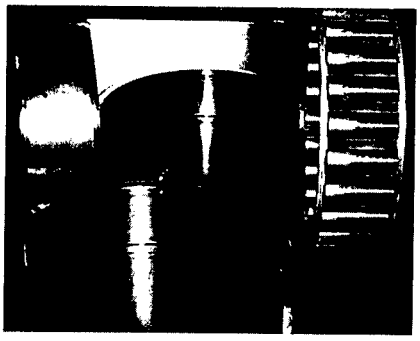


Figure 3. Defective bearing taken from the pinion gearbox of Figs. 1 and 2.

3.3 Inner Race Defect in Finish Mill Pinion Stand Gearbox: This pinion stand gearbox is separate from the example presented above. A separate data point was set in the database and data (PeakVue and normal) acquired on a scheduled basis beginning on Mar. 16, 1998. One of the trend parameters captured for trending was the peak g-levels in the PeakVue time waveform. Experience has shown this to be a key parameter for fault detecting and severity assessment.

The PeakVue peak g-level trend parameter for the lower output shaft and PeakVue spectra for last collection date of May 28, 2000 are presented in Figure 4. The alert and fault levels are set at recommended levels for this speed machine and type fault. From the spectra, the fault is an inner race fault which is side banded (amplitude modulated) at running speed which is indicative of fault going in and out of load zone at running speed. From Figure 4, it is obvious that the fault exceeded the "fault" level about 7 months prior to replacement, in July 2000.

The trend value for bearing fault over the same time interval for the normal vibration monitoring are presented in Figure 5. Here there are no indications of a bearing fault.

Based on the trend values in PeakVue, a work order was release in June 2000 to replace the bearing. The bearing was replaced in July 2000. A picture of the defective bearing is presented in Figure 6. The failure was clearly advanced and could have induced catastrophic failure easily by e.g., metal "chunks" interfacing with the gear teeth meshing.

3.4 Cracked teeth in a Precision Tension Bridle Gearbox: This gearbox was a single speed reduction gearbox with a dual shaft output. The slow speed shaft from the reduction gear set (40 teeth pinion gear driving a 158 teeth bull gear) was driving a second output shaft through a dual 90 tooth gear set. The input shaft was turning at 525 rpm and the output shafts turning at 215 RPM at the time the data presented below was acquired.

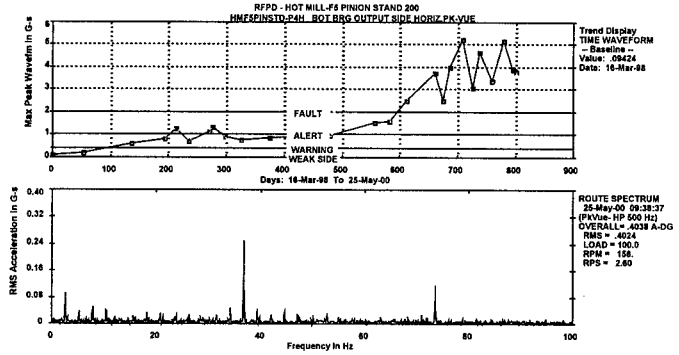


Figure 4. Maximum peak g-level (from PeakVue) trend from March 16, 1998 to May 25, 2000 and PeakVue spectra from May 25, 2000.

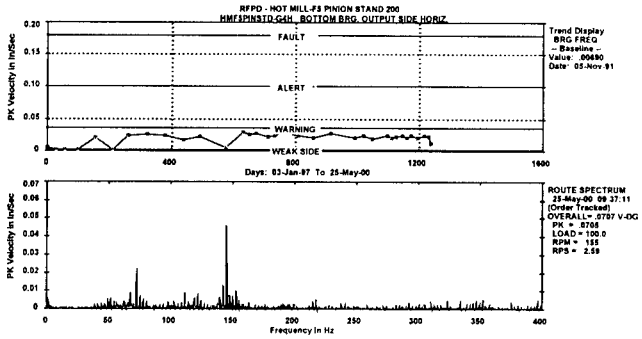


Figure 5. Normal vibration bearing fault trend and spectra for latest measurement over same time period as in Fig. 4.

Figure 6. Defective bearing taken from the pinion stand gearbox of Figs. 4 and 5.

The only accessible point for acquiring data was over the input shaft. The normal velocity vibration spectra and acceleration time waveform acquired on April 14, 1997 are presented in Figure 7. The speed reduction (input) gear mesh, 351 Hz, is dominant in the spectral data in Figure 7 and showing significant side banding (especially at 2 × GM). This pattern is indicative of gear wear and perhaps some misalignments. The P-P acceleration data in time waveform was less than 4 g's and not considered significant. This possible gear wear misalignment had been flagged with an action item to initiate a visual inspection at next opportunity.

It was decided to apply the PeakVue methodology on April 14, 1997. The PeakVue spectral and waveform data are presented in Figure 8. The only activity in the spectral data is the output shaft turning speed with many harmonics. The time waveform has two impacting regions per turn of the output shaft. The impacting levels exceed 40 g's. This signature indicates a gear with significant cracked teeth (at root) in two regions. One of the output shafts has the bull gear with 158 teeth and the pinion gear with 90 teeth driving

the second output shaft. The bandwidth in Figure 8 is not sufficient to encompass either gear mesh; therefore we cannot identify from this data set which gear set has the defective gear.*

From the spectral presented in Figure 7, one would be suspicious that the defective gear would be in the GM 1 set since nothing unusual is present relative to the GM 2 set. There was an additional PeakVue spectra taken with a bandwidth of 5000 Hz. In this PeakVue spectra set, the GM 2 activity was present and the GM 1 absent. This leads to the conclusion that the gears with the cracked teeth were most probably in the GM 2 set.

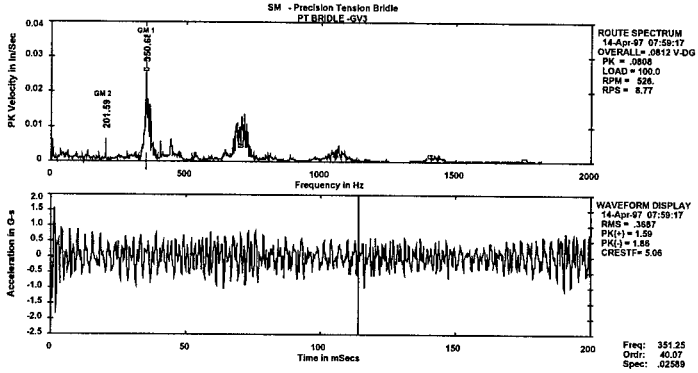


Figure 7. Velocity spectra and acceleration time wave form from the precision tension bridle gearbox.

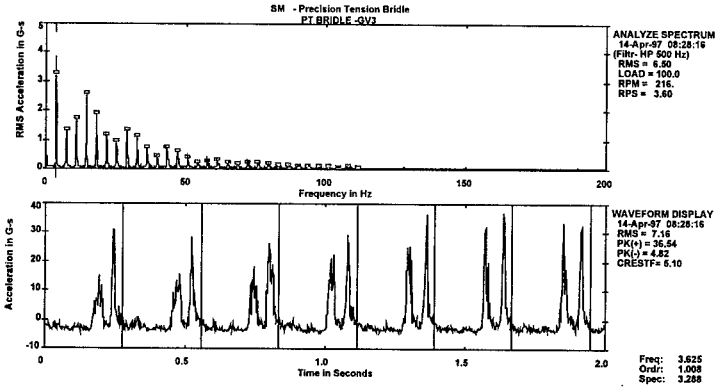


Figure 8. PeakVue spectra and time waveform from the precision tension bridle gearbox.

*The gear mesh where the defective gear is located would be present in the PeakVue spectra.

Following the acquisition of the PeakVue data, the gearbox was shortly shutdown and inspected. One of the gears in the GM 2 set was found to have two visible cracks.

3.5 Extruder Gearbox: The plan view of the extruder gearbox is presented in Figure 9. For the indicated input speed of 1840 RPM, the output shafts are turning at 316 RPM. The gear mesh frequencies are 1318 Hz, 659 Hz, 237 Hz, and 79 Hz. For a gearbox of this complexity, experience has shown a measurement point should be located at each bearing.

The monitoring of the gearbox using normal vibration and PeakVue methodologies identified worn gear sets, probable gear misalignment, defective bearings, and excessive driver (DC Motor) speed variations. The signatures identifying the driver speed variation are presented below.

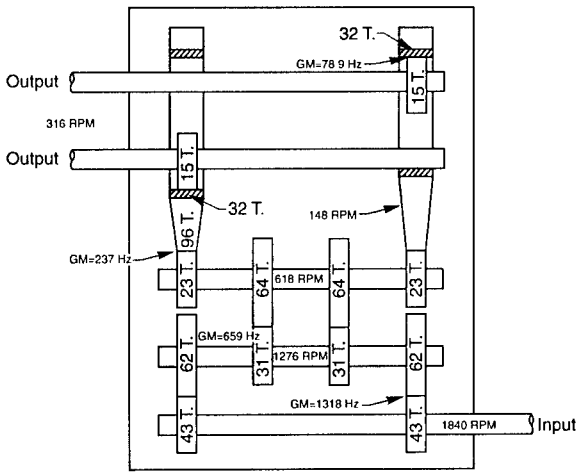


Figure 9. Plan view of the extruder gearbox.

The velocity spectra and acceleration time waveform for an input shaft speed of approximately 1825 RPM are presented in Figure 10. The dominant activity is the gear mesh for the 23T/96T set and the 43T/62T set. There are many harmonics of the 23T/96T set with the third being the largest (indicating looseness). The 43T/62T have reasonable shaft speed side banding at 2 × GM, which suggests some misalignment.

The PeakVue spectra and time waveform for inlet shaft speed of 1833 RPM are presented in Figure 11. There are activity at (1) the inlet shaft speed of 21.2 Hz (2) the first intermediate shaft speed at 31 Hz, and (3) the 43T/62T GM frequency and 2 times the 43T/62T GM frequency. The 2 times 43T/62T GM frequency is side banded with the shaft speed on which the bull gear is mounted. The impacting at 2 times gear mesh is indicative of significant back lashing which could be introduced with torsional resonance or (more probable) significant inlet shaft speed fluctuation.

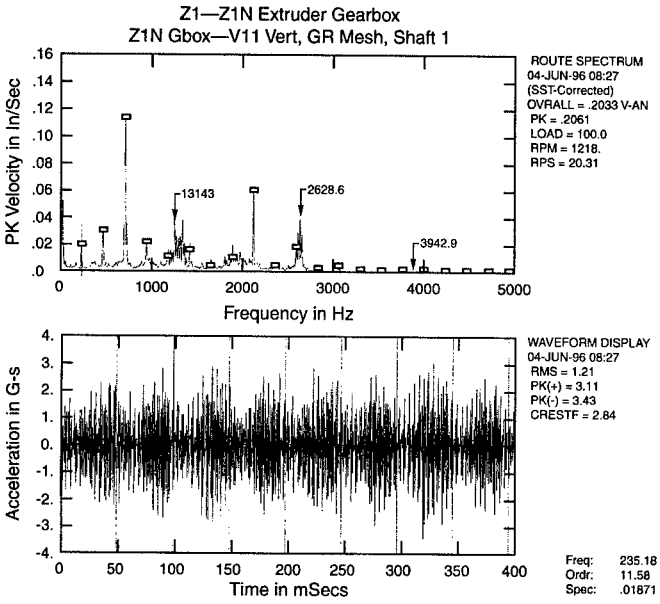


Figure 10. Velocity spectra and acceleration time waveform from extruder gearbox on June 4, 1996.

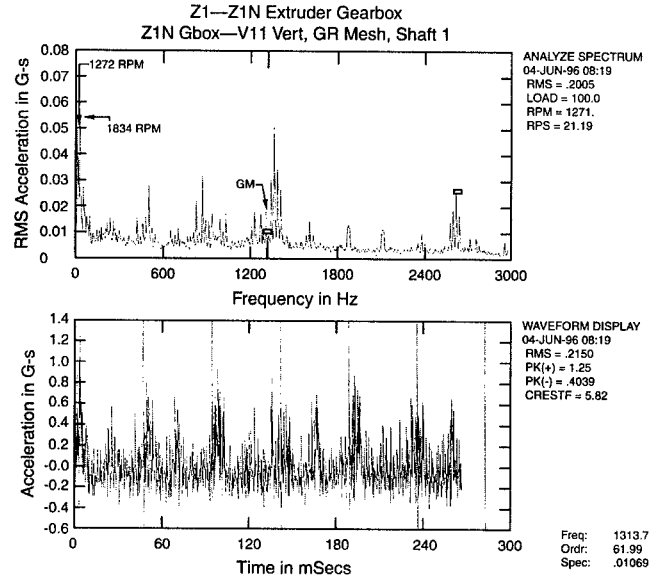


Figure 11. PeakVue spectra and time waveform form extruder gearbox on June 4, 1996.

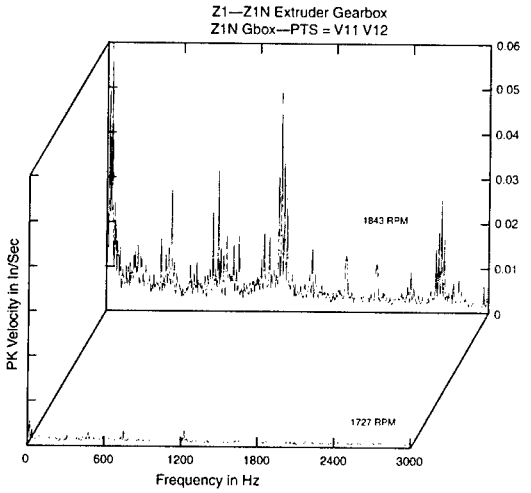


Figure 12. PeakVue spectra on extruder gearbox before DC motor speed adjustment (1843 RPM) and PeakVue spectra on extruder gearbox after DC motor speed adjustment (1727 RPM).

Spectra data on the inboard of the DC motor showed excessive* activity of 0.2 ips-peak at the SCR frequency (360 Hz) with amplitude variation (side banding) at the motor shaft frequency (inlet shaft to gearbox) and first intermediate shaft of the gearbox. The speed controller was adjusted and measurements on the gearbox repeated. The velocity spectra did not change, i.e., the 23T/96T set still showed signs of looseness and the 43T/62T set still showed probable misalignment. There were significant differences in the PeakVue spectra as shown in Figure 12. The spectra captured after adjustment of the speed controller (1727 RPM) shows no indication of impacting.

3.4 Crusher Gearbox: This gear box, driven by a 8-pole 2000 HP motor, drives a rock crusher at a mining facility. A plan view of the gearbox is presented in Figure 13. The gearbox is nominally 17′ × 12′ × 7′ in size. Normal vibration and PeakVue data was acquired on a scheduled basis on the motor and, as much as possible, at each bearing with the gearbox. The input shaft speed was in the proximity of 894 RPM. The first intermediate shaft was turning at 531.5 RPM or 8.86 Hz.

The PeakVue spectra and time waveform data taken at measurement point 3 (see Figure 13) are presented in Figure 14. The dominant activity in the PeakVue spectra are the intermediate shaft turning speed (8.86 Hz) and many harmonics with the 4^{th}, 8^{th}, 12^{th} etc. being dominant. In the PeakVue time waveform, Figure 14, there are four distinct impacts per turn of the first intermediate shaft (the vertical lines are spaced at time increments corresponding to the first intermediate shaft turning speed).

*Vibration exceeding 0.1 ips-peak at the SCR firing frequency on inboard of a DC Motor generally imply a problem in the controller circuit.

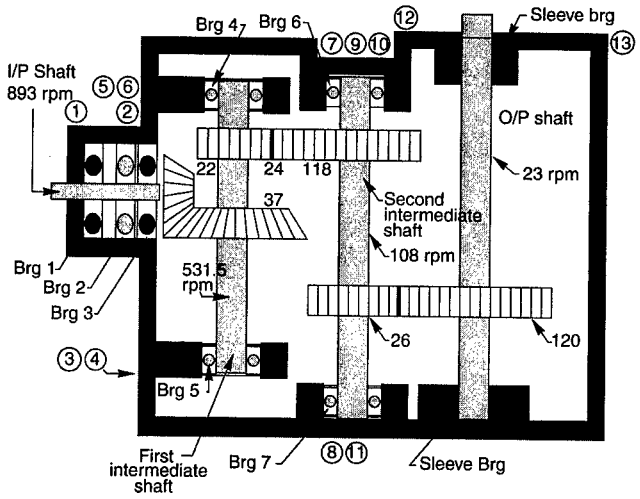


Figure 13. Plan view of crusher gearbox.

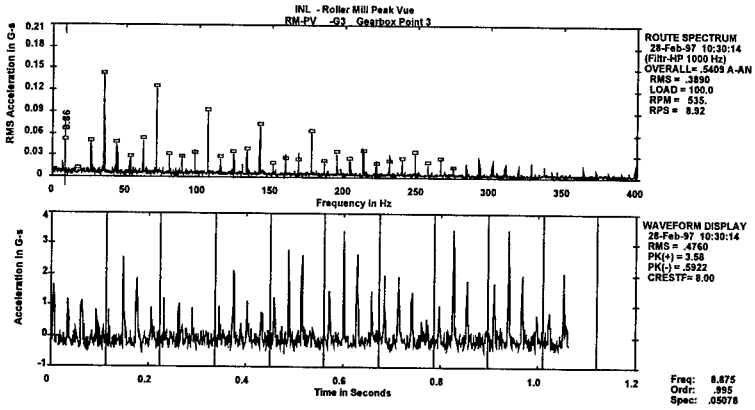


Figure 14. PeakVue spectra and time waveform taken on crusher gearbox at point 3.

The first intermediate shaft has a beveled bull gear with 37 T and a pinion gear with 22 T (see Figure 13). The first postulate was that the 24 T pinion gear had some fault at every 6th tooth since the total number of teeth (24) was divisible by 4. The bothersome fact with this postulate was the g levels of the impacts were greater at measurement point 3 than at measurement point 5 (the 24 T pinion was closer to measurement point 5). The impacts were clearly occurring at four equal intervals per rev of the first intermediate shaft and hence the 37 T bull gear was not considered to be the source where the impacting had occurred (37 not divisible by four).

The gearbox was disassembled and inspected. The 24 T pinion showed no indication of a problem. The impacting was occurring between the 37 T bull gear and the first intermediate shaft. An approximate 2 in. band of fretting was completely around the intermediate shaft at the top of the beveled bull gear.

The postulate then was the impacting was being introduced by a reasonable sharp (high Q) torsional resonance of the input shaft. Strain gauges were installed near the gearbox on the inlet shaft and torsional vibrations data acquired. The torsional resonance spectra showed dominant activity at 35.3 Hz, which is 4 times the intermediate shaft speed.

4. Conclusions: The capture and analysis of stress waves, which accompany many classes of faults experienced in gearboxes, has proven to be an effective diagnostic tool for fault detection and severity assessment in gearboxes. In this paper, five typical examples of faults within gearboxes were presented as case studies. In each case, the normal vibration analysis contributed very little to the fault detection and severity assessment.

The PeakVue methodology for the capture and analysis of the stress waves provide a very powerful trending capability. This is the case since the true amplitude of the specific faults in g-units is captured independent of the machine speed, the analysis bandwidth etc. This ability to capture the true impacting levels provides the knowledge to develop absolute levels from which alert levels and alarms can be set based on a broad case history library. Experience has shown these levels to be dependent on machine speed (in a predictable manner) and fault type (the same as in normal vibration analysis).

Assessment of Data and Knowledge Fusion Strategies for Diagnostics and Prognostics

Gregory J. Kacprzynski
Michael J. Roemer
Rolf F. Orsagh
Impact Technologies, LLC
125 Tech Park Drive
Rochester, NY 14623
(716) 424-1990

Abstract: Various data, feature and knowledge fusion strategies and architectures have been developed over the last several years for improving upon the accuracy, robustness and overall effectiveness of anomaly, diagnostic and prognostic technologies. Fusion of relevant sensor data, maintenance database information, and outputs from various diagnostic and prognostic technologies has proven effective in reducing false alarm rates, increasing confidence levels in early fault detection, and predicting time to failure or degraded condition requiring maintenance action.

The data fusion strategies discussed in this paper are principally probabilistic in nature and are used to aid in directly identifying confidence bounds associated with specific component fault identifications and predictions. Dempster-Shafer fusion, Bayesian inference, fuzzy-logic inference, neural network fusion and simple weighting/voting are the algorithmic approaches that are discussed in this paper. Data fusion architectures such as centralized fusion, autonomous fusion, and hybrid fusion are described in terms of their applicability to fault diagnosis and prognosis. The final goal is to find the optimal combination of measured system data, data fusion algorithms, and associated architectures for obtaining the highest overall prediction/detection confidence levels associated with a specific application. Evaluation of the fusion and diagnostic strategies was performed using gearbox seeded-fault and accelerated failure data taken with the MDTB (Mechanical Diagnostic Test Bed) at the ARL Lab at Penn State University.

Keywords: Fusion, Feature Extraction, Diagnostics, Prognostics

Introduction: The general objective of data or knowledge fusion is to combine information in the most efficient method possible such that the quality of the fused information is equal to or better than the sum of the parts. Specific to health management, this means reduced uncertainty in current condition assessment reduced (improving diagnostics) and better remaining useful life assessment. Multi-sensor data fusion refers to intelligent processing of an array of 2 or more sensors that have cooperative, complimentary and competitive qualities. As long as the sensor array does not contain independent sensors, arrays usually contain various levels of these three qualities. Cooperative sensors are those that work together to create a new piece of diagnostic information, while a complimentary array creates a more complete picture of a problem. Finally, a competitive array provides unrelated measurements of the same physical phenomena for improved reliability (Brooks, 97).

Fusion Application Areas: Within a health management system, there are three main areas where fusion technologies play a contributing role. These areas are shown in Figure 1. At the lowest level, data fusion can be used to combine information from a multi-sensor data array to validate signals and create features. One example of data fusion is combining a speed signal and a vibration signal to achieve time synchronous averaged vibration features.

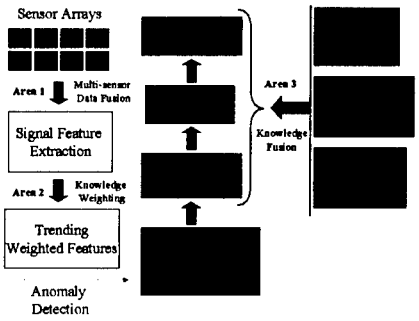


Figure 1 - Fusion Application Areas

At a higher level (area 2), fusion may be used to combine features in intelligent ways so as to obtain the best possible diagnostic information. This would be the case if a feature related to particle count and size in a bearing's lubrication oil were fused with a vibration feature such as kurtosis. The combined result would yield an improved level of confidence about the bearing's health. Finally, Knowledge Fusion (area 3) is used to incorporate experienced-based information such as legacy failure rates or physical model predictions with signal-based information.

One of the main concerns in any fusion technique is the danger of producing a fused system result that is actually performing worse than the best individual tool. This is because poor estimates can drag down the better estimates. The solution to this concern is to weigh the tools according to their capability and performance, which must be realized a priori. The degree of a priori knowledge is a function of the inherent understanding of the physical system and practical experience with the system. The ideal knowledge fusion process for a given application should be selected based on the characteristics of the a priori system information.

Fusion Architectures: Identifying the optimal fusion architecture and approach at each level is a vital factor in assuring that the realized system truly enhances health monitoring capabilities. A brief explanation of fusion architectures will be provided here.

The centralized fusion architecture fuses multi-sensor data while it is still in its raw form as shown in Figure 2. In the fusion center of this architecture, the data is aligned and correlated during the first stage. This means that the competitive or collaborative nature of the data is evaluated and acted upon immediately. Theoretically, this is the most accurate way to fuse data; however, it has the disadvantage of forcing the fusion processor to manipulate a large amount of data. This is often impractical for real-time systems with a relatively large sensor network (Hall, 97).

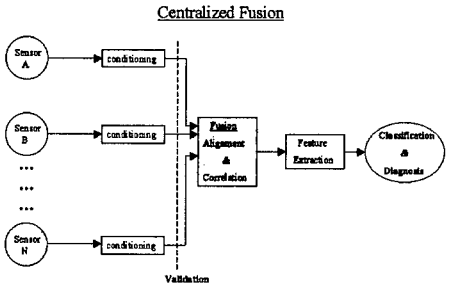


Figure 2 – Centralized Fusion Architecture

The autonomous fusion architecture shown in Figure 3 quells most of the data management problems by placing feature extraction before the fusion process. The creation of features prior to the actual fusion process provides the significant advantage of reducing the dimensionality of the information to be processed. The main undesirable effect of a pure autonomous fusion architecture is that the feature fusion may not be as accurate as in the case of raw data fusion because a significant portion of the raw signal has been eliminated.

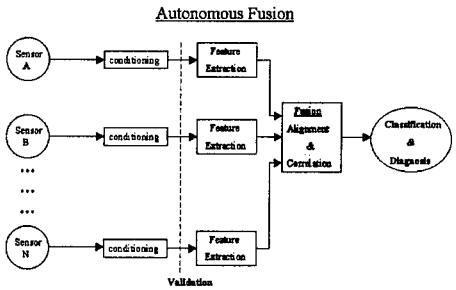


Figure 3 - Autonomous Fusion Architecture

A hybrid fusion architecture takes the best of both and is often considered the most practical because raw data and extracted features can be fused in addition to the ability to "tap" into the raw data if required by the fusion center (Figure 4).

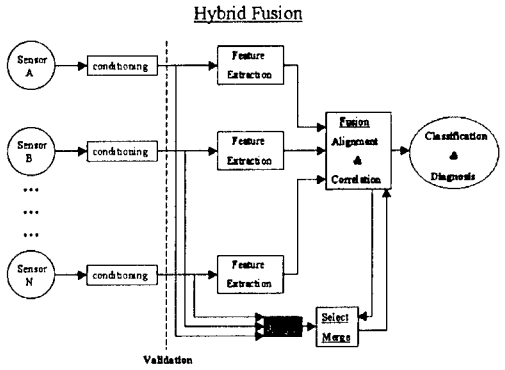


Figure 4 - Hybrid Fusion Architecture

Fusion Techniques: There are probably hundreds of techniques for performing data, feature or knowledge fusion. Because of this fact, sorting through which technique is best can be a daunting and involved task. In addition, there are no hard and fast rules about what fusion techniques or architectures work best for any particular application. The proceeding sections will describe some common fusion approaches such as Baysian Inference, Dempster-Shafer combination, Weighting/Voting, Neural Network Fusion and Fuzzy Logic Inference. A companion paper [3] describes a set of metrics for independently judging the performance and effectiveness of the fusion techniques within a diagnostic system.

Bayesian Inference

Bayesian Inference can be used to determine the probability that a diagnosis is correct, given a piece of a priori information. Analytically this process is described as follows:

$$P(f_1|O_n) = \frac{P(O_n|f_1) \cdot P(f_1)}{\sum_{j=1}^{n} P(O_n|f_1) \cdot P(f_1)} \qquad (1)$$

Where:

$P(f_1|O_n)$ = The probability of fault (f) given a diagnostic output (O), $P(O_n|f_1)$ = The probability that a diagnostic output (O) is associated with a fault (f), and $P(f_1)$ = The probability of the fault (f) occurring.

Bayes' theorem is only able to analyze discrete values of confidence from a diagnostic classifier (i.e. it observes it or it doesn't). Hence, a modified method has been implemented that uses three different sources of information. A-priori probability of failure at time t, ($P_{FO(t)}$) , the probability of failure as determined from the diagnostic classifier ($C_{D(i,t)}$) data, and feature reliability which is independent of time ($R_{D(i)}$). Care must be taken to prevent division by zero.

$$P_{f(t)} = \frac{\sum_{i=1}^{N} C_{D(i,t)}}{P_{FO(t)} \cdot \sum_{i=1}^{N} R_{D(i)}} \qquad (2)$$

The Bayesian process is a common and well established fusion technique, but also has some disadvantages. The knowledge required to generate the a priori probability distributions may not always be available, and instabilities in the process can occur if conflicting data is presented or the number of unknown propositions is large compared to the known propositions.

Dempster-Shafer Method

The Dempster-Shafer method addresses some of the problems discussed above and specifically tackles the a priori probability issue by keeping track of an explicit probabilistic measure of the lack of information. The disadvantage of this method is that the process can become impractical for time critical operations in large fusion problems. Hence, the proper choice of method should be based on the specific diagnostic/prognostic issues that are to be addressed.

In the Dempster-Shafer approach, uncertainty in the conditional probability is considered. The Dempster-Shafer methodology hinges on the construction of a set, called the frame of discernment, which contains every possible hypothesis. Every hypothesis has a belief denoted by a mass probability (m). Beliefs are combined in the following manner.

$$Belief(H_n) = \frac{\sum_{A \cap B = Hn} m_i(A) \cdot m_j(B)}{1 - \sum_{A \cap B = 0} m_i(A) \cdot m_j(B)} \qquad (3)$$

The technique can be best explained through the use of the following example.

Given:

A diagnostic classifier detects Fault A with the following probability and associated uncertainty:
$P_A = 0.80 +/- 0.15$

The a priori probability of Fault A occurring (based on current conditions and a priori information) is the following:
$P_B = 0.30 +/- 0.10$

Therefore: m(A) = 0.65 m(A') = 0.05
m(A,A') = 0.30
m(B) = 0.20 m(B') = 0.60
m(B,B') = 0.20

	B	B'	B,B'
A	0.13	0.39	0.13
A'	0.01	0.03	0.01
A,A'	0.06	0.18	0.06

And: $m(A) + m(B)\ \{True\} = (0.13 + 0.13 + 0.06)/(1-(0.01 + 0.39)) = 0.53$

This result is called the "belief" and it is the fused probability lower bound. The uncertainty in this result is the following:

$$m(A,A') + m(B,B') = 0.06 / (1-(0.01 + 0.39)) = 0.10 \text{ or } +/- 0.05 \quad (4)$$

Hence, the probability of Fault A having actually occurred given the diagnostic output and in-field experience is 0.58 +/- 0.05.

Weighting/Voting Fusion

Both the Bayesian and Dempster-Shafer techniques can be computationally intensive for real-time applications. A simple weighted average or voting technique is another approach that can be utilized. In both these approaches, weights are assigned based on a prior knowledge of the accuracy of diagnostic/prognostic techniques being used. The only condition is that the sum of the weights must be equal to one. Each confidence value is then multiplied by its respective weight and the results are summed for each moment in time. Weights can also change as a function of time.

$$P(F) = \sum_{n=1}^{i} C_{(i,t)} * W_{(i,t)} \quad (5)$$

Where i is the number of features, C is the confidence value, and W is the weight value for that feature. Although simple in implementation, choosing proper weights is of critical importance to highlighting the proper features under various operating modes.

Fuzzy Logic Inference

Fuzzy Logic Inference is a fusion technique that utilizes the membership function approach to scale and combine specific input quantities to yield a fused output. The basis for the combined output comes from scaling the developed membership functions based on a set of rules developed in a rulebase. Once this scaling is accomplished, the scaled membership functions are combined by one of various methodologies such as summation, maximum or "single best" techniques. Finally, the scaled and combined membership functions are used to calculate the fused output by either taking the centroid, max height or midpoint of the combined function.

An example of a feature fusion process utilizing fuzzy logic is shown below in Figure 5. In this example, features from an image are being combined to help determine if a "foreign" object is present in an original image. Image features such as tonal mean, midtones, kurtosis and many others are combined to give a single output that ranks the probability of an anomalous feature being present in the image.

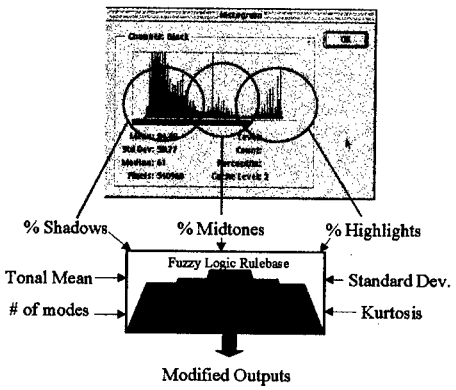


Figure 5 – Example of Fuzzy Logic Inference

Neural Network Fusion

A well accepted application of artificial neural networks (ANNs) is data and feature fusion. For the purposes of fusion, a networks ability to combine information in real-time with the added capability of autonomous re-learning (if necessary) makes it very attractive for many fusion applications.

Artificial neural networks (ANN) utilize a network of simple processing units, each having a small amount of local memory. These units are connected by "communication" links, which carry numerical data. The units operate only on their local data, which is received as input to the units via the connections. Most ANN's have some sort of *training* rule by which the weights of connections are adjusted based on some optimization criterion. Hence, ANN's learn from examples and exhibit certain capability for generalization beyond the training data (examples). ANN's represent a branch of the artificial intelligence techniques that have been increasingly accepted for data fusion and automated diagnostics in a wide range of aerospace applications. Their abilities to fuse features, recognize patterns, and to learn from samples have made ANN's attractive for fusing large data sets from complex systems.

A representative application of neural network fusion would be to combine individual features from different feature extraction algorithms to give a single representative feature. An example of this type of neural network fusion will be given in the following section.

Results

The fusion techniques previously described have been implemented on various vibration features extracted from a data set developed during a series of transitional run-to-failure tests on an industrial gearbox at Penn State ARL. In these tests, the torque was cycled from 100% to 300% load starting at approximately 93 hours. The drivegear experienced multiple broken teeth and the test was stopped at approximately 114 hours. The data collected during the test was processed by many feature extraction algorithm techniques that resulted in 26 vibration features calculated from a single accelerometer attached to the gearbox housing. The features ranged in complexity from a simple RMS level to a measure of the residual signal (gearmesh and sidebands removed) from the time synchronous averaged waveform. More information on these vibration features may be found in [Byington, 1997]. Figures 6 and 7 show plots of two of these features, Kurtosis and NA4, respectively. The smoothed line in each

of these plots is the "ground truth severity" or the probabiltity of failure as determined from visual inspections discussed next.

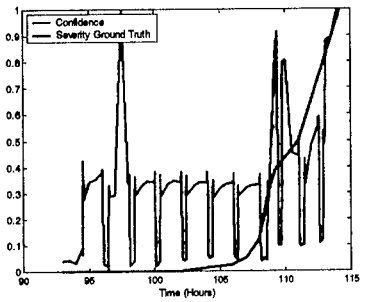


Figure 6 – Kurtosis Feature

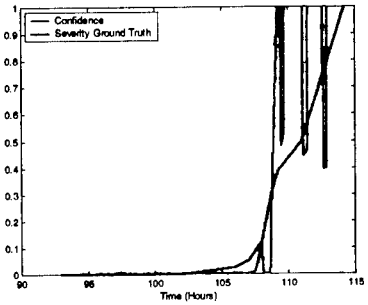


Figure 7 – NA4 Feature

Borescopic inspections of the pinion and drivegear for this particuler test run were performed to bound the time period in which the gear had experienced no damage and when a single tooth had failed. These inspection results, coupled with the evidence of which features were best at detecting tooth cracking prior to failure features (as determined from the diagnsotic metrics discussed later), was the a-priori information used to implement the Bayesian Inference, Weighting/Voting, Neural Network, and Dempster Shafer fusion processes.

The seven best vibration feature as determined by a consistent set of metrics described in [3] were assigned weights of 0.9, average performing features were weighted 0.7, and low performers 0.5 for use in the voting scheme. These weights are directly related to the feature reliability in the Baysian Inference fusion. Similarly, the best features were assigned the uncertainty values of (0.05), average performers (0.10) and low performers (0.15), for the Dempster Shafer combination. The prior probability of failure required for the Neural Network, Bayesian Inference and Dempster Shafer fusion were built upon the experiental evidence that a drive gear crack will form in a mean time of 108 hours with a variance of 2 hours.

Seven of the 26 total vibration features calculated are shown in Figure 8. Note that some of the features have little correlation to the actual tooth failure as defined by the ground truth inspection data. The results of the Dempster-Shafer, Bayesain and Weighted fusion techniques on all 26 features is shown in Figure 9. All three approaches increase in their probability of failure estimates at around 108 hours (index 269). Clearly, the voting fusion is most succeptible to false alarms, the Baysian Inference suggests a probability of failure increase early on but isn't capable of producing a high confidence level. Finally, the Dempster-Shafer combination provides the same early detection, achieves a higher confidence level, but is more sensitive throughout the failure transition region overall.

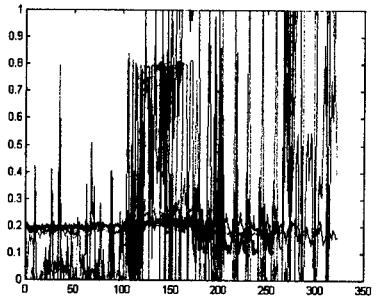


Figure 8 – Top Seven Vibration Features

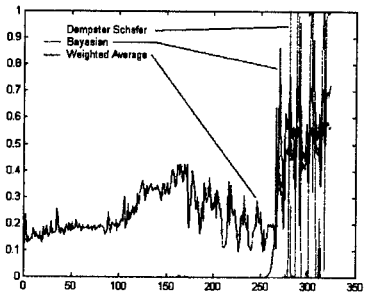


Figure 9 – Fusion of all Features

Next, the same fusion algorithms were applied to just the best seven features. The fusion of these seven features produced more accurate and stable results, which are shown in Figure 10. Note that the Dempster-Shafer combination can now retain a high confidence level with more robustness throughout the critical failure transition region.

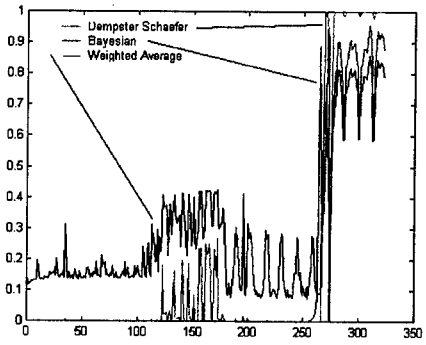


Figure 10 – Fusion of 7 best features

Finally, a simple back propagation neural network was trained on four of the top seven features previously fused (RMS, Kurtosis, NA4, and M8A). In order to train this supervised neural network, the probability of failure as defined by the "ground truth" was required as a-priori information as described earlier. The network automatically adjusts its weights and thresholds (not to be confused with the feature weights) based on the relationships it sees between the probability of failure curve and the correlated feature magnitudes. Figure 11 shows the results of the neural network after being trained by these data sets. The difference between the neural network output and the "ground truth" probability of failure curve is due to error that still exists after the network parameters have optimized to minimize this error. Once trained, the neural network fusion architecture can be used to intelligently fuse these same features for a different test under similar operating conditions.

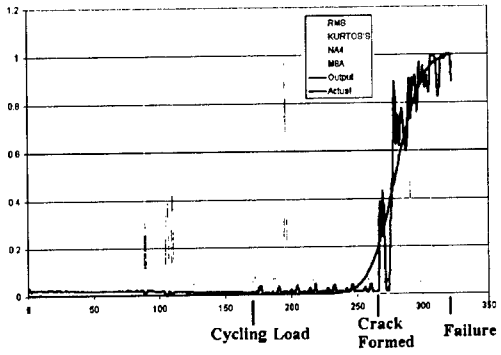


Figure 11 – Fusion with a Neural Network

Conclusion: This paper provides an in-depth discussion about many aspects of fusion including where fusion should exist within a health management system, the different types of fusion architectures, and a number of different fusion techniques. These fusion techniques were applied to vibration features extracted during a transitional failure test associated with an industrial gearbox. The results yielded conclusive evidence that fusion can be very valuable in the diagnostic process if chosen judiciously.

References

[1] Hall, D., and Llinas, J., "An Introduction to Multisensor Data Fusion", Proceedings of the IEEE, January 1997.

[2] Leferve, E., and Colot, O., "A classification method based on the Dempster-Shafer's theory and information criteria" Proceeding of the 2nd International Conference on Information Fusion, July 6-8, 1999.

[3] Orsagh R.F., Roemer M.J., et al "Development of Metrics for Mechanical Diagnostic Technique Qualification and Validation", COMADEM Conference, Houston TX, December 2000.

[4] Agosta, J. M., and Weiss, J. W., "Active Fusion for Diagnosis Guided by Mutual Information Measures", Proceeding of the 2nd International Conference on Information Fusion, July 6-8, 1999

[5] Brooks, R. R., and Iyengar, S. S, Multi-Sensor Fusion, Copyright 1998 by Prentice Hall, Inc., Upper Saddle River, New Jersey 07458

[6] Roemer, M. J. and Kacprzynski, G.J., "Advanced Diagnostics and Prognostics for Gas Turbine Engine Risk Assessment," Paper 2000-GT-30, ASME and IGTI Turbo Expo 2000, Munich, Germany, May 2000.

[7] Roemer, M. J., and Ghiocel, D. M., "A Probabilistic Approach to the Diagnosis of Gas Turbine Engine Faults" Paper 99-GT-363, ASME and IGTI Turbo Expo 1999, Indianapolis, Indiana, June 1999.

[8] Roemer, M. J., and Atkinson, B., "Real-Time Engine Health Monitoring and Diagnostics for Gas Turbine Engines," Paper 97-GT-30, ASME and IGTI Turbo Expo 1997, Orlando, Florida, June 1997.

[9] Byington, C.S., Kozlowski, J.D., "Transitional Data for Estimation of Gearbox Remaining Life", Proceedings of the 51st Meeting of the MFPT, April 1997.

MACHINERY DIAGNOSTIC FEATURE EXTRACTION AND FUSION TECHNIQUES USING DIVERSE SOURCES

Chongchan Lee, Ph.D.
Wyle laboratories
P.O. Box 077777
Huntsville, AL 35807-7777
chongchanlee@hnt.wylelabs.com

John Pooley
AMTEC Corporation
500 Wynn Drive, Suite 314
Huntsville, AL 35816-3429
jpooley@amtec-corp.com

Abstract: In order to optimize helicopter operational readiness a Joint Advanced Health and Usage Management System (JAHUMS) for helicopter must be highly reliable, minimize false alarms, and prevent catastrophic failures, while operating in real time. To achieve these goals, a fusion of features extracted from non-commensurate factors such as, vibration with oil temperature, oil pressure, and wear debris signatures was implemented via statistical fusion techniques.

This feature fusion of non-commensurate factors provides improved diagnoses capability and reduces false alarms. For example, there may be instances where one analysis factor indicates a fault while another has a contra indication. Clearly, fusion of non-commensurate features will be a very effective way to overcome these conflicts, thereby providing better diagnosis performance and improved flight safety of helicopters.

Another advantage of this feature fusion is significant data compression through dimensionless statistical discriminators, which is indispensable to efficient storage utilization and on-line real-time analysis. Therefore, data fusion of non-commensurate sources provides efficient machinery diagnosis and prognosis for both the military and commercial field.

Key words: Dimensionless discriminants; Feature extraction; Feature fusion; Machinery diagnostics; Normalization; Nominal/Anomalous diagnostic discriminator

Introduction: The JAHUMS operational system for helicopter must be highly reliable, minimize false alarms, and give sufficient advance warning to prevent catastrophic failures, while operating in real time, hence a high utilization rate for helicopter availability.

In order to achieve these goals, the fusion of features extracted from vibration signatures, with non-commensurate signatures derived from transmission oil temperature, transmission oil pressure, and wear debris was implemented via statistical fusion techniques. This fusion accomplishes nominal/anomalous diagnostic and prognostic health monitoring for JAHUMS. This idea fuses non-commensurate factors for JAHUMS, which provides increased diagnoses capabilities and reduces the number of false alarms. There

may be instances when one technique indicates a fault while another has a contra indication. For example, in applications where sliding wear is prevalent, chip detection might detect increasing rates of wear generation, while the vibration amplitude remains nominal. Another example, oil pressure and tail transmission oil temperature failure will not be detected by vibration features. Any anomalous condition indications could cause catastrophic failures. Therefore, features extracted from any single factor might not indicate an anomalous condition, whereas other feature co-factors may indicate an anomalous condition. Clearly, fusion of non-commensurate with vibration signatures provides better machine diagnosis performance and improved flight safety for helicopters.

Other distinguished features of JAHUMS system developed by AMTEC Corporation and Wyle Laboratories Incorporated are their data compression and dimensionless normalized features, which enables standardizing analysis regardless of helicopter torque or load conditions. This is achieved using normalized and dimensionless scaling transforms called statistical measures. Data compression is a must when sampling rates are high and data storage is limited. These techniques make real time and on-line a reality for JAHUMS.

Therefore, fusion techniques, which integrate data compression and dimensionless normalized features for the novelty detection of JAHUMS operational system provide a truly advanced and reliable, yet efficient machinery diagnosis and prognosis system for both military and commercial field.

Nominal/Anomalous Diagnostic Discriminator Plan for JAHUMS: A helicopter condition monitoring system most critical mission is to improve readiness and save crew's life, also it should be an efficient near real time on-line system. To meet these requirements, JAHUMS operational system employs novelty detection, to discriminate nominal/anomalous condition of a helicopter, which is a necessary requisite of fault diagnosis. This nominal/anomalous diagnosis initiates with discrimination analysis that produces normalized signatures suitable for transforming into dimensionless standardized statistical score elements, these elements can be individually fused into a feature vector.

To this purpose, the vibration data from seventeen sub-assemblies in the helicopter gearbox system are monitored. Each sub-assembly is paired with a single accelerometer, which is located at different locations on the gearboxes. In other words, conditions of crucial components such as, bearings and gears are being monitored by a single accelerometer mounted on a single sub-assembly. Whereas, only nine non-commensurate sensors are employed to monitor conditions of the gearbox system, meaning a single non-commensurate sensor will simultaneously monitor several sub-assemblies. Therefore, each accelerometer's vibration data contains signature information unique to its paired gearbox sub-assembly, while the same gearbox's non-commensurate data such as transmission oil temperature, oil pressure, and chip presence will include signature information common to the entire gearbox assembly. Table I summarizes a list of candidate components to be monitored.

Table I. Vibration Signal Sources from the Components Allocated in Sub-Assembly

Sub-Assembly	Components To Be Monitored		
	Bearings	**Gears**	**Gearbox Operating Condition**
Sub-Assy 1 (MGB: Port Ring)	Planet Carrier Sphr Roller 1, 2, 3, 4, and 5	Sun Gear; Planet Spur Gears; Ring Gear Planetary Assy.	Chip Input MDL LH Chip Input MDL RH Chip Main MDL Sump Oil Temperature Oil Pressure
Sub-Assy 2 (MGB: Stbd Ring)	Planet Carrier Sphr Roller 1, 2, 3, 4, and 5; 4th Hydraulic Pump Thrust, and Pump Preload	Sun Gear; Planet Spur Gears; Ring Gear Planetary Assy.	
Sub-Assy 3 (MGB: Stbd Main)	Stbd Main Bevel Pinion, and Pinion Roller; 4th Hydraulic Pump Thrust, and Pump preload	Main Bevel Gear	
Sub-Assy 4 (MGB: Port Main)	Port Main Bevel Pinion, and Pinion Roller; 4th Hydraulic Pump Thrust, and Pump preload	Main Bevel Gear	
Sub-Assy 5 (MGB: Port Input)	Port Input Pinion Roller, Ball, and INBC Roller; Port Input Gear Roller, and DBL; Port FWU Ball Input, Input Ball, and Output Ball; Port Gen Drive Ball; Port Hydraulic DR	Input Spiral Bevel Pinion, and Bevel Gear; Freewheel Spiral Bevel; Acc. Drive Bevel Pinion; Gen Spur Gear; Hyd Pump Spur Gear; Main Spiral Bevel Pinion	
Sub-Assy 6 (MGB: TTO Rad)	Main Rotor Roller, Timken Preload, and Timken Thrust; Main Rotor Swash Plate Bearing; Tail Take Thrust; Tail Take off Preload	Main Bevel Gear; TTO Main Bevel Gear; TTO Pinion Spiral Bevel Gear	
Sub-Assy 7 (MGB: Stbd Input)	Input Pinion Roller, Ball, and INBC Roller; Input Gear Roller, and DBL; FWU Ball Input, Input Ball, and Output Ball; Gen Drive Ball; Hydraulic DR	Input Spiral Bevel Pinion, and Gear; Freewheel Spiral Bevel; Acc. Drive Bevel Pinion; Gen Spur Gear; Hyd Pump Spur Gear; Main Spiral Bevel Pinion	
Sub-Assy 8	Hanger Bearing No. 1	N/A	Chip-Detection Oil Temperature
Sub-Assy 9	Hanger Bearing No. 2	N/A	
Sub-Assy 10	Hanger Bearing No. 3	N/A	
Sub-Assy 11	Hanger Bearing No.4	N/A	
Sub-Assy 12	Pillow Block Bearing	N/A	
Sub-Assy 13	Pylon Shaft Bearing	N/A	
Sub-Assy 14 (IGB: IGB Input)	IGB Input Thrust IGB Input Preload	Input Pinion Bevel Gear; Output Bevel Gear	
Sub-Assy 15 (IGB: IGB Output)	IGB Output Thrust IGB Output Preload	Input Pinion Bevel Gear; Output Bevel Gear	
Sub-Assy 16 (TGB: TGB Input)	TGB Input Thrust TGB Input Preload	Input Pinion Bevel Gear; Output Bevel Gear	Chip-Detection Oil Temperature
Sub-Assy 17 (TGB: TGB Output)	TGB Output Thrust TGB Output Preload	Input Pinion Bevel Gear; Output Bevel Gear	

The vibration signal of each sub-assembly and paired non-commensurate data will go through the feature extraction process using various diagnostic techniques to extract operational condition information on bearings, gears, and overall operating condition. The features from vibration and non-commensurate data are fused to form a feature vector,

which comprehensively represents SH-60 helicopter gearbox operational condition, thereby providing a more reliable and improved HUMS system. The overall schematic diagram for nominal and anomalous diagnostics of SH-60 helicopter gearboxes is illustrated in Figure 1.

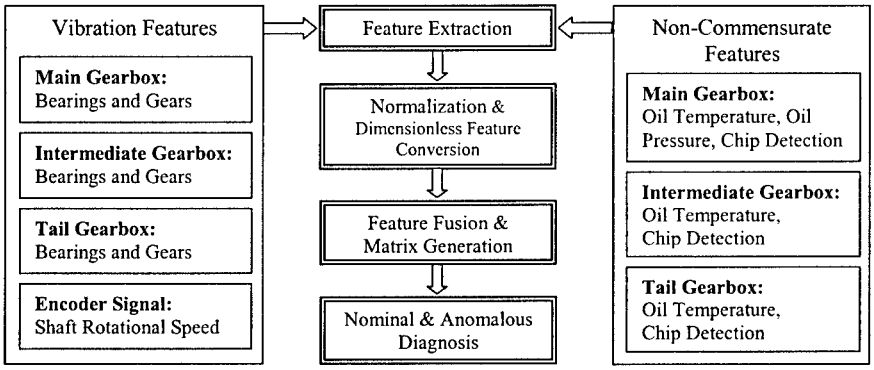


Figure 1. Overall Schematic Diagram for Nominal/Anomalous Diagnostic System

Information from a single accelerometer and non-commensurate sensors generate a feature matrix denoted by $\mathbf{FV}(i)_{m,n}$, comprised of a hundred feature vectors. An index *i* in the parenthesis denotes a sub-assembly or accelerometer number. A set of raw accelerometer data will be segmented into one hundred segments of equal length, and each segment is utilized to generate a row feature vector $\mathbf{FV}_m$ by using various nominal/anomalous diagnostic discrimination techniques, including statistics, DSP, frequency spectrum analysis, and classification from vibration and non-commensurate data. Therefore, there will be a hundred feature vectors collected into a feature matrix that has a dimension of 100 by 100 (Figure 2). To evaluate the condition of the entire SH-60 helicopter gearboxes, the same processes will be executed on all the 17 sub-assemblies of the SH-60 drive train.

Data Conditioning Process and Feature Vector Generation: Diagnostic algorithms are customized to each gearbox component to be analyzed. For a gear nominal/anomalous discriminator, synchronous averaging technique will be employed. In the averaging process, in-phase components will add together while the rest of the signal components will gradually cancel because of their random or non-synchronous relative phases. Therefore, the background extraneous noise and vibrations from other shafts and gears will be canceled because they are not phase synchronized an attenuate as the number of averages increase. Synchronous averaged signals initiates the gear nominal/anomalous discrimination analysis. Likewise, envelope analysis is utilized for bearing nominal/anomalous discrimination to diagnose rolling element bearing faults.

While vibration data analysis provides detailed gearbox components signature, non-commensurate chip detector data gives overall operating conditions of the gearbox that

can be characterized as fault or nominal. The gearbox operating condition discriminator module identifies the gearbox's operating condition as nominal or anomalous. Table II summarizes a list of various analysis techniques used by each nominal and anomalous discriminator module.

{Feature Matrix}=

$$FV(i)_{m,n} = \begin{Bmatrix} FV_1 \\ FV_2 \\ FV_3 \\ . \\ . \\ . \\ . \\ . \\ FV_m \end{Bmatrix} = \begin{Bmatrix} a_{1,1}\ a_{1,2}\ a_{1,3}\ a_{1,4}\ a_{1,5}\ a_{1,6}\ a_{1,7} \cdots\cdots a_{1,96}\ a_{1,97}\ a_{1,98}\ a_{1,99}\ a_{1,n} \\ a_{2,1}\ a_{2,2}\ a_{2,3}\ a_{2,4}\ a_{2,5}\ a_{2,6}\ a_{2,7} \cdots\cdots a_{2,96}\ a_{2,97}\ a_{2,98}\ a_{2,99}\ a_{2,n} \\ a_{3,1}\ a_{3,2}\ a_{3,3}\ a_{3,4}\ a_{3,5}\ a_{3,6}\ a_{3,7} \cdots\cdots a_{3,96}\ a_{3,97}\ a_{3,98}\ a_{3,99}\ a_{3,n} \\ . \quad . \quad . \\ . \quad . \quad . \\ . \quad . \quad . \\ . \quad . \quad . \\ . \quad . \quad . \\ a_{m,1}\ a_{m,2}\ a_{m,3}\ a_{m,4}\ a_{m,5}\ a_{m,6}\ a_{m,7} \cdots\cdots a_{m,96}\ a_{m,97}\ a_{m,98}\ a_{m,99}\ a_{m,n} \end{Bmatrix}$$

Figure 2. Feature Matrix and Its Elements

Table II. Nominal/Anomalous Diagnostic Analysis Techniques for Bearings and Gears

Component	Bearings	Gears	Operating Conditions
Detection Techniques	Skewness Kurtosis Crest Factor Impulse Factor Clearance Factor Shape Factor Signal-std (RMS) Signal-pk-pk Envelope Band Energy (BE-std-n) Envelope Band Energy (BE-mean-n) Envelop Band Kurtosis (BKv-n) Energy in the base band (EB-n) ORDI-n IRDI-n REDI-n OREI-n IREI-n REEI-n Peak Ratio RMS Ratio-A RMS Ratio-B	Skewness Kurtosis Fourth Root of Kurtosis Impulse Factor Clearance Factor Signal-std (RMS) Signal-pk-pk Signal-pk Peak ratio TEO-G Rice Frequency Harmonic Index SO1 SO2 FM0 FM1 FM2A, FM2B FM3 FM4A, FM4B NA4 1st Harmonic ratio	Classification

Feature Vector Sets and Their Elements: To reduce the errors due to inadequacies in the sensory system it is essential to select the best combination of input parameters. This

technique is called feature extraction, and a collection of these feature elements is called a feature vector, which is comprised of 100 elements.

This feature vector can be partitioned into three main categories; time invariant, time variant, and nominal bounding. A feature vector and its composition of elements is described in Figure 3.

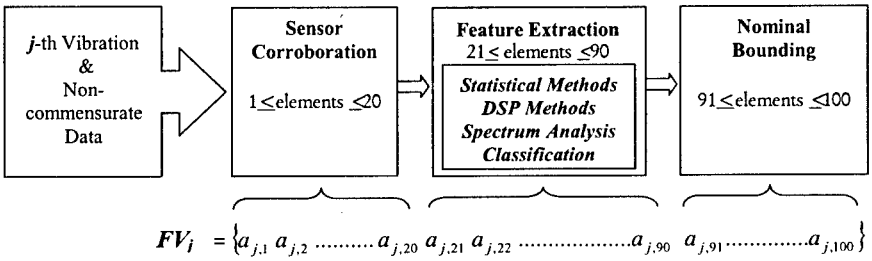


Figure 3. Feature Vector and Its Elements Generation from Modules

Time invariant feature elements ($a_{k,21}$ to $a_{k,90}$), are generated during raw accelerometer signal qualification, and consist of four major parts; sensor corroboration, rotational stability, randomness verification, and stationarity validation, which are necessary to validate correct sensor operation, to correct variations in shaft speed, to confirm data quality. An invalid accelerometer performance may cause misdiagnosis of the gearbox condition. Therefore, prior to using a series of data collected from an SH-60 mission, every accelerometer's performance must be corroborated to determine if their dynamic range is acceptable to support its successful analysis. If a sensor fails this time invariant analysis, the respective data from an accelerometer is discarded and corrective action will be taken.

Once the sensor performance is corroborated, the time variant feature elements, $a_{k,l}$ where $l = 21$ to 90, known as nominal/anomalous discriminant features, will be generated by using analysis techniques listed in Table 2. Discriminants extracted during machinery diagnostic analysis usually describe some specific local waveform property of the machinery's operational health. These discriminants are usually in spectral, time or amplitude domain and are of the dimensional variety. The SH-60 Seahawk helicopter drive train components are comprised of various gears and bearings. Conditions on bearings and gears, along with overall gearbox operating conditions are characterized by vibration, gearbox oil temperature, pressure, and wear debris signatures. The nominal/anomalous discriminator is composed of three major analysis algorithm modules, which are respectively bearings, gears, and overall internal gearbox condition nominal/anomalous discriminator algorithms (Figure 4).

This analysis should be conducted intra-dependently because signals generated by bearings are of different waveform and characteristics than that produced by gears and other non-commensurate data. The characteristic of gear vibration is its synchronicity of multiples of shaft rotational speed known as shaft order; therefore, a gear nominal/anomalous discriminator must be capable of isolating the gear components from all others. On the

other hand, bearings have been difficult to monitor due to its low energy distribution masked by extraneous noise from many sources. In order to detect bearing anomalies effectively, pre-signal processing techniques, known by either the term Enveloping or Demodulation, has to be employed. The gearbox operating condition discriminator module uses non-commensurate data to identify overall gearbox condition using detailed information such as oil temperature, pressure, and debris presence in MGB, IGB, and TGB.

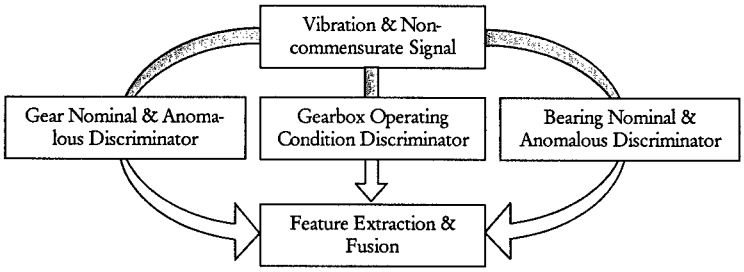


Figure 4. Nominal/Anomalous Diagnostic Discriminator Algorithm Structure

The final set of elements ($a_{k,l}$ where l =91 to 100) are generated that bound the classification of the helicopter drive train nominal operation. Nominal bounding shows how the nominal operational performance for drive train components can be bounded across the test group population of SH-60 helicopters. Magnitude and phase of the feature vector are calculated, and converted into standard statistical scores. All subsequent data can be statistically tested for its significance against the baseline to bound nominal operation for the population of SH-60 helicopters across the spectrum of their flight regime. Any subsequent anomalous data that is tested against these nominal criteria will be outside the acceptance region and will be classified as anomalous.

Nominal/Anomalous Diagnostic System and Its Benefits:

1. Dimensionless and normalized feature elements

Results from feature extraction techniques are not very useful when we want to make general comparisons across all fleets since an output level might not be the same from feature extraction to feature extraction, and helicopter to helicopter. Utilization of feature extracted outputs for comparison across two or more helicopters simply does not work if, as is usually the case, the helicopters have different running condition. Therefore, there is a need for global dimensionless discriminants that indicate the overall health condition of gearbox under analysis relative to its position in the population. Therefore, this procedure standardizes each feature value across helicopters so that their respective positions in population distribution can be evaluated. One way is to transform the values into scores within a universal scale. Standard scores, also called z scores, do this; they describe the relative position of a single score in the population distribution of scores. The benefits of converting the raw values into z scores are: 1) the shape of the distribu-

tion of standard scores is identical to that of the original distribution of raw values, 2) the mean of the distribution of z scores will always equal to zero regardless of the value of the mean in the raw value distribution. Therefore, dimensionless diagnostic discriminants' superiority to dimensional discriminants is predicated on their insensitiveness to amplitude and scale variations from helicopter to helicopter.

All feature vector element statistical scores will be sent to novelty detection diagnostic discriminator to make the final decision, nominal or anomalous. Additionally, these standardized scores known range afford significant data compression without loss of intelligence, to optimally diagnose the target machine's operating health.

2. Data compression

The resulting outputs from novelty detection diagnostic algorithms is transformed into statistical measures, which are of normalized and dimensionless statistical standard scores that provide relative positioning in the normal distribution yet retain individualized measurement scale, which results in significant data compression. Benefits of data compression are fast and efficient data calculation and reduced memory space, thus the system is able to be a real time and on-line system.

3. Data fusion

Normalized and dimensionless signature features enable the fusion of different characteristics such as vibration data and non-commensurate data. The major benefit from incorporating the non-commensurate data into nominal/anomalous features for the helicopter diagnostic system is to provide improved diagnoses capability and to reduce false alarms, thereby providing better diagnosis performance and improved helicopter operational health monitoring. For example, there may be instances where one analysis type indicates a fault while another has a contra indication, which is important because any abnormal condition indicators cause catastrophic failures. Another benefit from the implementation of data fusion is ability to simplify gauge display. Therefore, utilization of merged signatures increases diagnostic capabilities and results in less machinery mechanical trouble shooting.

Conclusion: JAHUMS must be highly reliable, and minimize false alarms. In order to achieve this goal, an advanced dynamic machinery health monitoring system that transforms of features extracted from raw data into normalized and dimensionless feature elements, has consistent output level via normalized dimensionless features, and standardize ranges regardless of different helicopter operational states of varying torque or load conditions. Implementation of significant data compression, utilization of data fusion, and integration of these two technique for the novelty detection of helicopter operational health monitoring provides an advanced, reliable, efficient, and robust diagnosis and prognosis system for both the military and commercial field. Also, this system gives sufficient warning to prevent catastrophic failures and a high utilization rate for helicopter availability, while operating in real time.

Bibliography:

[1] Cempel, C. (1991). Vibroacoustic Condition Monitoring. Chichester, England: Ellis Horwood Limited.

[2] Cempel, C. (1991). Diagnostically Oriented Measures of Vibroacoustical Process. Journal of Sound and Vibration, 73 (4), 547-561.

[3] Cempel, C. (1981). Amplitude and Spectral Discriminants of Vibroacoustical Processes for Diagnostic Purposes. STROJNÍCKY ČASOPIS, 32 (2), 171-179.

[4] Cleveland, G. & Trammel, C. (1996, June). An Integrated Health and Usage Monitoring System for the SH-60B Helicopter. AHS 52nd Annual Forum. 1767-1787.

[5] Decker, H. J. & Zakrajsek. J. J (1999, April). Comparison of Interpolation Methods as Applied to Time Synchronous Averaging. Proceedings of the 53rd Meeting of the SMFPT. 293-300.

[6] Hardman, W. J. & Frith. P. C. (1995, August). Interim Report: Analysis of a Severe IGB Tooth Fault Implanted In The 8W SH-60 Drive Train Rig (NAVAIRWARCENACDIVTRENTON-LR-PPE-95-7). Department of the Navy: Naval Air Systems Command.

[7] Hass, D. J., Spracklen, D., Baker, T., & Schaefer, Jr., C. G. (2000, May). Joint Advanced Health and Usage Monitoring System (JAHUMS) Advanced Concept Technology Demonstration (ACTD). AHS 56th Annual Forum.

[8] Hess, B. SH-60B IMDS Core Parameters (E-3538 Draft). BFGoodrich Aerospace Aircraft Integrated Systems.

[9] Howard, I. (1994, October). A Review of Rolling Element Bearing Vibration Detection, Diagnosis and Prognosis (DSTO-RR-0013). Melbourne, Australia: DSTO Aeronautical and Maritime Research Laboratory.

[10] Hinkle, D. E., Wiersma W., & Jurs, S. G. (1994). Applied Statistics for the Behavioral Sciences. Boston, MA: Houghton Miffin Company.

[11] Lee, C. & Liu T. I. (1998, November), On-Line Evaluation of Drill Conditions for Operational Safety and Product Quality. ASME IMECE 1998 Symposium on Advances in Tolling and Fixturing for Manufacturing.

[12] Mignano, F. (Date Unknown). An Advanced Method for Detecting Gear Teeth Problems. 655-659.

[13] NATOPS Flight Manual Navy Model SH-60B Aircraft (A1-H60BB-NFM-000). Authority of the Chief of Naval Operations and Under the Direction of the Commander Naval Air Systems Command.

[14] Stewart, R. M. (1997, September). Some Useful Data Analysis Techniques for Gearbox Diagnostics. Proceedings of the Meeting on the Applications of Time Series Analysis at the ISVR. 18.1-18.19.

STANDARDS DEVELOPMENT

Chair: **Mr. Henry C. Pusey**
MFPT Society

STANDARDS DEVELOPMENTS FOR CONDITION-BASED MAINTENANCE SYSTEMS

Michael Thurston and Mitchell Lebold
Applied Research Laboratory
Penn State University
P.O. Box 30
State College, PA 16804-0030

Abstract: An effort is underway to develop an Open System Architecture for Condition-Based Maintenance. The architecture development has focused on the definition of a distributed software architecture for CBM. The distributed software model was selected due to the recent emergence of enabling software technologies and the benefits of the approach. In particular, the availability of network connectivity provides a ready hardware backbone over which the software system may be distributed. The requirements for a general CBM architecture are defined, and the framework of the distributed architecture is provided.

Keywords: Condition-based Maintenance; Condition Monitor; Health Assessment; MIMOSA; Open System Architecture; Prognostics

Introduction: Condition Based Maintenance (CBM) is becoming more wide-spread within US industry and the military. The factors that have driven an increase in the use of CBM include the need for: reduced maintenance and logistics costs, improved equipment availability, and protection against failure of mission critical equipment.

A complete CBM system comprises a number of functional capabilities: sensing and data acquisition, signal processing, condition and health assessment, prognostics, and decision aiding. In addition, a Human System Interface (HSI) is required to provide user access to the system. The implementation of a CBM system usually requires the integration of a variety of hardware and software components. Across the range of military and industrial application of CBM, there is a broad spectrum of system level requirements including: communication and integration with legacy systems, protection of proprietary data and algorithms, a need for upgradeable systems, and limits implementation time. Most purchasers of CBM systems or equipment are also very sensitive to the costs.

Standardization of specifications within the community of CBM users will, ideally, drive the CBM supplier base to produce interchangeable hardware and software components. A non-proprietary standard that is widely adopted will result in a free market for CBM components. The potential benefits of a robust non-proprietary standard include: improved ease of upgrading for system components, a broader supplier community resulting in more technology choices, more rapid technology development, and reduced prices. This paper will describe an effort underway to develop an open system standard for CBM systems.

Open Systems and Standards: Openness is a general concept that denotes free and unconstrained sharing of information. In its broadest interpretation, the term "open systems" applies to a systems design approach that facilitates the integration and interchangeability of components from a variety of sources. For a particular system integration task, an open systems approach requires a set of public component interface standards and may also require a separate set of public specifications for the functional behavior of the components. The underlying standards of an open system may result from the activities of a standards organization, an industry consortium team, or may be the result of market domination by particular product (or product architecture). Standards produced by recognized standards organizations are called de jure standards. De facto standards are those that arise from the market-place, including those generated by industrial consortia. Regardless of the development history of the standards that support an open system, it is required that the standards are published, and publicly available at a minimal cost. An example of an open de jure standard is the IEEE 802.3 which defines medium access protocols and physical media specifications for LAN Ethernet connections. An example of a proprietary de facto standard is the Windows OS (operating system). Examples of open de facto standards are the UNIX OS and HTTP.

An open system standard that receives wide-spread market acceptance can have great benefits to consumers and also to suppliers of conformant products. The emergence of the IBM PC architecture as a market leading de facto standard stimulated the market for both PC hardware and software suppliers. It also led to price reductions due to market competition in the face of rapid technology advances. The emergence of a dominant proprietary standard for PC operating systems and software resulted in benefits to consumers in terms of application interoperability, but arguably at the cost of increased prices and reduced performance compared to the possibilities that an open system software standard might have offered. In addition to commercial issues, the interchangeability of components enabled by an open system architecture yields several technical benefits: system capability can be readily extended by adding additional components, and system performance can be readily enhanced by adding components with improved or upgraded capabilities.

CBM Architectures: A complete architecture for CBM systems should cover the range of functions from data collection through the recommendation of specific maintenance actions. The key functions that facilitate CBM include:

- sensing and data acquisition
- signal processing and feature extraction
- production of alarms or alerts
- failure or fault diagnosis and health assessment
- prognostics: projection of health profiles to future health or estimation of RUL (remaining useful life)
- decision aiding: maintenance recommendations, or evaluation of asset readiness for a particular operational scenario
- management and control of data flows or test sequences

- management of historical data storage and historical data access
- system configuration management
- human system interface

Typically, CBM system integrators will utilize a variety of COTS (commercial off the shelf) hardware and software products (using a combination of proprietary and open standards). Due to the difficulty in integrating products from multiple vendors, the integrator is often limited in the system capabilities that can be readily deployed. For some applications a system developer will engineer an application specific system solutions. When user requirements drive custom solutions, a significant part of the overall systems engineering effort is the definition and specification of system interfaces. The use of open interface standards would significantly reduce the time required to develop and integrate specialized system components. CBM system developers and suppliers must make decisions about how the functional capabilities are distributed, or clustered within the system. Due to integration difficulties in the current environment, suppliers are encouraged to design and build components which integrate a number of CBM functions. Furthermore, proprietary interfaces are often used to lock customers into a single source solution, especially for software components. An ideal Open System Architecture for CBM should support both granular approaches (individual components implement individual functions) and integrated approaches (individual components integrate a number of CBM functions).

A given CBM architecture may limit the flexibility and performance of system implementations if it does not take into account data flow requirements. To support the full range of CBM data flow requirements, the architecture should support both time-based and event-based data reporting and processing. Time-based data reporting can be further categorized as periodic or aperiodic. An event-based approach supports data reporting and processing based upon the occurrence of events (limit exceedences, state changes, etc...). A specific requirement that may be imposed on a CBM system involves the timeliness of data reporting. Timeliness requirements may be defined broadly as time-critical or non time-critical. The non time-critical category applies to data messages or processing for which delays have no significant impact on the usefulness of the data or processing result. The time-critical category implies a limited temporal validity to the data or processing result requiring deterministic and short time delays [1]. Two different messaging approaches may also be employed within a system, synchronous and asynchronous. In the synchronous model, the message sender waits for a confirmation response from the message receiver before proceeding to its next task. In the asynchronous model, the message sender does not wait for a response from the receiver and closes the communication. The asynchronous model is generally more applicable to time-critical communications.

A variety of communication models exist for communication between a network of components, they include: multicast, broadcast, and client-server. In the multicast model, the information supplier publishes his information to the network, addressed to a known list of recipients; an asynchronous approach. In the broadcast model, the information supplier publishes his information to all network listeners and the listeners must decide if they are interested in the content of the message. Finally, the c-s (client-server) model pairs a client (who initiates communication) with a server (who is designed

to respond to certain requests). The server implements interfaces that may be used by a client to request a service. A client can only request services available at the server's public interfaces. Data passing may be implemented by means of a single synchronous message, or through a pair of asynchronous messages.

Current PC, Networking, and Internet technologies provide a readily available, cost effective, easily implemented communications backbone for CBM systems. These computer networks are built over a combination of open and proprietary standards. Software technologies are rapidly developing to support distributed software architectures over the Internet and across LANs. There is a large potential payback associated with market acceptance of an open standard for distributed CBM system architectures. With the ready availability of network connectivity, the largest need is in the area of standards for a distributed software component architecture; that subject will be the focus of the remainder of this paper.

One model for distributed computing is Web-based computing. The Web-based model utilizes HTTP servers which function primarily as document servers. The most common medium of information transport on the Web is the HTML page; HTML is a format for describing the content and appearance of a document. An alternate format for information transport over the Web is becoming increasingly popular; XML (eXtensible Mark-up Language). In contrast to HTML, XML is focused on describing information content and information relationships. XML is readily parsed into data elements that application programs can understand and serves as an ideal means of data and information transport over the web. A simple model for data access over the web is a smart sensor which periodically posts new data in XML format to a Web page. Information consumers may access that updated data directly from the Web-page. HTTP servers also provide remote access to application programs by means of the Common Gateway Interface (CGI). In this model, the interface between the remote client and the Web server is by means of HTML pages or XML; the webserver utilizes the CGI to communicate with the application program. The web-based distributed computing model requires that each data server have HTTP server software. With the development of compact and embedded HTTP server software it remains a feasible approach.

With the growth of distributed computing, a class of software solutions are evolving which enable tighter coupling of distributed applications and hide some of the inherent complexities of distributed software solutions. The general term for these software solutions is middleware. Fundamentally, middleware allows application programs to communicate with remote application programs as if the two programs were located on the same computer. Current middleware technologies include: CORBA, Microsoft's DCOM, SUN's Java-RMI, and Web-based RPC (Remote Procedure Call). A brief discussion of the middleware options will be given here, those with a more detailed interest are referred to [2]. CORBA (Component Object Request Broker Architecture) is an open middleware standard developed and maintained by the Object Management Group (OMG) [3]. A number of companies have CORBA based product offerings for a variety of hardware and OS platforms. DCOM (Distributed Component Object Model) is the extension of Microsoft's object technology to distributed software objects[4]; DCOM is built into the Windows 2000 operating system and Windows NT 4.0. A number of software companies have ported DCOM to other OS platforms, however DCOM is just

one component of the complete solution for distributed computing which is provided by Windows 2000. The SUN solution for distributed computing uses JAVA RMI (remote method invocation) for managing calls to distributed Java objects [5]. JAVA RMI operates over IIOP (the Internet Inter-Orb Protocol)[6], the CORBA protocol for communication between distributed software components. This allows JAVA RMI based solutions some level of interoperability with CORBA based solutions.

A middleware technology using XML based RPC over HTTP is emerging as a possible solution for distributed software systems on the World Wide Web. XML based RPC utilizes XML syntax for the execution request and for returning the results of the remote program. The leading standard for XML based RPC is SOAP [7] (the Simple Object Access Protocol) being developed by Microsoft and DevelopMentor. The XML based RPC approach provides more transparent access to distributed applications than does the use of CGI scripts. The "full service" middlewares provide better application integration at the cost of higher upfront development costs, and should provide better system performance. However, there is a simplicity associated with the XML based approaches that is attractive.

Looking ahead, it is likely that the market will support a Web-based middleware and one or more full service middlewares. It is very difficult however, to project which (if any) of the current technologies will dominate, or when some new technology will come along and make the current technologies obsolete. A prudent approach would be to develop a core architecture standard which is technology independent, and extend the core architecture with implementation specifications for several of the current implementation technologies.

OSA/CBM: An industry led team has been partially funded by the Navy through a DUST (Dual Use Science and Technology) program to develop and demonstrate an Open System Architecture for Condition Based Maintenance. The team participants cover a wide range of industrial, commercial, and military applications of CBM technology: Boeing, Caterpillar, Rockwell Automation, Rockwell Science Center, Newport News, and Oceana Sensor Technologies. Other team contributors include the Applied Research Laboratory at Penn State, and MIMOSA (Machinery Information Management Open System Alliance). The focus of the team is the development and demonstration of a software architecture that facilitates interoperability of CBM software modules. This section will give an overview of the architecture being developed.

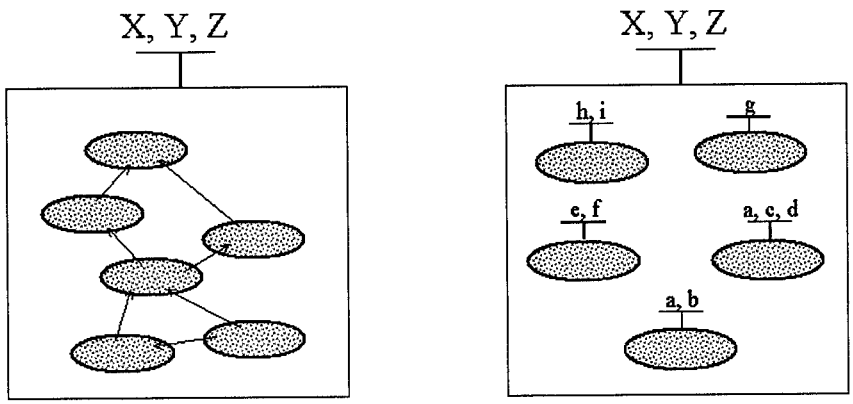
MIMOSA is a not-for-profit trade association founded in 1994 and incorporated in December of 1996. Their general purpose is the development and publication of open conventions for information exchange between plant and machinery maintenance information systems. The core of the MIMOSA development activity is the MIMOSA CRIS (Common Relational Information Schema). The second version of the CRIS (CRIS V2.1) was released in May 2000 and is publicly available at the MIMOSA website [8]. The CRIS defines a relational database schema for machinery maintenance information. The schema provides broad coverage of the types of data that need to be managed within the CBM domain:

- A description of the configuration of the system/equipment being monitored

- A list of specific assets being tracked, and their detailed characteristics
- A description of equipment functions, failure modes, and failure mode effects
- A record of logged operational events
- A description of the monitoring/measurement system (sensors, data acquisition, measurement locations, etc.) and the characteristics of monitoring components (calibration history, model number, serial number, etc.)
- A record of sensor data (and its characteristics) whether acquired on-line, manually logged, or manually acquired using hand held roving instrumentation.
- A means of describing signal processing algorithms and the resulting output data
- A record of alarm limits and triggered alarms
- A means of describing diagnoses of evolving equipment faults and projections of equipment health trends.
- A record of recommended actions and the basis of those recommendations
- A record of work requests from initiation through completion.

Based upon the CRIS, MIMOSA has also defined several system interface standards. A set of import/export file formats has been defined and released, followed by a set of SQL client/server interfaces. Currently under development is a complete set of XML interfaces for CRIS V2.1 defined by means of DTDs (XML Document Type Definition). These will be released after beta testing is completed.

The OSA/CBM development approach was formulated based on the assumption that the large body of work that constitutes the MIMOSA open standards would be used as a basis for development. The CRIS represents a static view of the data produced by a CBM system. The MIMOSA interface standards define open data exchange conventions for sharing of static information between CBM systems (openness at the intra-system level). The goal of the OSA/CBM project is the development of an architecture (and data exchange conventions) that enables interoperability of CBM components (openness at the inter-system level). Figure 1 illustrates the distinction being made. Figure 1a illustrates a CBM solution composed of components with proprietary interfaces, but open at the intra-system level. The proprietary system solution is enclosed in a MIMOSA compliant wrapper which exposes a set of public MIMOSA compliant server interfaces (X, Y, Z). The interface set (X, Y, Z) allows external clients open access to the information generated within the proprietary system solution. Figure 1b illustrates a CBM system open at the inter-system and intra-system levels. In this case the individual components all expose public server interfaces (a, b, c, ... , i). These component interfaces offer access to the data and services supplied by the component, and provide for open information flow between components during system operation. In addition, components may be readily replaced by components with improved capability as long as they follow the same public interface standards. If the interface standard is an open standard that has market acceptance, there should be available COTS components which may be readily integrated for the purpose of expanding or upgrading the system capability. If the open component interfaces are based on the same information model as the open system level interfaces, the mapping between the two sets of interfaces will be significantly simplified.



a) MIMOSA compliant CBM system with proprietary component interfaces

b) MIMOSA compliant CBM system with open component interfaces

Figure 1: Granularity of an Open System Solution

After evaluating a variety of architectural options, a decision was made to develop an object oriented architecture based upon a client-server approach to distributed computing. The initial architecture development direction was focused towards the loosely coupled Web-based approach and XML messages over HTTP. There were significant concerns that this approach would limit the usefulness of the resulting standard based on performance limitations. However, tying the architecture to a specific full-service middleware brought up concerns about being tied to proprietary middleware technology, or selecting a technology which might achieve limited market acceptance. The approach that was ultimately adopted was to develop a technology-neutral abstract design specification, which could then be mapped to implementations utilizing any of the implementation technologies discussed earlier. The decision was also made to utilize an object oriented design methodology as a core strategy. The components of the architecture will be discussed in more detail later.

In order to standardize an architecture for CBM components, the first step as suggested by Figure 1b is to assign the CBM system functions defined earlier to a set of standard software components. The software architecture has been described in terms of functional layers. Starting with sensing and data acquisition and progressing towards decision support, the general functions of the layers are given below. A complete description of the inputs and outputs required for a given layer is beyond the scope of this paper.

Layer 1 – Sensor Module: The sensor module has been generalized to represent the software module which provides system access to digitized sensor or transducer data. The sensor module may represent a specialized data acquisition module that has analog feeds from legacy sensors, or it may collect and consolidate sensor signals from a data bus. Alternately, it might represent the software interface to a smart sensor (e.g. IEEE

1451 compliant sensor). The sensor module is a server of calibrated digitized sensor data records.

Layer 2 – Signal Processing: The signal processing module acquires input data from sensor modules or from other signal processing modules and performs single and multi-channel signal transformations and CBM feature extraction. The outputs of the signal processing layer include: digitally filtered sensor data, frequency spectra, virtual sensor signals, and CBM features.

Layer 3 – Condition Monitor: The condition monitor acquires input data from sensor modules, signal processing modules, and from other condition monitors. The primary function of the condition monitor is to compare CBM features against expected values or operational limits and output enumerated condition indicators (e.g. level low, level normal, level high, etc). The condition monitor also generates alerts based on defined operational limits. When appropriate data is available, the condition monitor may generate assessments of operational context (current operational state or operational environment). Context assessments are treated, and output, as condition indicators. The condition monitor may schedule the reporting of the sensor, signal processing, or other condition monitors based on condition or context indicators, in this role it acts as a test coordinator. The condition monitor also archives data from the Signal Processing and Sensor Modules which may be required for downstream processing.

Layer 4 – Health Assessment: The health assessment layer acquires input data from condition monitors or from other health assessment modules. The primary function of the health assessment layer is to determine if the health of a monitored system, subsystem, or piece of equipment is degraded. If the health is degraded, the health assessment layer may generate a diagnostic record which proposes one or more possible fault conditions with an associated confidence. The health assessment module should take into account trends in the health history, operational status and loading, and the maintenance history. The health assessment module should maintain its own archive of required historical data.

Layer 5 – Prognostics: Depending on the modeling approach that is used for prognostics, the prognostic layer may need to acquire data from any of the lower layers within the architecture. The primary function of the prognostic layer is to project the current health state of equipment into the future taking into account estimates of future usage profiles. The prognostics layer may report health status at a future time, or may estimate the remaining useful life (RUL) of an asset given its projected usage profile. Assessments of future health or RUL may have an associated diagnosis of the projected fault condition. The prognostic module should maintain its own archive of required historical data.

Layer 6 – Decision Support: The decision support module acquires data primarily from the health assessment and prognostics layers. The primary function of the decision support module is to provide recommended actions and alternatives and the implications of each recommended action. Recommendations include maintenance action schedules, modifying the operational configuration of equipment in order to accomplish mission objectives, or modifying mission profiles to allow mission completion. The decision support module needs to take into account operational history (including usage and

maintenance), current and future mission profiles, high level unit objectives, and resource constraints.

Layer 7 – Presentation: The presentation layer may access data from any of the other layers within the architecture. Typically high level status (health assessments, prognostic assessments, or decision support recommendations) and alerts would be displayed, with the ability to drill down when anomalies are reported. In many cases the presentation layer will have multiple layers of access depending on the information needs of the user. It may also be implemented as an integrated user interface which takes into account information needs of the users other than CBM.

After defining the layers in the architecture, a decision was made to focus the standard on the specification of interfaces for layers 1 through 5. Although a general set of interfaces may be conceptually described for the decision support layer, the structure of the information that it serves is tied to the specific requirements of the targeted application to a degree that standardization of the interface is not feasible. Since the presentation layer acts only as a client in the c-s communication model there are no server interfaces that are required to be defined.

The components of the OSA/CBM architecture are shown in Figure 2. The primary inputs to the architecture definition are the functional description of the layers (as discussed above) and the MIMOSA CRIS, along with the general requirements described in the section on CBM Architecture. An object oriented data model has been defined (using Unified Modeling Language – UML - syntax) based upon a mapping of the MIMOSA relational schema to the OSA/CBM layers. For a given layer of the

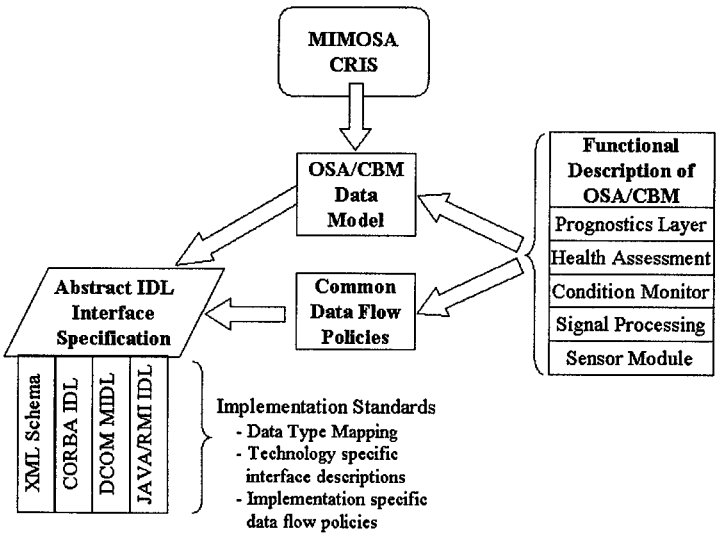


Figure 2: Outline of the OSA/CBM Architecture

architecture, the data model does not describe all of the object classes that would be required for a software implementation. The focus is on describing the structure of the information that might be of interest to clients of that layer. In fact, in the same way that the MIMOSA interface standard does not impose a structure on the components that comprise a MIMOSA compliant system, OSA/CBM does not impose any requirements on the internal structure of compliant software modules. The architectural constraints are applied to the structure of the public interface and to the behavior of the modules. This approach allows complete encapsulation of proprietary algorithms and software design approaches within the software module.

The common data flow policies were defined to standardize the approach to time-based and event-based messaging. In general, the information consumer (at the higher layer of the architecture) will initiate requests for information (a pull rather than a push messaging model). Requests for static information, information which is expected to be available upon request, (e.g. configuration information) will use synchronous messaging protocols. Requests for dynamic information (new sensor data, or updated processing results) will use asynchronous messaging protocols. The asynchronous messaging model requires that the supplier implement a "request()" interface and the consumer a corresponding "notify()" interface.

This messaging model defined above supports time-based or event-based pull of information but does not support event-based push from the information supplier. Event-based push may be implemented by the information supplier acting as a client and utilizing the information consumer's "notify()" interface to initiate communication. Implementation of this messaging model also requires that the supplier can associate a set of information consumers to a particular triggering event.

At the time of implementation, further design decisions may be necessary with respect to data passing. An information supplier may pass actual data elements, or alternately may pass remote references which the information consumer may use to "callback" to get the actual data. There are pros and cons to each approach, but a complete discussion is beyond the current scope. The further detail of the messaging approach will be called out as part of the implementation specific data flow policies.

For CORBA, DCOM and JAVA-RMI implementations, the software interfaces are formally described using their own specific Interface Definition Languages (IDL). In order to commonize the various implementations to the greatest extent possible, the architecture utilizes a common abstract IDL specification (the abstract IDL will be specified using the CORBA IDL syntax). The abstract IDL describes a set of common data structures and interface configurations. The OSA/CBM standard will include mappings to several of the implementation technologies listed in figure 2 (XML and DCOM implementation specifications are currently being developed). Although the abstract specifications provide a strong base of software commonality, software interoperability would not be possible without the detailed implementation specifications. Software applications implemented using different technologies will not be interoperable. In order to have interoperability between applications built in CORBA and DCOM, for instance, a software interface (typically called a bridge) would be required to implement a a mapping. The OSA/CBM architecture still provides strong benefits in this case. The efforts involved in designing and building the bridge will be significantly reduced due to

the common core design requirements and the existence of the associated implementation standards. This view of the architecture also illustrates its extensibility to new or evolving middleware technologies; the core architecture may be mapped to an implementation in the new technology. Since the core software architecture has not changed, existing applications should be able to be readily ported to the new middleware.

Current Status: The architecture that has been discussed is being applied in demonstrations across a variety of different CBM applications as a part of the DUST program. The demonstrations will cover several of the technologies which were discussed, and associated implementation standards will be developed. Lessons learned during the implementation process will be used to update and improve the core architecture. The architecture is also being evaluated for transition by both ARMY and NAVY programs. The ARMY is evaluating the use of the MIMOSA and OSA/CBM standards as a part of their maintenance architecture. Other NAVY research programs in the area of CBM are being directed to consider the OSA/CBM architecture in their system and component designs [9].

Acknowledgment: This work was supported by the Office of Naval Research through the *OSA-CBM Boeing DUST* (Grant Number N00014-00-1-0155). The content of the information does not necessarily reflect the position or policy of the Government, and no official endorsement should be inferred.

References:

[1] Thomesse, J.P., and Leon Chavez, M. "Main Paradigms as a Basis for Current Fieldbus Concepts," *Fieldbus Technology, Systems Integration, Networking and Engineering*. Dietrich, D., Neumann, P., and Schweinzer, H., eds. Vienna: Springer-Verlag, 1999.

[2] Serain, D. *Middleware*. London: Springer-Verlag, 1999.

[3] Object Management Group (OMG) Website, http://www.omg.org/

[4] Microsoft. "COM: Delivering on the Promises of Component Technology." http://www.microsoft.com/com/default.asp

[5] Sun Microsystems. "Introduction to the J2EE Architecture." http://developer.java.sun.com/developer/technicalArticles/J2EE/Intro/index.html.

[6] Sun Microsystems. "Remote Method Invocation over IIOP." http://java.sun.com/products/rmi-iiop.

[7] W3C Note 08 May 2000, "Simple Object Access Protocol (SOAP) 1.1." http://www.w3.org/TR/2000/NOTE-SOAP-20000508

[8] MIMOSA Website, http://www.mimosa.org/

[9] Roemer, M. J., et. al, "Prognostic Enhancements to Naval Condition-Based Maintenance Systems," *Improving Productivity Through Applications of Condition Monitoring*, 55th Meeting of the Society for Machinery Failure Prevention Technology, April, 2001.

Development of Performance and Effectiveness Metrics For Mechanical Diagnostic Technologies

Rolf F. Orsagh
Michael J. Roemer
Impact Technologies
125 Tech Park Drive
Rochester NY 14623

Christopher J. Savage
NAVSEA
5001 South Broad Street
Philadelphia, PA 19112

Kathy McClintic
Penn State ARL
P.O. Box 30
State College, PA 16804

Abstract: In recent years, numerous anomaly detection and diagnostic technologies have been developed for various military and industrial applications to aid in the detection and classification of developing faults. In many cases, significant reductions in machinery total ownership costs have been achieved through the judicious application of these technologies. However, there is currently no consistent methodology available for assessing both the technical and economic benefits of these machinery diagnostic technologies. In response to this need, a virtual test bench is under development by the Navy for assessing the performance and effectiveness of machinery diagnostic systems. The test bench utilizes a 'plug 'n play' interface that can readily accept standardized diagnostic/prognostic tools and link them to real and model-based transitional data from appropriate condition based maintenance (CBM) platforms. The assessment process relies on a standard set of mathematical ground rules and a statistical framework to directly identify confidence bounds, robustness measures, and various diagnostic thresholds associated with specific mechanical diagnostic technologies. Specific performance and accuracy of the diagnostic algorithms at the component or subsystem level are evaluated with performance metrics, while system level capabilities in terms of achieving the overall operational goals of the diagnostic system will be evaluated with effectiveness measures. This qualification and validation methodology is utilized to compare a variety of diagnostic tools that are commonly used to analyze gearbox vibration.

Key Words: Diagnostics; Prognostics; Metrics; Diagnostic Qualification; Diagnostic Validation

Introduction: The US Navy's operational goals include improving mission readiness, and crew safety while reducing the support requirements and costs associated naval platforms. To accomplish these objectives the Navy is adopting condition based maintenance (CBM) practices. CBM is based on the principle of monitoring the condition of machinery and repairing it just prior to failure or an unacceptable level of performance degradation. Mission readiness can be enhanced by CBM through the elimination of unnecessary preventive maintenance and by identifying impending failures so that corrective action can taken in an efficient manner. CBM procedures can also protect crewmembers by identifying impending machinery malfunctions with sufficient warning to avert a catastrophic failure. By avoiding unnecessary preventive maintenance and allowing a scheduled response to impending failures, CBM can reduce the support requirements and total ownership cost associated with many types of machinery.

The success of a CBM program in a given application depends to a great extent upon the availability of useful diagnostic and prognostic information. CBM practices are most beneficial when maintenance actions can be planned well in advance, and corrective measures are carried out just prior to failure. Such precise maintenance scheduling can only occur through the use of timely and accurate diagnostic, or better yet, prognostic information. However, a consistent methodology for evaluating the technical and economic benefits of mechanical machinery diagnostic technologies does not currently exist. In response to this need, a virtual test bench is

under development by the Navy for assessing the performance and effectiveness of machinery diagnostic systems.

Performance Metrics Development

The performance of specific detection and diagnostic algorithms or subsystems of a CBM system are measured with Performance Metrics[i]. The functionality of these diagnostic algorithms or subsystems directly contributes to the overall effectiveness of the entire system. However, the ability to assess the accuracy and robustness of particular algorithms is often more straightforward when the technologies making up the system are checked separately. Also, from a design and development point of view, it is often more logical to work on the improvements to specific algorithms or processes at the elemental level rather than the overall systems level. Metrics of performance for diagnostic/prognostic algorithms or subsystems are arranged into three categories (detection, isolation, and prognosis) as shown in Figure 1. Detection metrics measure the ability of diagnostic tools to correctly classify machinery operation as either normal or anomalous. Isolation metrics measure the ability of diagnostic tools to accurately identify the root cause and corrective action for a fault. Prognosis metrics measure the ability of prognostic systems to accurately forecast the future condition of a mechanical system. Scores from the individual performance metrics are combined according the hierarchy to produce summary scores for each category, and for overall performance.

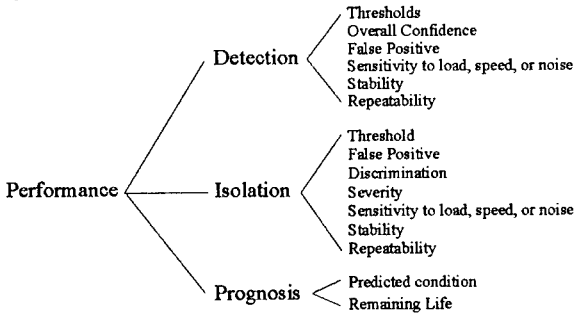


Figure 1 Performance Metrics

The ability of diagnostic/prognostic systems to detect and isolate faults or to predict failures is measured as a function of the fault severity. Figure 2 shows the confidence level reported by a hypothetical diagnostic tool and the corresponding fault severity level as functions of time. This could be the confidence that an anomaly exists or the confidence in a particular diagnosis. Varying operating conditions or noise could cause fluctuations in the diagnostic confidence level. The success function of the diagnostic tool is defined as the relationship between the average confidence and the average severity level. Note that this relationship may be used to assess either Boolean (0 or 1) confidence levels or continuous confidence levels within the same interval. The success function for the hypothetical diagnostic tool is plotted in Figure 3.

Fault severity must be established by objective and irrefutable measures to ensure that the assessments based upon it are accurate and impartial. This measure of severity will hereafter be referred to as the *ground truth severity level*. The ground truth severity of a system's condition may be assessed in a laboratory setting through the use of appropriate instruments or frequent inspections by nondestructive evaluation (NDE) techniques. Measurements of the fault severity are mapped onto the ground truth severity scale where zero represents a healthy operating condition, one represents an unacceptable level of performance degradation. Once the ground

truth is established, the anomaly detection threshold, isolation threshold, fault severity, stability, repeatability, and duty sensitivity metrics may be determined.

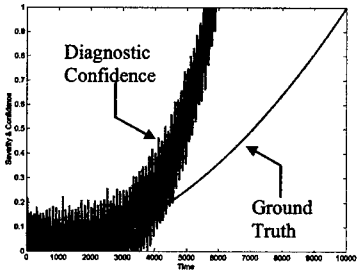


Figure 2 Diagnostic and Ground Truth Information

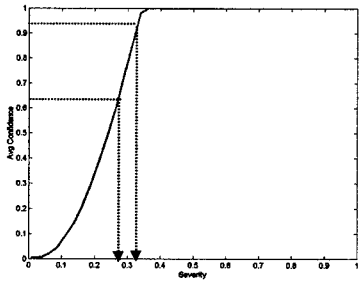


Figure 3 Success Function

Detection metrics

The ability of a diagnostic algorithm or overall system to detect anomalous machinery operating behavior is the most fundamental requirement for machinery health monitoring tool. For a diagnostic system to be useful it must detect anomalies associated with incipient faults so that corrective action may be taken in an efficient and timely manner. The *Detection Threshold Metric* measures a diagnostic algorithm or system's ability to identify anomalous operation associated with incipient faults with a specified confidence level. This metric is defined as the minimum ground truth severity corresponding to a designated confidence level on the detection success function as shown in Figure 3. Confidence levels of 67% and 95% corresponding to one and two standard deviations are used to calculate the detection threshold metric. Eq. (1) is used to calculate the detection threshold metric score.

$$\text{Detection Threshold} = 1 - S(c) \qquad (1)$$

where: S(c) = ground truth severity at a confidence of c

An assessment of the detection confidence level over the entire severity range for 0 to 1 is achieved with the *Overall Detection Confidence* metric defined in Eq (2). Graphically, the overall confidence score represents the area under the success function. An algorithm that detects an incipient fault with high confidence will receive a high Overall Confidence score, while an algorithm that does not report a fault until it becomes very severe would receive a low score.

$$OverallConfidence = \int_0^1 C(s)ds \qquad (2)$$

where: C(s) = The success function
s = severity

A confidence level that fluctuates wildly is difficult to interpret and therefore undesirable. For example, a diagnostic tool that produces a Boolean result of either no fault or fault may flicker as the fault severity approaches the detection level. The *Stability Metric* measures the range of confidence values that occur over the fault transition by integrating the peak to peak difference at each point on the success function as stated in Eq.(3).

$$\text{Stability} = 1 - \int_0^1 (C_H(s) - C_L(s))\ ds \qquad (3)$$

where: $C_H(s)$ = maximum value of the success function at severity s
$C_L(s)$ = minimum value of the success function at severity s
s = severity

Ideally, diagnostic systems should detect anomalies over the full range of operating (duty) conditions such as loads, speeds, etc. The *Detection Duty Sensitivity Metric* measures the difference between the success functions of a diagnostic tool under two duty conditions as stated in Eq.(4).

$$DutySensitivity = 1 - \sqrt{\int_0^1 (C_1(s) - C_2(s))^2\ ds} \qquad (4)$$

where: $C_1(s)$ = success function at duty condition 1
$C_2(s)$ = success function at duty condition 2
s = severity

A diagnostic tool that incorrectly reports anomalies is unacceptable because it reduces availability and increases maintenance costs for the equipment. The *False Positive Confidence Metric* measures the frequency and upper confidence limit associated with false anomaly detection by a diagnostic tool. Calculation of the false confidence metric is based on the false positive function that is stated in Eq.(5) and an example is shown in Figure 4.

$$F(c) = n(c) / N \qquad (5)$$

where: n(c) = number of false positive detection events with confidence > c
N = number of opportunities to detect a normal operating condition

Integration of the false positive function with respect to the confidence yields two parameters for assessing false anomaly detection by a diagnostic tool. The first parameter, α, represents the frequency of false positive anomaly detections and can be visualized as the area under the false positive function. The second parameter, β, is the confidence corresponding to 95% of α as shown in Figure 5. The mean value of the two parameters, α and β, helps determine the false confidence metric as shown in Eq.(6).

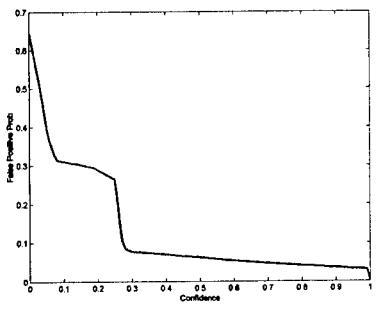


Figure 4 False positive function

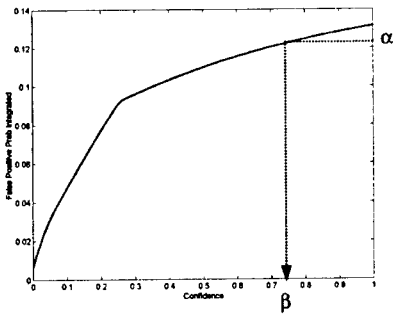


Figure 5 Integrated false positive function

$$FalseConfidence = 1 - \frac{\alpha + \beta}{2} \qquad (6)$$

In an operational environment sensor data is sometimes contaminated with noise that may interfere with the operation of diagnostic algorithms. The robustness of an algorithm to noisy data is measured by the *Noise Sensitivity Metric.* Two aspects of the diagnostic system's response, change in the success function and increase in the false positive score are combined to form the noise sensitivity metric. The difference between the success functions of a diagnostic tool when the sensor data is contaminated with two different levels of noise. Eq. (7).

$$NoiseSensitivity = \left(1 - \sqrt{\int_0^1 (C_1(s) - C_2(s))^2 \, ds}\right) * (\Delta FalsePositive) \qquad (7)$$

where: $C_1(s)$ = success function under noise condition 1
$C_2(s)$ = success function under noise condition 2
s = severity

Calibration of the performance metrics determine the weight that each individual metric carries in the category and overall composite scores. These weighting factors should reflect the specific requirements of the intended application, and therefore must be determined on a case by case basis. For example, when evaluating a gearbox diagnostic tool, knowledge of the gearbox's criticality (such as the main drive on helicopter vs. a redundant shipboard system) would determine the relative weight assigned to the detection threshold metric and the false confidence metric. The process of selecting weighting factors may be simplified by allowing the user to select a standard weighting scheme from a previously defined set or create a custom weighting scheme from scratch. A weighted average is used to calibrate and combine the individual performance metrics at the category level, and the category scores into an overall performance score as shown in Eq.(8).

$$CompositeScore = \frac{w_1 M_1 + w_2 M_2 + w_3 M_3 + \ldots + w_n M_n}{\sum w_i} \qquad (8)$$

where: Mi = metric scores
w_i = weight assigned to metric i

Effectiveness Metrics

The overall effectiveness of a diagnostic system in terms of achieving the desired CBM goal is measured with Effectiveness Metrics[i]. This could include the integration of all the monitoring and diagnostic systems on the entire platform or a single diagnostic system made up of several different diagnostic algorithms. In either case, the effectiveness metrics utilize many of the same metrics as defined for the performance metrics. However, the resulting scores of the metric may be calibrated and combined differently based on the scope of their application. Some metrics such as cost, speed, complexity, robustness, and resource requirements are unique to the overall effectiveness of the diagnostic system and are therefore only defined as effectiveness metrics.

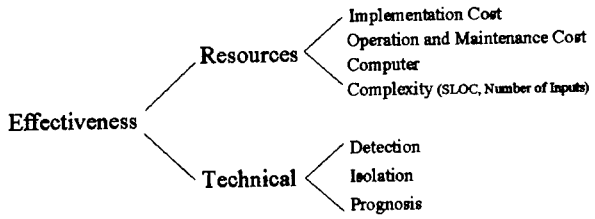


Figure 6 Effectiveness Metrics

Acquisition and implementation costs of the diagnostic system may have a significant effect on the system's cost effectiveness. The *Implementation Cost Metric* simply measures the cost of acquiring and implementing a diagnostic system on a single application. If the diagnostic system is applied to several pieces of equipment, any shared costs are divided among them. Operation and maintenance costs may also play a significant role in determining whether a diagnostic system is cost effective. The *O&M Cost Metric* measures the annual cost incurred to keep the diagnostic system running. These costs may include manual data collection, inspections, laboratory testing, data archival, relicensing fees and repairs.

The ability of the diagnostic algorithms or system to be run within specified time requirements and on traditional computer platforms with common operating systems is important when considering implementation on multiple machinery platforms. Therefore, a metric that takes into account computational effort as well as static and dynamic memory allocation requirements is necessary. The *Computer Resource Metric* computes a score based on the normalized addition of CPU time to run (in terms of floating point operations), static and dynamic memory requirements for RAM and static source code space, and static and dynamic hard disk storage requirements. Computer requirements may be a significant issue in some applications such as aircraft.

Complex systems are generally more susceptible to unexpected behavior due to unforeseen events. The *System Complexity Metric* measures the complexity of diagnostic systems in terms of the number of source lines of code (SLOCs) and the number of inputs required.

The individual effectiveness metric scores are combined to form an overall effectiveness score by means of a cost function. The benefits achieve through anomaly detection, fault isolation, and failure prediction are weighed against the costs associated with false alarms, inaccurate diagnoses, licensing, and resource requirements of implementing and operating a diagnostic tool. The simplified cost function in Eq. (9) states the *Technical Value* provided by a diagnostic system for a given fault. The value of a diagnostic tool in a particular application is the summation of the benefits it provides over all the failure modes that it can diagnose less the implementation cost, operation and maintenance cost, and consequential cost of incorrect assessments as stated in Eq.(10).

$$Value = P_f * (D * \alpha + I * \beta) - (1 - P_f) * (P_D * \phi - P_I * \theta) \quad (9)$$

where:

P_f = Probability (time-based) of occurrence for a failure mode
D = Overall Detection Confidence metric score
α = Savings realized by detecting a fault prior to failure
I = Overall isolation confidence metric score
β = Savings realized through automated isolation of a fault

P_D = False positive detection metric score
ϕ = Cost associated with a false positive detection
P_I = False positive isolation metric score
θ = Cost associated with a false positive isolation

$$TotalValue = \sum_{FailureModes} TechnicalValue_i - A - O - (1 - P_c) * \delta \quad (10)$$

where:
A = Acquisition and Implementation Cost
O = Life Cycle Operation and Maintenance Cost
P_c = Computer Resource Requirement score
δ = Cost of a standard computer system

CBM Metrics Database

One of the most significant aspects associated with the development and implementation of diagnostic system metrics is having well-documented fault data sets. Initial fault/failure data sets were obtained primarily from previously acquired test bed (including accelerated loading and run to failure tests) and simulation data sets with actual in-service data being applied later in the program. The Penn State ARL Mechanical Diagnostics Test Bed (MDTB) was utilized in this program as the basis for the diagnostic system metrics evaluation, testing and verification. The MDTB represents a wealth of well-documented data sets and information on gear, shaft and bearing faults and failures critical to Naval aircraft carrier day-to-day operations. The database of fault scenarios already developed under existing Multi-disciplinary University Research Initiative (MURI) provided an excellent basis and source of data from which the fault data sets utilized in this program were built upon. Identified metrics that require additional or more specific seeded fault or failure test data sets can be acquired from this test bed configuration or Penn State ARL's other test beds (Bearing Test Rig, Diesel Enhanced MDTB) throughout and after the duration of this program.

The metrics evaluation process is currently being implemented within the framework of a Test Bench that will utilize this database of sensor data from carefully constructed tests of selected CBM platforms as a basis for evaluating diagnostic/prognostic systems. Each test documents the transition of a mechanical system from a normal operating condition to failure or significantly degraded performance. Use of transitional data is necessary for the assessment of diagnostic/prognostic tools that rely on trending, and for evaluating the response of diagnostic/prognostic algorithms as a function of fault severity. Potential future sources for data of this type include the manufacturer of the equipment, Naval laboratories, and independent testing facilities. Contributions to the database should be screened to ensure data integrity and that the data remains unbiased toward any particular diagnostic/prognostic approach. The review process should include Naval engineers who will use the Test Bench to evaluate diagnostic/prognostic tools, and Naval maintenance officers who possess an intimate knowledge of the machinery reliability issues in the fleet.

Specifics of the MTBD Test Bed at ARL

The MDTB at Penn State was built as an experimental research station for the study of fault evolution in mechanical gearbox and power transmission components. It consists of a motor, gearbox, shafts, bearings, and a generator on a rigid steel platform. Gearboxes, shafts and bearings are instrumented with 52 sensors including accelerometers, thermocouples, acoustic emission sensors, and oil debris sensors. Tests are run at various load and speed profiles while

logging measurement signals for later analysis. Duty cycle profiles can be prescribed for any speed and load.

CBM Metrics Test Bench Web Application

Implementation of a standardized process and associated metrics for efficiently evaluating CBM information systems could potentially enhance the quality of diagnostic/prognostic technologies in two ways. First, doing so will allow the Navy and other users of diagnostic/prognostic tools to select the most appropriate algorithms for their application and verify the advertised capabilities of candidate systems. Second, developers of diagnostic systems may use the metric-based evaluation process to assess and improve their algorithms. To encourage participation, developers will have the option to evaluate their algorithms without creating any permanent record of the results.

In order to provide easy access to the CBM metrics developed under this program, a WEB-based prototype application called the **CBM Metrics Test Bench** has been developed to evaluate diagnostic technologies. Users of the site will upload algorithms to the server for evaluation and an e-mail will be issued to them indicating that their results are complete. The site will also provide access to a limited set of the maintained databases. However, a comprehensive set of data will only be accessible to Naval and other relevant DOD personnel for official use in qualification and validation of diagnostic tools.

On the Log-in page shown in Figure 7, the user can access the "Motivation and Evaluation Criterion", the "New User Registration", and the "User Log-In" links. Users who are not registered to use the web-site, may do so by clicking the "New User Registration" link. After successfully logging in, users may choose links that will allow them to obtain data, submit an algorithm, or view the results of an evaluation. Some of the transitional machinery failure data used in the evaluation will be available to facilitate the development of algorithms. Users will be able to download sample data sets from the web-site, or request a full data set to be mailed to them.

Figure 7 Log-in page

Figure 8 Database and sensor selection

To submit an algorithm, users will begin by uploading it to the server by either typing in the path and file name of the file containing their algorithm, or select it using "Browse". An algorithm description field is provided to allow users to identify their algorithms. After entering a file name, the algorithm will be assigned a Job ID that will be used to identify the algorithm within the Test Bench. Users may also choose the platform, faults, and sensor data on which their algorithm is evaluated. As the database grows, the user will be able to select a variety of failure modes for each platform. Information about the conditions under which each data set was collected is available through the links under the heading "Database Development and Specifications".

The weighting factors that are used to combine and calibrate the metric scores are accessible to the user on the metric weighting page shown in Figure 9. Users may view the definition of each metric by clicking on its name. When the user is satisfied with their choices, they may choose to perform the evaluation on either an official or a confidential basis. Algorithms that are evaluated on an official basis will have their scores added (anonymously) to a publicly accessible database.

Evaluation results are accessible on two levels. The lower level shows the scores earned by an algorithm while evaluating one particular fault on a platform. Users may view the definition of each metric by clicking on its name. The higher level results page presents the combined results for the algorithm against all of the selected faults. In the case of performance metrics, the scores are averaged, and for the effectiveness metrics reflect the sum of the technical values achieved by the algorithm for each fault type.

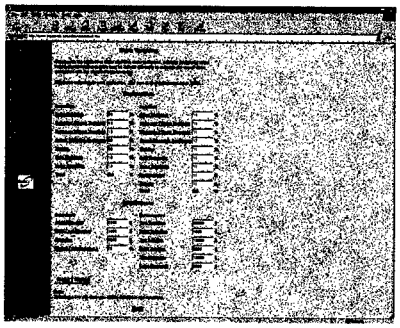


Figure 9 Metric weighting page

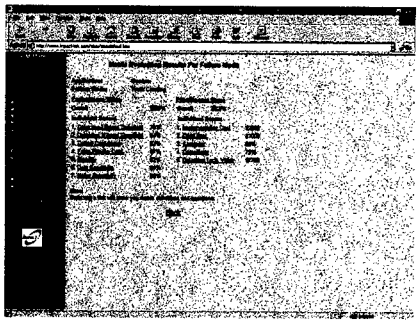


Figure 10 Evaluation results page

Results

The CBM Metrics Test Bench was used to evaluate the performance of ten anomaly detection algorithms for a gearbox. Gearbox failure data collected on the MDTB was used to evaluate the ability of the selected algorithms to detect gear tooth failures. During the test, cyclic loads as high as three times the rated load for the gearbox accelerated gear tooth failure rates. All of the algorithms utilize the same time domain vibration data, but process it in different ways.

Table 1 shows selected scores for each of the algorithms. For all of the metrics, a low score indicates an undesirable result, and high score indicates a desirable result. For example, a high Computer resource requirement score is awarded to algorithms that use a small portion of the computer's resources. Calculation of Detection Technical Value, Overall Performance, and Overall Effectiveness are based on weighting factors described in Eqs (8), (9), and(10). The factors used to calculate these results are stated in Tables 2 and 3. Evaluations of three diagnostic algorithms (RMS, Wavelet, and FM4) are described in detail.

Table 1 Metric Scores

Metric	RMS	Kurt	Wavelet	FM0	NA4	M6A	Dempster Shafer	FM4
Detection 1σ Threshold	22	19	27	68	64	77	73	76
Detection 2σ Threshold	0	0	12	52	19	64	64	64
Overall Confidence	51	39	44	75	64	84	79	82
False Positive Conf.	44	57	99	53	99	60	92	87
Stability	36	45	55	52	48	75	81	82
Duty Sensitivity	47	60	74	73	59	75	78	84
Noise Sensitivity	95	99	100	98	97	96	99	98
Implementation Cost $	1500	1500	1500	2000	2000	2000	2500	2000
O&M Cost $	700	700	700	1000	1000	1000	1000	1000
Computer	99	99	88	65	65	65	47	65
Complexity	99	99	97	79	79	79	78	79
Detection Tech. Value $	3255	2551	4387	6010	6374	7076	7640	7798
Overall Performance	42	46	61	65	66	75	82	82
Overall Effectiveness $	1052	333	1820	1021	1325	1879	1433	2448

Table 2 Performance Weighting Factors

Metric	Weight
Detection 1σ Threshold	10 %
Detection 2σ Threshold	10 %
Overall Confidence	20 %
False Positive Conf.	20 %
Stability	20 %
Duty Sensitivity	10 %
Noise Sensitivity	10 %

Table 3 Effectiveness Weighting Factors

Factor	Weight
Probability of Fault	20%
Cost of False Alarm	$4000
Benefit of Detection	$50000
Cost of Std. Computer	$2000

RMS is a simple and commonly used technique for detecting anomalous machinery operation. The RMS based algorithm calculates the root mean square value of the time domain vibration signal. The RMS level of a signal x consisting of N samples is calculated using Eq.(11). Figure 11 shows the diagnostic confidence reported by the RMS algorithm as compared to the ground truth severity level. The low performance scores assigned to RMS reflects the fact that RMS does not respond well in the early stages of gear damage and that the RMS level increases significantly with load. However, the low costs and low complexity (high complexity score) of the RMS algorithm make its overall effectiveness comparable to more sophisticated algorithms.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i} \qquad (11)$$

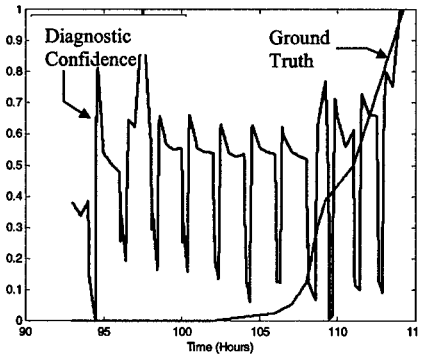


Figure 11 Diagnostic confidence reported by the RMS algorithm

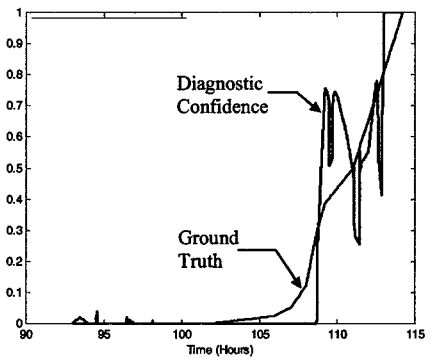


Figure 12 Diagnostic confidence reported by the Wavelet algorithm

The Wavelet algorithm uses a wavelet transform to analyze the nonstationary characteristics of vibration signal. The continuous wavelet transform of a time function *f(t)* is defined in Eq. (12) where *g(t)* is a given "mother wavelet" wavelet. The Morlet wavelet was chosen for *g(t)* and is defined mathematically by Eq. (13).

$$G(a,s) = |a|^{-1} \int_{-\infty}^{\infty} f(t) g[(t-s)/a] dt \qquad (12)$$

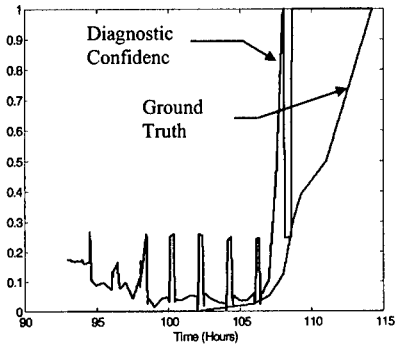
$$g(t) = \exp(-j\omega_0 t) \exp(-t^2/2) \qquad (13)$$

Where t is time and ω_0 is the fundamental (radian) frequency of the wavelet. Eq. (13) shows that the (complex) Morlet wavelet can be interpreted as a "modulated Gaussian." The actual Morlet wavelet chosen for the analysis is given by $\omega_0 = 5$ in Eq. (13) above. An adaptive IIR thresholding/tracking filter for processing wavelet output (at 550 Hz.) was also introduced. This kind of filter design is particularly robust against false alarms. The features resulting from the CWT processing include the number of detection counts (threshold crossings), and the peak amplitude and frequency obtained by a peak search of the CWT power spectral density near the frequency of interest (usually one of the shaft frequencies).

Figure 12 shows the diagnostic confidence reported by the Wavelet algorithm as compared to the ground truth severity level. Inspection of the Wavelet's diagnostic confidence will confirm that it warrants the high False Positive Detection score that it received. Furthermore, Wavelet shows very little load dependence as indicated by the Duty Sensitivity metric score.

The FM4 based algorithm uses the difference signal to detect changes in the vibration pattern resulting from damage on a limited number of teeth[ii]. FM4 is calculated for a difference signal d consisting of N samples according to Eq. (14). shows the diagnostic confidence reported by the FM4 algorithm as compared to the ground truth severity level. After calculating FM4, an empirical load correction was applied to reduce the load-induced fluctuations in the output. As a result of the load correction, the Duty Sensitivity metric score is higher (indicating that the confidence reported by the corrected algorithm is less dependent on the applied load. The same load correction technique was also applied to the M6A and Dempster Shafer (fusion) algorithms,

but not to the others. As expected, these load-corrected algorithms receive the highest duty sensitivity scores.



$$FM4 = \frac{N\sum_{i=1}^{N}(d_i - \bar{d})^4}{\left[\sum_{i=1}^{N}(d_i - \bar{d})^2\right]^2} \qquad (14)$$

Figure 13 Diagnostic confidence reported by the FM4 algorithm

Conclusion:

The metric-based process developed during this program clearly demonstrates the feasibility and potential benefits of a comprehensive system for evaluating the performance and effectiveness of diagnostic/prognostic tools. The principal achievements include the development and verification of diagnostic system metrics for evaluating and comparing the benefits advertised by system developers, and the eventual demonstration of these metrics in the assessment of various diagnostic tools. These achievements have been demonstrated through a comprehensive and easy-to-use internet-based software tool. The next necessary steps must include demonstration of the metrics software capabilities for various machinery diagnostic applications.

References:

[i] Essawu, M.A., Zein-sabatto, S, "Measures of Effectiveness and Measures of Performance for Machine Monitoring and Diagnostic Systems", Maintenance and Reliability Conference, May 1999.

[ii] Lebold, M., McClintic, K., Campbell, R., Byington, C., Maynard, K., "Review of Vibration Analysis Methods for Gearbox Diagnostics and Prognostics", 54th Meeting of the MFPT, May 2000

SMART SENSORS

Chair: **Dr. Kang B. Lee**
National Institute of Standards & Technology

THE CLIENTS' VIEW OF CBM IN 2001

Lewis Watt
President, RLW, Inc.
1346 South Atherton Street
State College, PA 16801
lwatt@rlwinc.com

Abstract: Advances in condition-based maintenance (CBM) are being driven by an array of technologies, including: speed and miniaturization of signal processing hardware; improvements in power supplies and sources; and smaller, lower-cost RF transmitters. As the set of industries and organizations that are developing and integrating these technologies into CBM components and systems move forward, they should insure that a balance is achieved between technology-push and customer-pull. The existing and potential customers span a wide range of technical sophistication. Many have a thorough understanding of the science and engineering behind the CBM systems that are evolving. Others, just as important as customers, look to us, to provide them with the appropriate tools to achieve the savings and productivity they have been led to expect.
Failure to provide not only excellent engineering, but also good fit, will result in black eyes for all of us. Less noble, but more obvious, we will fail as business people if we don't listen to our customers.
The author has had the good fortune to deal with CBM clients ten years ago and to find himself back in that community again. Comparisons are available, and trends stand out. The customers are, indeed, different.

Key words: Condition-based maintenance; customers; open systems; wireless.

PICK YOUR CUSTOMER: The people responsible for financial results are becoming the customers and supporters of CBM. Efforts to educate and markets to the power generation industry in the late Eighties all to often ended with such statements as: "Your CBM stuff requires approval as a capitol purchase; if we had any funds available for capitol equipment, we would spend it on revenue-generating equipment". The company from which that statement came is now a leader in CBM applications, pulling our industry to provide the maintenance cost-saving tools they need. The difference appears to be in the substantially improved tracking of the costs of maintenance. Perhaps deregulation of the electric power generating industry has opened eyes, to the benefit of those who are reading this paper.

Lagging in that industry, and in many manufacturing facilities as well, is acceptance of CBM by the people who are responsible for the day-to-day status of machinery. Several explanations are available. If we do our jobs, some of them lose theirs. When we hear cost cutting on the business news, it means payroll reductions. The obvious virtues of CBM include replacing the human monitor with a device, but more threatening is substantial reductions in both preventive maintenance hours and repair hours following run-to-failure events. Emergency repairs are all to often the makers of heroes. Taken to

its criminal extreme, sabotage is the result. The author, like many of the readers, has witnessed examples of this behavior. The point for our industry is that we should be aware of this psychology and know whom we are talking to in our client base. The marketing manager who is handed off to the client's maintenance manager and told that that is who he must sell, has his work cut out for him. Given the choice, find the person who will be promoted if maintenance costs are reduced, emergencies eliminated, and uptime is improved.

A large and important set of potential clients traditionally makes a nice profit selling replacement parts. In industries like aviation that are, of necessity, conservative, CBM may be a harder sell than common sense would indicate. The author has watched as the message sunk in: CBM is inevitable, and we will no longer laugh all the way to the bank as we sell spares that are not needed. Power by the hour and similar programs should be on the tip of your tongue. It is certainly costing someone dollars per hour to fly or operate, and the human who knows, to the penny, what that cost is, is a good starting point for briefing CBM in organizations which are sellers of spares for their own equipment.

In recent discussions with a small company that manufactures equipment for chemical and environmental applications, another consideration and a very positive indicator of the strength of CBM was offered by the client: CBM as a product discriminator. The VP of Marketing, with little background in maintenance, recognized the marketing importance of being the first in their industry to offer a CBM approach to maintaining their equipment. The set of questions that followed included the possibility of exclusivity, and lengthy discussions about warrantees, service and all that goes with the "Who is going to watch the scope?" set of issues. Those of us who are systems integrators need to ensure that we understand just where the boundaries of those systems are, and those of us who want to sell elements of systems need to have the larger picture as well. Perhaps we all need our virtual teams, bench strength and all, in place as we describe CBM to the universe of potential users. New applications are numerous, and the CBM story is grasped very, quickly. The questions that follow will test the best of us before we sell systems or system elements.

DATA FUSION: A monitoring system was installed in a nuclear electric power generation plant in 1990, which contributed significantly to bottom line improvements within weeks of installation. The system monitored vibration; the facility already had thermocouples on the bearings of the generator sets. Two thermocouple indications of problems occurred; both subsequently proved to attributable to faulty/failed thermocouple wiring. The human operators made the comparisons between the heat and vibration sensing systems, and determined which was the correct indication. Various lessons are supported by this event. Data fusion need not be complex; conversely, it should be part of most systems. Systems we deliver in this decade should sort out such ambiguities without human help or intervention. Beyond such simplistic multi-data events, powerful data fusion tools are appearing.

Beware the sensor that measures one variable and purports to tell the remaining useful life of a component. Be it vibration, lube spectroscopy, temperature, or pressure, get a confirming "second opinion". "Is it a faulty fire warning system, or am I really on fire?" the pilot asked before ejecting from his single-engine jet. Similarly, the plant manager wants to be very sure before he orders parts, schedules a shut down and replaces a perfectly good bearing. Data fusion is powerful stuff that will help put CBM on the map

BE WIRELESS: That 1990 monitoring system described above sent data by wire to the signal processing hardware, and the information generated by wire to the human interface screens. The wiring installation cost more than the rest of system. Although the wiring had been done during a planned outage, the facility management became aware of the hours incurred during the installation. Although the thermocouple occurrences saved the facility more than the cost of the monitoring system, citing system cost, the management could not be convinced to purchase another vibration monitoring system. Real reasons, one could conclude: recognition that the wiring became a maintenance burden of its own as it had with the thermocouple; and wiring more than doubled the initial cost of the system. We must have wireless system in our bag of products.

Ongoing dialogs with several companies confirm the importance of wireless systems. The managers of an automotive parts manufacturing facility recognize the need for monitoring and diagnostics at various choke points. Wiring will not stand up to the environment. A manufacturer of test cells had been unwilling to discuss wireless systems in his cells until his customers recently inquiring about wireless technologies, citing electrical problems as accounting for 85% of down time. A manufacturer of plant equipment for several industries is now eager to include CBM as a part of his systems, citing his own experience with wiring maintenance costs, and reluctance of his customers to deal with more wire. A heavy equipment manufacturer recently revealed that he had a sensor design on the shelf, with intellectual property locked in, because the intended environment was hostile to wires. Wireless systems are now in demand, are what will make CBM work, and are what will sell.

OPEN SYSTEMS: Closed systems have limited the growth of CBM, and, in many cases, given our community a bad name. Educate your clients as to the meaning and importance of open system architecture, support the development of the needed standards, and spent your intellectual property dollars elsewhere. In Marketing I parlance, we can grow the pie faster than any of us can keep up. Don't hurt us all by trying to defend your slice.

LEGACY SYSTEMS OR OEMS?: The author was recently asked whether his marketing plan was aimed at retrofit on legacy systems or at embedded sensors sold to the OEMs for new, smart machines. "All of the above" is not an easy answer to defend. Beware the assumption that the same systems are good fits for both sets of application. None of us will ignore the huge OEM market, and the bright future it holds for CBM, but, likewise, we note that the average US Air Force aircraft is twenty years old, and the US Navy has ships on the seas that will still be there forty years from now. Manufacturing equipment and other CBM candidates split in similar proportions. The recommendation

is: polish up two stories, and be tuned to the distinctions. Back to Marketing I, they are both huge pies. Perhaps some of us should feed from one and not the other.

SUMMARY: CBM is becoming accepted at an accelerated rate. The folks who count the beans see the payback both in organizations that previously rejected CBM, and in new applications. As power by the hour, and improved methods of tracking maintenance costs grow, the acceleration of CBM will continue. Those of us positioned to influence the direction of CBM should stoke these flames with open system standards, data fusion approaches, and, where warranted, wireless systems. One size does not fit all: ensure excellence for both our legacy system clients and the OEM market.

CONDUCTIVE POLYMER SENSOR ARRAYS—A NEW FRONTIER TECHNOLOGY FOR CBM

Jeffrey N. Schoess
Honeywell Laboratories
3660 Technology Drive
Minneapolis, MN 55418

Abstract: Today's commercial and military aircraft require significant manpower resources to provide operational readiness and safety of flight. Aging aircraft fleets are much in need of new and innovative health-monitoring methods to prevent catastrophic failure and reduce life-cycle costs. The key needs of characterizing in situ structural integrity characteristics of corrosion and barely visible impact damage (BVID) to determine the "damage susceptibility" must be addressed. This paper presents a new concept for performing onboard real-time monitoring using conductive polymer sensor array technology.

Using conductive polymer thick film (PTF) technology and elastomer materials, Honeywell is developing a family of low-cost sensor-on-film technology (SOFT) capable of sensing temperature, moisture, vibration, structural impact and, strain quantities. These sensors conform to surface profiles (6 to 10 mils thick) adding little weight and can be easily replicated to provide deeply distributed and highly redundant web architecture solutions. The SOFT approach is based on the novel idea of directly integrating sensory, control, and data processing electronics into the system of interest (vehicle, space-borne structure, etc.). The polymer sensory system is proposed to conform to the shape of the platform into which it would be integrated, or in other words, be "conformal," which by definition means to "have the same shape or contour". The technical approach defines the novel idea of using a polymer film as a flexible substrate, on the backside of which electrical interconnects, sensory functions, and data processing electronics would be directly integrated. The sensory functions are defined by incorporating polymer thick-film patterns on the film surface which can then be bonded to the platform of interest to perform failure prevention diagnostics.

Key Words: Conductive polymer sensor, sensor arrays, conformal sensor, condition-based maintenance

Background: Both commercial and military service personnel currently employ "walk-around" structural inspection as a cornerstone for performing condition-based maintenance. This means that a hierarchy of inspections is required to ensure that fleet readiness and availability requirements are met. Structural inspection includes daily inspection, phased maintenance based on aircraft operating time, conditional inspection based on the mission and location of the aircraft, and calendar-based inspection.

Although condition-based maintenance inspection is mature and performed reliably in most cases, its application in future military and commercial systems has significant drawbacks:

- High Cost—Currently, the cost to maintain a Navy aircraft is up to $200,000 per year. A 1996 Naval Center for Cost Analysis AMOSC report indicates that the direct cost of maintaining Navy aircraft and ships is at least $15.0 B per year. As much as 25% to 30% of operating revenue is spent on maintenance activities for commercial air carriers.
- Manpower-Intensive Effort—According to a 1995 study performed by the office of the Under Secretary of Defense, 47% of the Navy's active duty enlisted force (173,000 sailors) and 24% of the Marine Corps (37,600 marines) are assigned to maintenance functions. The mandate to reduce manpower while performing duties faster, cheaper, better, and with increased reliability is a reality in both military and commercial transportation segments.

In addition to these issues, problem areas exist specifically for maintaining structural integrity, including:

- BVID—The increased use of composite materials in aircraft structures introduces the potential for BVID, a maintenance-induced damage effect. At least 30% of all maintenance performed is related to structural repair due to tool dropping, in-service damage, etc.
- Hidden and Inaccessible Corrosion—A significant amount of structural integrity loss is due to hidden corrosion as well as corrosion located in inaccessible areas (wheel wells, landing gear areas, fuel tank, etc.). The practice of applying surface treatments of various types to provide adequate protection, in some cases overcoating the surface with several layers, causes considerable weight increase. This increase results in loss of fuel savings and aircraft performance.

Technical Approach:

Trade Study Results: This section summarizes a trade study performed to identify and assess potential aircraft inspection areas that could benefit from conductive polymer sensor array technology. The trade study involved the identification of seven key areas of a generic fighter aircraft (F-18 or equivalent). The areas addressed in the study were external wing structure, internal wing and fuselage structure, including landing gear and cockpit canopy, communications, external stores, and empennage structure. Figure 1 is a drawing of the F-18 aircraft showing the functional layout of the seven aircraft sensing areas for possible future technology insertion. The sensing areas are mapped to the aircraft geometry, labeled by area, and keyed with the full-scale trade study chart shown in Table 1. For each sensing approach, three packaging options exist: (1) a conformal sensor array, which would cover a larger surface area such as an external wing area over several square feet; (2) a conformal sensor applique to provide sensing coverage in a smaller area (a few square inches, possibly with significant contour shapes); and (3) a conformal boot assembly. The conformal boot design would involve the fabrication of a

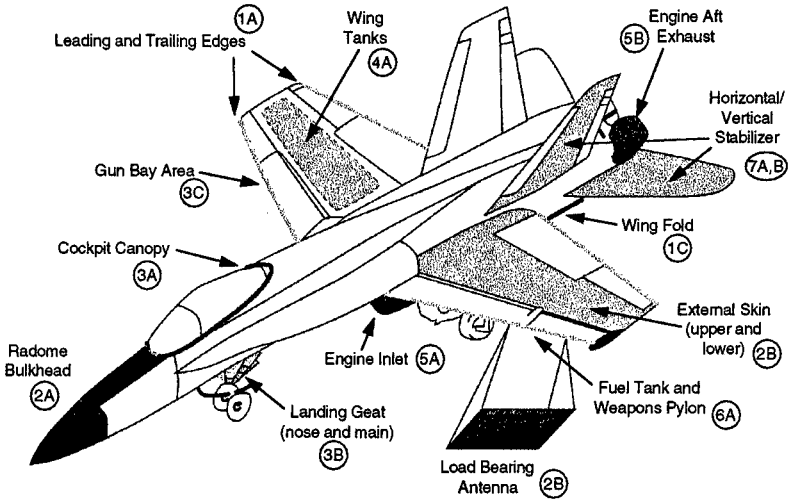


Figure 1. Key Sensing Locations on Aircraft

Table 1. Aircraft Trade Study Chart

Aircraft Area		Part/Assembly	Problem Definition	Sensing Approach*	Sensing Configuration
1	Wing External	• Leading edges • Trailing edges	• Flap and drive assembly • Impact (BVID) • Corrosion—wing attach fitting • Erosion	• M/C • ID	• Conformal array • Conformal boot
		• External skin (upper and lower)	• Impact (BVID due to maintenance/ repair) • Corrosion (fastener area)	• M/C • ID	• Conformal array • Conformal applique
		• Wing fold	• Corrosion in hinge area • Wing attachment fatigue	• M/C	• Conformal tape
2	Communications Support	• Radome bulkhead • Wing antenna	• Corrosion (dissimilar ± F-galvanic) • Phased-array antenna	• M/C • LBA	• Conformal boot • Conformal applique
3	Fuselage	• Cockpit canopy	• Corrosion—dissimilar interface (galvanic)	• M/C	• Conformal applique
		• Landing gear	• Corrosion in wheel well area, main landing gear assembly	• M/C	• Conformal applique
		• Gun bay area	• Corrosion—dissimilar interface	• M/C	• Conformal applique
4	Wing Internal	• Wing tank	• Fuel leakage in web area (wet bay) • Electrical connector/ground straps	• M/C	• Conformal applique
5	Engine	• Engine inlet • Aft engine exhaust area	• Impact (BVID) to debris/bird strike • Corrosion—moisture	• ID • M/C	• Conformal applique
6	External Stores	• Fuel tank pylon • Weapons pylon	• Corrosion—dissimilar interface • Erosion	• M/C	• Conformal applique
7	Empennage	• Horizontal stabilizer • Vertical stabilizer box	• Pivot shaft corrosion • Corrosion	• M/C	• Conformal applique

* M/C = Moisture/Corrosion; ID = Impact Detection; LBA = Load-Bearing Antenna

preformed structure—a sensory boot that fits the spatial constraints of the aircraft contour shape. An example of this configuration would be a preformed boot fit over the leading edge or radome bulkhead assembly.

Sensor Development: This section describes details of the conductive polymer sensor array design [1] that provides the capability for performing multifunction conformal sensing. Also presented are the sensory applications for using the linear sensor array to detect corrosivity characteristics as an outer aircraft skin or below floorboards, impact forces that can cause BVID and electromagnetic energy.

Polymer Sensor Array Design: Honeywell has developed polymer sensors to sense moisture (i.e., electrolyte) conditions and the presence of moisture/fluids over an extended surface area. A primary maintenance concern is the need to sense and quantify moisture trapped between the protectant system layer and aircraft surface that could cause corrosion to occur. Typically, the moisture is an electrolyte, an electrically conducting fluid that has ions in solution. The polymer sensor array has been designed to detect the "presence" of an electrolyte, which can be seawater, acid rain, lavatory fluids, fuel, hydraulic fluid, chemicals, or cargo by-products.

The basic design is implemented by printing on a flexible substrate material with a specific pattern design, curing it, and layering it with a pressure-sensitive adhesive. A typical pattern developed for electrolyte sensing is a transducer design with alternating electrode pairs. Figure 2 illustrates the pattern layout for a polymer sensor array. The figure shows a set of dedicated electrode pairs, each of which operates as a sensory element. The sensor is designed to function as a linear 2-D array that measures the "location" where the electrolyte is sensed and the "amount" of electrolyte based on exposure across the sensor array.

Detection of Corrosivity: Four conditions must exist before corrosion can occur: (1) presence of a metal that will corrode an anode; (2) presence of a dissimilar conductive material (i.e., cathode) that has less tendency to corrode; (3) presence of a conductive liquid (electrolyte); and (4) an electrical path between anode and cathode. A corrosion cell is formed if these four conditions exist due to the electrochemical effect, as shown in Figure 3. In future aircraft, paintless appliques will be applied to the surface of the metal to act as a moisture barrier to protect the bare metal from being exposed to the electrolyte. The applique film layer (6.0-mil-thick fluoropolymer film) prevents the corrosion cell from functioning by separating the electrolyte from the anodic and cathodic sites on the metal surface. If this layer is damaged due to erosion, heat exposure, or aging, the cell is activated, which causes corrosion to occur.

Figure 3 also highlights the concept of using a polymer sensor array to detect corrosive susceptibility. A polymer sensor array is patterned on the backside of applique film layer using standard ink-jet printing techniques. The applique is then bonded to the aircraft surface via a pressure sensitive adhesive (PSA) layer. The sensor array then operates to sense the "conductivity" of the trapped fluid by conducting a current through the fluid located between IDT electrode pairs. The fluid's conductivity property is, by definition, "the ability to act like an electrolyte and conduct a current, or a measure of its corrosivity."

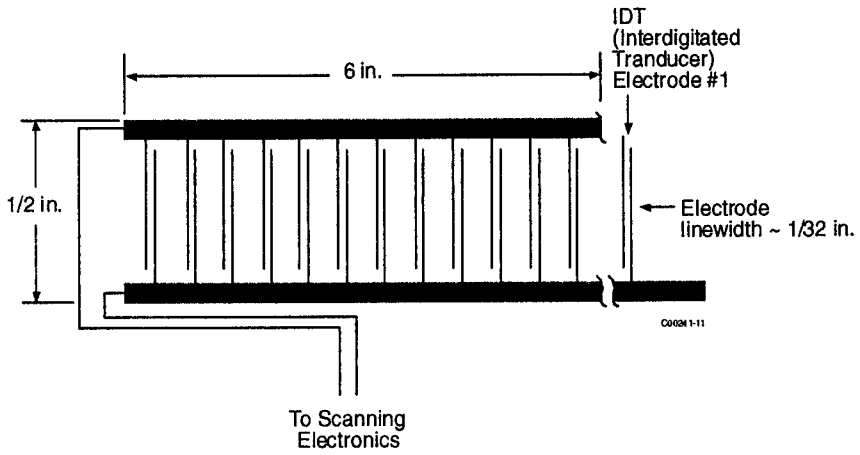


Figure 2. Pattern Layout of Polymer Sensor Array

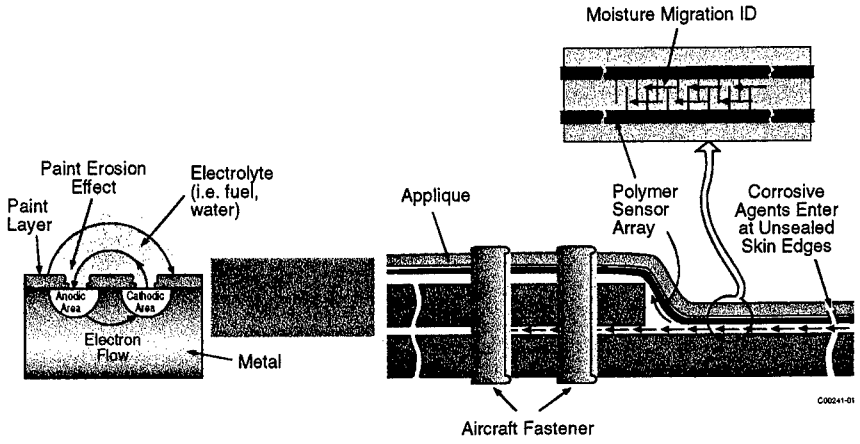


Figure 3. Smart Corrosion Sensing Concept

The concept of performing corrosive environmental "exposure susceptibility" index monitoring to minimize scheduled inspections and provide direct cost savings is shown in Figure 4. The basic idea is to continuously monitor the actual exposure of each aircraft to corrosive environmental factors (moisture ingress into protective coating, type of corrosive agent, etc.) and then schedule corrosion inspections based on these measurements, rather than on preset rules that are only loosely related to corrosion. Typical preset rules that an exposure susceptibility index would replace are calendar-based (i.e., inspection every 30 days) or usage-based (i.e., inspection every 10 hr of operation) inspections. One can think of the system as a "corrosion odometer" with a

readout that steadily increases according to the corrosiveness of the environment to which the aircraft is exposed. Maintenance personnel can intermittently check the odometer and perform inspections as needed.

$$I = \int dt \cdot F(W, Cl, C, T)$$

W ≡ Humidity, Wetness
Cl ≡ Concentration Level of Index
C ≡ Corrosivity
T ≡ Temperature

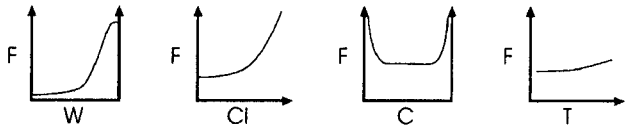


Figure 4. Exposure Susceptibility Index

The sensor array approach is capable of sensing and calculating an exposure index to ingress of an electrolyte (i.e., water) and the "wetness" effect of the electrolyte. The wet/dry cycle of exposure is a strong indicator of how susceptible an aircraft is to corrosion, with wetness being a basic requirement for corrosion to occur. The wetness exposure index is defined as the integral over time of the function $F_W(W)$. Here W is the time-varying output of a "wetness" sensor (1 = wet, 0 = dry) and quantifies the total corrosive effect of wetness. F_W is a simple function that gives the exposure index in a convenient scale, so an abbreviated inspection is called for each time the index passes through a multiple of 100, for example. Thus, for severe environments such as in Puerto Rico, an increase by 100 every 15 days could occur, as compared to an increase by 100 every 90 days in Denver.

Further improvement to the exposure susceptibility index can be obtained by adding other environmental factors that can influence corrosion. These include the concentration level of the electrolyte, temperature, and conductivity (corrosivity factor).

Figure 5 illustrates the index calculation concept, showing the maintenance cost savings concept in detail. The design approach is set up to collect and analyze the environmental factors related to structural health (moisture ingress, impact forces, etc.) that could lead to loss of structural integrity. These factors are collected and integrated as a "cumulative index" to determine (1) the level of "susceptibility" to failure and (2) whether maintenance is required at a given location in the aircraft. The cumulative index value is envisioned to be represented as simple whole number from 0 to 100 (which indicates the level of susceptibility, with a higher number indicating more potential for damage may exist) that could be read out by maintenance personnel from the aircraft maintenance debriefing interface at scheduled inspection intervals. The crew could then make a decision to perform scheduled maintenance or bypass the action, which reduces overall operating cost by reducing inspection time.

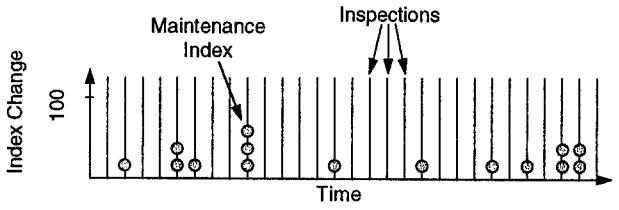


Figure 5. Maintenance Cost Savings Tutorial

Impact Detection: The polymer sensor implemented for moisture/corrosion sensing is also capable of sensing impact forces caused by maintenance-induced damage or operational servicing. To provide sensing for impact forces, the polymer sensor array is configured with an additional semiconductor polymer layer, as shown in Figure 6. The design approach is set up to operate as a force-sensing resistor (FSR). An FSR operates on the principle of converting force applied via a structural impact event to an equivalent voltage output. As pressure is applied, individual electrode pairs are shunted, causing a decrease in electrical resistance. The measurement of impact force magnitude, impact direction vector along the sensor array, and impact surface area can be quantified depending on polymer composition, shunt pattern and shunt shape, and the method for applying pressure (hemispherical vs. flat). Figure 7 shows the typical curve of sensor response. The figure is a plot of electrical resistivity vs. applied force with an active sensing region of two to three orders of magnitude from low impedance (kilohms) to high impedance (megohms). Over a wide range of applied pressure, the sensor response is approximately a linear function of force. The first abrupt transition that occurs is at low pressure. This point is called the "breakover point" where the slope value changes. Above this region, the force is approximately proportional to 1/R until a saturation region is reached. When force reaches this magnitude, applying additional force does not decrease the resistance substantially.

Figure 8 is a photo illustration of a commercially available off-the-shelf FSR product called Uniforce, which has an operating range of 0-1000 psi.

Another type of conductive polymer sensor is a polymer matrix sensor, which consists of electrically conducting and nonconducting particles suspended in a matrix binder material. Figure 9 shows a cross-sectional view of a polymer matrix sensor. Typical design construction includes a matrix binder and filler. The choice of matrix binder materials can include polyimides, polyesters, polyethylene, silicone, and other nonconducting materials. Some typical filler materials include carbon black, copper, silver, gold, and silica. Particle sizes typically are on the order of fractions of microns in diameter and are formulated to reduce temperature dependence, improve mechanical properties, and increase surface durability. Applying an external force to the surface of a sensing film causes particles to touch each other, decreasing the overall electrical resistance.

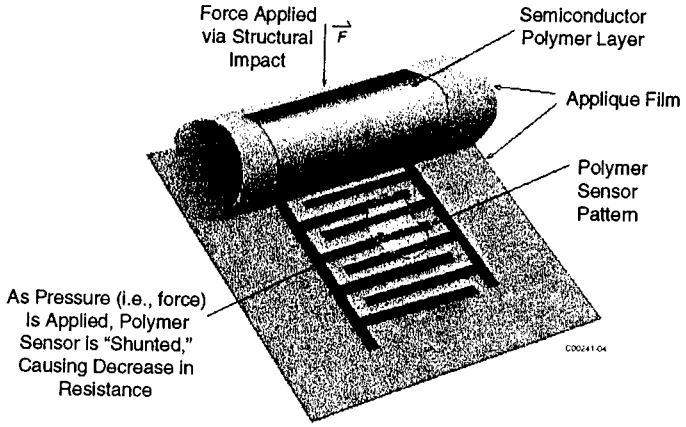


Figure 6. Force-Sensing Resistor (FSR)

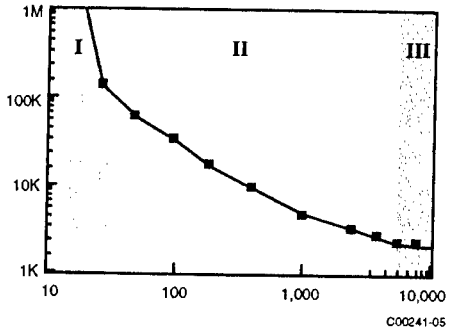


Figure 7. FSR Response vs. Applied Force

Figure 8. Example of Off-the-shelf FSR Product

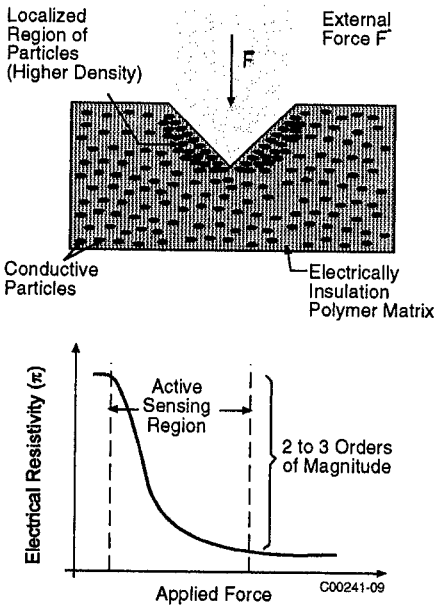


Figure 9. Polymer Matrix Sensor

Table 2 illustrates the typical performance properties for polymer thick-film (PTF) resistor technology and other resistor technologies. The table includes a summary for thin films, semiconductor, and continuous metal films. The significant advantage of PTF resistor technology over all other resistor sensing is the cost to fabricate devices. The PTF cost factor is achieved by the ability to print resistive material via stencil, screen printing, and ink-jet printing techniques.

Table 2. PTF Resistor vs. Other Resistor Technology

Resistor Type	Gauge Factor (G)	TCR (ppm/°C)	Application Method	Relative Cost
Continuous metallic films	2.0	20.0	• Spin cost	High
Thin film	50.0	20.0	• RF sputter • Evaporation	High
Semiconductor	50.0	1500.0	• Diffused • Implanted	Medium
Thick Film (PTF)	10.0	50.0–500.0	• Screen print • Stencil • Spin cost	Low

Source: G. Harsanyi (Ed.), *Polymer Films in Sensor Applications—Technology, Materials, Devices and Their Characteristics,* Technomic Publication, 1995.

A prime example of how FSR technology could be used for aerospace sensing is structural integrity monitoring. Today's commercial and aerospace structures incorporate a large amount of composite materials to reduce structural weight and increase load-bearing properties. Composites are susceptible to damage due to impact forces experienced in operation, including debris picked up from runways and maintenance-induced damage caused by tool dropping. Figure 10 illustrates the system-level concept of impact-damage-detection-based applied force vs. damage for a composite aircraft panel. A matrix array of FSR elements is shown integrated into the aircraft panel. Panel construction involves printing FSR elements directly on the panel surface or on a film layer, which is then bonded to the panel via a pressure-sensitive adhesive layer. The polymer patterns incorporated on the panel include a combination of sensor elements and electrical interconnects implemented with conductive polymer materials.

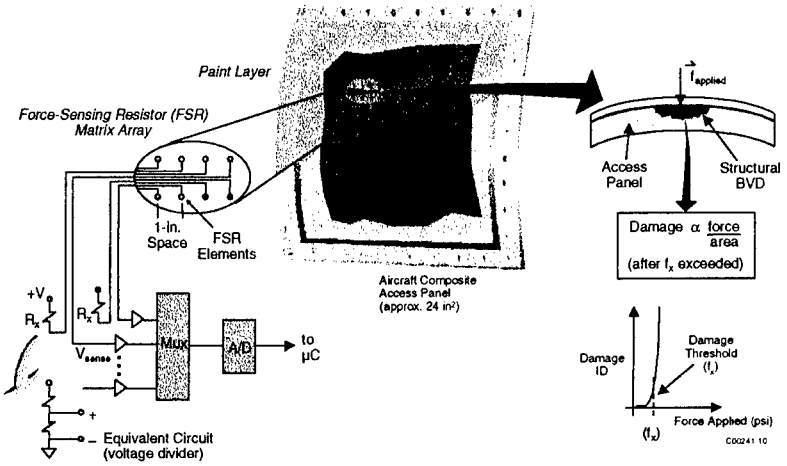


Figure 10. Structural Impact Damage Tutorial

To measure and record impact forces in real time, the output of each FSR element is converted to an equivalent voltage via a simple voltage divider circuit and provided as input for a dedicated data acquisition system. Each FSR element output is routed to an analog multiplexer. An analog-to-digital converter sequentially digitizes each FSR value into an equivalent digital word to be processed by a dedicated system controller. The illustration on the right-hand side of Figure 10 shows what happens if structural damage occurs. An external force event (i.e., tool dropped on the surface) causes an impact to occur. Structural damage usually consists of multilayer delaminating or microcracking of individual composite layers. In composite structure applications, the curve for quantifying structural damage is an exponential relationship and is detected by setting a force threshold value. Exceedance of the threshold value f_x indicates that barely visible structural damage has occurred. The effects of detected damage can be read out by

maintenance personnel on a periodic basis to determine if structural repair is needed or marked as suspect and the vehicle returned to active service. A set of damage identification threshold values could be retained for each major structural component of the aircraft in a 3-D map database to perform maintenance on demand.

Conformal Antennas: A significant feature of polymer sensor array technology is the arrays' ability to operate as a low observable (LO) conformal antenna. A collocated antenna could be used to debrief sensory data to a central maintenance database in ground-support applications. The polymer sensor has been tested in laboratory conditions to detect broadband frequencies of several megahertz without any optimization of the polymer circuit pattern. The conformal antenna capability offers a significant benefit of increasing detection of "bad guy" signature threats. Tests performed by aircraft primes have indicated that conformal load-bearing antennas improve detection by a factor of 6X to 14X. In addition, the conformal polymer construction makes it suitable for phased-array antenna design for munitions and guided projectiles. Figure 11 illustrates the feasibility of using the polymer design for antenna functions.

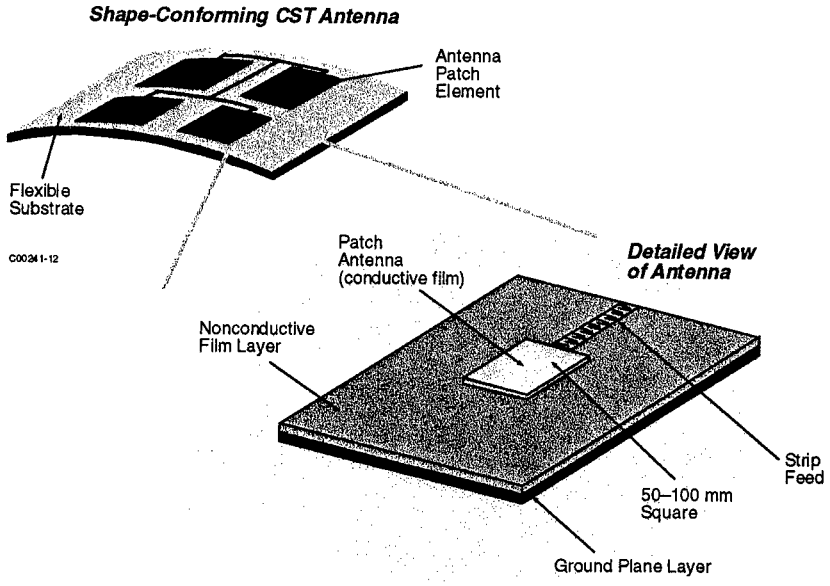


Figure 11. Example of Conformal Antenna

Figure 12 illustrates a CBM application of conductive polymer sensor arrays for machinery health monitoring. The figure highlights a PTF FSR circuit on a polymer film substrate configured as a low-cost vibration sensor. A small inertial mass is shown placed on the top of the polymer circuit which inertially loads the sensor, applying an external force related to operation of the machinery component (i.e. pump). The vibration sensor is shown mechanically bonded to pump with a pressure-sensitive adhesive tape layer. The

vibration output signal is conditioned by the co-located electronics module and transmitted via a wireless transmitter to a central maintenance database for detailed analysis and determination of pump health status. The key advantages of this type of condition monitoring are: 1) ease of placement- the conformal vibration sensor can be placed at any physical location on the pump to improve vibration pickup characteristics, and identify any structural modes of interest, 2) low cost of implementation- the polymer sensor offers significantly lower acquisition cost (x 10) than a conventional vibration sensor which presents opportunities to adding additional sensors to increase health "awareness" and improve overall system level redundancy performance.

Conclusions and Summary: The details of conductive polymer sensor arrays and their applications for structural health monitoring have been addressed. The applications of corrosion susceptibility, impact damage, and conformal antennas were presented. A conceptual view of wireless sensor web communications for field operation to support decision making and maintenance was presented.

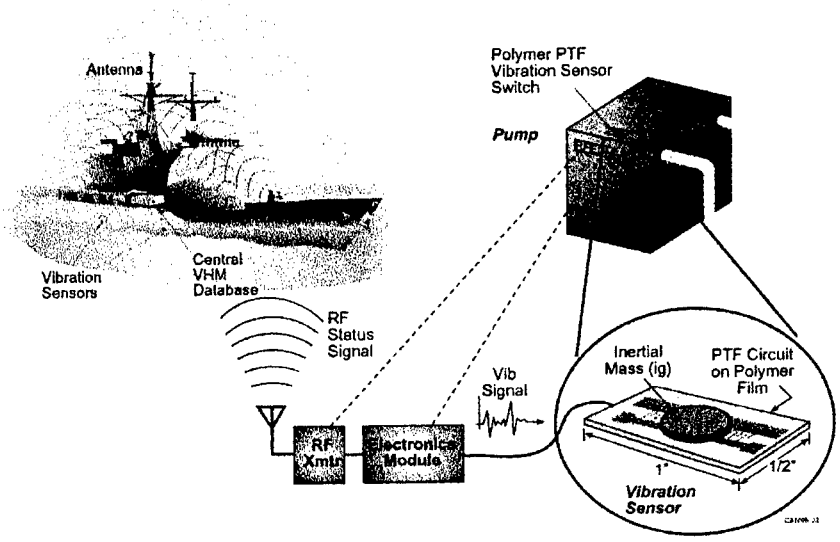


Figure 12. CBM Application of Conductive Polymer Sensor Arrays for Machinery Health Monitoring

References:

1. J.N. Schoess, "Aerospace Applications of Smart Materials: A Sensing Perspective Using Conductive Polymer Sensor Arrays," *Encyclopedia on Smart Materials*, (1st ed.). New York: John Wiley & Sons (2001).

INTELLIGENT SENSOR NODES ENABLE A NEW GENERATION OF MACHINERY DIAGNOSTICS AND PROGNOSTICS

Fred M. Discenzo	Kenneth A. Loparo	Dukki Chung	Allen Twarowski
Rockwell Automation Cleveland, OH	Case Western Reserve University Cleveland, OH	Rockwell Automation Cleveland, OH	Rockwell Science Center Thousand Oaks, CA

Abstract: Compelling economic, competitive, and technological factors are changing the way many companies view machinery maintenance, repair, and overhaul (MRO) activities. This shift toward a new Maintenance Management Paradigm has implications in many areas of the business including manufacturing scheduling, control, finance, inventory, quality, and asset management. Implementation of the new Maintenance Management Paradigm will require three fundamental building blocks. First, is a framework that enables the efficient re-use of best-in-class diagnostic and prognostic software, hardware, and sensor modules. An open-system architecture will be fundamental to meeting this objective. Second, is the ability to rapidly deploy needed hardware and software elements in a reliable and cost-effective manner across distributed system components. Wireless, intelligent sensor nodes will play an important role in the deployment of future systems. And third, is the infrastructure and that will permit system level integration of an ensemble of distributed intelligent system elements to develop actionable diagnostic and prognostic information. Higher-level diagnostic and prognostic information will drive critical decision making to insure maximum system reliability, lowest operating cost, maximum revenue generation or mission success for example. This paper provides specific examples of elements in the areas of Framework, Distributed Intelligent Modules, and Infrastructure for system-level integration.

Key Words: Agent; diagnostics; distributed intelligence; failure prediction; intelligent sensors; maintenance management paradigm; open systems architecture for condition based maintenance (OSA/CBM); prognostics

I. Introduction: Machinery health assessment is becoming critically important across a broad spectrum of shipboard, industrial, and commercial applications. The operational demands and high procurement costs of today's systems in these applications requires a high degree of uptime and high reliability. Unexpected failures are costly to correct at best, at worst; such failures will be catastrophic.

At a recent workshop on Intelligent Devices various industry representatives consistently said that their top priorities were 1. machinery prognostics and 2. overall process or system-level health [1]. Similar priorities also emerged following a two-day NIST

workshop on Condition-Based Maintenance on November 17-18, 1998. The barriers or technology requirements identified at this workshop were 1. continuous monitoring, 2. accurate prediction of remaining useful life, and 3. generate adaptable and actionable recommendations.

The benefits available through effective diagnostic/prognostic system are now becoming a business necessity for many organizations. Labor, material, and energy costs have already been minimized for many plant operations. Maintenance expenses and the operational impact of unexpected failures represent the largest remaining controllable expenditure. In some cases, maintenance expenses may exceed the profit from a plant. For other organizations, industry-leading machinery reliability and maximum uptime are absolutely essential to suceed at various contemporary business strategies such as JIT, TQM, supply-chain management, and OEM service outsourcing among others.

New developments in algorithms, industry standards, communications, and software architectures promise to accelerate the deployment effective diagnostic and prognostic systems. These developments will enable new technologies to be readily integrated with existing systems for near-term, observable business impact.

II. Background: The diagnostic and prognostic needs expressed by a broad range of organizations may be met in a cost-effective manner and with minimal risk by leveraging specific developments occurring in three critical areas. First is a framework that provides the ability to efficiently integrate re-usable algorithms. Second are developments in communications and in particular wireless technologies. Third is the development of an infrastructure for the system-level integration of distributed intelligent system elements. System-level integration provides the foundation for robust systems which are capable of generating actionable diagnostics and prognostic plans., hardware platforms, sensor modules, database information, and intelligent devices in an automation system.

The integration of these three factors together with advanced prognostic algorithms will form the cornerstone of a new Machinery Maintenance Paradigm. This paradigm employs targeted, task-specific sensors and algorithms integrated in a framework that is readily expanded and adapted to changing operational needs. Open, industry-standard systems and interfaces are fundamental to this framework. Such systems will then be readily integrated throughout the plant equipment and coupled with various IT, operational planning, finance, and control systems.

III. Framework: Various components and functional capabilities of diagnostic and prognostic systems may be organized as a framework of system components or building blocks. Such a framework provides a scheme for identifying and organizing essential information about a system and also provides an structure for defining standard interfaces between system elements and functional requirements for important system elements. The requirements for capturing machinery health information, interpreting the data, and acting on the results of the analysis may be arranged in a hierarchy (Table I). This hierarchy ranges from simply capturing data from a particular process or machine to

interpreting sampled data to detecting a fault has occurred or what fault or faults will occur. Higher levels of the hierarchy provide a range of capabilities for automatically reacting to novel faults or reacting in advance of an anticipated failure. Higher levels of this hierarchy beginning with diagnosis and prognosis will typically require the integration of multiple data sources as well a knowledge of the process equipment and operating state or context.

Table I

HIERARCHY OF INTELLIGENT MACHINES

1. Data Acquisition
2. Monitor
3. Detect
4. Diagnose
5. Prognosis
6. Prognostics & Control
7. System-Level Prognosis & Control
8. Dynamic optimization / multi-objective control
9. Adaptive / Reconfigurable

Order of increasing complexity / cost / economic benefit

There is clearly a move to intelligent devices and distributed intelligence. Intelligent components may occur at the structural level (e.g. smart materials), at the sensor level, or at the device level (e.g. embedded intelligence). One effort to establish a standard transducer (sensor / actuator) interface is the IEEE 1451 standards effort. This standard seeks to move data acquisition, distributed sensing, and control to more of an open system by establishing a framework and data elements for "smart" tranducers. Included in this standard is a specification to facilitate sensor identification, calibration, documentation, sensor replacement, and network integration among other features [2]. Research in self-validating sensors (SEVA) at the University of Oxford over the last 12 years have been directed at defining intelligent sensors which dynamically sense their own condition and provide information regarding the quality or validity of the sensor value returned [3]. This is an important, emerging area currently being proposed at a draft standard by BSI for Data Quality Metrics [4]. Information on data quality becomes critically important as we move to higher levels of the hierarchy of intelligent machines shown above toward automatic control, decision-making, and autonomous machines.

Data critical to establishing the health of equipment may be captured and stored in an application-specific manner to accommodate essential memory or timing constraints. Preferably, machinery data should be organized in an open, industry standard format such as defined by MIMOSA (Manufacturers Information Management Open Systems Alliance). The data format and definitions established by MIMOSA are the result of many years work by an international team of MIMOSA sponsors and members. This standard is open and accessible to the public [5].

More recently, a group of industry and academic partners have teamed together to develop an Open Systems Architecture for Condition Based Maintenance (OSA/CBM). This program is part of the Dual Use Science & Technology program (DU&ST) program with joint industry-government funding (BAA 98-023). This effort, leveraging off the work of MIMOSA, has resulted in an operational framework for machinery diagnostics in an open-system, layered framework as shown in Fig. 1 [6].

This open system architecture implements a middleware interface to support a broad range of operational models including COM/DCOM, CORBA, and XML/HTTP client-server architecture. This model was demonstrated on a laboratory pumping system in December 2000 and will be demonstrated on aircraft, off-road vehicle, and shipboard applications during the next year.

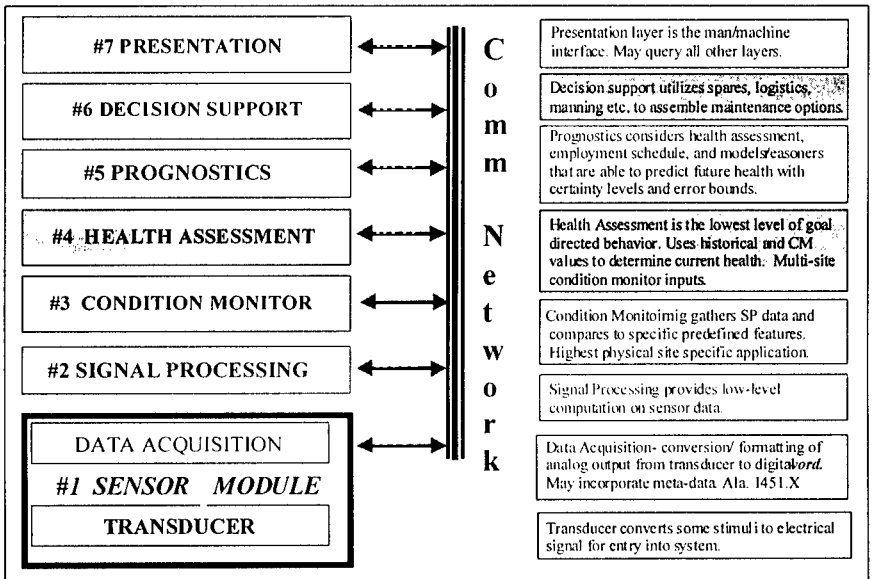


Figure 1
Open Systems Architecture for Condition-Based Maintenance (OSA/CBM)

The integration of intelligent sensors and self-validating sensors in an open, operational framework as specified by MIMOSA - OSA/CBM promotes the development and deployment of distributed intelligent sensors across a broad range of applications.

IV. Distributed Intelligent Sensors

We continue to see a rapid pace of development in intelligent sensors and open systems. These developments are leveraging off developments in software architectures, sensor technologies, and networks which are moving toward open, public standards. The wide-scale deployment of intelligent devices in manufacturing and commercial operations remains limited due to the high cost of installation and the cost and complexity associated the processing and analysis of the massive amount of real-time data received from the hundreds or thousands of sensors. Typically only data is received from remote sensors, as opposed to information (e.g. actionable information or health information).

Developments are occurring in wired networks and wireless networks that promise to reduce or virtually eliminate the wiring costs for distributed sensors. Bit-oriented networks such as DeviceNet are effective for local-scale integration of plant sensors. For higher bandwidth and extended networking TCP/IP becomes more attractive particularly when the higher cost of network access and transport may be justified by many remote intelligent nodes.

Wiring costs are often significant a cost component in the installation of many manufacturing and commercial systems. These costs may be eliminated with the use of wireless communications technology although at the expense of needed radio links. A variety of radio links exist and are selected based on needs for various bandwidth, reliability, distance, interoperability, and network architecture. Various government funded programs from the Department of Energy, Department of Commerce and DARPA seek to support the development of wireless technology for smart devices [7].

Recently there has been significant interest in low-cost wireless networks. Much of this interest is driven by the significant commercial potential for wireless consumer products. The development of the Bluetooth communications standard driven by the huge commercial potential may provide very low cost, 2.4 Ghz wireless data links for local sensor networks and data acquisition [8].

Wireless sensor networks promise to enable numerous applications for sensing and control including machinery monitoring for intelligent maintenance management. Rockwell Science Center has created a development environment for testing wireless sensing applications [9]. The Wireless Integrated Networked Sensing (WINS) platform includes

- bi-directional RF communications hardware and sophisticated networking protocols,
- processing and memory with a multi-tasking, real-time operating system for sensor data acquisition, signal conditioning and algorithmic processing of the data, and
- support for multiple sensor inputs including wide bandwidth accelerometers for vibration monitoring.

The integration of these technologies allows the easy installation of remote access sensors that can be configured in a variety of ways. The individual units, or nodes on the network, receive sensor data from attached sensors, process the sensor data and send back messages to the end-user, informing him/her about the condition of the equipment or process that is being monitored. Implicit in this architecture is a degree of distributed processing that can range from individual WINS nodes processing the data from their sensors and interpreting that data, to higher level algorithms that perform system-wide diagnostics using pre-processed sensor data from other WINS nodes [10].

The current WINS communications network operates in the 900 MHz RF band that is regulated for unlicensed use in the United States for transmission powers under 1 watt (spread spectrum) or 1 mwatt (single frequency). Because of the limited transmit power required for operation at these frequencies, the range of each unit is nominally 100 meters. Added to this the possibility of other path losses from absorbers or reflectors

between the WINS nodes and the end-user gateway node, and the possibility that these may be dynamic, the need for a robust network system is apparent.

Rockwell Science Center has launched a commercial product, HiDRA (Highly Deployable Remote Access), that is based on the WINS technology [Fig. 2]. Included in this product is a network protocol that supports broadcast and uni-cast, multi-hop communication links. The multi-hop protocols allow routing of messages from nodes that are out of the RF communications range of the end-user gateway node through intermediate nodes that are within RF range of each other. The system is self-configurable, so that the user does not have to spend time setting up the network, it is done automatically at startup. It is also dynamic, so that if conditions change in the RF environment of the HiDRA node, the routing table for relaying a message through several nodes to arrive at the desired destination will also change.

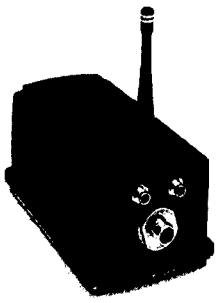


Figure 2
Rockwell Science Center
Hidra node

The WINS system has been deployed for over a year monitoring the bearing health in the HVAC facility that serves the two plants at Rockwell Science Center in Thousand Oaks. The 10 WINS nodes are mounted on 50-75 hp motors, housed in the cooling fluid pump room, that drive pumps supplying the cooling fluids for the buildings at Rockwell Science Center. Accclerometers are mounted to the motor casing over the bearing locations of the motors and measure the vibration at these locations. A temperature sensor monitors the temperature of the motor case close to the position of the accelerometer. Each WINS node does a spectral analysis of the vibration signals that it receives and computes bearing health status indicators that it transmits through RF link to a base-station WINS node connected to an internet server.

V. Infrastructure for Intelligent Systems

Distributed intelligent sensor nodes capable of processing data from multiple sensors provide unique opportunities for data fusion and for cooperative processing. Our objective is to put as much information into each node and leverage the capabilities for processing complex algorithms in parallel and in collaboration with other sensor nodes. This permits establishing accurate, dynamic, and robust models for diagnostics and prognostics. The following outlines recent developments in model-based and non-linear analysis methods. These new diagnostic / prognostic and modeling tools are considered foundational and provides a basis for future self-organizing sensor networks and dynamically reconfigurable systems.

There are a variety of approaches for the development of fault detection, diagnosis and prognosis algorithms with each having unique advantages and limitations when implemented as distributed processing nodes.

Model-based approaches combine physical modeling of the system with experimental data to determine a mathematical relationship between the occurrence of a fault and the characteristics of a measured quantity in the system. Extended model-based techniques may employ a family of models that relate each of the individual faults and their severity. The residuals from each of the model-based observers are combined and then integrated with other information available from casual modeling, signal processing, expert systems, etc. to arrive at a decision regarding the current operating status of the system. A diagram of the fault detection and diagnosis system implemented for rotating machinery is shown in Figure 3. The core of this technology is an array of nonlinear filter/model-based observer blocks that combine the output from a suite of observers. For details of this structure and operation refer to [11][12].

An important aspect of the system shown in Figure 3 is the ability to integrate information from a variety of different sources such as multiple sensor nodes, into a comprehensive fault detection and diagnosis decision. We have also implemented novel algorithms for the detection and diagnosis of faults in rolling element bearing. These algorithms are particularly well-suited for implementation on distributed sensor nodes. Bearing fault detection has been demonstrated using model-based techniques with a fault-detection filter. The ability to isolate the specific fault (e.g. ball, inner and outer race defects) a time-frequency analysis method was developed. This was demonstrated using experimental data collected from an induction motor system, refer to [13]. Also, for a novel approach that integrates sliding mode observers and fault detection filters for the detection and isolation of faults in rolling element bearings, refer to [12].

A new method has been developed for the detection and diagnosis of defects in ball bearings using the wavelet transform [14]. The signature produced by damage on the DB2 wavelet is used for the wavelet decomposition of the preprocessed vibration signals. A set of feature vectors are then defined based on the wavelet decompositions.

Finally, we present a method for the detection and diagnosis of mechanical faults in rolling element bearings using vibration data and knowledge of the bearing defect frequencies. For a particular bearing geometry, inner raceway, outer raceway and rolling element faults generate vibration spectra with unique frequency components. The bearing defect frequencies are linear functions of the rotating speed of the shaft. Outer race and inner race frequencies are also linear functions of the number of balls in the bearing. The operating speed changes with load and is often unknown and/or unmeasurable. In addition, even if the type of bearing in the machine is known, the number of balls in the bearing may be unknown. Thus, estimating the running speed and the number of balls in the bearing are required for failure detection and diagnosis methods that rely on knowledge of the defect frequencies of the bearing. We have developed and implemented separate algorithms for estimating the rotational speed and the number of balls in a bearing from vibration data.

Spectral information obtained from the Fast Fourier Transform (FFT) of the vibration data is used to obtain these estimates and this information is used to calculate the bearing defect frequencies. The estimation algorithms have been tested using experimental data that consisted of vibration signals gathered from a transducer mounted on the drive-end bearing of an induction motor. The induction motor was operated under four different load conditions (four different running speeds), and three different types of single point defects (inner race, outer race and ball) were introduced into the drive-end bearing. The test results proved the algorithms to be very reliable and when integrated with an envelope detection algorithm reliable fault detection and diagnosis were obtained. Refer to [15] for more details. These core capabilities are well suited to be implements in a distributed agent-based architecture.

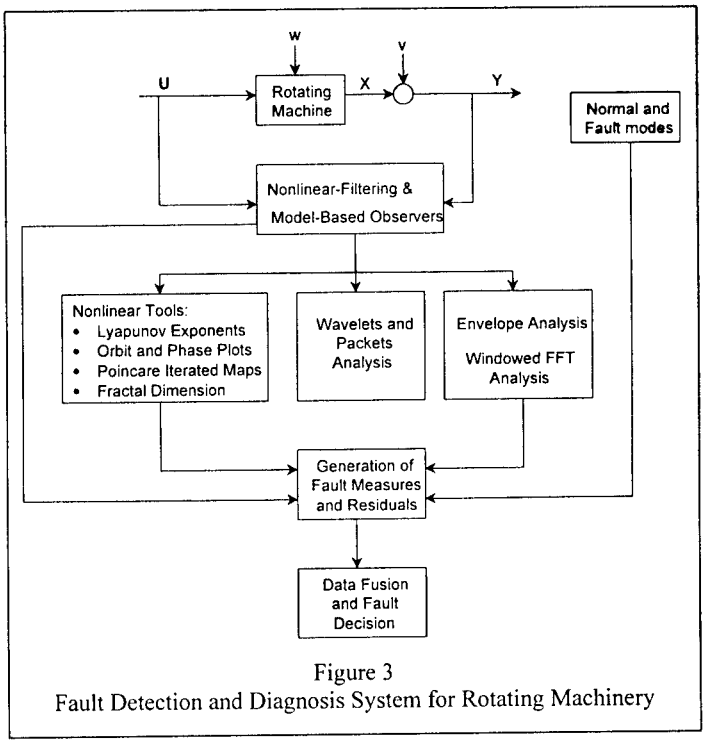


Figure 3
Fault Detection and Diagnosis System for Rotating Machinery

Agent technology expands the notion of distributed computing which may partition a problem to distribute the computation load to one in which the solution method is both localized (autonomous) and also collaborative and adaptive (cooperative/goal oriented). Within the collection of software agents each will have a local goal or agenda. In this sense, each module is autonomous. In addition, there is an overarching goal or system objective that each individual module must accommodate and through the collective efforts of the multiple agents. The collection of such agents is termed a Holonic System. This is a concatenation of the term *Holos*, meaning total or whole, and *on*, as in a neutron or elemental body [16][17].

The multi-agent approach provides an extremely powerful framework for integrating partial solutions into more complex and more sophisticated diagnostic, decision making and control structures [18][19]. This concept is particularly beneficial to machinery diagnostics and prognostics. For example, remote intelligent sensor nodes may efficiently monitor critical system components such as bearings. In the event an excessive vibration level or temperature level is observed, information regarding the degraded operation and reduced component lifetime may be relayed back to a central information system. In addition, specific data on abnormal operation may also be exchanged with neighboring smart sensor nodes. This will permit the nodes to collaborate and each to exchange data and analysis to jointly establish a more accurate, complete hypothesis of the root cause of the fault, such as a bent shaft. This will also permit maintaining a more robust and accurate system model essential for accurately predicting the future operating state of the machinery and estimated time until failure. Wireless technologies will permit the widescale deployment of smart sensor nodes across many system components. This will lead to more accurate system models and prognostic estimates at a much lower cost. It will also enable very flexible and easily reconfigured monitoring and diagnostic system.

The availability of complete and accurate process information and superior failure prediction accuracy directly addresses the key concerns expressed by a broad range of major manufacturers to know 1) overall process health and 2) accurate prognostic information. The integration of this new, accurate information into existing control, information systems, plant monitoring, scheduling, and maintenance system will provide unprecedented levels of plant performance and economic value from installed equipment.

V. SUMMARY

The developments described above provide new and important capabilities for reusable software and hardware modules. International efforts toward standards and open system specifications will provide new opportunities for intelligent sensors and distributed smart adaptable sensor nodes. In this infrastructure, model-based and observer-techniques implemented on smart sensor nodes may be readily integrated into a broad range of critical manufacturing processes. We anticipate future low-cost wireless solutions to further propel the deployment of distributed sensor nodes. With an infrastructure to effectively integrate the massive, parallel, distributed computing power there will be a new era of monitoring and managing even the most complex systems. These new, more powerful tools will form the cornerstone of the new Maintenance Management Paradigm.

VII. REFERENCES

The work of K.A. Loparo was supported in part by EPRI, Palo Alto CA., CAMP and First Energy Corp., Clevland OH, the US Office of Naval Research under agreement #:N00014-98-3-0012 and the National Science Foundation under Grant # ECS-9906218.

[1] Discenzo, F. M., DelVecchio, P., Schaefer, R., Tompkin, E.; *First Intelligent Motor Design Conference*, Conference Proceedings, January 17-18, 1996, Cleveland, Ohio

[2] O'Mara, "Designing an IEEE 1451.2 – Compliant Transducer", Sensors, August 2000, Vol. 17, No. 8, pp. 46-51.
[3] Henry, M. "Self-Validating Digital Coriolis Mass flow Meter", Computing & Control engineering Journal, October 2000, ISSN 0956-3385, pp. 219-227.
[4] Wood, G., "UK Activities in Measurement Validation and Data Quality", Computing & Control engineering Journal, October 2000, ISSN 0956-3385, pp. 214-218.
[5] See www.mimosa.org
[6] F. M. Discenzo, W. Nickerson, C. E. Mitchell, K. J. Keller, *Open Systems Architecture Enables Health Management for Next Generation System Monitoring and Maintenance,* Development Program White Paper from Open System Architecture for Condition-Based Maintenance,
[7] Manges, W., et.all., "Intelligent Wireless Sensors for Industrial Manufacturing", Sensors, April 2000, Vol. 17, No. 4, pp44-55
[8] Browne, J., "Wireless Sensors Leverage Bluetooth Communications Standard", Wireless Systems Design, July 2000, Vol. 5, No. 7, ISSN 1089-5566, Penton Media Inc.
[9] Agre, J.R., Clare, L.P., Pottie, G., Romanov, N., "Development Platform for Distributed Microsensor Networks", Proceedings of SPIE's 13th Annual International Symposium on Acrospace / Defense Sensing, Simulation, and Controls Conference, Orlando, FL., April 1999
[10] Clare, L.P., Pottie, G., Agre, J.R., "Self-Organizing Distributed Microsensor Networks", Proceedings of SPIE`s 13th Annual International Symposium on Aerospace / Defense Sensing, Simulation, and Controls Conference, Orlando, FL., April 1999
[11] Loparo, K. A. and Adams, M. L., *"Development of Machinery Monitoring and Diagnostic Methods"*, Proc., 52nd Meeting of the Society For Machinery Failure Prevention, April 1998, Virgina Beach.
[12] Loparo, K.A., Adams, M.L, Lin, W., Abdel-Magied, M.F. and Afshari, N., *"Fault Detection and Diagnosis of Rotating Machinery"*, IEEE Transactions on Industrial Electronics, Vol. 47, No. 5, October 2000.
[13] Lou, X., Loparo, K.A., Discenzo, F.M., Yoo, J. and Twarowski, A., *"A Model-Based Technique for the Detection of Bearing Faults"*, ICANOV2000, August 2000.
[14] Lou, X., Loparo, K.A., Discenzo, F.M., Yoo, J. and Twarowski, A., *"A Wavelt-Based Technique for Bearing Daignostics"*, ICANOV2000, August 2000.
[15] Ocak, H., Loparo, K.A., Discenzo, F.M., Yoo, J. and Twarowski, A., *"Estimating the Running Speed and Bearing Defect Frequencies of an Induction Motor from Vibration Data "*, ICANOV2000, August 2000.
[16] Koestler, Arthur, 1989, The Ghost in the Machine, Arkana Books
[17] Van Brussel, Hendrik, 1994, Holonic Manufacturing Systems, The Vision Matching Problem, First European Conference on Holonic Manufacturing Systems, Hannover, Germany, December 1, 1994
[18] Balasubramanian S., Maturana, F., Vasko D., and Lenner J., A Hybrid Autonomous Control Architecture, ISA-Tech/1999 Conference, Philadelphia, PA, October 5-8,1999.
[19] Maturana, F., Balasubramanian S., and Vasko D., An Autonomous Cooperative System for High Speed Sorting Systems, Accepted for publication and presentation in the HOLOMAS 2000 workshop on Industrial Applications of Holonic and Multi-Agents Systems.

SENSWEB: A WIRELESS SELF-ORGANIZED COOPERATIVE SENSOR NETWORK TOPOLOGY

Jeffrey N. Schoess and Sunil Menon
Honeywell Laboratories
3660 Technology Drive
Minneapolis, MN 55418

Abstract: Honeywell has been developing an exciting new Web-based architecture called *SensWeb*. This totally new approach to information sharing and decision-making for vehicle health monitoring and condition-based maintenance applications consists of numerous (hundreds to thousands of) sensors. These low-cost sensors (i.e., polymer thick film) are highly redundant in nature and have some processing capability and memory capacity. Sensors can be configured to have individualized passive or active sensor webs to increase coverage and improve overall sensor network performance. The SensWeb architecture is logically organized into three layers of processing: the individual sensor, web clusters of sensor nodes, and collections of sensor web clusters.

Key Words: Distributed sensor architecture, Power Aware, self-organized sensor, sensor web, wireless sensors

Background: NASA and all DoD agencies currently employ "walk-around inspection" as a cornerstone for performing vehicle and asset health monitoring. This means that a hierarchy of inspections is required to ensure fleet readiness and guarantee that availability requirements are met. The DoD inspection schedule includes daily inspection, phased maintenance based on operating time, conditional inspection based on mission and location of the DoD assets, calendar-based inspection, and limited on-board vehicle monitoring.

Some of the key issues related to vehicle and asset health monitoring include:

- Key Damage/Failure Modes—Fatigue cracking due to thermomechanical cycling, impact damage and delamination, thermal overstress, thermal fatigue, or corrosion.
- Damage Assessment—Interpretation of damage in composites is challenging due to lack of adequate sensor coverage and sensor reliability.
- Harsh Environmental Exposure—The operating environment has a direct impact on sensor reliability, calibration, and operation.

Technical Approach: This section summarizes the key technical approach for SensWeb [1,2]. A quick summary is provided, including an overview of how SensWeb can offer the benefits of sensor array construction without adding significant weight or volume. SensWeb consists of numerous sensors (hundreds to thousands). These low-cost sensors (i.e., polymer thick film) are highly redundant in nature and have some processing capability and memory capacity. Sensors can be configured to have individualized passive or active sensor webs to increase coverage and improve overall sensor network

performance. Figure 1 shows the SensWeb data flow hierarchy. The architecture is logically organized into three layers of processing: the individual sensor, web clusters of sensor nodes, and collections of sensor web clusters. Since a large amount of data is collectively generated by the sensor nodes, the information processing approach is that of progressively increasing the algorithmic complexity when progressing from the individual sensor to the sensor node cluster and, finally, among sensor node clusters. At the same time, the data content that is transmitted through the SensWeb layers is decreased due to local preprocessing performed at the sensor level. In this way, the sensor nodes can effectively reduce the data communication bandwidth and improve power management capabilities.

Low-Cost Polymer Sensors: Using polymer thick-film technology (PTF) and elastomer materials, Honeywell is developing a family of low-cost sensor-on-film technology capable of sensing temperature, moisture, vibration, structural impact, and strain quantities. These sensors conform to surface profiles (6 to 10 mils thick), adding little weight, and can be easily replicated to provide a deeply distributed and highly redundant sensor web architecture solution. The PTF approach is based on the novel idea of directly integrating sensory, control, and data processing electronics into the coating protectant system of interest (vehicle, asset, spaceborne structure, etc.). The polymer sensory system is proposed to conform to the shape of the surface into which it would be integrated, or in other words, to be "conformal," which means to "have the same shape or contour."

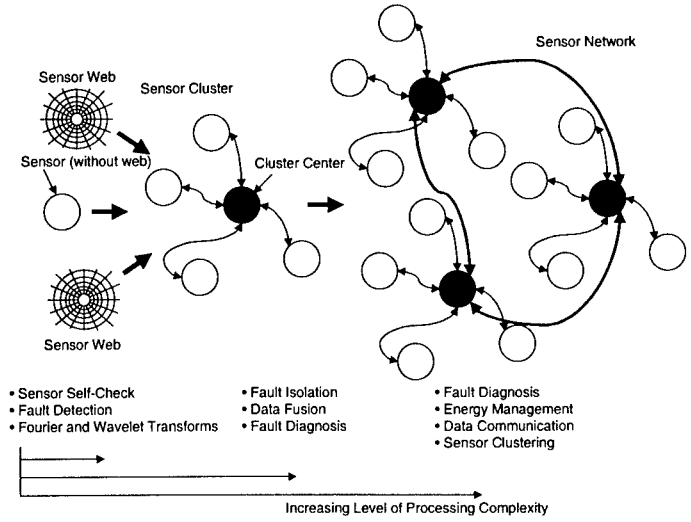


Figure 1. SensWeb Data Flow Diagram

Honeywell has been working under internal research funds for 2-1/2 years to develop conformal sensing, which integrates polymer films with PTF-based "built-in" sensory

functions. The technical approach is based on the novel idea of using a polymer film as a flexible substrate, on the backside of which electrical interconnects, sensory functions, and data processing electronics would be directly integrated. The sensory functions are defined by incorporating polymer thick-film patterns on the film surface bonded to the surface of the test structure.

Figure 2 illustrates a concept for moisture/corrosion sensing as an edge seal detection scheme. The figure shows a film panel peeled back to reveal a set of PTF sensing elements located on the backside of the film. The polymer sensing elements are organized as a linear array to detect the integrity conditions of the structural panel edge seal. The linear array elements are shown in the detailed view of the figure and positioned near the edge of the panel to detect penetration of moisture and fluid ingress. Each array element is designed as a "built-in" sensory function to detect the presence of moisture ingress from the edge of the panel as a conductivity measurement. The array element is organized to sense moisture based on a unique polymer film circuit pattern, which is printed on the backside of the film using standard inkjet processing techniques.

Honeywell has performed additional internally funded development that demonstrates the feasibility of using PTF sensing to perform functions for impact force (i.e., barely visible impact damage, or BVID) and conformal antennas with the potential of also sensing airflow (i.e., air data), temperature, and vibration parameters.

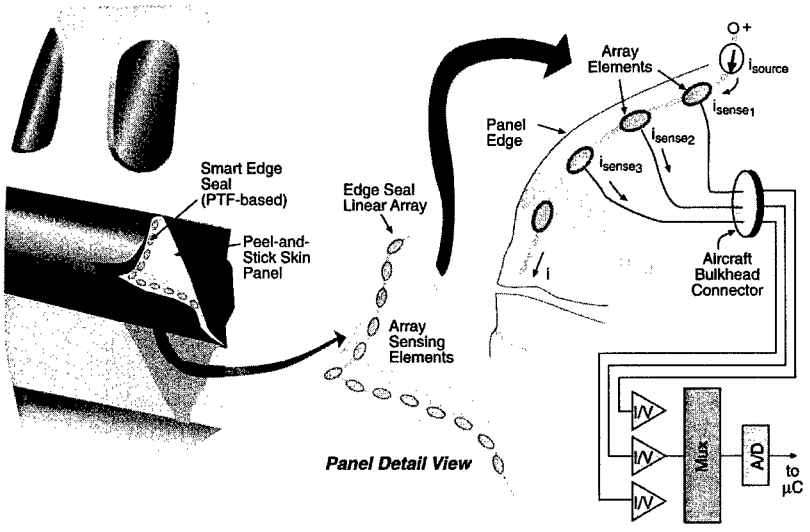


Figure 2. Smart Edge Seal Concept

A key benefit of the polymer sensor array is its multifunction sensing. The polymer circuit pattern implemented for moisture/corrosion sensing is also capable of sensing impact forces caused by maintenance-induced damage or operational servicing. To provide sensing for impact forces, the linear sensor array is configured with an additional semiconductor polymer layer, as shown in Figure 3. The design approach is set up to operate as a force-sensing resistor (FSR). An FSR operates on the principle of converting force applied via a structural impact event to an equivalent voltage output. As pressure is applied to the sensor pattern, individual pairs are shunted, causing a decrease in electrical resistance.

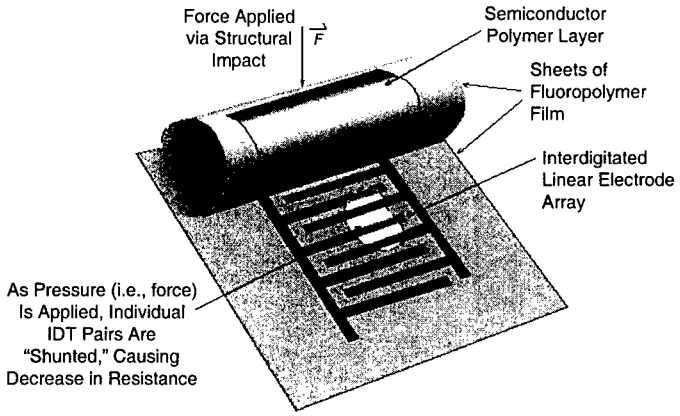


Figure 3. Force-Sensing Resistor (FSR)

Sensor Web Topology: An exciting area of research for asset monitoring applications currently being investigated by Honeywell is sensor web architecture topologies. The term "sensor web architecture" refers to the methodology by which sensor resources are allocated and physically organized in monitoring applications. A smart structure/vehicle element is illustrated in Figure 4. The figure highlights a conceptual view of several sensor webs distributed uniformly on the surface of the smart structure. Each sensor web would be implemented to sense structural parameters of interest, such as vibration or acoustic emission, when embedded into the structure or structurally bonded as a conforming polymer skin.

The advantages of having this kind of sensor distribution proposed in SensWeb are:

- Coverage of a large area by unifying the coverage areas of a multitude of individual sensors.
- Greater sensitivity to system faults by using all the available collective sensor information.

- The information resolution available is very broad—from raw data from single sensors to fault information about a region covered by a group of sensors to whole system information.
- The algorithms for fault detection, diagnosis, and isolation can be specialized at different locations. This allows for a more powerful system analysis methodology.
- Different sensors can be distributed as part of the common sensor network. Data fusion algorithms can process data from different sensor types, and more accurate fault diagnosis is possible.

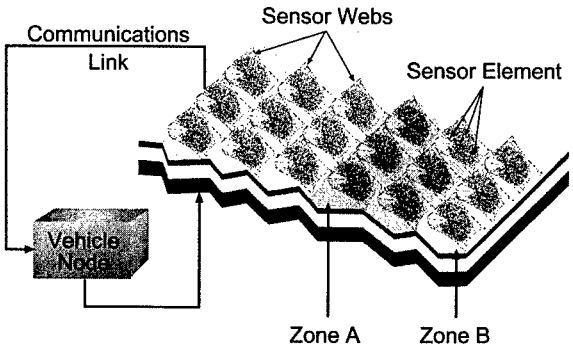


Figure 4. SensWeb Topology

A specific focus of Honeywell's recent research in sensor web technology is investigating alternative methods for constructing sensor layouts to provide optimal sensor coverage and performance. The unique construction pattern of a "spider web" offers several significant benefits to a potential user, including:

- Improved Sensor Coverage—The centralized hub web design implemented by a spider allows much-improved coverage. The typical "orb-style" design facilitates detection of a friend or food for the spider by using the web as a "sensor antenna," as the vibration signature transmitted via the web is detected as a unique signature. The same principle of operation could be applied for structural monitoring applications where a dedicated polymer sensor element could be located at the center or hub of a polymer-like sensor web to detect structural vibrations, achieving 100% coverage at minimum cost.
- Best Sensor Mapping—The radial design web structure provides an optimal method for mapping sensor resources on complex shaped surfaces.
- Power-Aware Wireless Communications—PTF sensors are envisioned to communicate and share data via seamless peer-to-peer communications. Bluetooth and related wireless network protocols will be implemented to provide key benefits of wiring weight savings, reduced sensor installation and maintenance costs, and

sensor placement optimization. Power-aware architecture techniques (i.e., clock gating, frequency scaling, power vs. performance constraints, sensor correlation, virtual sensing) will be incorporated to adaptively turn power on and off according to processing needs and minimize power to complete mission tasks. To accomplish this, power and energy will be treated as independent system-level variables to be optimized via real-time assessment.

Conclusions and Summary: The details of sensor web architecture design and polymer PTF sensors have been presented. The SensWeb architecture is logically organized into three layers of processing: the individual sensor, web clusters of sensor nodes, and collections of sensor web clusters.

References:

1. J.N. Schoess, "Aerospace Applications of Smart Materials: A Sensing Perspective Using Conductive Polymer Sensor Arrays," *Encyclopedia on Smart Materials* (1st ed.). New York: John Wiley & Sons (2001).

2. J.N. Schoess, Sunil Menon, "Conductive Polymer Sensor Arrays—A New Frontier Technology for CBM," to be published, 55th Meeting of Machinery Failure Prevention Technology (MFPT), April 2001, Virgina Beach, VA.

INDUSTRIAL CASE HISTORIES

Presented by: Nelson L. Baxter, ABM Technical Services, Inc.
Kevin R. Guy, CJ Analytical Engineering, Inc.

Industrial Case Histories

Kevin R. Guy

C J Analytical Engineering, Inc
R.R. #1 Box 353
Francisco, Indiana 47649
cjinc@cjanalytical.com

Abstract: This paper is a series of case histories encountered over twenty – two years of performing vibration analysis. While each case history is not necessarily outstanding in its own right, they do show the type of problems one sees in an industrial environment. The paper will cover cases that are fairly routine to in-depth problems that required rotor dynamic modeling, structural modeling or both. Equipment from paper machines to turbines and pumps (vertical and horizontal) are presented. Both sleeve bearing and rolling element bearings machines are discussed.

This paper is an attempt at providing a resource of information with examples for the inexperienced analyst. The goal of the paper is to help point the analyst in the right direction for the analysis or to jog their memory to look further into the problem. Each history will include a brief discussion on what equipment was required for the analysis. In today's proactive maintenance society, it is thought the data collector is the all encompassing equipment for analysis, in many of the cases, mention of equipment required for each analysis will helpfully dispel this theory.

Key Words: Imbalance; Blade Pass; Whirl; Whip; Structural Resonance; Variable Frequency Drives; Bearing Defect Frequency; Gearmesh

Case #1 – Balancing of 9.5 Mw Turbine after a Catastrophic Failure

Problem: This unit was repaired after suffering catastrophic shut down during a hurricane. The turbine is owned by a government utility on a small island nation in the West Indies. The utility had utilized diesel generators until a far eastern country built their first steam plant. The utility ran into problems during Hurricane Hugo. Operations personnel not being experience with steam turbine failed to maintain the DC battery system. When the storm struck the island; the electric distribution system was destroyed causing the turbines to trip. The turbine generator train (Figure #1) went from operating speed to a dead stop in 45 seconds without lubricating oil. The turbine journals and bearings were destroyed. The bearings in the gearbox and generator were also completely wiped; however, there was no damage to the shaft journals.

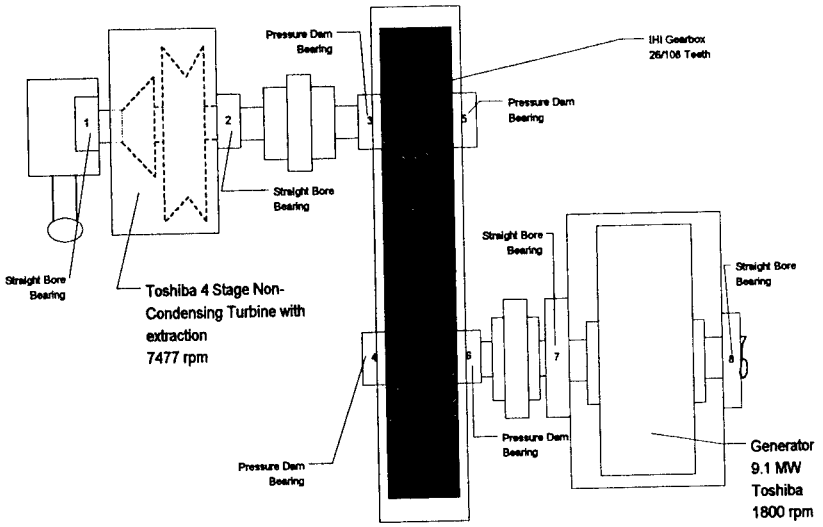


Figure #1

The unit was pre-warmed and run at slow speed. Once the slow rolling was completed and the oil system checked out, the turbine was to be brought up to speed. During this initial run the unit tripped on high vibration on the number one turbine bearing. The unit tripped below the first critical. The first critical is at 3400 rpm. The turbine supervisory system uses units of mm/100. All analysis and balancing data was collected utilizing casing mounted (magnet) velocity sensors. The turbine rotor sat in wood cradle for eighteen months after the journal repairs.

Symptoms: Vibration data indicated all the vibration was at running speed (1X). Since the unit was tripping at the first critical (first mode) and the rotor had sat for eighteen months in the cradles, there was a concern of a bow (first mode) in the rotor. The number one (turbine inlet) bearing balance ring was external to the turbine shell. This coupled with the turbine tripping on bearing #1 vibration it was decided to balance the unit with a single plane shot in the number one balance ring.

Test Data and Observations: Balance data was collected with a tracking filter and the analysis was performed with a Hewlett Packard Dual Channel FFT Analyzer. Velocity sensors were used throughout the balancing and analysis. Because no previous balance data was available; it was decided to perform a measured effect balance. This balance method was both ineffective and inclusive (figure #2).

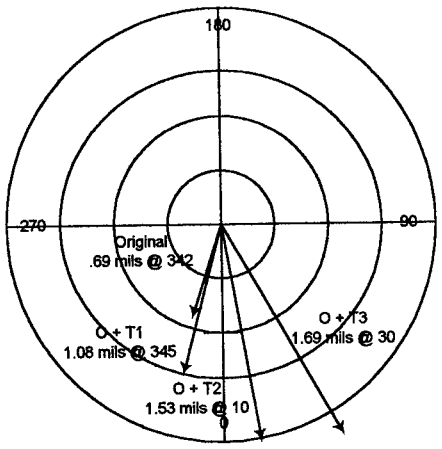


Figure #2

Since the effect was fairly decent, it was decided to again try a measured effect balance in conjunction with a four run no phase balance method. This proved to also be fruitless. Looking over the turbine schematic (figure #3) it was determined the #1 balance plane was too far from the center of the rotor (first mode) balancing on the #1 bearing balance ring.

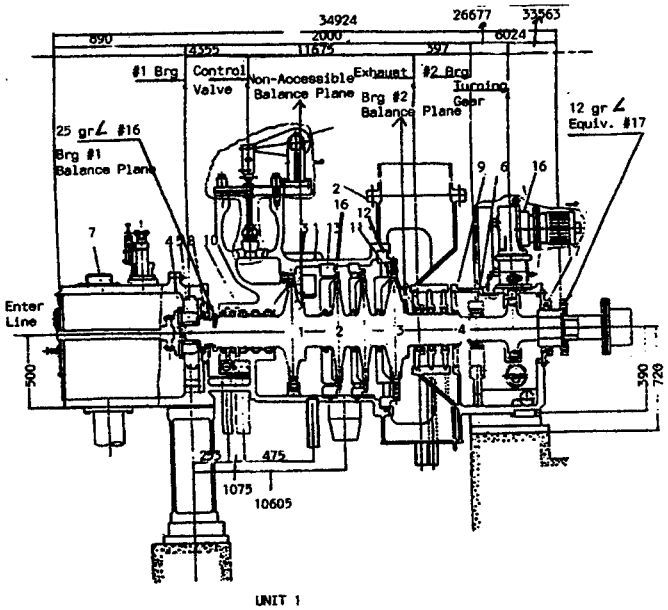


Figure #3

Corrective Action: Since the second balance plane (exhaust end) was closer to the mid span and we were trying to balance a first mode it was decided to place weights in this ring. Also, since this ring was closer to bearing #2 and the highest vibration was on bearing #1 a two plane balance program was entered. The unit was successfully balanced below the first critical; however, the unit again tripped at the second critical. A second two plane balance program was started below the second critical. During the balancing at the second critical two other problems developed. The unit suffered a rub and a bow was detected in the coastdown data (figures #4 & #5). The rub was not initially detected and several balance runs were lost to the rubbing.

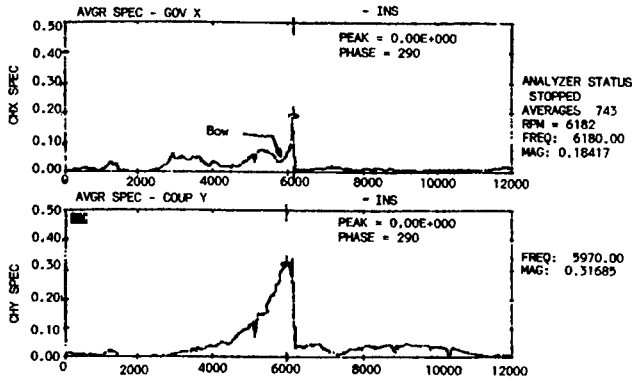


Figure #4

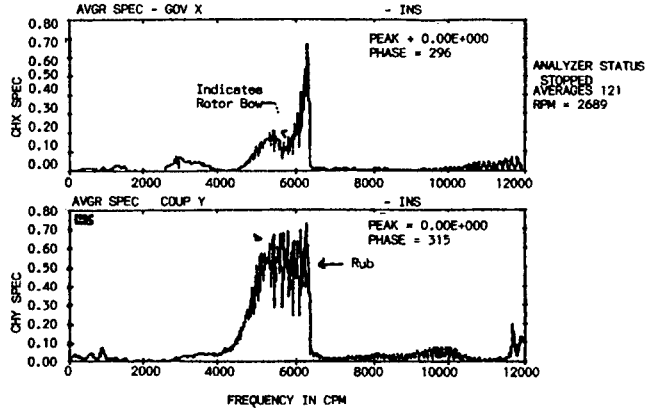


Figure #5

Results: Once the rub issue was eliminated balancing progressed without problem. The unit did have to be balanced at running speed with a two plane procedure even with balancing at the first and second critical. The ineffective results with the single plane balance on the #1 bearing really did not add much to the balance time.

The final data is listed in Table #1:

Bearing #	Control Room Data mm/100	Turbine Cap Data Mils (pk-pk)
1	1.6	.63
2	1.6	.63
3	.9	.35
4	.1	.04
5	.2	.08
6	.7	.28

Table #I

Conclusions: The balancing was hampered by several problems. The biggest problem; the bearings was designed with a clearance of three mils for every inch of journal diameter. This did not provide enough damping for the critical speed vibration. The clearances were designed over size because of not having spare parts on site and not wanting to wipe any bearings.

The rub problem was not detected early on because of data acquisition. The control system was not adequate in controlling speed to get good data during the balancing at the critical speeds. The speed would wander 200 to 400 rpm. Also, running the unit below rated speed caused the exhaust end to become hot. This in turn caused the bow in the rotor; therefore, the unit was run for very short times. While this would allow the collection of phase and amplitude data, it did not allow collection of analysis data. This caused those involved to miss the initial rubbing indications.

Case #2 – Induced Draft Fan ½ Order of Blade Pass

Problem: Vibration problems were suffered immediately upon running a new fan wheel. An Induced Draft Fan on the bag house of a cement plant was retrofitted to increase fan efficiency. This fan had a spare wheel and shaft. The new fan utilized the same shaft; however, the wheel design was changed. This fan was retrofitted with a new fan wheel using the existing shaft. The number of fan blades was increased from twelve (12) blades to eighteen (18) blades. The new shaft wheel system is approximately 250 pounds lighter then the old design. One

other difference in this new fan was the addition of a turning vane to increase the efficiency of the fan.

Symptoms: Vibration problems were occurring when the fan is cooling down from 450 degrees F operating temperature to 220 degrees F operating temperature, during the plant start-up conditions. During the start-up process, the fan gas temperatures exceed 425 degrees due to some of the process duct work being by-passed. As these other parts of the process are brought in service the fan temperatures drop to a normal operating temperature of 220 degrees. Vibration amplitudes in the control room exceed 5.0 mils during the high temperature conditions.

Plant engineers balanced the fan from above 4.0 mils (pk-pk) to below 1.0 mil (pk-pk); however, within one day the amplitudes were back up to over 3.5 mils. Since this was essentially a new fan, the owners were concerned about the reliability of the new fan. One item noted in plant operation logs was how fast the vibration dropped when the fan was shut off.

Some initial data collected by the Cement Plant indicated there might be a natural frequency close to operating speed. The fan utilizes straight bore sleeve bearings and operates at 1196 rpm (19.94 Hz). The motor has been replaced several times and due to these replacements the motor and fan pedestals have been modified extensively.

During the initial field balancing by the fan manufacturer the turning van had to be cut back to get steady phase readings.

Test Data and Observations: The test plan was to look at the possibility of a natural frequency in the operating speed range of either the shaft or pedestal. Also; operating data would be collected from the casing using accelerometers and the shaft using prox probes. The accelerometers were powered by an external integrating power supply. A dual channel FFT analyzer was used for the impact tests. A three (3.0) pound force hammer was used for the impact tests. Start-up data, coastdown data and operational data was analyzed with either a dual channel data collector or a eight channel FFT analyzer.

Impact tests were performed first on both the shaft and fan wheel. Table #II contains the information from the impact tests. The concern was a natural frequency around operating speed and one close to blade pass (358.9 Hz). While the response at blade pass looked favorable for no problems, the frequencies at 23.0 Hz were a concern.

Once this testing was completed operational data was collected. Figure #6 contains a layout of the test points.

Point	Direction	Freq.	COH.	Critical Damping	"Q" Factor	Problem
Shaft	H	23.0	.970	.097	3.4	**Yes**
	V	30.5	.983	.087	5.7	
Shroud	A	333.75	.983	.014	35.2	?
	A	385.0	.974	.009	54	?
Blade	H	340	.893	.007	74	

Table #II

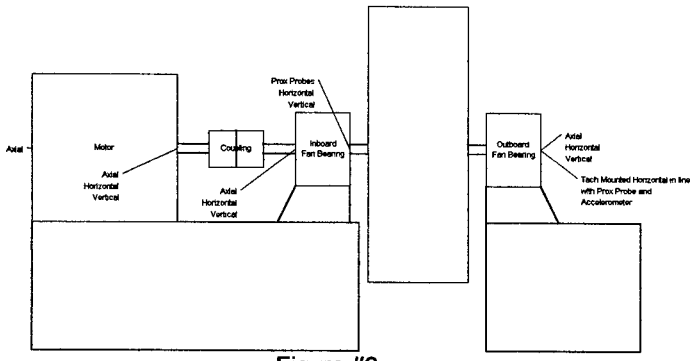


Figure #6

Phase and amplitude data showed thermal vectors as high as 6.8 mils (Figure #7). Coastdown data indicated the vibration dropped off almost immediately as the power was cut (Figure #8). This is a definite indication the fan is running very close to a natural frequency. In 52.5 cycles, the shaft vibration drops by over 76%

Analysis of the vibration data indicated there was an alignment issue on the fan bearings. Vibration data indicated the fan bearing to bearing alignment was off while the shaft to shaft alignment was correct. The axial vibration on the fan bearings was one-half the radial vibration; however, the axial vibration on the motor bearings was acceptable.

One other item that showed up in the vibration data collected during operation was the presence of a ½ X of blade pass (figure #9). When an orbit analysis was performed this 1/2x of blade pass also was present.

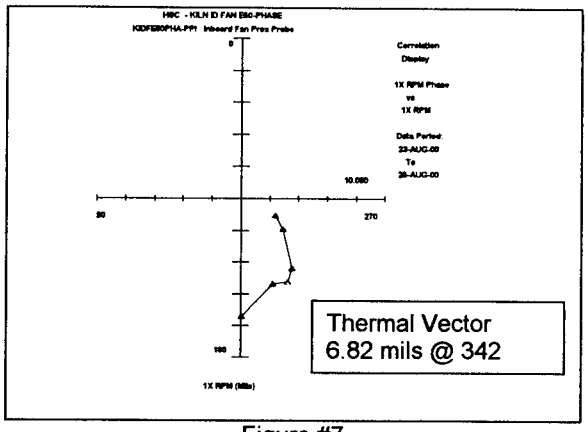


Figure #7

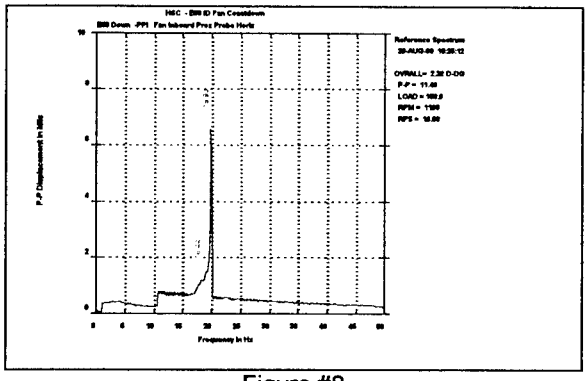


Figure #8

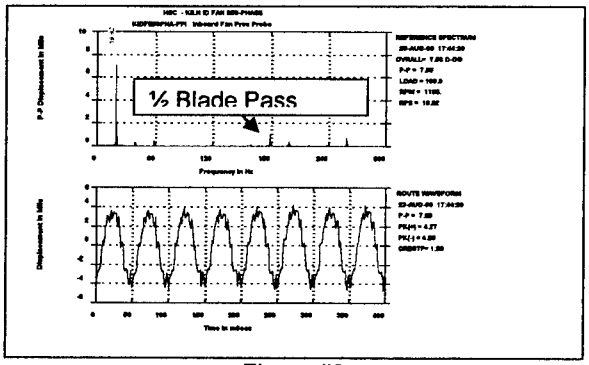


Figure #9

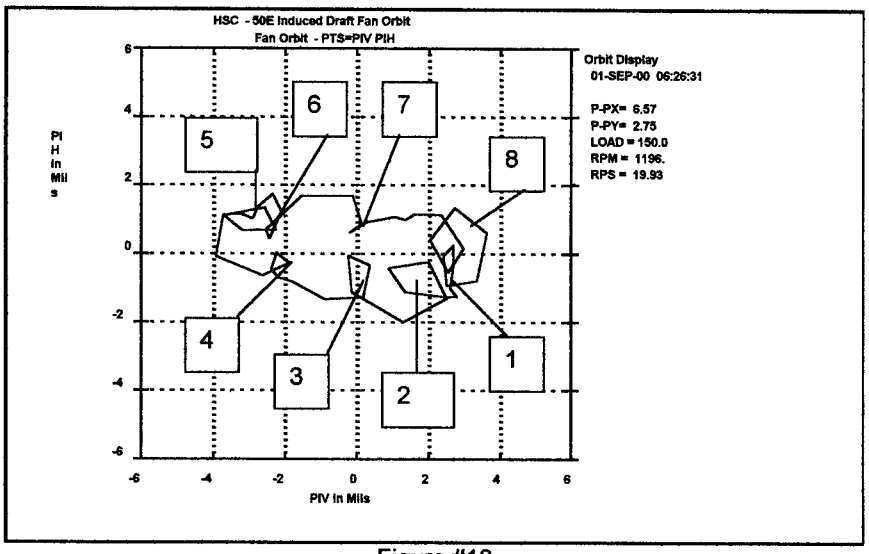


Figure #10

The internal loops indicate a forward whirl of the shaft. The frequency of the vibration is found from the following formula:

$$Frequency = \left[\frac{1}{n+1}\right]$$

n = number of internal loops

This means the period of the vibration is .111 (1/9). The frequency of the vibration is the whirl period divided by the shaft period or .111 divided by the shaft period (.0502). The period of the vibration is .0056 sec or 179.5 Hz. This frequency is one half blade pass.

Corrective Action: Due to the past history of balance problems with the turning vane, it was recommended to removed it. Also, recommended was to balance the fan with prox probes. The shaft vibration was 1.5X to 2X higher then the casing vibration. All previous balancing was only performed with seismic readings.

Improving the balance condition would lower the thermal effect due to the temperature extremes incurred from start-up to normal operating conditions.

It was also strongly recommended the fan bearing to bearing alignment be corrected as soon as possible.

Results: Vibration amplitudes dropped significantly when the balancing was completed. The alignment correction lowered the axial vibration. The plant eventually removed the turning vane. This removed the ½ blade pass vibration and cleaned up the orbit.

Conclusions: The fan was definitely running on a critical speed. Since all the operational data and transient data was recorded on a sixteen channel digital tape recorder, bode and nyquist plots were made of the coast downs. This data showed the fan was rolling off a critical speed.

Improving the balance condition and correcting the fan bearing to bearing alignment lowered the 1X forcing function reducing the vibration. This improved the vibration to the point, the thermal effects of the temperature excursions during start-up became a non issue. It should be noted that during the site data collection a production supervisor commented the thermal effects were present from the first day he started with the company and he was employed at this site for twenty – five years.

Also, this fan was always running close to a critical speed. The new shaft and fan wheel was 650 pounds lighter then the original rotor. This reduction in weight was not enough to change the shaft critical significantly; therefore, the fan has always operated close to a critical speed.

Finally the 1/2 Blade Pass was caused by a force on the fan wheel generated by the turning vane.

Case #3 – Cooling Tower Blade Pass Resonance

Problem: This unit had a history of running high vibration amplitudes in the 20 Hz frequency range. The operating speed of the "A" Cell Fan was lowered to 2.65 Hz from 3.33 Hz. This was done per recommendations from Cooling Tower Consultants who also changed the pitch of the fan blades to improve efficiency. The "A" Cell motor has an input speed of 1790 rpm (29.83 Hz) and the gearbox lowers this to 2.65 Hz. The blade pass frequency (6 Blades) is 15.9 Hz. All the vibration of the cooling tower is at the blade pass frequency. This has been the history of these fans.

Symptoms: Plant personnel had performed a shaker test on the "A" cell and found vibration response was high in the 15 - 17 Hz range (Figure #11). The dominant vibration frequency from the data plots (motor to gearbox) is blade pass. This is found in the axial, horizontal and vertical directions. Once the review of the past data was completed a test plan was developed.

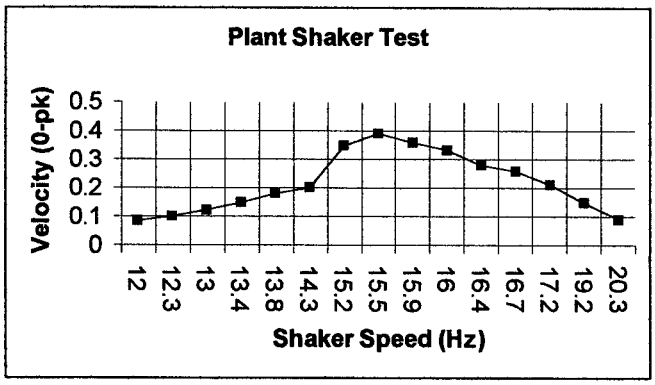


Figure #11

Test Data and Observations: A complete set of vibration data was collected with a dual channel data collector (Figure #12). Impact tests were performed with a dual channel FFT analyzer. A three (3.0) pound instrumented hammer was used for excitation. Along with the impact tests; another set of shaker tests were performed throughout the support structure of the motor, gearbox and fan system (Figure #13).

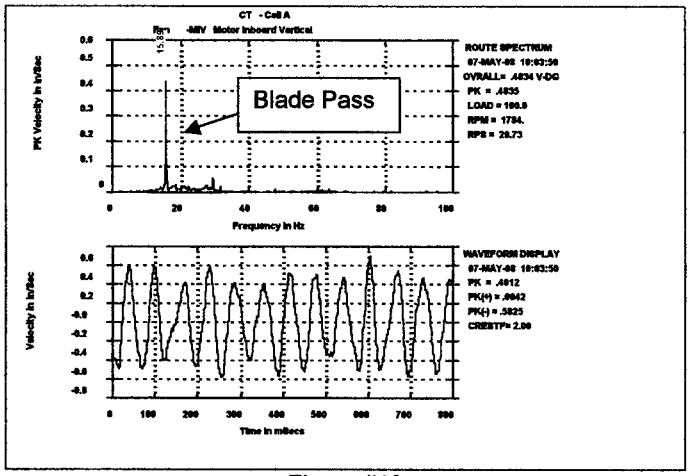


Figure #12

The Shaker tests indicated distinct natural frequencies around blade pass Figure #13). The amplification (Q Factor) from the tests were from 5.3 to 10.5 plus. Impact test (Figure # 14) and Shaker Test results can be found in Table III.

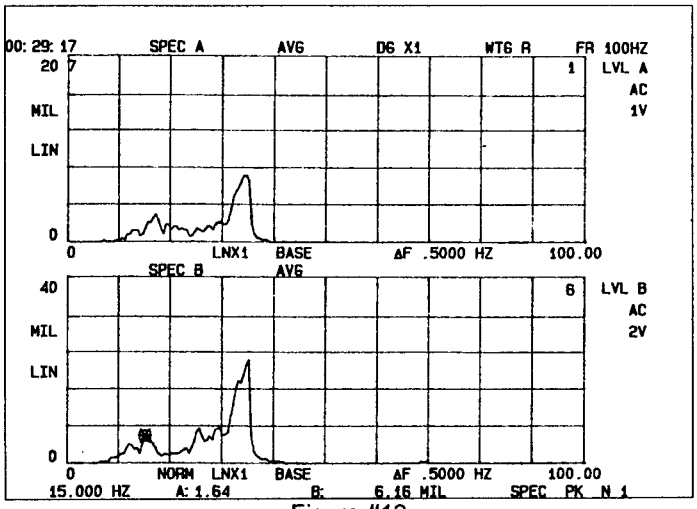


Figure #13

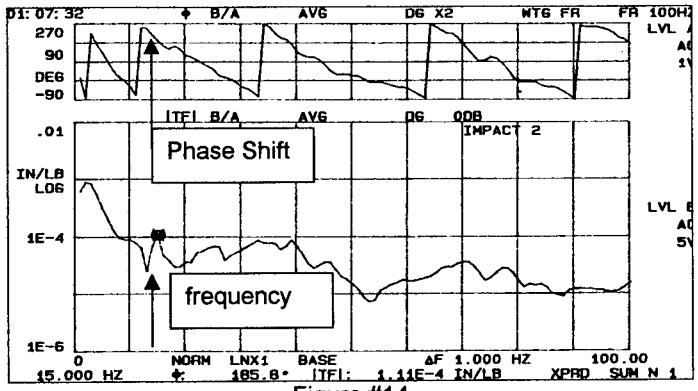


Figure #14

	South Beam (H)	South Beam (V)	North Beam (H)	North Beam (V)
Impact	12.0	15.0	12.0	15.0
	16.0	25.0	20.0	26.0
	19.0			
	26.0			
Shaker	15.0	30.0	12.5	15.0
	30.5		30.0	

Table III

Corrective Action: The data definitely indicates a natural frequency excited by the blade pass frequency of the fan. The vibration problem being experienced is caused by the six blades on the fan exciting the natural frequencies.

The natural frequency of a system is based on the mass and stiffness of the system. No one component is usually the fault; but, rather all the components of the system tied together cause the problem. Breaking the system up (detuning) by placing a neoprene material between the joints of connected parts may break the system up and eliminate the problem, however, this may or may not work. The final step would be a redesign of the system by replacing the six blade fan with a fan of five or seven blades.

Because this is a natural frequency problem the solution is to remove the forcing function which is the blade pass. Figure #15 shows an interference diagram for the different pieces of the system tested by impact hammer and shaker. The plot clearly shows that a seven blade fan operating at 2.65 Hz will have a blade pass frequency that will fall between natural frequencies and will have little or no excitation from the blades.

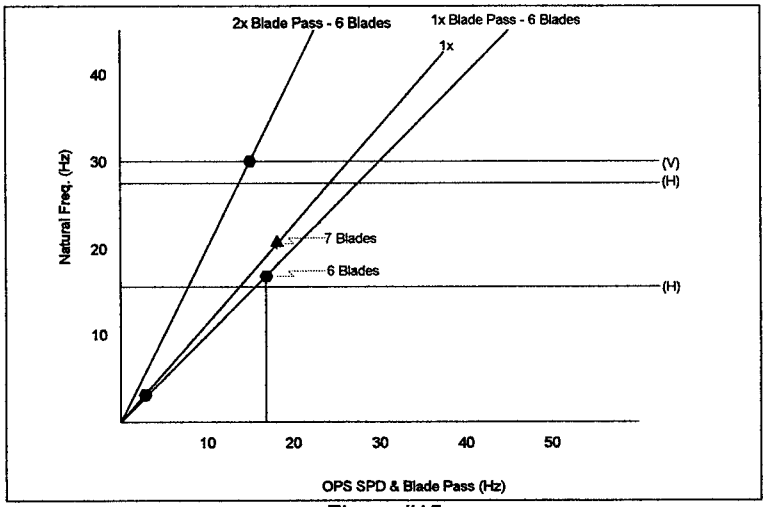


Figure #15

Since the fan has operated for so long (over eighteen months) with high vibration and no significant problems have been experienced the best option maybe to run as is and not change anything in the system. Based on the fact that no serious problems have occurred in the present condition the recommendation was made to run the cooling tower without any modification. If problems arise in the future; Febreeka should be placed under the gearbox and motor to detune the system,

lowing the vibration. This detuning of the system should be done while a study on the effect of efficiency due to changing the fan from six blades to seven blades is performed. The change to seven blades is the long term fix.

Results: The unit was run for several months until the vibration amplitudes could no longer be tolerated. The system was detuned by installing Febreeka between the gearbox and mounting frame. This detuning provided enough damping so the fan could be run while a study on the effect of efficiency due to changing the fan from six blades to seven blades was performed. This study was completed and the fan was change to seven blades and now operates without problems.

Conclusions: The changing of the system operating speed; having been lowered to 2.65 Hz from 3.33 Hz, was the cause of the problem. This was done per recommendations from Cooling Tower Consultants who also changed the pitch of the fan blades to improve efficiency. However; no one looked at the possibility of excitation from the blade pass. Essentially the plant took a good running system and created their own problem. If the time had been spent performing resonance checks before making the change, this problem would have been identified and avoided. The ironic part of this case is plant engineering wanted to perform the resonance checks; however, they were deemed unnecessary by the corporate engineers.

Case #4 – Shaft Critical Primary Air Fan

Problem: This plant has had a history of recurrent vibration problems on their primary air fans. Looseness problems show up several times per year. The fan is an overhung design (figure #16). The replacement of the straight bore babbitt bearings is followed by a balancing of the fan using seismic readings. The fan is easily balanced below 1.0 mils (pk-pk). However; the vibration is always back into alarm with in a few weeks.

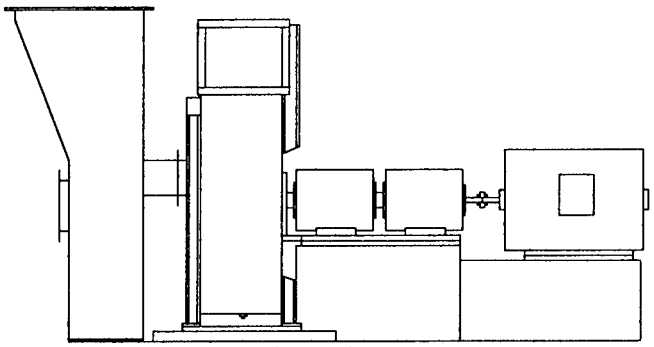


Figure #16

Symptoms: The symptoms remain consistent. The spectrum shows multiples of 1X and the timewave shows clipping (Figure #17).

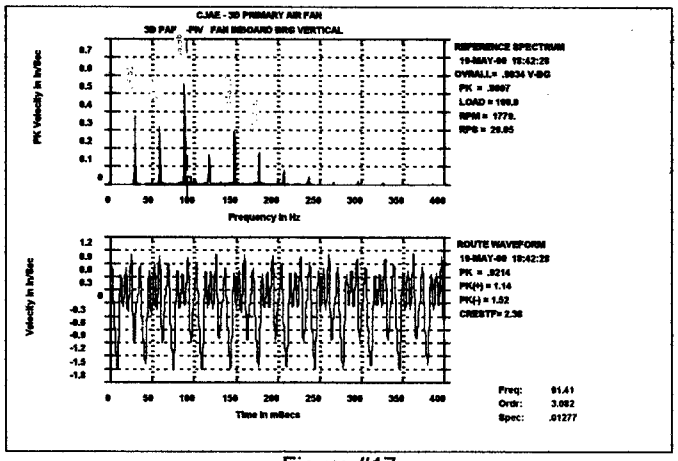


Figure #17

Test Data and Observations: This unit is on a monthly monitoring schedule; so it's easy to see when problems are developing. Once the looseness is apparent, the adjustment nut on top of the bearing has become loose. This nut is tightened down until the looseness is eliminated. This unit went through several cycles of the bearing nut becoming loose and having to be readjusted. The day after the last adjustment the bearing (outboard) closest to the fan failed. Failures of this nature have also occurred in the past. The failure required the bearing liner and bearing housing to be replaced. Data collected immediately after the fan was put back in service indicated the fan needed to be balanced (Figure #18).

The fan was balanced with seismic sensors. This has been the balancing procedure since the fan was put in operation in the early 1980's. The fan was balanced in one shot using past sensitivities and lag angles. Vibration levels were below .80 mils (pk-pk) at operating speed (Figure #19).

While the casing reading were more then acceptable, the cooling lines for the bearing were visibly moving. Vibration readings were taken on the cooling lines. Amplitudes were as high as 1.0 in/sec (0-pk). All the vibration was at the running speed of the fan. The first thought was the cooling lines could be resonant. The concern of the vibration analysts was the shaft vibration. Could the shaft vibration still be high even though the seismic readings were acceptable.

Prox probes were externally installed to read the shaft vibration between the bearings. Shaft vibration amplitudes (Figure #20) were still 15.0 mils (pk-pk) plus.

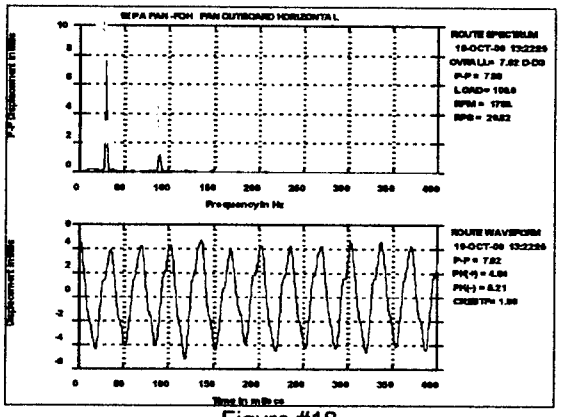


Figure #18

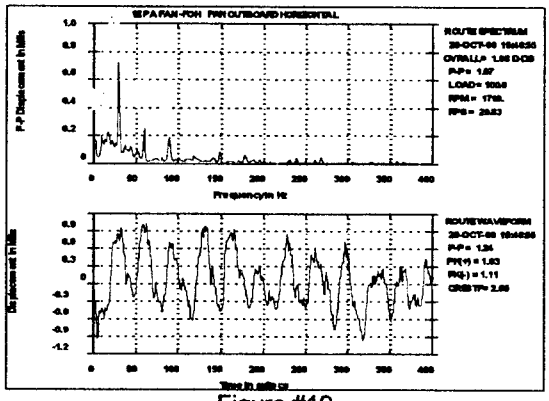


Figure #19

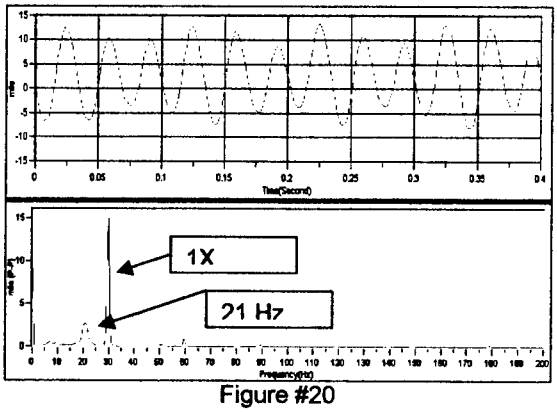


Figure #20

Also, the data showed a frequency at 21.0 Hz (1260 cpm). It was decided to collect a bode plot on a coast down (Figure #21). The bode plot shows a critical speed just below 1200 rpm. It should also be noted it appears the fan rolls off a critical as soon as the power is cut.

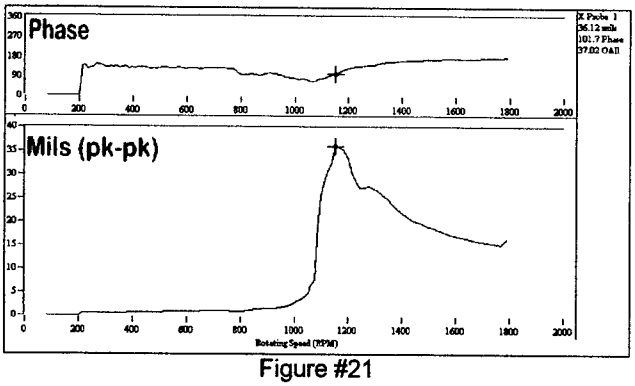


Figure #21

Corrective Action: Since this plant has six fans that behave this way, all the fans will be balanced using prox probe readings. It was recommended to perform an in-depth analysis of the fan bearing - pedestal system. A modal analysis of the fan movement in the axial, horizontal and vertical directions will be performed. Impact tests of the shaft and bearing caps will be done for validation of a rotor dynamic computer model of the shaft and bearing system.

Results: The modeling and shaft rotor dynamic study are presently in progress. Initial indications are the fan is running on a critical speed and the vibration is also effected by the critical speed just below 1200 rpm. Modifications to the bearing may provide the needed relief to allow for long term reliable operation.

Conclusions: The long term problems experienced with these fans are the result of bearing design. Balancing with seismic data was fruitless because the bearing liner is not in contact with the bearing housing. The bearing liner rests on a vertical support that has no horizontal support even though the bearings are held in place by the torque nut at the top of the bearing cap. The overhung design of the fan puts no load on the inboard bearing. This is evident the inboard bearing rarely needs changing and seldom shows any wear pattern.

Case #5 – Coupling Lockup Generating Gearmesh

Problems: Random occurrences of vibration started occurring throughout the load range of a turbine driven boiler feed pump. Vibration problems were experienced on both the turbine drive and the pump. The turbine drive is a six stage 16,700 horsepower turbine and the feed pump is a six stage double

suction pump. The normal operating speed range of the system is 4000 – 5200 rpm.

Symptoms: The vibration occurrences had been monitored for a couple of weeks before they became steady. Once they became steady, the vibration occurred throughout the load range. Since the vibration amplitudes were high on both the pump and turbine it was felt there may be a problem with the coupling. Past documented information showed the coupling was very good at isolating the vibration into the pump or the turbine not allowing any cross effect.

Test Data and Observations: The initial request was to find if the unit could be relied upon for load sales the next day. Vibration data collected on the turbine indicated a ratty vibration spectrum with most of the vibration below running speed. Figure #22 shows the outboard pump bearing with over 1.40 in/sec of vibration at 3390 cpm. The inboard pump bearing had high vibration (>.75 in/sec) at 2900 cpm. This data indicate two possible problems. First the 3390 cpm frequency was a problem caused by rubbing of the rotating and stationary balance components of the pump. This was determined from past documented data on boiler feed pump vibration problems. Second, the spectrum from the inboard end of the pump indicated a frequency that was too low to be a rubbing problem in the pump.

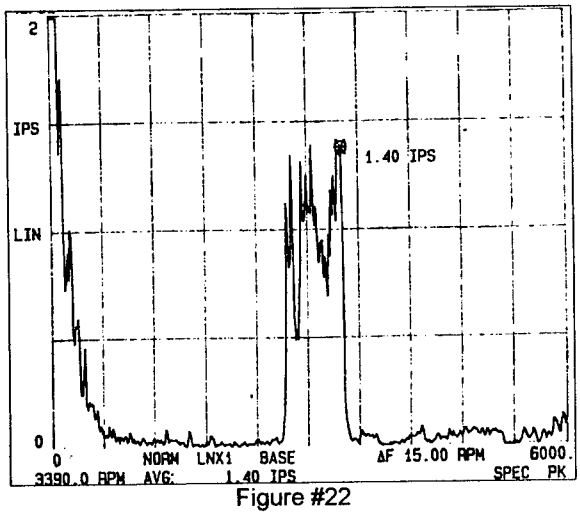


Figure #22

Next vibration data was analyzed from the turbine. Figure #23 is the outboard end of the turbine. The data is take off the shaft riders installed on the TSI system. The data shows ratty sub synchronous vibration. The peak amplitude at 2940 cpm has side bands at + and – 720 cpm. The 720 frequency is ten times

the number of teeth on the coupling hubs. It was concluded from this data that the pump and coupling both had serious problems.

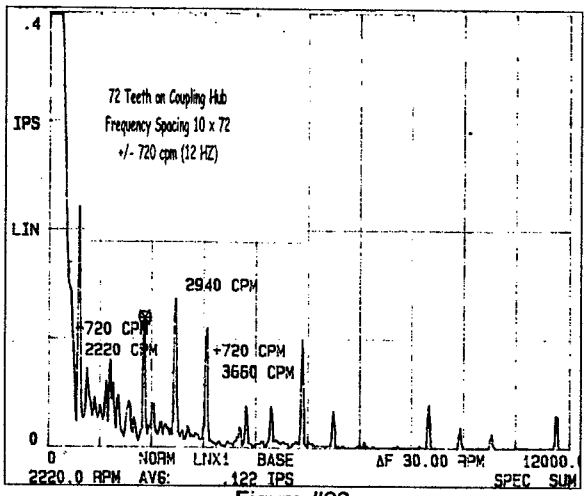


Figure #23

Corrective Action: It was recommended that the unit be shutdown and the coupling be inspected for wear. The total thrust of the pump was also to be measured. This would indicate if the pump and coupling were the problem or if further analysis would be required.

Results: The coupling inspection found severe wear and it had to be replaced. The thrust check found the pump had severe wear and would have to be rebuilt. This pump had been in operation over eleven years and was the last original pump in the plant. The coupling had been in service for at least seven years and may have been the original.

Conclusions: The documented data from past pump rubbing problems led to a fast analysis of the pump problem. The coupling problem was the result of a long operational life. If better maintenance had been done on the coupling, it may have been replaced at an earlier date. When the pump was put back in service it ran with vibration levels below 1.25 mils. The pump has operated without problems since this repair.

It is felt that the rubbing in the pump was the result of the coupling becoming locked in one position pulling the pump into the stationary balance components. If the coupling had not been worn, it may have allowed the pump to run another year; however, the pump repair showed we would have had to perform an overhaul within the next twelve months.

Case #6 – Oil Whip 150 MW Turbine Generator

Problem: This unit returned to service in November 1999 following a bearing inspection and realignment of the turbine generator set. The unit was subsequently balanced with all bearings below 2.7 mils (pk-pk). Vibration was very stable over a 48 hour period of operation; however, vibration started a steady climb while holding steady at 90 megawatts. Vibration amplitudes increased to over 7.0 mils (pk-pk) on both LP bearings. This symptom had been experienced in the past on this unit; however, no data was collected during these excursions.

This unit is a General Electric (Model D5R) 150 Megawatt Turbine Generator operates at 3600 rpm. The HP/IP turbine is opposed flow with a single LP rotor. A gearbox drives the exciter at 1200 rpm. This unit is instrumented with shaft riders for monitoring of bearing vibration. Velocity sensors are on the tops of the shaft riders. The subject unit has a history of vibration problems that center around running speed. Problems had been encountered in 1996 balancing this unit.

Symptoms: It was initially thought this unit might have developed a rub. In the past this type of problem had occurred. Temporary vibration instrumentation had been left on site to collect final data before clearing the unit for full operation. Before any testing was performed; trend data from the control room instrumentation was reviewed and analyzed.

Control room trend data indicated vibration amplitudes on the turbine LP section were very erratic. Amplitudes were instantaneously going form below 3.0 mils (pk-pk) to over 7.0 mils (pk-pk). If the problem were a rub the vibration amplitude should have been somewhat steady and not erratic. Since analysis equipment was still hooked up to all turbine bearings it was decided to perform a complete analysis. This problem occurred late on a Sunday afternoon and the utility was scheduled for a week long power sale Monday morning at 6:00 AM and the plant needed to know if they could meet the requirement.

Test Data and Observations: Plant personnel had dropped load when the vibration appeared. This did lower the amplitudes below the alarm settings; however, the amplitudes were not below the levels achieved with balancing. Test data was recorded with a sixteen channel digital tape recorder and analyzed with a dual channel data collector. Real time data was viewed on a Dual Channel FFT.

The data indicated the vibration was all subsynchronous (Figure #24). An oil whip problem was detected during the over speed testing of the unit (Figure #25); however, this was only present during the running of the over speed tests. During the over speed testing; the subsynchronous vibration was locked on 24.75 Hz

(1485 cpm) and appeared instantaneously. The onset of this whip during overspeed testing was at a shaft speed of 3510 rpm (58.5 Hz).

Normally; oil whirl or whip problems are caused by alignment problems or trying to operate the equipment with oil temperatures that are too cold.

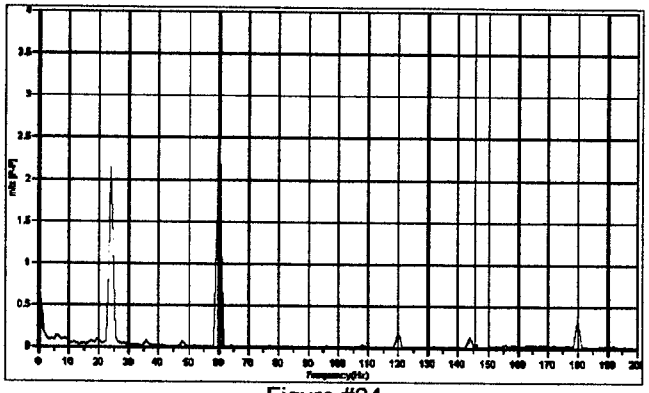


Figure #24

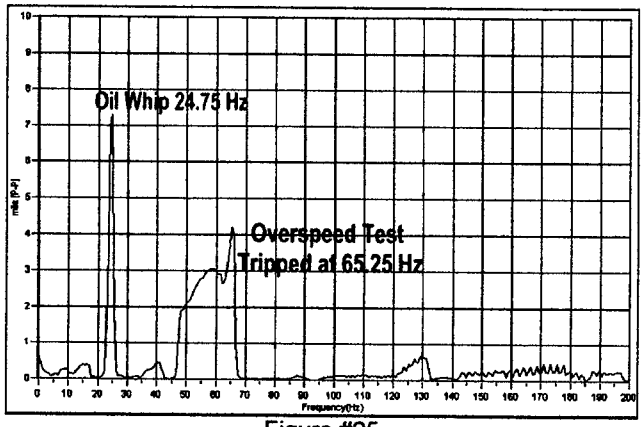


Figure #25

When this problem developed at 90 megawatts the oil temperature was at 105 degrees Fahrenheit. It was decided bring the unit back to 90 megawatts and raise the oil temperature from 105 degree F up to 110 degrees F. As the oil temperature was raised the subsynchronous vibration dropped. A three degree F temperature rise effected the subsynchronous vibration (Figure #26).

This unit was only on line for three days. During the outage the turbine was realigned. The outage lasted four weeks so the unit had not grown into alignment. The oil whip vibration amplitudes were highest on bearing #3.

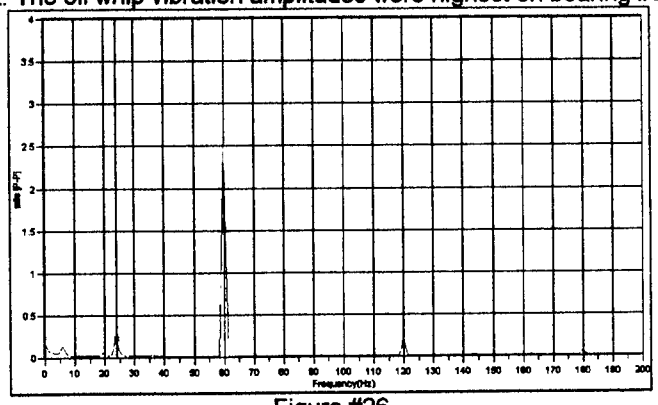


Figure #26

Corrective Action: Operations personnel at this plant had a history of running the oil temperature below 105 degrees F. Normal operating oil temperature main turbine generators is above 115 degrees F. Operating procedures were initialed to ensure the oil temperature was kept above 115.

Results: The unit has operated since this time without problems related to subsynchronous vibration. The operations personnel keep the oil temperature above 115 degrees F and will not roll the unit to speed with the oil temperature below 100 degrees F.

Conclusions: The oil whip problem was a combination of two factors. First the unit had not grown into alignment. This unit is aligned so that bearing #3 is loaded heavy during the start up. This is done because it is susceptible to whirl and whip conditions. The oil being cold and the unit not having grown into alignment allowed the whip to develop. Since this problem; questions have arisen about the alignment of this unit. Optical alignment data shows that bearing #3 may be unloaded. The alignment procedures used during the outage have also been brought into question.

Case #7 – Oil Whip 30 MW Turbine Generator – Alignment Generated

Problem: This unit is located in a power house at a paper mill. This unit is a General Electric Turbine Generator set. The turbine is a single rotor with extraction steam (Figure #27). Operating speed is 3600 rpm and the generator is rated at 30 megawatts. The unit had recently been overhauled. A spare rotor had been installed on the turbine. Within a few days of being put back in service the

turbine bearings started to experience vibration excursions. The unit was not able to be run above 10 megawatts without vibration amplitudes above 7.0 mils (pk-pk).

During the rotor change out all bearings had been sent out for repair and the unit had been realigned. Realignment of the turbine took longer then planed. The generator had to be lowered to get the coupling alignment correct with the turbine.

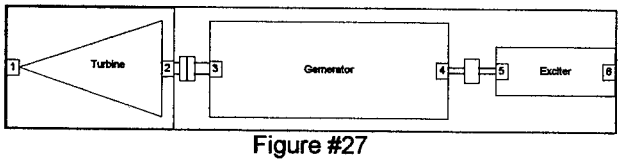


Figure #27

Symptoms: The vibration would cycle from low amplitudes to high amplitudes over a ninety minute cycle. This repeated consistently when the unit was operating above 20.0 megawatts. Below 20 megawatts this cycling would still happen; however, the pattern was not as repeatable. It was also noted that when the operators tried to bring the unit to full load when the turbine vibration cycle was at a low point the first generator bearing (bearing #3) would have a sudden increase in vibration that was erratic.

Test Data and Observations: Test data collected by the plant personnel indicated the vibration was at 1X (Figure #28).

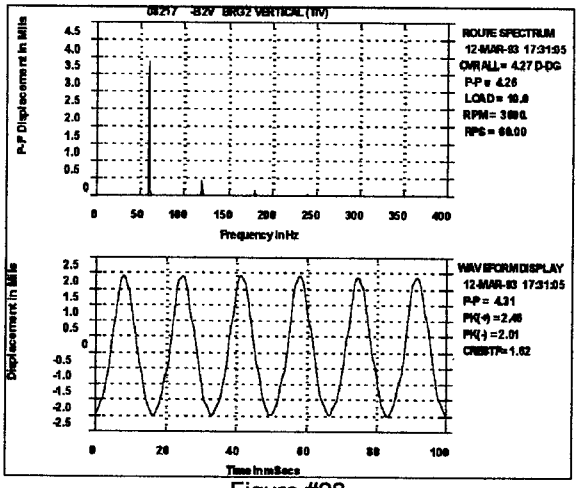
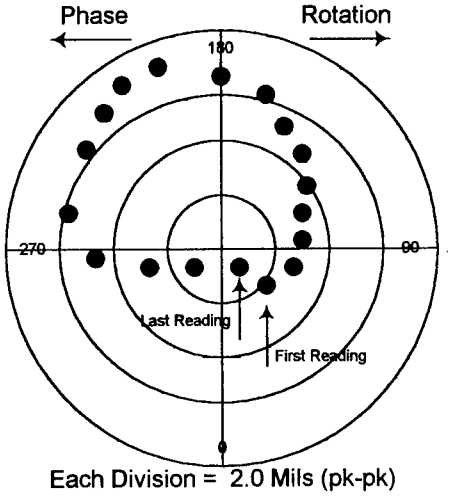


Figure #28

To get a better picture of the cycling of the vibration 1X amplitude and phase data was collected every five minutes for 90 minutes. This showed the vibration was moving opposite rotation over time (Figure #29).



Each Division = 2.0 Mils (pk-pk)

Figure #29

Phase movement with 1X vibration is an indication of a rub. When the unit was at a low cycle of vibration an attempt was made at running the unit at full load. Vibration spectra data was collected on bearing #3 to see what the cause of the erratic vibration was. When the unit reached 20 megawatts a subsynchronous vibration appeared (Figure #30).

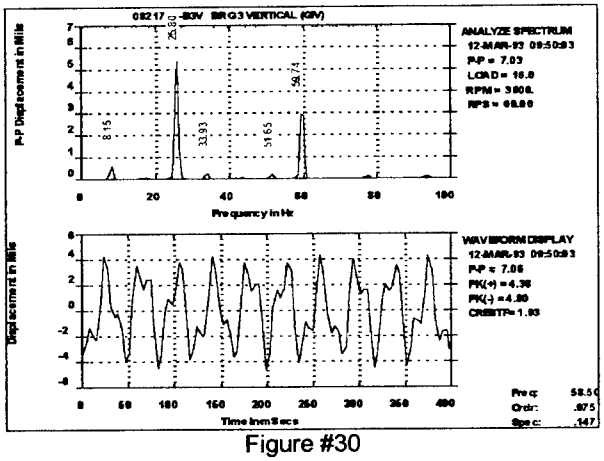


Figure #30

Once it was determined the turbine vibration was due to a rub and the generator problem was due to an oil whip, the cause of the problem had to be found. Looking over the operating parameters indicated the steam operating temperatures were all in the correct range. The oil temperature was also in the correct range.

The next thing to look at was the work done during the turbine outage. Since the unit was completely apart; realignment of the unit was a good place to investigate. If problems with the alignment developed, a bearing could become unloaded and develop a oil whip condition. Alignment could also cause a rub by closing up the clearances between a seal and shaft.

Corrective Action: It was recommended the alignment setting be verified and changed as necessary to relieve the situation. This needed to be done as soon as possible; because the plant generated its own power and could not run at full production without the turbine.

Results: It was found that during the calculation of alignment settings a dial indicator was read incorrectly. This resulted in the generator being set too low unloading the first generator bearing. In turn, bearing #2 was rubbing due to the seal clearances being closed up.

Conclusions: The problem could have been avoided with better quality control during the reassembly and realignment of the turbine generator set. Due to extenuating circumstances the repair work took longer then expected. In the rush to overcome the delays, mistakes were made during the alignment. If two people had been involved in the realignment calculations this problem may have been caught and problems avoided. The whip problems had damaged the first generator bearing to the point that it had to be replaced. Steam seals on the #2 turbine bearing also had to be replaced.

Case #8 – Oil Whip Bearing Starvation on Vertical Pump

Problem: Motors installed on flood control pumps suffer sudden increases in vibration and had an abnormal shaft orbit. This is a synchronous motor with a variable frequency drive (VFD). The design motor speed is 514 rpm (8.57 Hz). The motor and pump bearings are instrumented with dual prox probes (X & Y). . Also, dual seismic accelerometers are installed on each bearing. This motor is rated at 4865 horsepower and has tilting pad bearings in the top and bottom guide bearings. While this type bearing provides stability for whirl and whip conditions on horizontally mounted equipment; in vertically mounted equipment these bearings have problems controlling vibration due to the lower damping of the bearing. This is a six pad bearing (pads 3.0"W x 6.0'L) with zero preload.

The formula for determining excitation and rated speed is as follows:

$$Excitaion = \left(\frac{Line\ Frequency}{Rated\ Speed} \right) x\ Shaft\ Speed \quad (1)$$

Representatives of the motor manufacturer felt the vibration problem was due to balance problems. Specific vibration specifications were to be met by both the pump and motor manufacturers before the owners would accept the equipment from construction.

Symptoms: Motor vibrations were above 5.0 mils (pk-pk) and as high as 10.0 mils plus. Vibration trips set on this motor pump train are set at 8mils (pk-pk) and have a ten (10) second trip delay. It had been noted on test runs the vibration suddenly increases when the excitation is at about 39 Hz. This equates to about 334 rpm shaft speed. The motor field service also identified an abnormal orbit with external loops.

Test Data and Observations (1): All initial test data was collected with the motor uncoupled from the pump. Because of the sudden vibration trips when the excitation is at an excitation of 39 Hz to 40 Hz, data was recorded to determine the cause of the vibration. Data was collected from both the installed proximity probes and extra mounted seismic sensors. Trying to minimize the problem with tripping, vibration data was collected at 5 Hz to 10 Hz excitation intervals. Since the data was primarily 1X, Plant personnel and equipment vendors wanted the motor balanced (Figure #31).

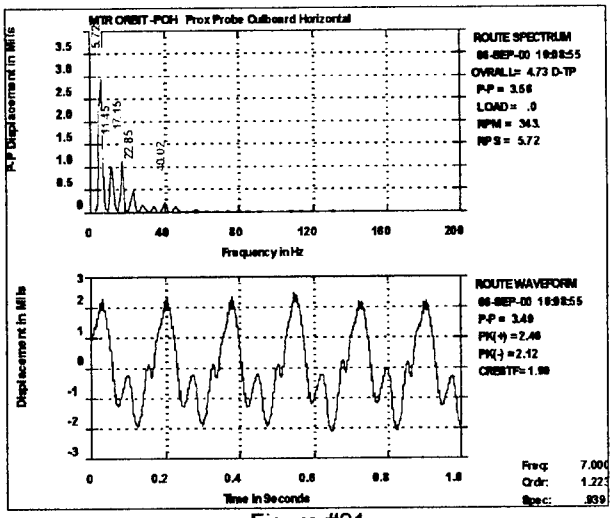


Figure #31

Balancing was able to somewhat lower the vibration in the uncoupled condition; however, when coupled up the vibration was still high at 40 Hz excitation and would trip the motor. There was also a thermal effect noted in the balancing process. During the step by step progression of collecting data at several points as the motor excitation was increased a poor mans bode plot was constructed (Figure #32). This data is from the outboard motor horizontal proximity probe.

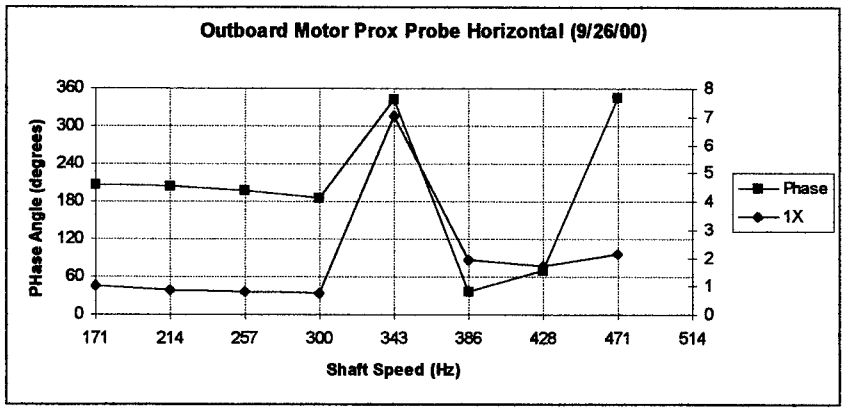


Figure #32

The increase in vibration at 343 rpm (39 Hz excitation) with the associated phase shift is showing the shaft is going through a critical speed.

Corrective Action (1): Recommendations were made to replace the tilting pad bearings with an elliptical or straight bore bearing. It was additionally stressed that a rotor dynamic study be performed to determine the best bearing for the situation. This recommendation was met with stiff resistance by the motor manufacturer. A second consultant was contracted to look over the analysis and balancing data along with the "poor man bode plot." This consultant agreed with the first conclusions and recommendations.

The motor manufacturer contacted the bearing supplier and asked for recommendations on how to make the installed bearing work. They recommended cutting off .75 inches from the length of the each tilting pas. This was to be removed from the trailing edge of the pad. The motor is running counterclockwise; therefore, the right side of the pad was removed (Figure #33).

Once this modification was performed the equipment owner wanted the motor rebalanced in the uncoupled condition. Immediately upon bringing the motor up to rated speed another problem developed that was not expected. A subsynchronous vibration appeared when the motor was running at 40 Hz excitation.

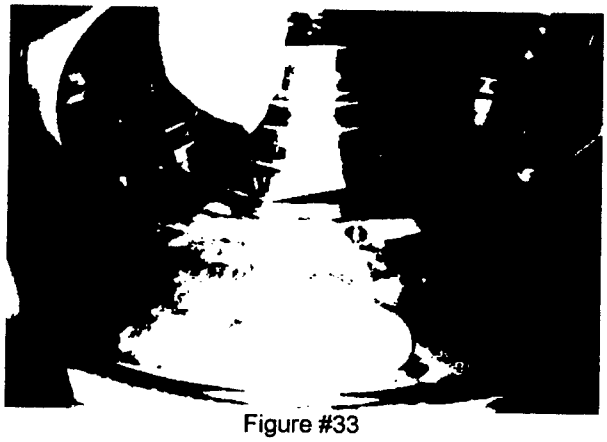


Figure #33

Test Data and Observations (2):

The subsynchronous vibration at 2.37 Hz is 47% of running speed (Figure #34). The subsynchronous vibration was only present on the outboard motor bearing. This vibration would only appear when the unit was at 40 Hz excitation and above. Below 40 Hz excitation this vibration was not present. It had been noted during the balance runs; a subsynchronous vibration was present; however, in amplitudes below .2 mils (pk-pk).

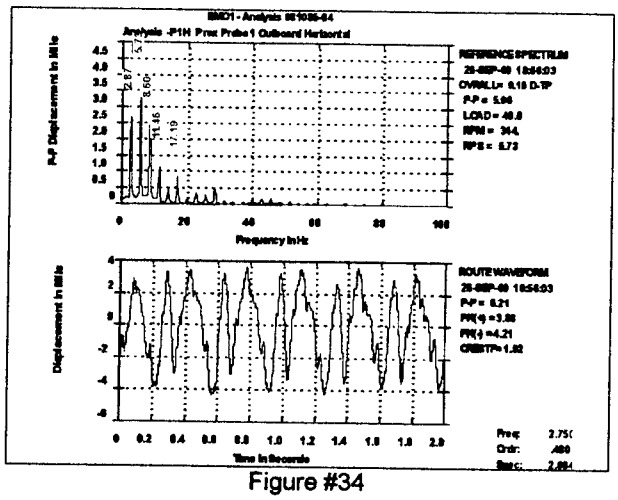


Figure #34

When moving extra seismic sensors at the top of the motor; it was noted the two oil site glasses on the top motor bearing had low levels of oil. The motor was shut down and oil was to be added. However; when the motor was shut down the oil

levels were correct. Now the question became; was the oil level was correct when the unit was operating. Additional oil was added to the sump. However; when the motor was again run, the subsynchronous vibration was still present. The additional oil now had one site glass at the correct level and the second oil site glass with a low level. It was decided the oil sump oil baffle system was not distributing the oil evenly throughout the bearing.

Results: The oil baffle system was redesigned and the extra oil that was added was subsequently drained out. This corrected the situation. The bearing modification was made to all the motors at this installation and they have since operated with out problem.

Conclusions: The initial vibration problems were due to bearing design. Removing the trailing edge of the bearing loaded the bearing and added stability. The subsynchronous vibration was due to starving the bearing on one side. This allowed an oil wedge to develop causing the oil instability problem. If a rotor dynamic study had been performed during the motor design stage this problem may have been avoided. It is felt that even with the rotor dynamic study the oil starvation problem due to the sump baffles would have still been a problem.

Case #9 – Coupling Induced Externally Generated Whip

Problem: Excessive vibration was first reported by Plant Operations on the feed pump and then following pump repairs, vibration was reported on the feed pump turbine drive. The pump is a DeLaval six stage, double suction pump. The pump is direct coupled (gear tooth with spacer) to the drive turbine. Normal speed range is between 4000 and 5000 rpm depending on feed water requirements of the boiler. The first indication of a problem came when Operations reported high pump vibrations while bringing the pump up for morning load.

After the pump was overhauled, abnormal vibration appeared on the high pressure bearing (outboard turbine bearing) of the pump turbine drive. This vibration started around 4000 rpm on the initial run up after the pump overhaul. The vibration, at first, was random; but, after several hours, it was constantly present. The turbine is a General Electric six-stage, 16700 horsepower turbine with a maximum speed of 5200 rpm.

Symptoms (1): Operations reported that at approximately 4200 rpm the feed pump started to vibrate severely . The higher in the load range the pump was run, the more violent the vibration amplitudes. On-line vibration monitors showed 10+ mils (pk-pk) in radial direction (shaft riders mounted 45 from the horizontal). A single axial thrust probe (non-contact pick-up) went out of limits when the severe vibration started. Normal vibration levels were the observed on the turbine drive during this pump vibration (1.5 – 2.8 mils).

Test Data and Observations (1): The first decision that needed to be made was would it be safe to run this pump to gather test data for analysis. Normally; when the on-line axial position monitor goes out of limits, internal metal to metal contact has occurred in the pump. This would entail the rotating element coming in contact with the balance ring assembly. Since this had previously been documented on another pump, it was determined to immediately take the pump off as soon as the back-up pump could be placed in service. A tape recorder was used to record any data that could be acquired in the limited time the pump was being taken off the line.

With the operating speed on the pump at 4350 rpm a subsynchronous vibration is present at 3225 rpm (Figure #35); the running speed vibration is non-existent. Only the sub-running speed vibration is present. The subsynchronous vibration at 74-78% of running speed indicates hard, continuous contact between the rotating and stationary parts of the pump.

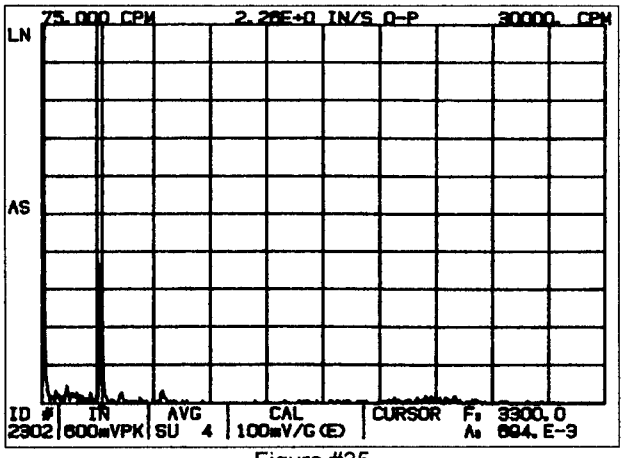


Figure #35

Corrective Action (1): Once Operations had shut the pump down, no emergency action was needed. Mechanical maintenance personnel were instructed to inspect for severe rubbing on the balance components. Mechanical maintenance mechanics found severe rubbing wear and abnormally large clearances on pump internals. It was decided to completely overhaul the pump at this time. A complete new pump barrel, balance components, and bearings were installed.

Symptoms (2): After the pump overhaul was completed, the pump and the turbine were warmed up through normal procedures for the initial roll-off to 1200 rpm. When this was completed, the unit was brought to 4000 rpm, abnormal vibration started to appear on the high pressure bearing (outboard inlet bearing)

of the turbine. This vibration kept increasing as pump flow increased. Since this unit is direct coupled, pump flow is directly related to rpm of the turbine. During the first several hours, this vibration would come and go. Finally; after four hours, the vibration was constantly present. At 4200 rpm, vibration was 4.0 mils and steadily increased as speed was increased.

Test Data and Observations (2): Since the pump was just overhauled, our first thought was the turbine vibration was the result of some procedure done during the overhaul; even though the vibration was in the turbine and not the pump. Because this is a capability related piece of equipment, it was recommended that we go into a two part analysis program. First, a decision was needed to determine, if the pump could be used for operational needs; second, find the cause and solution to the problem. Before any decisions were made, a full set of data was collected to determine if the unit could be run without severe damage. Spectrum data shows approximately 7.9 mils of subsynchronous vibrations present at 1920 rpm and 3840 rpm (Figure #36).

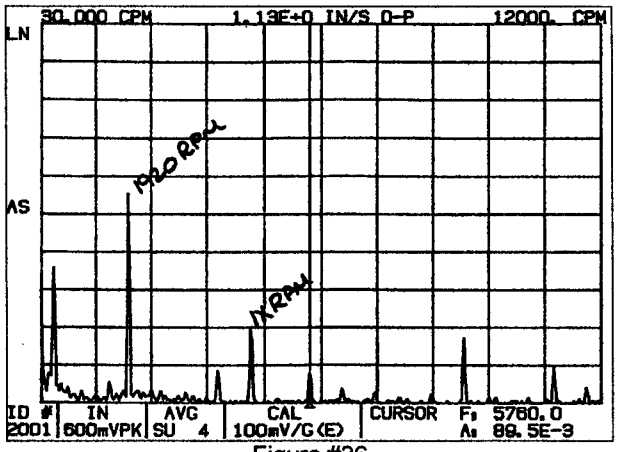


Figure #36

The predominate frequency is 1920 rpm (42.3% of running speed) with an amplitude of .627 in/sec (6.24 mils). Subsynchronous vibration should never be allowed to exceed 30% of the bearing clearance without immediate shutdown, 20% shutdown should be considered, and 10% would be allowable." Since we were over the 20% rule and being under the 30%, it was decided to run the turbine; but, limit its speed range except for testing purposes so it would fit into the afore mentioned preset rules.

The turbine speed was varied up and down the load range (speed range) to see if this subsynchronous vibration "tracked" running speed or "locked on" one

frequency. This test indicated the vibration was locked on to 1920 rpm. Therefore; the cause of the vibration was some type of "whip".

Next, the turbine was brought up to 4700 rpm. Coast down curves (Figure #37) in peak average (Bottom Plot) were going to be plotted vs. instantaneous spectrums (Top Plot) to determine the speed this whirl develops. This data plots the vibration with the turbine running 4065 rpm. Note the subsynchronous vibration is present at 1920 rpm and at 3840 rpm.

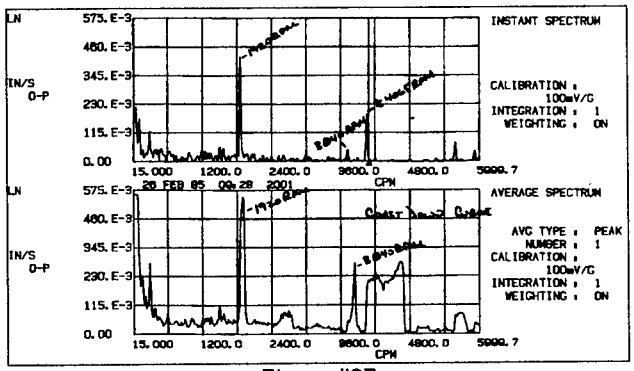


Figure #37

Figure #38 shows the vibration with the turbine running at 4010 rpm with no subsynchronous vibration. This indicates that the factor causing this vibration is not present until the turbine is in the last 20% of its speed range, the point at which steam flow is the highest.

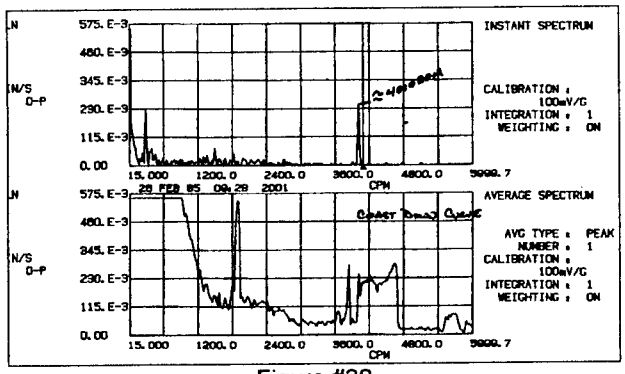


Figure #38

While the testing was progressing, discussions were taking place with Plant Operations and Mechanical Supervisors. The intent was to find if any abnormalities were encountered putting the turbine and pump back in service.

One statement stood out. A unit operator stated the critical around 1900 rpm was rougher than normal. The only critical should be at 2400 rpm. If the critical speed is around 1900 rpm, this would account for the subsynchronous vibration at 1920 rpm. The vibration being locked onto the first rotor critical at 1920 rpm means the first rotor critical would have been lowered from 2400 rpm.

Basic engineering tells us that critical speeds are based on mass, stiffness and geometric properties of the rotor. Vibration data did not show any high 1X, so it could be concluded no weight loss was suffered by the shaft. Also; if there were a lowering of the mass due to weight loss, the critical speed should have increased. The only way to lower the critical speed is to add mass or lower the bearing stiffness. Since the turbine bearings had not been changed it is unlikely the stiffness was lowered.

$$f_n = \sqrt{\frac{576EIg}{WL^3}} \quad (2)$$

Natural Frequency in Radians

E= Modulus of Elasticity (lbs/in^2)

I = Moment of Inertia (in^4)

G =, Gravitational Acceleration (ft/sec^2)

W = Total weight of Shaft (**lbs**)

L = Effective Length of Shaft (**in.**)

In this formula you have four (4) fixed parameters: the modulus of elasticity, the moment of inertia, gravitational acceleration, and the rotor weight. Using this formula and plugging in the 1920 rpm present critical speed and the 2400 rpm should-be critical speed; then solving for the shaft length, you determine the shaft needs to be eighteen inches longer to lower the critical speed to 1920 rpm.

Since the simple model showed the shaft had to be lengthened to lower the critical speed, it was decided to remove the coupling spool piece from the coupling and do a start-up and coast down curve looking for the critical speeds. When this was performed, only one critical speed was found at 2400 rpm proving the theory of a coupling problem was correct.

Corrective Action(2): Maintenance was instructed to inspect the complete coupling for abnormal wear or evidence of coupling torque lock.

Results: Operations Personnel correctly diagnosed a severe pump problem when on-line monitor indicated abnormal axial position and high radial vibration. Past case histories helped solve this problem. Previous data and reports can cut analysis time down when good, neat, documented data can be referenced. Another pump had the same problem eighteen months earlier. Maintenance personnel found that after the pump was overhauled, no grease had been packed into the turbine half of the coupling. The turbine coupling hub had

cracked from heat generated by no lubrication. The complete coupling had to be replaced. When the turbine was put back in service, no subsynchronous vibration appeared, and the only critical speed was at 2400 rpm.

Conclusions: The low pressure turbine bearing (inboard turbine) did not experience this whirl vibration because, when steam flow lifted the shaft enough to lock the coupling; the shaft lifted off the bearing (4010 rpm). This was evidence because the bearing metal temperature dropped. Once steam flow dropped off, the coupling was able to unlock and slide, eliminating the whirling problem. If this problem had gone unchecked, a catastrophic failure of the coupling and turbine was a very real possibility.

A problem such as the friction whirl should never have occurred. The last step in an overhaul should be to grease the coupling. This type of externally excited friction whirl locks on the frequency of excitation. It is both harmonic and sub-harmonic. A problem like this does not occur without some cause; usually a maintenance oversight. The lack of grease in the turbine coupling half did not allow for spacer movement.

Case #10 – Structural Pedestal Resonance – New Installation

Problem: Immediate problems were suffered during the initial running of two new DC motors. The motors were a retrofit of older drive motors on an aluminum rolling mill. The new motors (two motors in series – Figure #39) were supplied by General Electric with a maximum operating speed of 1500 rpm. The motors are a complete retrofit including the motor frame and pedestal. It should be noted the steel pedestal for the system is hollow. A sister installation of this system is installed in another plant and operating without vibration problems throughout the speed range.

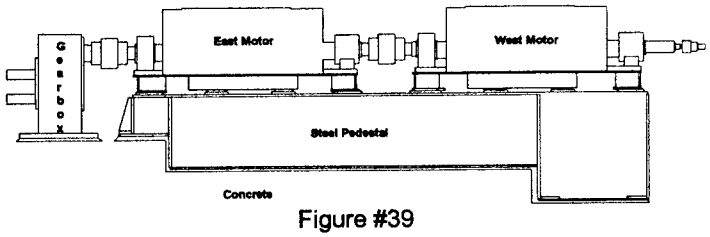


Figure #39

The only difference between the motors in this plant and the sister plant; is the problem motors utilize rolling element bearings. The sister plant is using straight bore babbitt sleeve bearings while the problem motors were retrofitted with rolling element bearings against the manufactures recommendations.

Symptoms: The motors were still not in operation; however, during trial runs high vibration amplitudes were found on the west motor while running around full rolling mill speed, 1260 rpm (21 Hz). While there was vibration on the east motor the amplitudes were acceptable. During the setup of equipment for testing the motor accidentally turned on and went up to just above operating speed (Figure #40). This data indicated the vibration increases suddenly when the motor speed is at the top end of the operating range.

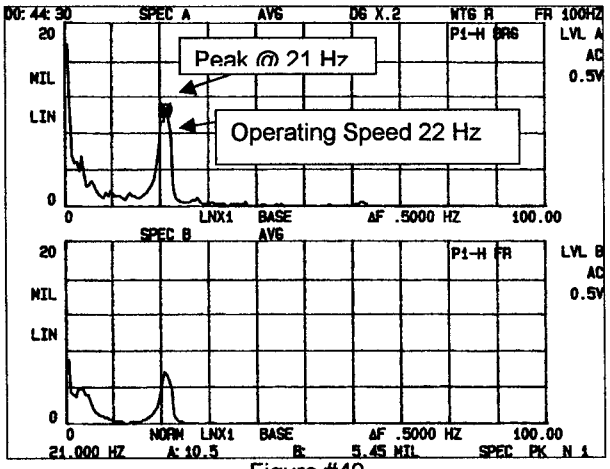


Figure #40

The operating speed reached 22.0 Hz with the peak vibration at 21.0 Hz.

Test Data and Observations: A test plan was developed to look for a natural frequency of the shaft and bearing pedestal system. Impact tests would be used to look for natural frequencies around 20 Hz. These tests would also show if there were other natural frequencies in the running speed range or if natural frequencies could be excited by multiples of operating speed. Impact tests would be conducted in the axial, horizontal and vertical directions. Tests would be conducted on the motor bearings, frame, steel pedestal and foundation. Included in the impact tests would be information on stiffness of each equipment component and information for amplification calculations.

Equipment used for the testing included: instrumented hammers for the impact tests, a dual channel FFT analyzer for the impact tests and the startup / coast down data, a sixteen channel digital tape recorder along with amplifying and integrating power supplies for the accelerometers used in the data collection.

Impact test data indicated natural frequencies at the upper end of the operating range (Table VI).

Test Position	Horizontal	"Q" Factor	H- (K)(lbs./in)	V- (K)lbs./in)
West Motor Outboard Shaft	**23.0 Hz** **40.0 Hz** **62.0 Hz**	**3.5** **7.0**	1,049,723	1,037,736
West Motor Outboard Bearing	**23.0 Hz** 31.0 Hz 62.0 Hz	**8.4** 13.3 13.8	787,735	1,432,836
West Motor Outboard Frame			3,957,115	264,957
West Motor Inboard Bearing	**23.0 Hz** 62.0 Hz	**6.9** 8.7		
West Motor Inboard Frame	40.0 Hz 64.0 Hz	16.7 10.4	4,134,420	113,600
East Motor Outboard Bearing	**23.0 Hz** 31.0 Hz 62.0 Hz	**4.0** 17.6 9.5		
East Motor Inboard Bearing	31.0 Hz	14.1 Hz		

Table #VI

Operating data showed the majority of the vibration to be in the horizontal direction and a drastic drop in vibration as the motor speed lowered form full speed operation (Figure #41). The vibration amplitude drops 3.0 mils in 2.0 Hz.

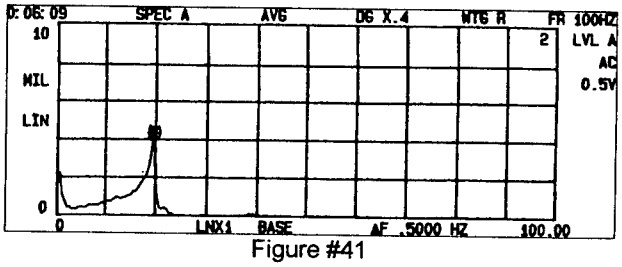


Figure #41

This equipment train is definitely operating at a natural frequency when the motor speed is at 1322 rpm (22.03 Hz). There are two ways to essentially move a natural frequency and they are to either change the mass or the stiffness of the system. Presently; there is approximately 2,425,000 lbs./in of stiffness in the system. Additional supports will probably not add to this total. Calculations were made to see if filling the steel pedestal with concrete would add enough mass to lower the natural frequency enough, so the equipment could be run at full speed operation. The addition of mass; likewise, will have little effect on the natural

frequency. This leaves the options of adding additional support or installing a Dynamic Vibration Absorber.

Corrective Action: The idea behind a Dynamic Vibration Absorber (DVA) is to have the absorber vibrate instead of the equipment. The dynamic absorber is designed to have the same natural frequency as the offending machine. This equipment train would require a DVA to have a natural frequency around 22.0 Hz. This would then absorb the vibration energy and allow the machine to run normal vibration amplitudes. Since the west motor had the highest vibration amplitudes, the absorber would be designed for this motor.

The design of the DVA dictates that it must be at least 10% of the total weight of the machine. The west motor, frame and pedestal weighted roughly 25,000 pounds. This means the absorber must have at least 2500 pounds of weight. To design an absorber one has to determine the length of the vertical support based on the weight to be supported (Figure #42).

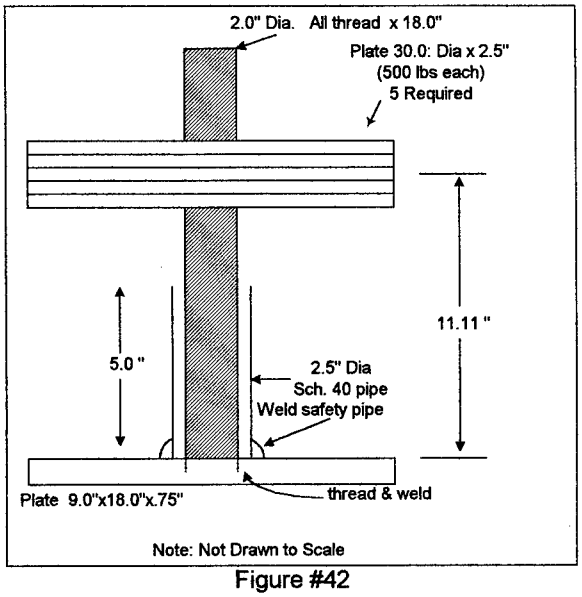


Figure #42

Results: This design was intended to be a temporary short term solution. The final recommendations were to retrofit the bearing to a sleeve design. In fact the motor manufacturer had the bearings and bearing pedestals in stock since the original design was for sleeve bearings. The intention was to install the absorber for a short time so the plant could get the new mill up and running. Once the mill

was operating correctly the bearings would be changed out. The absorber dropped the vibration by 75% and has continued to operate without a bearing change back to the sleeve bearings.

Conclusions: The cause of the vibration was the type of bearings chosen by the plant. The sister plant chose to use babbitt bearing for their motors. Babbitt bearings were what the motor was designed to utilize. However; this plant decided to use rolling element bearings. This decision was made because plant personnel felt that the operating speed of 1260 rpm was too fast for the babbitt bearings. The rolling element bearing did not provide any damping; therefore, allowing the high vibration amplitudes.

Case #11 – Electrically Excited Structural Resonance

Problem: While collecting baseline data on a new rolling machine vibration was detected in one speed range on a variable speed machine. The motor drive is variable frequency rated at 1195 rpm (11.91 Hz), 600 volt and 350 horsepower. The motor drives a single reduction gearbox with a gear ratio of 15 to 1.

Symptoms: During the equipment baseline check out and acceptance process high vibration was experienced on the input shaft of the gearbox in the horizontal direction. Vibration amplitudes of over .60 in/sec 0-pk were documented. The plant had set overall alarm settings of .10 in/sec 0-pk. The vibration appeared to occur in only one area of the speed range. Figure #43 is a schematic of the gearbox layout.

Production personnel noticed floor vibration when the mill was operating at 2350 feet per minute (fpm). A plot of motor speed versus mill operating speed is in figure #44.

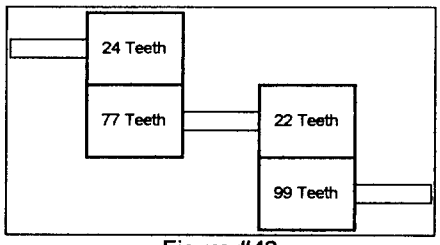


Figure #43

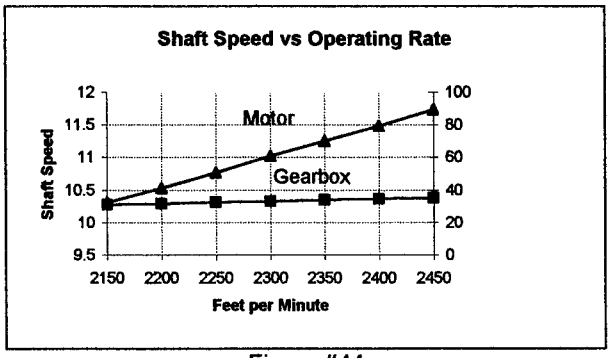


Figure #44

Test Data and Observations: The motor was operated through out its speed range and the overall vibration was recorded from 2100 fpm to 2450 fpm. Figure #2 shows us this is from 10.3 Hz to 11.73 Hz on the motor . The overall vibration data plotted against motor speed is in figure #45. Figure #45 illustrates how the vibration peaks when the motor is operating at 11.3 Hz. Vibration spectral plots were taken at various speed setting. Figure #46 shows the vibration at a motor speed of 9.88 Hz and Figure #46 shows the data from a motor speed of 11.25 Hz. The Data at 11.25 Hz shows a vibration at 3X running speed of the motor. The time data is essentially a pure sine wave.

This data points to a resonance problem because it is only in one area of the speed range. Generated frequencies of the bearing in the motor, gearbox and roll are all well above the operating speed of this frequency at 33.81 Hz. Since this is a variable speed motor with a maximum speed of 1195 rpm (19.91 Hz) the excitation frequency is equal to line frequency divided by rated speed times shaft speed.

$$Excitaion = \left(\frac{Line\ Frequency}{\text{Rated Speed}}\right) x\ \text{Shaft Speed} \qquad (3)$$

$$Excitaion = \left(\frac{60}{19.91}\right) x\ 11.25 = 33.9\ \text{Hz}$$

The excitation frequency at 33.9 Hz is equal to the dominant frequency of 33.81 Hz in figure #47. This is a definite indication that the excitation frequency of the motor is exciting a natural frequency when the motor speed is 11.25 Hz.

An impact test was attempted on the gearbox to confirm the natural frequency; however, because of the system mass and close quarters it was very difficult to swing a hammer. Because this frequency was being generated by the excitation frequency of the motor, it was decided to shut the motor down and observe the vibration when the power was cut. Immediately upon the cut off of power,

the vibration dropped out. This conclusively pointed toward a natural frequency excited by the excitation frequency of the variable speed motor.

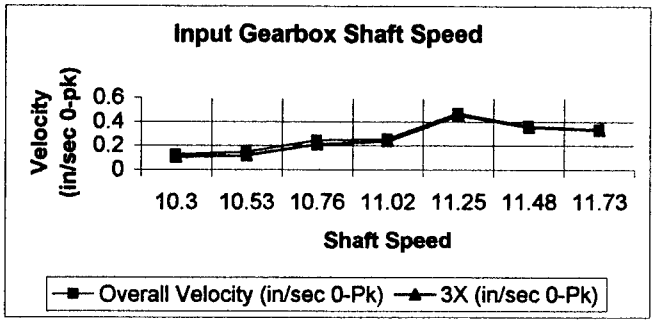


Figure #45

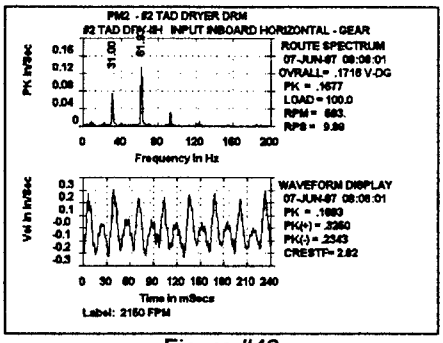


Figure #46

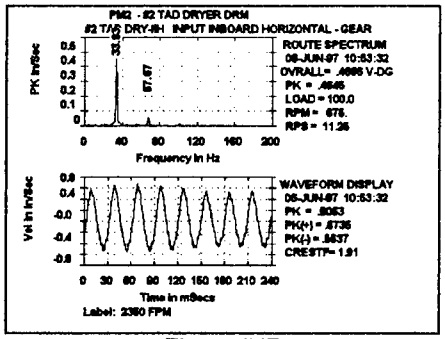


Figure #47

Corrective Action: The plant turned over the data to the equipment supplier and instructed them to redesign the equipment foundation.

Results: The equipment supplier designed isolators and snubbers on the gearbox to help relieve the vibration caused by the natural frequency. Figure #48 shows data after the isolators and snubbers were installed.

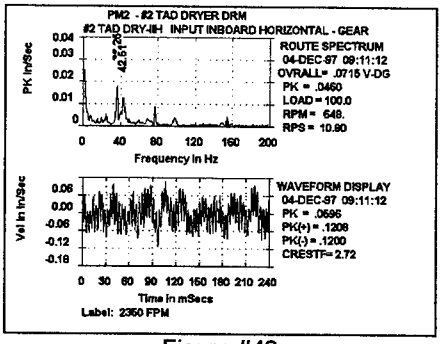


Figure #48

Case #15 – Rolling Element Bearing Fundamental Train Frequency

Problem: Trend vibration indicated the possibility of a bearing vibration on a Vertical Boiler Circulating Water Pump. The motor had been rebuilt in the last year and the pump radial bearings had been replaced. The plant suspected a bearing problem; however, they could not determine if the problem was in the pump or motor.

Symptoms: Vibration trend data showed a high frequency vibration at 3643.6 Hz (Figure #49). The overall amplitude of vibration had been trending up over the past several months.

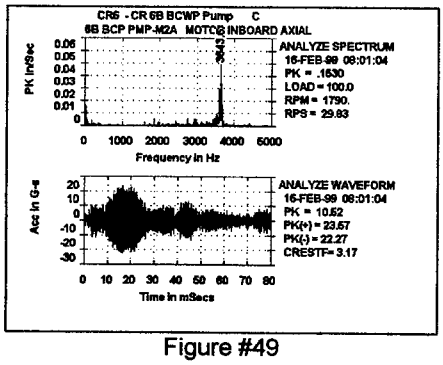


Figure #49

This data was collected with 400 lines of resolution with an Fmax of 5000 Hz. The resolution for this plot is 37.50 Hz per bin. Because of the bin width, frequency identification would be difficult. Timewave data shows peak acceleration at 23. 57 g's. Looking at the same data on a smaller frequency range shows the 1X vibration along with a frequency at 5.58 Hz (Figure #50). A walk down of the unit reveled a high pitch squeal around the pump. The motor pump train operates at 29.58 Hz with radial bearings in the motor, top, and bottom of the pump. Below the upper pump bearing are two thrust bearings.

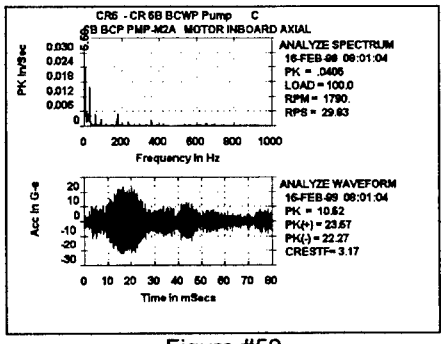


Figure #50

These thrust bearings were not replaced during the last pump overhaul. Utilizing a data collector to get an overview of the data, vibration on the pump casing around the location of the thrust bearings (Figure #51) was excessive (1.0 in/sec 0-pk +).

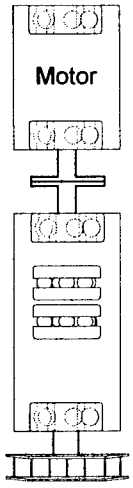


Figure #51

Based on these observations a test plan was developed.

Test Data and Observations: Since the previous data was collected with only 400 line of resolution, frequency identification was a problem. Also, the time plot was based on 1024 bits of data (400 line) and provided .08 seconds of data. This only showed 2.4 rotations of the shaft, not enough data to determine if the frequency of the energy bursts in the time data (Figure #49 or #50). The new data would be collected with 3200 lines of resolution and have a time plot of 4096 bits (1600 lines). This would provide a time plot of .32 seconds. Data was collected from all points available on the machine and on the pump housing in line with the thrust bearings in the horizontal and vertical directions.

Calculation of the bearing fault frequencies provided the data in Table #V.

	SKF 6226	SNR 7230	SKF 6324	SKF 29434	SKF 6326
FTF	12.18	12.82	11.58	**13.14**	11.58
BSF	81.24	83.56	64.87	79.63	65.0
BPFO	109.6	205.2	92.63	**210.3**	92.67
BPFI	156.6	268.2	144.0	263.1	143.9

Table #V

The new plots with the longer time data shows the bursts of energy are at 1X (Figure #52).Looking at the same point in Velocity (Figure #53) shows essentially the same information as figure #2; 1X and a subsynchronous vibration between 5.0 and 6.0 hertz. None of the velocity or Acceleration plots indicate any bearing frequencies. The time data is typical of bearing frequencies; however, no distinct frequencies can be identified. Since the goal of this analysis was to identify if the problem was a pump or motor bearing it was decided to try peak view analysis. The peak view data plots showed two distinct bearing problems on the thrust bearing in the pump (Figure #54).

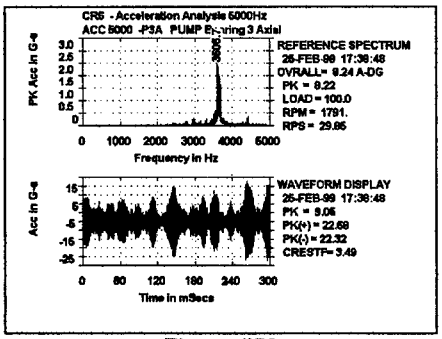


Figure #52

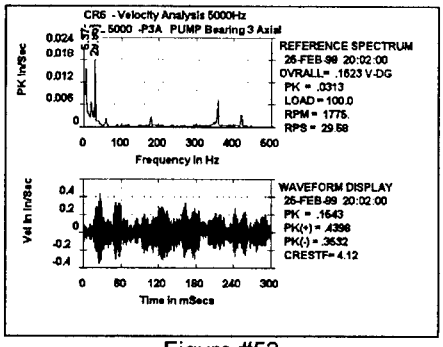


Figure #53

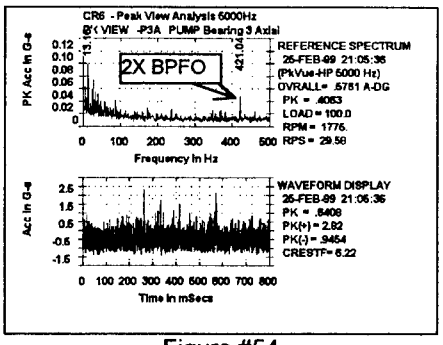


Figure #54

Corrective Action: The recommendation was to replace the thrust bearings.

Results: Both the cage and outer race of the bearings were cracked.

Conclusion: The decision to not replace the thrust bearing in the pump when the radial bearings were replaced was inappropriate. Replacing the thrust bearing required removing both radial bearings. In a two week time the radial bearings had to be replaced twice. This decision to not replace the bearing was made totally on the cost of the thrust bearing. In the end the thrust bearing replacement ended up costing replacement power in addition to labor and parts.

TROUBLESHOOTING VIBRATION PROBLEMS
A COMPILATION OF CASE HISTORIES

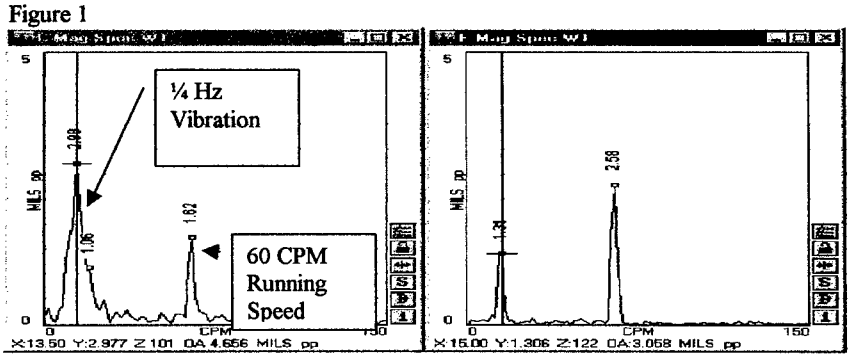
Nelson L. Baxter
ABM Technical Services, Inc.
8529 E. CR. 300 S
Plainfield, In. 46168
ABM@SURF-ICI.COM

ABSTRACT: Vibration analysis can be utilized in the solving both rotating and non-rotating equipment problems. This paper presents several case histories where vibration analysis was utilized to troubleshoot a wide range of problems.

Key Words: Analysis; fans; motor; pumps; turbines; vibration

CASE 1- ¼ Hz vibration in Hydroelectric Dam. During the operation of a hydro-electric turbine, the whole structure would begin to shake in a certain load range. The problem was identified as Rheingan's influence. Rheinghan's influence is caused by spiral vortex filaments that rotate at a speed lower than the turbine RPM. The load range at which the Rheingans influence occurred varied in this case, depending upon the up and down stream water level. The problem was that the vibration supervisory system did not register any excessive levels during the transient, so the operators who were located at a remote location, did not know what load ranges to avoid. The spectra shown in Figure 1 were taken from shaft proximity probes while the problem was present.

Figure 1



The above plots taken with a spectrum analyzer that was DC coupled showed that there was no difficulty in detecting the vibration. The reason that the supervisory did not detect the 15 cpm vibration caused by the Rheinghan effect was that its AC coupling capacitor filtered out the vibration. Since nothing can be done to prevent the Rheinghan effect, modifications to the supervisory system are being considered to allow the operators to at least avoid the unstable load ranges.

CASE NO. 2- VIBRATION OF STEEL STRIP IN STEEL MILL CAUSED BY INDUCTION FURNACE.

EQUIPMENT: An induction furnace in a steel mill was used to heat and diffuse galvanize into the steel strips.

SYMPTOMS: During the induction process, a loud high pitch frequency would radiate from the steel plate. When the sound would begin, vertical stripes would also appear on the plating. The stripes were causing the steel to be rejected by the customer of the steel mill.

TEST DATA AND OBSERVATIONS: An FFT analyzer was set up to determine the frequency of the sound, along with the vibration on the induction furnace and the current being supplied to the induction coils. The frequency being detected in all three cases was at 7250 cycles/second. This frequency corresponded to the operating frequency of the induction furnace. To determine if a change in frequency would have an effect on the problem, the furnace frequency was increased to 9000 Hz. The stripes did not disappear at the higher frequency, but merely moved closer together

CONCLUSIONS AND RECOMMENDATIONS: It was determined that the induction furnace was exciting the natural frequencies of the plate, creating standing waves, which resulted in the stripes being formed. Since the thin plate had several natural frequencies within the normal operating range of the furnace, changing from one frequency to another did not help. Increasing the frequency made the stripes closer together, decreasing the frequency resulted in the stripes being further apart. When turbine generators are brought to operating speed, there are several speeds which match the natural frequencies of the various blade lengths. Due to thermal stress considerations, it is necessary to operate in the range of speeds where problems can occur, while the rotors get thermally stabilized. To prevent damage from occurring to the blades, the speed on the turbine is continually varied, rather than staying at one speed, during the thermal soak periods.The above analogy was used to approach the striping problem. Rather than operating at one frequency, as it had in its past, the induction furnace's control circuit was designed to continuously vary the frequency . When this modification was made, the striping problem was eliminated.

CASE NO. 3 VIBRATION OF MICROSCOPE IN MICROSURGURY ROOM OF A HOSPITAL

Equipment: A microscope mounted from the ceiling of a surgery room. The special microscope was used by the surgeon during operations which involved replanting limbs.

Symptoms: The chief surgeon complained that the image was jittery and that it was very tiring to operate under those conditions, particularly when the scope was set for its maximum magnification.

Test Data and Observations: The scope was set to its greatest magnification and printed material was placed on the operating table. Vibration was clearly noticeable, just as the surgeon had indicated. Vibration spectra were taken on both the table and the microscope. The levels on the table were very low across the spectrum. However, the levels on the microscope were significant. A view of the vibration spectra revealed that peaks were present at 225 CPM and 435 CPM. To trace the source of the vibration,

levels were measured on the top of the microscope's isolator and on the structural steel supporting the isolator . It was discovered that the levels on the isolator were seven times higher than on the steel support. This meant that instead of isolating the microscope from the structural vibration, the isolators were actually amplifying the vibration which was present on the I-beam. To determine the cause of the amplification, an impact test was performed on the microscope to determine its natural frequencies. It was found that the natural frequency of the scope on its isolation system matched the vibration which was present on the I-beam.

Conclusion and Recommendations: Isolators perform their isolation function by creating a system with a natural frequency tuned much lower than the expected disturbing frequency. This in turn creates a mechanical low pass filter, which will not pass the higher frequencies. A problem can, however, occur if a low frequency is present near the low tuned natural frequency of the isolated system. Instead of isolating the frequency, the isolators will then actually amplify the levels. The solution in this case was to ground out the isolators. When the isolators were grounded out, the levels on the scope dropped to acceptable levels. The frequencies which had been present were due to isolators on the roof fans being tuned to the same frequencies as the microscope. Flow excitation in the fans excited the fans' isolated natural frequencies, which were transmitted through the structural steel. Grounding of isolators should only be tried if nothing else works. When the isolators are grounded, higher frequencies, if any are present, will obviously pass through. In this particular case, grounding out the isolators didn't introduce any significant higher frequency vibrations.

CASE NO. 4- TORSIONAL VIBRATION ON A RECIPROCATING PUMP
Equipment: 66 RPM Reciprocating Water Pump driven by a gear case and belt reduction.

Problem: Excessive torsional vibration at the pump speed was being picked up at the gearcase.

Background on Torsional Testing:
Torsional testing generally is performed using either one of two methods. The first is the use of a strain gauge to measure the alternating torsional strain. The second method involves measuring the change in the passing frequency of equally spaced gear teeth or equally spaced reference marks. The change in passing frequency of equally spaced marks on a shaft is an indication of corresponding changes of angular velocity. This data can therefore be integrated to produce angular displacement.

Test Data and Observations:
For this test, both the strain gauge and equally spaced reference mark techniques were used so that a comparison between the two methods could be made. A strain gauge was mounted at a 45 degree angle on the drive shaft between the gearcase and the belt driving the reciprocating pump. An FM transmitter and a battery were also mounted on the drive shaft to transmit the strain information to an antenna and FM receiver-demodulator.
The above set up was calibrated by putting one end of the drive shaft in a vice and applying 100 ft-lb of torque to the other end. While the 100 ft-lb of torque was applied,

the output of the demodulator was measured with a volt meter. The calibration constant from this test was then input into an FFT analyzer, resulting in readings directly in ft-lbs at each frequency present. The second test involved putting equally spaced photo-reflective tapes at twenty locations around the output hub of the gearcase. A photocell device was then mounted to pick up the pulse train of the reflective tapes. The output from the photocell was then input to a torsional-demodulator-integrator which produced an output of 200 mv/degree peak to peak. With the above combination, it was therefore possible to measure the torque being fed back to the gearcase from the pump and the amount of angular displacement it produced. When the pump was operating against no appreciable back pressure, there was 1.5 degree of angular displacement present at the gearcase hub at 66 cycles/minute. The torque from the strain gauge at this condition was 27 Ft-lb. As the back pressure on the pump was increased, the above values also increased. When the output pressure from the pump was 180 PSI, the torsional vibration had increased to 8.79 degrees and the alternating torque to 123 ft-lb, both at the pump speed of 66 cycles/minute.

Corrective Action:
The above data showed that the alternating torque peaks from the pump were too high. To correct this problem, a flywheel was added to level out the torque peaks by absorbing energy during one half of the cycle and return it to the system on the other half cycle.

Final Results:
The alternating torque values were reduced by a factor of three by the addition of the flywheel.

Conclusions and Recommendations:
The above case history shows how two different techniques led to the same conclusion. The test method used will depend on what the investigator needs to know, the test equipment available and the accessibility of the machinery to be tested. Another interesting note concerning the above is that when the coherence was measured between the two signals with a dual channel analyzer the level was .98. This indicates a direct correlation between the alternating torque on the drive shaft and the displacement on the gearcase hub, which is as would be expected.

CASE No. 5- GHOSTS
Symptoms of Problem: In a small Midwestern town, the residents complained that they felt movement of their houses, particularly at night. One of the residents stated that her sister would no longer come and visit her because she thought the house was haunted with ghosts. She felt this way because lamp shades would move, pictures would rattle and rocking chairs would rock without anyone being in them.

Test Procedure: In an attempt to identify the cause of the above problem, vibration measurements were made at several of the houses in the community, along the sidewalks and at a factory located near the houses. The testing was performed with an SD-380 analyzer utilizing a 1000 mv/g low frequency seismic accelerometer to convert the mechanical motion into an electronic signal.

Test Results:
The vibration signature taken at one of the houses where complaints had been registered showed a level of 1.22 mils at 300 cpm. It was observed that the amplitude of the vibration would oscillate, indicating that more than one frequency might be present. (Beat frequencies are generally produced by two or more closely spaced frequencies adding and subtracting as they go in and out of phase.) To determine if there was more than one frequency present, the zoom feature of the SD-380 analyzer was utilized. A 16:1 zoom plot of the vibration at one of the residences clearly showed the presence of several frequencies, all close to 300 CPM. The next step in the investigation was to make a survey of the vibration present at the nearby factory. The factory was a foundry, which contained several vibratory conveyors that moved parts from one area to another. Zoom plots were taken at each conveyor to determine their frequencies as close as possible. Exact matches were found between the frequencies present at the houses and several of the vibratory conveyors in the factory. Additional testing confirmed the correlation between the operation of the conveyors and the vibration at the houses.

Conclusions: Rather than having ghosts, the residents of the small town were experiencing low frequency vibrations from the vibratory conveyors in a nearby factory. The 5 hz (300 cpm) frequency is easily perceived by the human body, particularly at night when other motion and noise is at a minimum. It also can excite low stiffness structures (i.e. lampshades and rocking chairs).

CASE NO. 6

VIBRATION OF NUCLEAR MAGNETIC RESONANCE MACHINE

Equipment: Nuclear Magnetic Resonance instrumeny used to test chemical samples.

Symptoms: Following the movement of the unit from a second floor location to a third floor room, the NMR instrument performed poorly. A test technician determined that everything was operating properly within the unit. The technician thought that vibration might be the cause of the problem, so tests were performed.

Test data and Observations: A signature taken at the probe of the NMR unit in the new location, showed a level of 8,190 micro g's. The floor beside the NMR unit had a level at the same frequency of 1,550 micro g's. The readings meant that the vibration on the NMR unit was 5.2 times higher than the level measured on the floor at the 26 Hz frequency of the vibration. To determine the cause of the amplification, a resonance check was made. The floor was impacted and the response was measured on the detector of the NMR unit. The transfer function clearly showed a peak at 26 Hz, indicating that the Unit was resonant at the frequency which was present on the floor.
Conclusions and Recommendations: It was concluded that fans in an HVAC room near the third floor location were providing the 26 Hz forcing function. The resonant condition was significantly amplifying the vibration that was present. It was recommended that the NMR unit be installed on isolators with a 95% efficiency in attenuating the 26

Hz vibration. The 95% reduction resulted in levels which were lower than those the unit had been exposed to in its original location, where it had operated satisfactorily. Following the installation of the isolators, the NMR unit performed well.

CASE NO. 7-VIBRATION INDUCED BY SOUND

Equipment: Rotary casting conveyor in a foundry.

Symptoms: After the installation of a rotary casting conveyor, vibration of the walls and particularly on the windows in the control room of the foundry experienced high levels of vibration.

Test Data and Observations: The locations with the highest levels of vibration were the windows in the control room. A plot of the vibration measured on the foundry windows showed a level of 39.2 mils near the center of one of the windows. A frequency of 885 cycles per minute was predominant in the spectrum. The 885 CPM vibration on the windows was also found to be present on the walls of all the offices in the foundry. This frequency matched the frequency of a large rotary casting conveyor. Vibration measurements next to the conveyor were, however, low. The conveyor was mounted on springs and also was fitted with dynamic absorbers, which were apparently working as designed, considering the low levels observed on the floor next to the conveyor. The next test involved taking measurements with a microphone. The output from the microphone was analyzed on an FFT analyzer and it was found that the sound level at 885 CPM (14.75 Hz) was over 100 dB. Since this was below the normal hearing range for humans, the sound level did not seem bad, however, it could be felt and a sheet of paper held in front of the conveyor would move noticeably. The final test involved performing a resonance check on the window. A plot of the response of one of the control room windows showed that the natural frequency of the window was very close to the frequency of the pressure waves being emitted by the rotary casting conveyor.

Conclusions and Recommendations: The vibration problem undoubtedly was caused by the rotary conveyor. The transmission path was through the air rather than through the structure. The windows being resonant near the operating frequency of the conveyor were further amplifying the problem. It was recommended that the windows be fitted with cross braces to move their natural frequencies away from the operating frequency of the conveyor and that a sound absorbing enclosure also be built around the conveyor.

CASE 8 - HIGH ALMOST 2X VIBRATION ON A HIGH PRESSURE CORE INJECTION PUMP AT A NUCLEAR PLANT

Equipment: Steam Turbine Driven High Pressure Core Injection Pump on a Boiling Water Nuclear Reactor.

Symptoms: Operators of Nuclear Power Plants are required to comply with section XI of the ASME code regarding In service Inspection. The code states that base line readings are to be taken in Mils Displacement and that if the base line readings double, then action

must be taken. On this particular pump, the base line level had nearly doubled, so an investigation was initiated

Test Data and Observations:

The unit consisted of a turbine driver, a high pressure pump, a gear case and a low pressure booster pump. Vibration spectra were obtained for all the bearing locations on the pump train. The readings were at normal levels everywhere except on the High Pressure pump in the horizontal direction. A level of 1.4 inches per second was present at what appeared to be twice the running speed of the unit at that location. A cascade plot of vibration on the HP pump in the horizontal direction from the inboard end to the outboard end showed that the vibration was high at each end of the pump, but was nearly zero in the center. Due to the large difference between the vertical and horizontal readings and the presence of an apparent rigid body pivoting mode, a horizontal resonance was suspected. An impact test was performed on the pump. A natural frequency at near twice running speed was found. This mode matched the response found while the pump was running (i.e. high on the ends and low in the middle). Since the vibration was predominately at what appeared to be twice running speed, it was suspected that misalignment might be the source that was exciting the horizontal structural resonance. The alignment was checked and found to be out of specifications. The alignment was corrected and a test was run on the unit. There was no improvement in the level of vibration. In fact it was slightly higher on the test run than it had been previously. This change in the level of vibration followed a pattern that showed that the vibration in the 2X cell would vary from .9 to 1.6 in/second from one test run to another. In an attempt to determine the phasing of the vibration, a once per revolution pulse was used as a reference trigger for the FFT analyzer and a signature was taken in the synchronous time average mode. As the number of averages increased, the vibration at what appeared to be twice running speed disappeared. This was one of the breakthroughs in the analysis of the problem. It meant that the vibration was not phase locked to the high pressure pump shaft.

The drawings were again examined on the pump and it was found that the gearcase had a 1.987:1 reduction. In addition, it was discovered that the low pressure pump had a four vane impeller. The pieces were beginning to fall together. The pump manufacturer was contacted with the above information. The pump man recalled that there had once been a case of an acoustical resonance with a similar pump. To determine if this could be contributing to the problem, the piping between the low pressure and high pressure pump was measured. It was found that the length of pipe connecting the pumps was equal to one half wave length of the low pressure pump blade pass frequency. Since the low pressure pump had 4 vanes and the gear reducer had a 1.987:1 reduction, the vane pass frequency looked exactly like twice the running speed of the High Pressure Pump. As a final test to confirm the above theory, a tach pulse was put on the low pressure shaft. The vibration pickup was placed on the high pressure pump. The vibration on the HP pump was found to indeed be phase locked to the low pressure pump.

Conclusions and Recommendations:

The problem on the High Pressure pump was not at twice the HP pump running speed, as it had appeared. It was actually at the vane pass frequency of the low pressure pump. It appeared to be at twice the HP pump running speed because of the 4 vane impeller in the LP pump and the almost exact 2:1 speed reduction between the two pumps. The problem was amplified by the acoustical resonance of the pipe connecting the discharge of the LP pump to the suction of the HP pump. The vibration was further amplified by the horizontal structural resonance of the HP pump casing. The major clue to the solution of the problem was that the vibration on the HP pump was not phase locked to the shaft on that pump and was therefore coming from another source. The final clue was the past acoustical resonance problem encountered by the pump manufacturer. The recommenced solution was to change the 4 vane impeller out to a 5 vane design. This change entirely solved the problem.

CASE NO. 9-MACHINE TOOL VIBRATION

Equipment: Machine that cut chamfer on wrist pin hole of rods
for automotive engines.

Symptoms: A loud high pitch sound was produced during the backside chamfer operation. Examination showed that the surface that was machined was very rough and completely unacceptable to the customer.

Test Data and Observations: The manufacturer indicated that the problem might be due to bearings or gearing within the chamfer head. The test data proved that this was not the case. During the cutting operation, a spectrum was taken on the machine being used to cut the chamfer, while it was operating at 900 RPM. The data was taken on the chamfer head in the vertical direction. The frequency of the vibration was at approximately 115,000 cycles per minute (1916 Hz). Data was then captured in the digital buffer of a spectrum analyzer during the transient. It could be determined from the captured data that the vibration would build up, then abruptly quit when the cut was completed. The peak level of vibration measured on the part being machined, reached a level of .82 inches/second. Another test was performed with the machine operating at 1200 RPM, which was a 33% increase in speed. The vibration frequency which was produced on the machine and the part was the same frequency which had been present when the machine operated at 1200 RPM. Based upon the fact that the frequency did not change when the tool RPM was varied, it was suspected that a natural frequency was being excited. To test this theory, an impact test was performed on the cutting tool, which was the most likely source of the problem. A resonance peak clearly showed up at 114,750 CPM. A mode shape of the first natural frequency of the tool was produced by profiling the imaginary components of the transfer functions taken along the tool's surface . The 114,750 CPM mode was found to be the first cantilever mode of the cutting tool. To complicate matters, it was discovered that the part in its holder also had a natural frequency near that of the tool.

Conclusions and Recommendations: It was recommended that the tool be modified to separate its natural frequency away from that of the part. It was also recommended that

the basic process be reviewed. Since the problem only occurred when the back side cut was made, the problem may have originated from trying to make the cut when the tooling bar was under tension rather than in compression.

CASE 10- EXTREMELY HIGH LEVELS OF VIBRATION ON THE END CAP OF A LARGE PIPE.

SYMPTOMS: A large pipe (37" DIAMETER)at a refinery had very high levels of vibration on an end cap after an expansion joint. The levels were over 4.0 inches/second. There had been failure of several of the retaining rods which spanned an expansion joint.

TEST DATA AND OBSERVATIONS: A vibration spectrum was taken on the end cap. The vibration was found to occur at 4500 cycles/minute with a level of 4.47 inches/second. The area surrounding the pipe was checked for rotating equipment operating at that frequency. No machinery was found which operated anywhere close to that speed. A visual exam of the pipe showed that there was a large butterfly valve up stream of the end cap. The end cap was on a dead end section of a tee. Conversations were held with plant personnel to determine when the vibration had started and what, if anything, had been done to the piping prior to the high vibration and retaining rod failures. The first response was that work had been performed on the expansion joint, however, "nothing had been changed". Further discussions and examination of the piping drawings did show that one thing had indeed been changed. A baffle had been removed just upstream of the expansion joint. The baffle was a thick flat plate with two small holes in it. The plant personnel didn't think that it served any purpose. An overall review of the system showed that it was very important. What was occurring was that the butterfly valve was causing flow disturbances. The valve was found to be operating only 30% open. This resulted in pressure pulsations in the pipe. The baffle was acting like a low pass filter which allowed the static pressure to equalize, but would not let the dynamic pressure pulsations pass. The pressure pulsations were probably small, however, there were two design features that caused the vibration levels to be high. The first was the amount of area on the end cap. A 37" diameter end cap has 1075 square inches of surface area. Therefore even a small pressure pulsation can generate a large force, when acting on such a large area. The second reason was the length of the retaining rods. The retaining rods were 20 feet in length, so the stiffness value was low.

CONCLUSIONS AND RECOMMENDATIONS: The vibration problem was the result of pressure pulsations within the pipe acting on a large area with low stiffness. The removal of the baffle had been a key element in the problem. The plant was advised to install short bolts across the expansion joint until the unit was brought down for its next outage. The purpose of this action was to stiffen up the system and provide a backup to the long bolts that had been failing. This action was possible because the piping was already at its maximum temperature. During the following outage, the baffle was reinstalled in the pipe and the short bolts across the expansion joint were removed. The vibration problem was eliminated.

CASE 11-LARGE ALIGNMENT CHANGE ON BOILER FEED PUMP CAUSING FAILURE OF PUMP AND CRACKING OF TURBINE PEDESTAL.

EQUIPMENT: Boiler Feed Pump on a 500 Megawatt turbine generator.

SYMPTOMS: After being in operation only a few months, the boiler feed pump on at a new generating station failed. The inboard seals were wiped out and the stage next to the coupling destroyed. Due to the type of failure, the plant initiated an alignment study.

OBSERVATIONS AND TEST RESULTS: In order to determine the amount of movement from the hot to cold condition, Dynalign bars were mounted between the pump and the turbine driver, while the unit was in operation, following repairs. When the unit was brought off line, there were no significant changes recorded. However, a few minutes later, the Dynalign system was found to be entirely out of range. When the operators were questioned as to what had happened during that period, they replied that the only thing they had done was to break vacuum on the main turbine. To determine if the vacuum had anything to do with the apparent alignment change, the bars were reset and long range probes were installed. The vacuum was then reapplied to the system. Everything appeared normal until the vacuum reached 11" of Hg. At that point, the changes in the gap voltages started to be recorded. By the time full vacuum was achieved, the relative motion between the turbine and the pump was over .100". The vacuum was then released and the readings moved .100" in the opposite direction. The test was repeated with identical results. An examination of the system was then performed. The boiler feed pump turbine was found to be connected to the main turbine condenser by a large pipe. On the end of the pipe was an end cap. Three expansion joints were installed between the main condenser and the feed pump turbine. The purpose of the expansion joints was to isolate the boiler feed pump turbine from stresses induced by thermal growth of the condenser pipe. Thrust canceling rods were installed between the end cap and the main condenser. The thrust canceling struts transmitted the atmospheric load on the end cap to the condenser. The source of the problem was that threaded studs in the thrust canceling struts were sliding into the struts. This was the result of the welds on the large nuts on the back side of the plates the ends of the struts failing. The net result of the failure was that atmospheric pressure which was being applied to the 6' diameter end cap was pushing on a 20' vertical run of pipe attached to the bottom of the feed pump turbine. The large force applied to the 20' lever had the capability of generating nearly 1 million ft. lbs of torque to the turbine. Examination of the concrete turbine base showed that it had several cracks from the bending torque. The thrust canceling struts were repaired and the alignment changes were rechecked as vacuum was reapplied. Following the repair, there were no significant changes in alignment as a result in variations in vacuum.

CASE NO. 12- Large fan that could not be balanced.

Equipment- A 5000 HP 720 RPM fan at a power plant.

Symptoms: Above normal levels of unbalance. Several attempts were made to balance the unit, all of which were unsuccessful.

Test Results: The casing vibration levels were around 2-3 mils. Since there was difficulty in balancing the fan, shaft stick measurements were taken to determine the absolute motion of the shaft. The shaft movement was discovered to be over 17 mils. Since the bearing clearance was only 8 mils, there was a strong indication that the bearing was moving in the housing. A large plunger bolt at the top of the bearing was tightened. After tightening, the casing vibration increased to 21.5 mils. The fan was then easily balanced to below 1 mil. What had been occurring was that with the bearing moving in the housing a non-linear system was present. This made balancing all but impossible.

CASE NO. 13- Incorrect selection of isolators.

Equipment: Several air handler units

Symptoms and test results: A number of air handler units in a large city all had the same symptoms. The fans operated with acceptable levels of vibration, however, the motor drivers all had high levels of vibration in the vertical direction. When the spectra were examined, it was discovered that the primary component of the motor vibration was at the fan's operating speed. An investigation resulted in finding that the isolators under the motor were sized improperly. The isolators were too stiff. This resulted in the natural frequency of the isolated system matching the operating speed of the fans. The isolators were therefore acting as amplifiers of the fan vibration rather than isolators of the motor vibration. What appeared to have occurred is that the same size isolators were used for the motor as were used for the heavier fans.

Case No. 14- High axial vibration on a fan due to a disk wobble natural frequency.

Equipment: A belt driven exhaust fan operating 1200 RPM

Symptoms and test results: The fan's axial vibration was always high. The fan would be balanced and then a couple of weeks later, the axial vibration would go back up. Due to the sensitivity of the fan to unbalance, a resonance was suspected, so natural frequency tests were performed. There was no natural frequency match when the fan was struck in a lateral direction. However, when the fan was impacted axially, there was a match with running speed. To gather more information, the mode shape was measured. It was found that the shaft was the node point and that the opposite sides of the fan were out of phase. This is commonly called a disk wobble natural frequency. Fans that have this problem commonly exhibit sensitivity to unbalance, particularly in the axial direction.

The solution in this case was to simply change the fan's speed. If this is not an option, then stiffening of the back plate of the fan wheel may be necessary.

CASE NO. 15- Cavitation destroying impellers on large circulating water pumps.

Equipment: Large low RPM 156,000 gpm circulating water pumps.

Symptoms: The impellers on the circulating water pumps on a cooling lake at a large power plant were failing. The failure mode was cavitation. The impellers looked like they had been attacked by metal eating termites. When a vibration spectrum was taken, the spectrum contained a large amount of broad band energy with no distinct peaks.
The key to the analysis, as is the case with a good percentage of pump problems was to look at the flow head curve. The flow head curve indicated that at the design flow of 156,000 gallons per minute that the back pressure would be 30 ft. When the back pressure was measured, it was only 10 ft. What had happened was that during cold weather when the lake was cold, operations was operating with only one pump to reduce power consumption. The system was designed to operate against the back pressure produced by two pumps in parallel. When only one pump was on, the system back pressure dropped and the one pump that was on line went into cavitation.

CASE NO. 16- Low pump flow destroying antifriction bearings in a pump.

Equipment: Double suction single stage pump

Symptoms and test results: Three identical pumps sat in a row at a power plant. The bearings were failing on one of the pumps every few months. The other two pumps had no failures. Alignment was checked and different bearings were tried. Nothing helped.
While the pump was in operation, it was noticed that the shaft vibrated in the axial direction. This is called axial shuttling. It can occur when a pump is operating against too much back pressure. The suction and discharge pressure were therefore measured and compared to the flow head curve. It was discovered that the pump was operating at its shutoff head. Based upon this observation, the system was examined. The pump in question, pumped water from a tank in the basement up seven stories to another tank. The tank on the upper floor had a level switch that shut a control valve when it was full. When this occurred, water from the pump flowed through a bypass line back to the tank in the basement. The flow through this recirculation line was controlled with an orifice plate. The orifice plate said that it had a 2" hole, however, specifications said that the hole should have been 3". The surprising thing that was discovered was that there was only a 1" hole in the plate. The result being that when the control valve to the upper tank shut, the pump was effectively operating at its shut off point. This caused the axial shuttling which in turn destroyed the bearings.

CASE NO. 17- Large vertical pumps were cracking shafts every few weeks.

Equipment: Vertical service water pumps

Symptoms and test results: Due to the severity of the problem, underwater proximity probes were installed on one of the pumps. In addition, casing probes and torsional instrumentation was also installed. When the pump was put into operation, high levels of sub-synchronous vibration were observed. Natural frequency and mode shape measurements determined that the sub-synchronous vibration was centered around the shafts 1st lateral natural frequency. The cause of the problem was traced to a maintenance superintendent purchasing impellers from a non OEM source. The design of the impellers varied significantly from the original design. This caused high levels of turbulents that excited the shaft's natural frequency. Since non-synchronous vibration causes stress reversals, this caused the shafts to fatigue.

CASE NO. 18- Boiler feed pump would operate successfully for several months, then the running speed levels would start trending upward.

Symptoms and test results: Two feed pumps had a history of vibration problems. Following their overhauls, they would operate smoothly, but after a few months, the vibration levels would start to increase. The increase was due to higher levels of running speed vibration. When tests were finally run on the pumps, it was discovered that they were operating near a critical speed when fully loaded. This was determined when changes in speed resulted in large amplitude changes and shifts in the phase angles. A newly overhauled pump did not show these same traits. It was determined that when the seals were wearing, this reduced the Lomakin stiffening effect allowing the shaft natural frequency to drop into the operating range.

CASE No. 19- Oil whirl in a chiller unit

Equipment: A steam turbine driven chiller at a University campus heating facility.

Symptoms and test results: The inboard bearing of the turbine driving a chiller experienced repeated failures. Examination of the bearing showed that the top half had been fatigued to the point that babbit was separating from the base metal. Vibration spectra contained high levels of sub-synchronous vibration. Since the bearing was located next to the coupling and oil whirl was present, misalignment was suspected.
A series of alignment measurements were made across the coupling and it was found that there was significant movement during the first hour of operation. To determine which machine was moving, a laser was set up with a receiver mounted on an I-beam. It was determined that the chiller was moving. The root cause was determined to be the suction line on the chiller shrinking as the unit cooled. This shrinking caused the chiller to rock back. This in turn lifted the shaft unloading the turbine bearing causing it to go unstable.

CASE 20: 500 megawatt generator destroyed when phase lead insulation failed. Due to 120 Hz resonance.

Equipment: 500 megawatt 2 pole generator at coal plant

Symptoms and test results: A massive failure resulting from a phase to phase short led to testing the natural frequencies of phase leads on generators of a particular manufacturer. The testing showed that the phase leads on this style of generator normally have resonances just above 120 HZ. This is very dangerous because 120 Hz is the rate in the U.S. at which the magnetic poles pass by a stationary structure, thereby providing strong excitation at that frequency. It was discovered that the phase lead natural frequencies tend to drop as the phase leads loosen up with operating time. It was determined that yearly testing was required in order that the problems be found and corrected prior to any future failures.

CASE No. 21: Coupling lock up of nuclear steam generator feed pump turbine

Equipment: Steam turbine driven feed pump.

Symptoms and test results: A feed pump turbine experienced high levels of twice running speed vibration. The orbits indicated that the problem was misalignment. According to the plant personnel, the unit had been aligned per the specifications. When the unit was brought down for an outage, the coupling was examined. Its teeth were severely worn, the lubricant had failed and it had evidently locked up. Based upon this evidence, a study of the operating alignment was made. It was determined that the original specification was wrong. Originally the pump had been set high relative to the turbine. The final setting required that the pump be set .020" lower that the turbine. The change was due to two main factors. The first was vacuum draw down of the turbine. The second was the amount of growth of the pump.

CASE No. 22: Oil Whip in a 500 Megawatt Turbine

Equipment: 500 Megawatt Turbine

Symptoms and test results: A large steam turbine had very peculiar behavior characteristics. It would operate with no problems for months at a time. If it then had to come off line for a few hours, it could not be started back up, due to high vibration from oil whip in the fisrt Lp bearing. However, if the unit was off line for a day or so, it would start back up with no problem. It would also start up OK if it was restarted immediately after being brought off line. Such a situation has all the signs of a thermally related alignment problem. Since normal alignment equipment cannot be used on an operating turbine, a special system was developed to measure the elevation changes of the bearings. This system showed that when the vacuum was drawn on the unit, the low pressure rotor bearings dropped significantly. When this effect was combined with differential thermal shrinkage as the unit cooled down, it resulted in the first LP bearing

being unloaded enough to cause oil wirl. As the unit came up to speed, the oil whirl locked onto the rotor's first natural frequency and developed into oil whip. The solution was to install a tilt pad bearing in the first LP position. After the installation of the tilt pad bearing, the hot startup problem was eliminated.

CASE NO. 23- Large 3600 RPM Motor with a thermal vector

Equipment: 4000 Hp 3600 RPM Feed Pump Motor

Symptoms and test results: A large feed pump motor was sent out for a normal inspection and cleaning. After returning from the motor shop, it was put into operation and after 45 minutes, high vibration destroyed its bearings. It was returned to the motor shop where it was rebalanced. When it was returned to the plant, it again destroyed the bearings. It was then sent back to the manufacturer and put in a high speed pit and balanced at speed. When it was reinstalled back at the plant, it destroyed the bearings for a third time. Due to the nature of the problem, proximity probes were installed on the motor. When it was first started, the vibration was normal. However, as the motor was loaded, the vibration level increased to the point that the motion was nearly equal to the bearing clearance.
It was determined that the motor had a thermal vector. The solution was to balance the motor in the loaded condition. It operated with the thermal compromise shot installed for several years. It was discovered that the motor shop had dropped the rotor on its first visit. This had damaged the laminations causing a hot shot to develop. This hot spot on one side caused the rotor to bow as it heated up thereby producing the sensitivity to load.

CASE NO. 24: Cracked Rotor Bars

EQUIPMENT: 1800 RPM 250 Hp Service water motor

Symptoms and test results: While in operation a noticeable variation in the sound pattern was evident. The current meter also showed oscillations. Based upon these symptoms, the motor was connected to a dynamometer and spectra of the current were obtained.
The spectra which were taken at various loads showed the presence of side bands spaced at the number of poles times the slip frequency in both the current and vibration spectra. Since this is a sign of broken rotor bars, the motor was disassembled and the rotor was re-barred. Following the repairs, the sidebands disappeared and the sound and current oscillations also went away.

CASE NO. 25: Over heating of D.C. Motor

Equipment: D.C. Motor on press roll at a paper mill

Symptoms and test results: A D.C. motor had repeated failures due to overheating. In addition, the vibration levels were high at a frequency of undetermined origin.
Analysis of the current indicated that there was instability of the drive. The drive was trying to speed up, then slow down the motor at a rapid rate. This is like hitting the

accelerator then the brakes rapidly on a car. The result being high vibration and heat. The drive was retuned and both the vibration and the heat problem disappeared.

Case No. 26- Process related problem picked up on D.C. Motor

Equipment: D.C. Motor on couch roll at paper mill

Symptoms and test results: A unidentified vibration was detected on the motor driving a couch roll at a paper mill. The frequency of the vibration did not match any known source. A current spectrum was taken on the motor and it was discovered that the same frequency appeared in the current spectrum. What was occurring was that the motor was being loaded and unloaded at that particular frequency. The source of the loading oscillations was traced to the fan pump blade pass frequency. The fan pump was generating pressure pulsations that caused oscillations in the head box pressure. This in turn resulted in variations in the pulp thickness. When the thicker areas passed the vacuum rolls, the suction pulled harder against fabric. This increased tension in the fabric caused the tangential force to increase on the couch roll which increased torque demand on the motor and thus the amount of current draw.

AUTHOR INDEX

Author Index

Altmann, J. 37, 297
Aslanimanesh, M. 127
Babich, G. 237
Banks, J. C. 27, 47, 93, 237
Baxter, N. L. 467
Begg, C. 47, 93
Bloor, G. 247
Byington, C. 27, 47, 93, 151, 165, 237, 261, 275
Caguiat, D. E. 139
Campbell, J. 177
Campbell, R. 27, 47, 93, 237, 261, 275
Champagne, V. K. 81
Chung, D. 405
Craciun, G. 285
Cuc, A. 69
Dempsey, P. J. 307
Discenzo, F. M. 405
Essawy, M. A. 227
Gaberson, H. A. 3
Galie, T. R. 139, 151
Garga, A. K. 165, 237
Ghiocel, D. M. 37
Giurgiutiu, V. 69, 285
Goodenow, T. 205
Goodman, P. 69
Groover, C. 217
Gumina, M. P. 139
Guy, K.R. 423
Hay, T. 165
Hess, A. 205
Hines, J. 93
Kacprzynski, G. .139, 151, 247, 261, 341
Karchnak, M.F. 205
Kozlowski, J. D. 165
Lebold, M. 93, 151, 363
Lee, C. 351
Lee, J. 17
Li, C. J. 17
Loparo, K. A. 405
Mathew, J. 297
Matzelevich, W. W. 57
Maynard, K. 217
McClintic, K. 237, 375
McGroarty, J. J. 139
McInerny, S. 319
Menon, S. 415
Nwadiogbu, E. 247
Orsagh, R. F. 341, 375
Palladino, A. 151
Pepi, M. 103
Pooley, III, J. 351
Rekers, A. 285
Riahi, M. 127
Robinson, J. C. 329
Roemer, M. J. 139, 151, 247, 261, 341, 375
Savage, C. J. 375
Scavuzzo, R. J. 193
Schoess, J. N. 393, 415
Thurston, M. 237, 363
Trethewey, M. 217
Twarowski, A. 405
Walker, Jr., A. 117
Wang, J. 319
Watson, M. J. 165
Watt, L. 389
Wolf, K. 185
Yukish, M. 275
Zakrajsek, J. J. 307
Ziegler, W. 117