CIRCUITS SYSTEMS SIGNAL PROCESS
VOL. 13, NO. 4, 1994, PP. 411–433

DEVELOPMENT AND APPLICATION OF A STATISTICALLY BASED FEATURE EXTRACTION ALGORITHM FOR MONITORING TOOL WEAR IN CIRCUIT BOARD ASSEMBLY*

R. W. Barker[1], *G-A. Klutke*,[2] *M. J. Hinich*,[3]
C. N. Ramirez[4] *and R. J. Thornhill*[4]

Abstract. Vibroacoustic signals of rotating machinery are composed of sums of modulated periodicities, broadband random components, and occasionally a set of transient responses. These signals are not ergodic as the modulated periodicities are partially coherent. Progressive wear of the rotating machine causes the nonlinear structure of the received signal to intensify, and nonlinearity results in transfer of energy between harmonics of the signal's periodic components. Statistics developed from bispectrum and second-order cumulant spectrum estimates of the measured signal are combined with power spectrum amplitudes as feature inputs for standard multivariate classifiers. The higher-order statistics measure, respectively, the extent of nonlinearity and intermodulation of the received signal. Classification results of simulated and actual incipient wear data collected from a controlled experiment drilling circuit boards illustrate the potential of this novel statistical signal processing approach.

Subject Classifications: Statistics: Nonparametric time series estimation for pattern analysis. Industries: Health monitoring and durability of rotating machinery. Reliability: Incipient failure inspection/quality control/system safety.

* Received March 27, 1992; accepted August 30, 1992; revised May 21, 1992.

[1] Office of Secretary of Defense, Joint Tactical Missile Signatures, 1431 McGuire Street, Lackland AFB, TX 78236-5532.

[2] Graduate Program in Operations Research, Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712.

[3] Applied Research Laboratories, The University of Texas at Austin, P.O. Box 8029, Austin, TX 78713.

[4] IBM Corporation, 11400 Burnet Road, Austin, TX 78758.

1. Introduction

Circuit boards are by now a common component in many of the products we use on a daily basis. Automobiles, telephones, personal computers, and even toasters contain circuit boards that provide the electronic control for their specific functions. Circuit board manufacturing is a major industry, with a production of roughly $24.12 billion worldwide, of which, according to [1], the United States makes up roughly $6.37 billion. This market is projected to grow to $31.9 billion by the year 2000 [2].

Construction of circuit boards is dependent on the application requirements. Cost, operating temperature, moisture resistance, electrical characteristics, shock and vibration, and other environmental variables all affect the choice of materials used in the boards. Military applications, for example, often require ceramic boards for their thermal properties and electrical performance. A common material for most applications, including computers and telecommunications, is a glassy-epoxy composite known as FR-4. Circuit boards consist of many sheets of electrically insulating layers of composite material alternating with conductor layers (usually copper). Figure 1 illustrates the typical construction of FR-4 circuit boards. The sandwich is built up a layer at a time, with patterning and etching steps performed at each layer to create the circuitry. Insulating composite layers are created by first impregnating glass cloth with epoxy. Several layers of this material are then used to form the thicker insulating layer. The assembly is compressed by a heated platen, then goes through a final curing, and is trimmed and drilled. The number of conductor layers can vary from one to several dozen, and conductor thicknesses also vary due to required resistances.

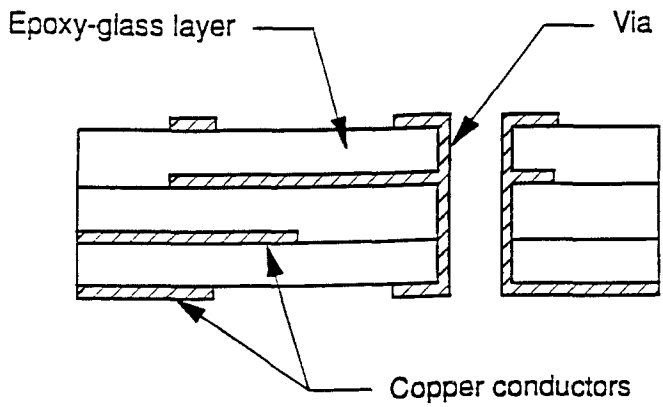

Figure 1. FR-4 circuit board construction.

To produce the requisite functional complexity, electrical interconnections between layers are provided by copper-plated through holes called "vias." Vias are

made by drilling the board on high speed machines with miniature drill bits. Standard vias extend completely through the board, intersecting one or more conductor layers (see Figure 1). Blind vias extend partially through the board thickness and are created by drilling to a predetermined depth or by drilling a subassembly before final lamination. Buried vias are holes that interconnect internal layers of the board, but do not extend to either surface. Drilling of all via types exposes the internal conductors, and electrical connection is established by electrolytic plating of copper after the exposed conductor edges of the drilled board are desmeared by chemical or plasma etching to guarantee a good connection. Taper and other dimensional irregularities can degrade the uniformity of the plating process which increases the probability of open circuits. Most circuit board manufacturing defects are known to be due to problems in the drilling process [3].

The drilling process must be carefully controlled as hole quality significantly affects the plating quality of the finished board. Many defects are avoided by careful optimization of speeds and feeds, drill design modifications, increased drilling spindle stability, and manufacturing material improvement. Nevertheless, ragged and protruding glass fibers due to tearing by a dull drill bit degrade the hole plating quality. A worn drill bit must be changed before significant damage occurs and renders an expensive circuit board unusable. The value of finished boards can be hundreds of dollars, and poor hole quality is a leading cause of board scrap.

Present industrial drill replacement strategy is conservatively based on the shortest observed useful life of a particular drill, and the cost of replacing drill bits is typically a large portion of the total cost of drilling in the manufacturing operation. A medium-sized factory producing 50,000 finished boards yearly with 3,000 holes each would use roughly 30,000 drills annually if each drilled 5,000 holes on average. At $2 per bit, yearly expense for bit replacement would be $60,000. Significant drill life is lost when drills are changed prematurely. We felt strongly that operations research techniques would be quite valuable in developing a reliable monitoring methodology. On-line methods of monitoring drill condition are preferred in the circuit board fabrication industry as they offer the potential for maximizing drill usage, and minimizing machine down time due to drill replacement, without sacrificing hole quality.

Drilling machines used in circuit board production are quite different from those used in metal cutting operations. The most significant difference is that metal drilling is generally done at speeds not more than several thousand revolutions per minute (RPM) while FR-4 board drilling requires speeds up to 200,000 RPM to achieve the desired surface contact speeds with the smallest drills. Drilling spindles are carefully constructed to avoid damaging shaking forces caused by dynamic imbalance, and spindle bearings must be capable of sustained operations at these high speeds. Newer drilling machines use externally pressurized air journal bearings because of their stability, long life, and low friction at very high speeds (see Figure 2).

The drill is held in the spindle by either a split or centrifugal collet. The spindle is driven rotationally by either a reaction air turbine or electric motor, and thrust is

also carried by a two-sided air bearing. Infeed motion that is both programmable and tightly controlled is supplied by a cam. To maximize throughput of a drilling machine, several circuit boards are drilled simultaneously by assembling a panel stack. The stack consists of several identical boards with an entry and exit sheet (see Figure 3).

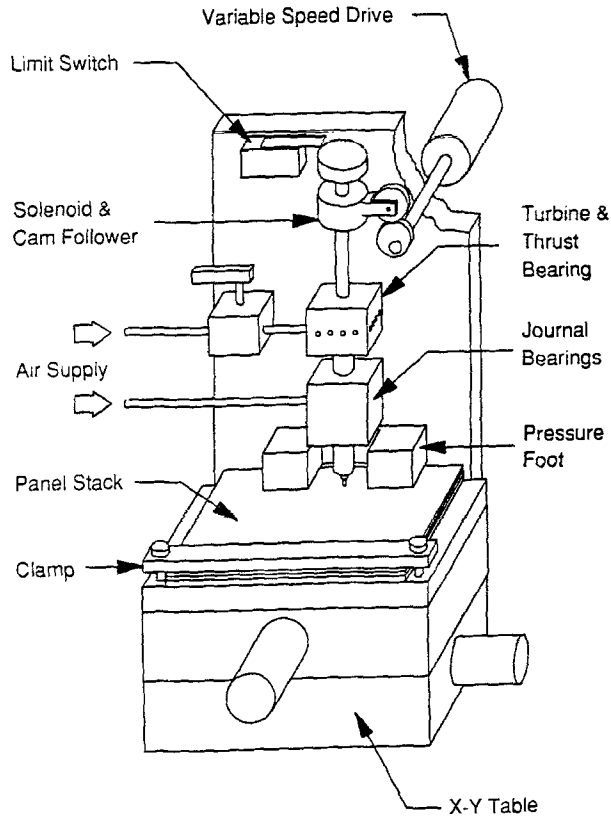

Figure 2. Industrial drilling machine.

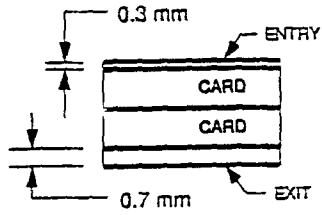

Figure 3. Panel stack configuration.

The stack is clamped to an X–Y table which positions the stack at the appropriate via locations under computer control. The stack is compressed by a pressure foot to control the formation of burrs between the panels and to avoid seizing of the drill. Drilling debris is collected by a vacuum system (not shown). Drilling hit rates are usually several holes per second. A single circuit board may contain approximately 5,000 via holes. Some drilling machines can operate a number of spindles so that several panel stacks on multiple X–Y tables can be drilled.

In this paper, we present an on-line, statistically based algorithm that addresses the random elements of the drilling process and provides a considerably more accurate classification scheme than is currently available. This work is a novel application of statistical signal processing to more accurately capture the characteristics of a nonlinear, nonstationary, and nonergodic wear signal. The paper is organized as follows. After this introductory section, Section 2 gives some details on how data was obtained on drill bit condition. Section 3 gives the pattern recognition formulation of the machine health monitoring problem. Section 4 defines relevant spectral representations of the stochastic processes viewed as providing important feature information for the random fault mechanisms of rotating machinery, and describes the Higher-Order Statistical (HOS) feature extraction method used. Section 5 discusses the performance of the algorithm using a testbed of simulated data and data from the controlled drilling experiment. Section 6 describes the projected implementation of this new monitoring approach, and Section 7 presents our conclusions.

2. Monitoring drill bit condition

Developing reasonable criteria for monitoring drill bit condition involves first defining tool failure. One definition which has received much attention in the literature is tool breakage (see references [4] through [10]). However, catastrophic failure of the tool is rarely the point at which machining should be terminated. In fact, breakage of the miniature carbide drills is rare when drilling FR-4; they are usually removed from operation because circuit board hole quality degrades as a result of worn drills. Hence, a better criterion for tool failure in FR-4 drilling is based on measurements of drill condition which leads to classification of *incipient* failures of drill bits; these are early failures that result in degraded performance but may not be noticeable to the naked eye.

Direct monitoring methods use actual tool wear measurements to determine the remaining tool life. Indirect monitoring methods use other parameter measurements to infer the wear state and remaining tool life. Direct methods are generally preferable if the measurements are reliable and do not interfere with machine productivity. Since these two requirements are rarely met, indirect methods are frequently investigated as alternatives. Monitoring techniques are also categorized with regard to implementation strategy: on-line techniques are used while machining, and off-line techniques require the machine to be stopped. Naturally, on-line

techniques are desired for use in high volume manufacturing environments or when machine interruption might affect finished part quality. Direct monitoring methods are very difficult to use on-line because of continuous contact between the tool and the workpiece. There are also instrumentation difficulties with measurements on rotating tools [11]. Many monitoring methods have been studied and reported in the literature but very few have actually been implemented in industry because they are generally not robust enough to detect all types of tool wear under all conditions [12]. Measurements, whether direct or indirect, have high degrees of uncertainty and are sensitive to cutting conditions and other environmental variables that complicate the proper selection of a sensing approach. Consequently, recent research focuses on finding ways to use multiple sensors and/or signal features simultaneously. Collectively called *sensor fusion methods*, these techniques attempt to reduce monitoring system sensititivity to variables other than tool condition by combining different indicators to classify a tool as acceptable or unacceptable.

While there are examples of models using a combination of sensors and signal features for monitoring rotating tool wear (see, for instance, references [13] through [17]), little published work has dealt specifically with high speed circuit board drilling. Ramirez [18] was the first to investigate an on-line, indirect wear monitoring approach using shaft speed, shaft displacement, and spindle vibration in the manufacturing of electronic circuit boards from epoxy-glass composite. His monitoring methods are the source of data used for the classification algorithms discussed in this paper. The potential responses and instrumentation are shown schematically in Figure 4.

Time series signals were collected, filtered, digitized, and then sampled using a multi-channel transient recorder with 12-bit resolution. Signal capture was triggered using an optical switch that sensed the drill spindle height and began sampling just before the drill entered the panel stack. In this manufacturing environment, drills are typically replaced at 8,000 drilled holes, a conservative estimate of the minimum drill life. A 1.09 mm diameter drill with a nominal spindle speed of 47,000 RPM was used. While torque, displacement, and acceleration results are reported in [19], only the acceleration responses are of focus in this article. Accelerometers were placed at three locations, measuring the spindle vibration along the X, Y, and Z axes.

As is typical when measuring the vibration response of physical systems, the signal contains spurious information (noise) that is not of interest to measuring drill wear. The cutting forces and resulting spindle vibration time series generated during high speed drilling of FR-4 produced good signal-to-noise ratios in the Z or thrust axis accelerometer; of the three accelerometers, this one was shown to be the most useful for monitoring drill wear [20]. This article consequently focuses on 240 Z accelerometer time series: 30 new (0 holes) and 30 slightly used (8,000 holes) drill bit signals for each of four stack material/cutting load conditions.

Typically, time series data are analyzed in the frequency domain using the power spectrum, the Fourier transform of the autocovariance function. While vibration power spectra observed in metal cutting and drilling are usually dominated by a few

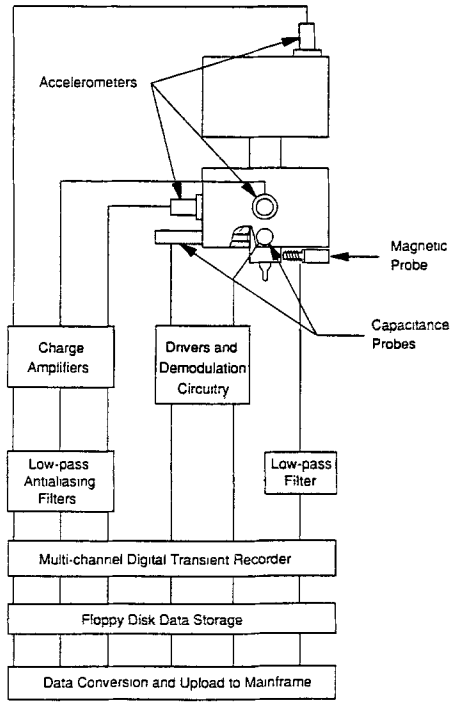

Figure 4. Instrumentation of drill spindle.

structural resonances, FR-4 power spectrum responses were strongly influenced by the internal structure of the composite. Cards are manufactured in the same facility with the same resin system but have different glass cloth and layer thicknesses. Figure 5 shows two different types of boards. Because the glass fibers (oval disks in diagram) cut during each hole are not uniformly configured in the board layers, the vibration signals have both periodic and aperiodic characteristics and reflect the effects of many different cutting geometries randomly encountered by the drill. The cutting forces and energy represented by the vibration measurements change within a certain layer of the board, and also for each revolution of the drill. The measured signal is thus composed of a sum of modulated periodicities, broadband random components, and occasionally a set of transient responses. The modulated periodic components are partially coherent, and thus the signal is not ergodic or mixing, i.e., there is a periodic "memory" in the signal. Transients are nonstationary, and the broadband signal may contain nonlinear structure. When the drills start to wear, the nonlinear structure intensifies in the received signal. Vibration measurements carrying wear information of the drill cutting edges should be sensitive to spectral correlations among the harmonics of the periodic components that are not present in the stationary case (see Figure 6). Consequently, subtle differences in energy transfer during the wear process are not sensed in the power spectrum alone, as

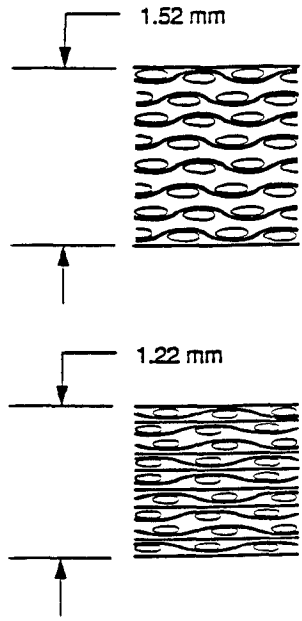

Figure 5. Internal view of composite circuit boards.

the power spectrum measures energy in a stationary and linear stochastic process.

Weaknesses of a power spectrum monitoring approach and the physics of the FR-4 cutting process motivated Barker [21] to investigate higher-order statistical (HOS) methods, which extend the power spectrum estimation techniques for early or incipient (rather than severe) fault detection. The circuit board drilling application became the vehicle for testing the validity and marginal discrimination power of a HOS approach. The algorithm developed combines information from several types of spectral measurements (power spectrum, second-order cumulant spectrum, and bispectrum) from a single sensor to improve a monitoring system's classification performance. Evaluated first with simulated data and then with actual accelerometer data, this new type of sensor fusion approach led to significant improvements in the classification of drill condition. Although only one sensor is used in this application, the approach is adaptable to multiple sensors.

3. Pattern recognition formulation of the machine health monitoring problem

In investigating multivariate group differences in a statistical setting, we have several objectives in mind. First, can we identify significant differences between groups with respect to their multivariate descriptions? This question is answered

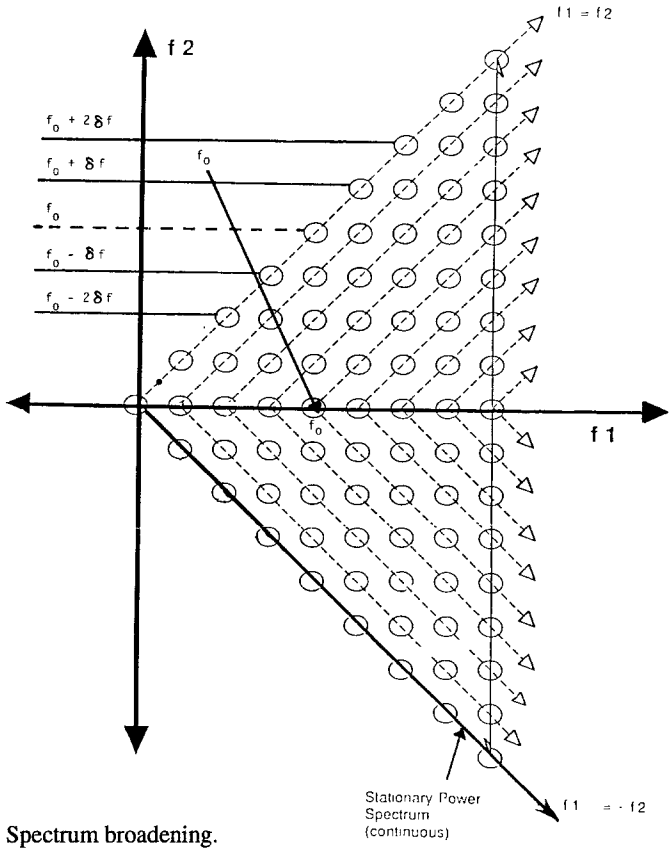

Figure 6. Spectrum broadening.

by performing the multivariate equivalent to the (univariate) t-test on population means. A sample mean vector, or centroid, for each population is formed, and the null hypothesis of equal population centroids is tested using Hotelling's T statistic, or equivalently, Wilks' Lambda statistic when considering only two groups (see [22]). Second, what role do the measurement variables play in separating the groups? A *discriminant function* which may be a linear, quadratic, or other function of the measured variables is used to answer this question. Simply put, a good discriminant function yields similar values for samples from the same population and different values for samples from different populations. Developing a discriminant function involves determining which measurement or feature variables are most important in separating the groups and corresponds to searching for a vantage point that provides a view of the data with maximum separation between groups or populations. We began this research by conjecturing that higher-order forms of spectra combined with power spectra will provide a better vantage point for separating data (in this case time series records) into groups. As will be seen, our conjecture was strongly supported by the signals collected and analyzed. And

finally, can we use the methods of detecting group differences to determine the group that generated a new time series record? This is the multivariate classification problem. In many applications of discriminant analysis, classification is the major objective.

Many examples of machinery monitoring systems in the literature can be categorized as using a general pattern recognition approach. One excellent commercial example is the statistically based system developed at Oak Ridge National Laboratory for continuous, on-line, unattended surveillance of dynamic reactor signals [23]. This monitoring system is based on identification of changes in the power spectrum of measured variables where change is detected using discriminant functions formulated to emphasize relevant features. As is typical of most pattern recognition systems, this system uses a heuristic feature extraction process to feed the raw time series data into Bayesian discriminant functions. "Features" are the properties of the time series used for classification purposes, and the choice of a feature set as well as its statistical properties affect monitoring performance. Additional details on the Bayesian approach used to select appropriate features in this study are included in [21].

4. Spectral representations of time series and the HOS approach

4.1. Preliminaries

Analyzing and interpreting time series measurements can be made in the time domain or the frequency domain. Spectral or Fourier analysis decomposes functions representing fluctuating phenomena in space or time into sinusoidal components that have varying frequencies, amplitudes, and phases. Spectral analysis has an inherent consistency and efficiency in its application because all spectral density functions use the estimates provided by the direct discrete Fourier transform (DFT) of the raw time series. Definitions of terms used in *cumulant* spectral analysis follow.

The characteristic function of a random variable X is given by

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)dx, \quad t \in R,$$

where $f(x)$ is the density function of X. The rth cumulant, K_r, is defined as the coefficient of the rth term in the Taylor series expansion of $\log \phi(t)$. Special relationships exist between moments and cumulants of X; Hinich has developed a general formula for the expansion of moments in terms of cumulants (see [24]).

Let $\big(x(t_1), \ldots, x(t_N)\big)$ be a collection of N random variables with zero-mean marginals ($E[x(t_i)] = 0$ for each $i \in (1, \ldots, N)$). Let $\text{cum}(t_1, \ldots, t_N \mid N)$ denote the Nth order cumulant of $\big(x(t_1), \ldots, x(t_N)\big)$. The Nth *order cumulant spectrum*,

denoted by Cum $S_N(f_1, \ldots, f_N)$, is the Nth order discrete Fourier transform of the Nth order cumulant:

$$\text{Cum } S_N(\underline{f}) = \sum_{\underline{t} \in \mathcal{L}} \text{cum}(\underline{t} \mid N) \exp[-i2\pi(\underline{t}' \underline{f})],$$

where $\underline{t} = (t_1, \ldots, t_N)$, $\underline{f} = (f_1, \ldots, f_N)$, and $\mathcal{L} = \{n: n = 0, +1, -1, +2, -2, \ldots\}$.

A stochastic process is called *stationary* if the probability law of the process is invariant to shifts in the time axis. For a stationary time series, the second-order cumulant is a function only of the lag ($\tau = t_1 - t_2$); i.e.,

$$c_2(t_1, t_2) = c_2(\tau) = \text{Cov}[x(t_1), x(t_1 + \tau)], \quad t_1, t_2, \tau \in R. \tag{1}$$

The *power spectrum* is defined as the Fourier transform of (1). The third-order stationary spectrum (called the *bispectrum*) reflects the cumulant structure over a particular pair of frequencies and provides information about nonlinearity in a stationary stochastic process. For nonstationary processes, however, the *second-order cumulant spectrum* contains information beyond that contained in the power spectrum. Similarly, the *third-order cumulant spectrum* contains information on the process covariance structure beyond the bispectrum. Taken as a whole, the techniques used to estimate spectral densities beyond the power spectrum are known as higher-order statistics (HOS). While a little more computational effort is required to estimate these densities, they provide the foundation for more powerful feature characterization of measured signals.

While the power spectrum represents the contribution to the second moment over a particular *range* of frequencies, the bispectrum represents the contribution to the third moment over a particular *pair* of frequencies. Nikias and Raghuveer list a wide range of bispectrum applications in [25]. Proceedings from the 1989 Workshop on Higher-Order Spectral Analysis [26] contain some recent developments in bispectrum theory with applications to signal processing. The latest developments are given in the IEEE Proceedings of the 1991 International Signal Processing Workshop on Higher-Order Statistics [27]. In this research, in addition to the power spectrum, the second-order cumulant spectrum and the bispectrum are estimated from the measured signal so that *intermodulation* and nonlinear effects of random fault mechanisms of rotating machine systems are captured in the physical process representations.

4.2. Classification via HOS feature extraction

Because of the number of spectral estimates generated (several thousand) and the highly non-Gaussian and nonlinear nature of the data collected in this study, a feature extraction algorithm was needed to select a finite subset of the spectral estimates for input to the classification algorithm. Although it is possible that power spectrum estimation and feature extraction may provide sufficient classification accuracy, performing HOS estimation and feature extraction will be worthwhile if it improves the overall classification performance of the power spectrum-based

approach. The HOS feature extraction approach is not only multivariate, but also multispectral. Algorithmic approaches for determining an appropriate feature space are categorized into two major areas: selection and transformation. Feature variable selection is appropriate if cost or other factors present prevent all of the original set of features to be measured and used; it is thus a combinatorial analysis problem. When all the variables can be measured, variable transformation is performed, but increased reliability occurs if a lower dimensional space is used. Variable transformation approaches include linear and nonlinear techniques. Both approaches assume that the number of potential features is much less than the number of training samples. This was not the case with the time series cases analyzed in this research, which typically had training samples of at most 60 time series records and many more potential features. A hybrid approach was necessary in this work.

The HOS feature extraction algorithm consists of three stages. First, visual plots of ensemble averaged spectra and their differences between groups are generated after each spectral estimation process to obtain a rough idea of which estimates to use as possible feature variables. This is the *graphical variable selection* stage. Second, univariate analyses of variance are performed for the estimates of each spectrum type and the top 10 F-statistics are chosen as potential feature variables. This step accomplishes a *dimensional reduction* of the feature space. Third, a conventional stepwise variable selection algorithm available in SAS 6.0[5] is applied using the 30 spectral variables (10 each from the power spectrum, bispectrum, and second-order cumulant spectrum feature sets), identified from the second stage. Stepwise discriminant analyses are performed to obtain the best reduced and combined spectral feature space. The spectral feature sets composed of 10 similar feature variables (e.g., all power spectrum) are found to be average discriminators by themselves. However, when the different spectral feature sets are combined according to certain statistical criteria, their discrimination and classification power significantly increase. Feature selection and classification results are given next.

5. Performance of the approach in simulated and actual wear data.

Having developed the methodology, we needed data to thoroughly exercise our approach. Simulation provided a basis for generating an extensive testbed of problems to test classification performance. It allowed us to generate, in a controlled way, a vast database of time series records that represented a range of difficulties in classification problems. The simulation was developed to mimic the physics and dynamics of the drilling process as it is currently understood (see [18]), and allowed us to study performance in a much broader range than is available exper-

[5] SAS is a registered trademark of the SAS Institute, Inc.

imentally. The following paragraphs give a brief description of the physics of the process that we simulated.

Rotating machine signature signals received by sensors are represented by a harmonic process:

$$V_c(t) = \sum_{n=0}^{k} A_n \cos(2\pi n f_c t + \phi_n) + n(t), \tag{2}$$

where $V_c(t)$ represents the voltage of the cosine wave carrier signal, A and ϕ are the amplitude and phase terms of the driving force mechanism (drill rotation) or carrier signal, f_c is the fundamental carrier frequency determined by the period of the driving force function and vibration characteristics of the machine system, and $n(t)$ is the corrupting noise generated by other vibratory sources and distortional effects. Noise is assumed to be Gaussian and independent of the emitted vibration signal and k is the number of interacting sinusoids.

The rotating drill spindle does not generate a signal that is a pure harmonic tone, $\cos(2\pi f_c t) = \cos(\omega_c t)$, but rather is an amplitude and phase modulated representation of (2):

$$V_{ampm}(t) = k[1 + m_a f(t)] \cos(\omega_c t + \phi_c + m_p g(t)) + n(t),$$

where $V_{ampm}(t)$ is the amplitude and phase modulated cosine wave carrier signal, m_a is the amplitude modulation index, $f(t)$ is the amplitude modulating signal, ϕ_c is the carrier signal phase, m_p is the phase modulation index, and $g(t)$ is the phase modulating signal. In this study, $f(t) = \cos \omega_a t$ and $g(t) = \cos \omega_p t$ were simple cosine waves. As the tool wears, small cracks and jagged edges develop and influence the signal signature. The primary interest for incipient fault detection, we felt, lies in classifying changes in the *phase* modulation index parameter, although we generated time series records over a broad range of changes in amplitude modulation as well.

5.1. The simulation study

The simulation experiments consisted of 250 independent classification runs using 3 alternative feature extraction methods for 14 different detection problems. Three parameters (amplitude and phase modulation indices and standard deviation of an independent Gaussian noise term) were adjusted. The 14 problems are collected in the seven scenarios shown in Table I. Each scenario has two treatments that differ in phase modulation index at fixed levels of amplitude modulation and noise standard deviation. Numbers in parentheses are the simulation parameters: amplitude modulation index (AMI), phase modulation index (PMI), and standard deviation of Gaussian noise. Amplitude modulation levels were fixed for each scenario as they tend to represent changes in environment rather than changes in process state; however, amplitude modulation was varied between scenarios for

sensitivity and verification purposes. Three phase modulation levels were considered; .7 was chosen as the base reference to represent a new process condition, and .71 and .72 represented an increasingly worn condition. Note that zero values for the modulation indices represent the pure cosine wave carrier frequency.

Table I. Seven incipient fault detection scenarios for simulated wear data.

Scenario Number	Simulation Parameters (AMI, PMI, Noise Standard Deviation)
1A	(.3,.7,.4) vs. (.3,.71,.4)
1B	(.3,.7,.4) vs. (.3,.72,.4)
2A	(.3,.7,.8) vs. (.3,.71,.8)
2B	(.3,.7,.8) vs. (.3,.72,.8)
3A	(.3,.7,1.4) vs. (.3,.71,1.4)
3B	(.3,.7,1.4) vs. (.3,.72,1.4)
4A	(.3,.4,.4) vs. (.3,.41,.4)
4B	(.3,.4,.4) vs. (.3,.42,.4)
5A	(.3,.4,.8) vs. (.3,.41,.8)
5B	(.3,.4,.8) vs. (.3,.42,.8)
6A	(.3,.4,1.4) vs. (.3,.41,1.4)
6B	(.3,.4,1.4) vs. (.3,.42,1.4)
7A	(.5,.7,.4) vs. (.5,.71,.4)
7B	(.5,.7,.4) vs. (.5,.72,.4)

The experimental design is a randomized complete block. Three strategies are employed to eliminate bias in the measurement of the two major response variables, probability of false alarm and probability of detection. First, since classification performance is directly related to the trained discriminant rule, 10 different training functions were calculated for each classification treatment. Each of the training rules was constructed from a randomly selected sample of 30 of the 250 signal ensembles for each class. A jacknife estimation procedure reported in [28] was applied. Misclassifications were tallied after the jacknife procedure was performed for all 30 time series. Third, paired comparison t-tests eliminated any classification performance variability due to different capabilities of the 10 training discriminant rules.

Alternative spectral feature extraction approaches are compared by examining two performance components that define the rate of correct classification: probability of detection and probability of false alarm. These classification performance measures are reported as relative comparisons via paired t-tests for each scenario/classification treatment. Two types of classification performance are reported: discriminant or training classification and test classification. Classification results revealed no significant statistical difference in classification performance between the alternative feature extraction methods for the four treatments with a high (1.4) level of noise standard deviation. However, some interesting results

were obtained for the other 10 treatments shown in Table I. Extracted feature sets for all simulated scenarios were composed of 28 power spectrum, 51 bispectrum, and 25 second-order cumulant spectrum feature variables. For each scenario, the HOS features were most significant in increasing discriminating and classification power. One type of feature extraction vector by itself (power spectrum, bispectrum, or second-order cumulant spectrum) was never as powerful as a combination of feature types. Our simulation results were very encouraging. The substantial improvement gained by including HOS features clearly justified an investigation of the actual data.

Table II shows the marginal contribution to *training* classifications of combining second-order cumulant spectral features (SCUM) to power spectra features (PS), and of combining bispectral features (B) to second-order cumulant and power spectra features. The data reported are relative performance differences of 10 different discriminant rules over 30 classification runs. The statistical significance level reported is for rejecting the null hypothesis of equal performance means. Clearly, combining nonstationary feature information and stationary features improves both false alarm probability and detection probability with extremely high levels of statistical significance. Furthermore, the inclusion of nonlinear information gives better training classification performance with a high level of statistical confidence.

Table II. HOS vs. power spectrum training classification for simulated wear data.

Feature Extraction Method	Percent Change in: False Alarm Probability	Significance Level	Detection Probability	Significance Level
PS and SCUM vs. PS	− 4.9	.0004	+ 4.8	.0018
PS, SCUM, B vs. PS	−10.4	.0001	+11.5	.0001

Clearly, combining HOS features with power spectrum features improves training classification of the simulated scenario data. More important to the evaluation of the feature extraction methods is an estimate of the actual test classification error rate. This measures a feature extraction method's capability to classify *future* time series samples. Tables III and IV, respectively, show the marginal contributions to test classification performance by combining second-order cumulant spectra features with power spectra features, and combining bispectra features with second-order cumulant and power spectra features. The numbers reported represent relative performance difference over 250 runs per classification problem. As before, the significance level reported is for rejecting the null hypothesis of equal performance means.

Table III. Second-order cumulant spectrum and power spectrum vs. power spectrum test classification for simulated wear data.

Simulation Parameters	Percent Change in: False Alarm Probability	Significance Level	Detection Probability	Significance Level
(.3,.7,.4) vs. (.3,.71,.4)	−0.3	.85	+2.7	.0001
(.3,.7,.4) vs. (.3,.72,.4)	−4.5	.0001	+4.6	.0001
(.3,.7,.8) vs. (.3,.71,.8)	+2.7	.01	+2.7	.01
(.3,.7,.8) vs. (.3,.72,.8)	+1.2	.21	+3.5	.0006
(.3,.4,.4) vs. (.3,.41,.4)	−0.1	.90	+2.8	.002
(.3,.4,.4) vs. (.3,.42,.4)	−3.1	.006	+4.5	.01
(.3,.4,.8) vs. (.3,.41,.8)	+3.6	.01	+3.6	.01
(.3,.4,.8) vs. (.3,.42,.8)	+0.2	.97	+2.4	.003
(.5,.7,.4) vs. (.5,.71,.4)	+1.5	.16	+3.1	.0001
(.5,.7,.4) vs. (.5,.72,.4)	+0.8	.45	+5.9	.02

Table IV. Bispectrum, second-order cumulant spectrum and power spectrum vs. power spectrum test classification for simulated wear data.

Simulation Parameters	Percent Change in: False Alarm Probability	Significance Level	Detection Probability	Significance Level
(.3,.7,.4) vs. (.3,.71,.4)	−1.0	.45	+3.4	.0001
(.3,.7,.4) vs. (.3,.72,.4)	−4.9	.0001	+6.8	.0001
(.3,.7,.8) vs. (.3,.71,.8)	+0.2	.88	+0.3	.81
(.3,.7,.8) vs. (.3,.72,.8)	+1.2	.30	+3.9	.001
(.3,.4,.4) vs. (.3,.41,.4)	−1.5	.32	+1.8	.002
(.3,.4,.4) vs. (.3,.42,.4)	−6.5	.0001	+6.2	.002
(.3,.4,.8) vs. (.3,.41,.8)	+3.0	.005	+3.0	.005
(.3,.4,.8) vs. (.3,.42,.8)	+0.07	.01	+1.0	.0009
(.5,.7,.4) vs. (.5,.71,.4)	−0.4	.73	+4.6	.0001
(.5,.7,.4) vs. (.5,.72,.4)	−0.5	.69	+7.0	.004

When the feature information set includes both stationary (power spectrum) and nonstationary (second-order cumulant spectrum) components, better test classification performance is obtained. Significantly, detection performance is increased for all treatments. Additionally, within each pair of treatments, or scenario, a greater change in the phase modulation index parameter is accompanied with an increased false alarm and detection capability. Thus, HOS features appear sensitive to greater changes in phase modulation, which implies that they have an increasing ability to detect more severe wear condition states. Noise does impact classification performance, but the HOS approach still maintains its superiority

over the power spectrum approach. Combining nonlinear (bispectrum) feature information further improves test classification performance. Therefore, modulated signal simulations revealed an increasing marginal benefit for conducting HOS estimation for subsequent feature extraction and input to a linear classifier.

5.2. *Analysis of actual tool wear data*

Results from the simulations were quite encouraging, and we were anxious to see whether the method worked as well on actual rotating machine wear data. For computational purposes, each 3 mil/rev (4 mil/rev) accelerometer time series was divided into appropriate record lengths to avoid the effects of leakage when performing spectral estimation procedures. Examples of "typical" raw accelerometer time series are shown in Figures 7 and 8.

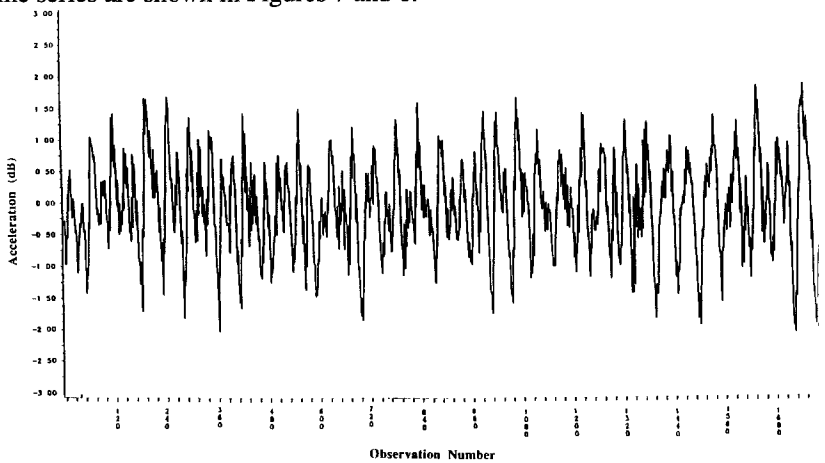

Figure 7. Accelerometer time series for a typical new drill bit.

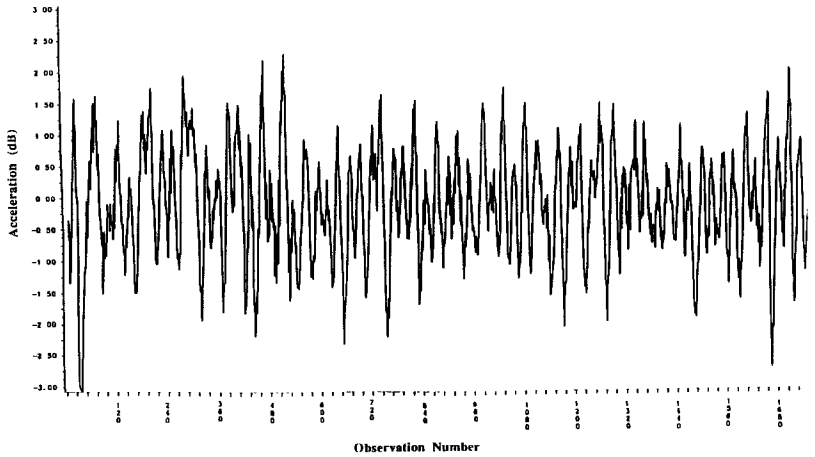

Figure 8. Accelerometer time series for a typical slightly used drill bit.

Power spectrum, cumulant spectrum, and bispectrum estimates are computed over blocks within each time series record. Each spectral estimate is averaged over the block length, and then incorporated into the ensemble average over all samples in its particular class. Visual inspection of each type of ensemble averaged spectral plot gives a preliminary look at which frequency variates are bit class distinguishable. Figures 9 and 10 show, respectively, typical power spectrum and bispectrum estimates (these for NIP stack chip load at 3 mil/rev with new and slightly used bits) that are supplied to the feature extraction algorithm.

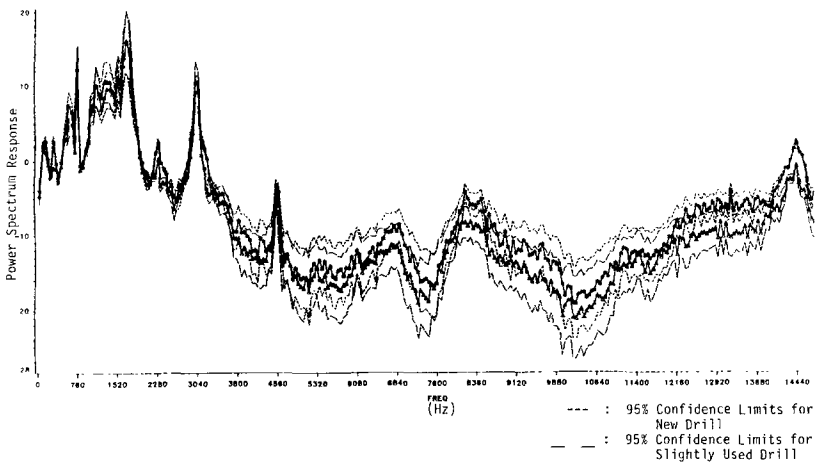

Figure 9. Ensemble averaged power spectrum differences—NIP stack load 3.

Differences in ensemble averaged power spectrum plots for all four stack/chip load cases were quite similar to those shown in Figure 9. Power spectra exhibited the presence of strong spectrum peaks at frequencies below 5 kHz. These peaks occurred at the shaft spindle rotational frequency f_0 and its harmonics, $2f_0$ and $4f_0$, and reflect the periodic cutting forces due to hardness differences of the glass and epoxy material in the circuit board layers. Most of the signal content occurs at these harmonic frequencies; however, there are two frequency ranges that appear to be much more useful as potential wear indicators: frequency values near one-half the fundamental rotational frequency of the drill spindle, $.5f_0$, and between 14–14.5 kHz. Other researchers have noted this subharmonic structure with journal bearings in high speed turbomachinery, sometimes referring to it as a whirl frequency [12], and the differences observed in the power spectra may be due to different frictional forces of a worn drill bit. The higher range of frequency values is near a *torsional resonant* frequency of the drill spindle. It appears that the spindle torsional mode is more strongly excited by new drills than worn drills [18].

The bispectrum estimate in Figure 10 is in the form of difference of chi-square test statistics (this procedure is described in [29]). The bispectrum difference plots clearly show the general regions and the particular frequency interaction *pairs* that are class distinguishable. Significantly different frequency interactions were observed between the harmonics of the fundamental rotational frequency and frequencies greater than 14 kHz. This significantly different frequency structure may be due in part to parametric coupling of the torsional resonant frequency with each of its lower harmonics. Also significant is that for all the stack/load cases, bispectrum estimates are most significantly different in the outer triangle region of the bispectrum principal domain. This provides evidence and motivation for further investigation with cumulant spectrum estimation and feature extraction as it demonstrates that nonstationary generating sources are present in the time series records [30].

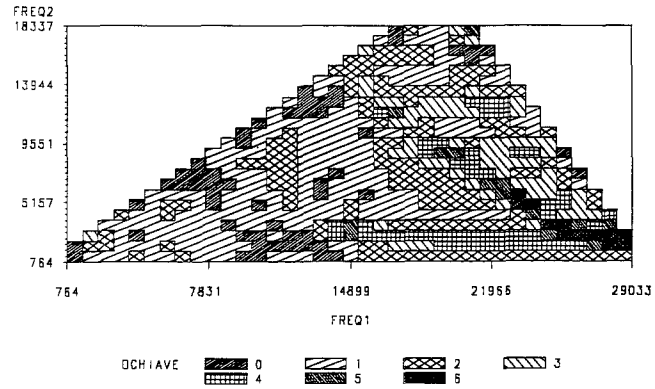

Figure 10. Ensemble averaged bispectrum differences—NIP stack load 3.

Actual feature extraction performance results were obtained for both discrimination and classification. Of the features extracted, second-order cumulant features were the most statistically significant. Further breakout of the actual feature variables is given in [21]. Table V shows total classification performance results (percentage of correctly classified time series). The "combined load" category tested the impact of stack variation, the "combined stack" category tested the impact of load variation, and the "load/stack" category is averaged over all drilling process parameters. PS denotes power spectrum estimates only, 2C denotes the addition of second-order cumulant spectrum estimates to the feature set, and B denotes the addition of bispectrum estimates. The classification schemes used are labelled LDF and QDF, respectively, for linear and quadratic discriminant functions, and NN for a non-parametric classifier based on the nearest-neighbor distance measure.

Table V. Total classification averages (percentage of correctly classified signals) for actual incipient wear data.

Features and Classification	Combined Comb. Load	Datasets Comb. Stack	Homogeneous Datasets Load and Stack
PS and LDF	75.1	79.1	84.5
PS, 2C and LDF	77.9	82.6	83.8
PS, 2C, B and LDF	78.3	84.6	87.8
PS and QDF	74.5	80.5	79.7
PS, 2C and QDF	77.9	81.7	84.0
PS, 2C, B and QDF	86.7	86.7	86.3
PS and NN	75.4	71.6	83.5
PS, 2C and NN	85.4	85.4	86.8
PS, 2C, B and NN	84.1	87.5	90.6

Table V clearly demonstrates the increasing marginal benefit of HOS information for each of the classifiers. Average stack/load QDF classification is degraded with power spectrum features and has no consequential impact on either averaged load or stack classification. However, the impact on averaged total QDF classification accuracy using a full HOS feature set is significant, especially in the combined stack and combined load datasets. For the stack/load case, the false alarm rate increased more than the corresponding increase in detection capability so there was a slight decrease in overall QDF classification performance for this database partition. However, the masking of wear effects due to variations in board construction is not as great with HOS features and QDF classification. In fact, variations in board construction and chip load had no wear masking effect when full HOS feature sets were used with a quadratic classifier. Direct comparisons of 4-nearest neighbor classification results clearly show the increased total classification power of the HOS feature extraction approach.

6. Implementation as an on-line tool

This study clearly demonstrated the ability to correctly classify worn tools using HOS information. In addition to establishing the feature extraction and classification methods, instrumentation must be designed and installed and the modifications made for commanding the tool change when required. IBM has begun work on a prototype monitoring system based on the results obtained in [21] and [20]. Figure 11 shows schematically the elements of a circuit board drill monitoring system based on the techniques presented in this paper.

Accelerometers mounted to each drill spindle measure thrust axis vibration and their signals are selected for processing using a switch. In this example, samples are collected from each sensor once every fourth hit, or once every 1.33 seconds at a peak drilling rate of 3 hits/second. Since drill lifetimes are generally several

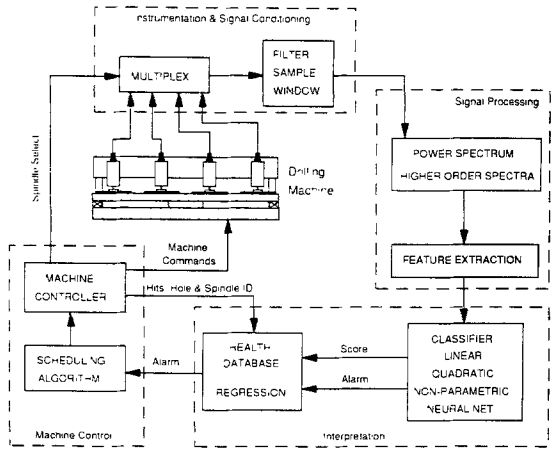

Figure 11. Drill monitoring system.

thousand hits or more, a large number of signals can be obtained for each drill for wear classification purposes. Each sampled signal, tracked by a cell controller to identify spindle location and hole location on the board, is pre-processed to remove noise and to avoid leakage. The power and higher-order spectrum estimation is then performed. The features relevant to wear detection are extracted and input to a classifier. The output of the classifier is either an alarm or a score. To improve the reliability of an alarm, a series of such classifications are combined by regression methods. Finally, a scheduling algorithm determines the next available stopping point for a drill change if the immediate response would unnecessarily slow production. All signal processing, classification, and scheduling steps are accomplished using a single small computer. While computational restrictions might have been significant in the past, modern workstations are capable of performing these steps quite rapidly (for example, calculation of a 2,048 point power spectrum and HOS using a Fast Fourier Transform algorithm can now be accomplished by digital signal processing chips in milliseconds.)

7. Conclusions

Evidence obtained from both simulated and actual incipient wear data clearly supports the hypothesis that HOS features are significant in reflecting class differences. Moreover, the data confirmed that second-order cumulant spectrum estimates *off* the diagonal support lines (i.e., that measure periodic correlations) best characterized incipient faults in the rotating machinery we analyzed. A HOS approach for incipient fault detection has increased discrimination and classification power, and is less sensitive to process and noise conditions than solely a power spectrum approach. Selecting and combining HOS features that capture the nonstationary, aperiodic, and nonlinear characteristics of the cutting forces as the drill bit pen-

etrates the circuit board layers clearly enhances the total classification capability. The evidence clearly advocates for the adoption of a HOS feature fusion approach in a condition monitoring scheme for rotating systems.

This paper is the first work to provide a rigorous study of the HOS approach in detecting incipient faults. The methodology presented in this paper is applicable to many other monitoring systems where the time series records contain nonlinear and nonstationary elements. We are currently applying these techniques to monitoring newly developing cracks in the turbine blades of jet aircraft for the United States Air Force, and other applications come to mind as well.

Most incipient fault detection methods are tested against one specific alternative monitoring system. Clearly, a statistical and experimental design framework provides a more thorough investigation of a monitoring approach. In this paper, we have demonstrated the consistency of the HOS approach through both simulated and actual wear data. A new tool is now available for investigating and exploiting a much wider range of time series characteristics generated by random fault mechanisms. These techniques have great potential for improving maintenance decisions and increasing product quality while reducing operational costs.

Acknowledgments

We are grateful to IBM Corporation for its interest in and contributions to this research.

References

[1] M. Hannon, in *Les Echos*, June 5, 1991.

[2] *Chemical Marketing Research*, January 7, 1991.

[3] J. P. Block, Part I: The ideal hole, *PC Fab*, pp. 33–40, February 1989.

[4] Y. Altinas, In-process detection of tool breakage using time series monitoring of cutting forces, *International Journal of Machine Tools and Manufacture* **28**, pp. 157–172, 1988.

[5] Y. Kashimura, Study of detection of drill wear and breakage (1st report), *Bull. Japan Soc. of Prec. Eng.* **20**, pp. 159–164, 1986.

[6] T. Moore and Z. Reif, Using vibration data to detect drill breakage in high volume, multistation transfer machines, *Society of Manufacturing Engineers Technical Paper No. MS84-908*, 1986.

[7] J. W. Powell, In-process tool sensing, *Carbide and Tool Journal* **20**, pp. 17–23, 1988.

[8] J. Tlusty and Y. S. Tarng, Sensing cutter breakage in milling, *Annals of the CIRP* **37**, pp. 45–50, 1988.

[9] Y. S. Tarng, Study of milling cutting force pulsation applied to the detection of tool breakage, *International Journal of Machine Tools and Manufacture* **30**, pp. 651–660, 1990.

[10] E. Brinksmeier, Prediction of tool fracture in drilling, *Annals of the CIRP* **39**, pp. 97–100, 1990.

[11] S. Jetly, Measuring cutting tool wear on-line: some practical considerations, *Manufacturing Engineering*, pp. 55–60, July 1984.

[12] S. Braun (ed), *Mechanical Signature Analysis*, Academic Press, London, 1986.

[13] Y. Yao, D. Fang, and G. Arndt, Comprehensive tool wear estimation in finish-machining via multivariate time-series analysis of 3-D cutting forces, *Annals of the CIRP* **39**, pp. 57–60, 1990.

[14] M. A. Elbestawi, J. Marks, and A. Papazafiriou, Process monitoring in milling by pattern recognition, *Mechanical Systems and Signal Processing* **3**, pp. 305–315, 1989.

[15] T. I. Liu and S. M. Wu, On-line detection and drill wear, *ASME Journal of Engineering for Industry* **112**, pp. 299–302, 1990.

[16] S. Rangwala and D. Dornfeld, Sensor integration using neural networks for intelligent tool condition monitoring, *ASME Journal of Engineering for Industry* **112**, pp. 219–228, 1990.

[17] S. Braun and E. Lenz, Machine-tool wear monitoring, *Mechanical Signature Analysis*, Academic Press, London, 1986.

[18] C. N. Ramirez, On-line drill condition monitoring by measurement of drill spindle vibration and dynamics, Ph.D. Dissertation, Department of Mechanical Engineering, The University of Texas at Austin, 1991.

[19] C. N. Ramirez and R. J. Thornhill, Drill wear monitoring in circuit board manufacturing I: motivation, methods, and challenges, Submitted for publication, *ASME Journal of Electronic Packaging*, 1991a.

[20] C. N. Ramirez and R. J. Thornhill, Drill wear monitoring in circuit board manufacturing III: an experimental study, Submitted for publication, *ASME Journal of Electronic Packaging*, 1991b.

[21] R. W. Barker, Incipient fault detection using higher-order statistics, Ph.D. Dissertation, Department of Mechanical Engineering, The University of Texas at Austin, 1991.

[22] M. Kendall, *Multivariate Analysis*, Charles Griffen and Co., Ltd., London, 1980.

[23] C. M. Smith, A description of the hardware and software of the power spectral density recognition (PSDREC) continuous on-line reactor surveillance system, *ORNL/TM-8826/VI*, October, 1983.

[24] M. J. Hinich, Higher order cumulants and cumulant spectra, Working Paper, Applied Research Laboratories, The University of Texas at Austin, 1990.

[25] C. L. Nikias and M. R. Raghuveer, Bispectrum estimation: a digital signal processing framework, *Proceedings of the IEEE* **75**, pp. 869–891, 1987.

[26] C. L. Nikias and J. M. Mendel, Workshop on higher-order spectral analysis, IEEE Societies: Controlled Systems, Geoscience and Remote Sensing, Acoustics, and Speech and Signal Processing, Vail, CO, 1989.

[27] J.-L. Lacoume, M. A. Lagunas, and C. L. Nikias (eds), *Proceedings of the International Signal Processing Workshop on Higher-Order Statistics*, Chamrousse, France, July, 1991.

[28] P. A. Lachenbruch, *Discriminant Analysis*, Hafner Press, New York, 1975.

[29] M. J. Hinich, Testing for Gaussianity and linearity of a stationary time series, *Journal of Time Series Analysis* **3**, pp. 169–176, 1982.

[30] M. J. Hinich, Detecting a transient signal by bispectral analysis, Technical Paper ARL-TP-88-99, Applied Research Laboratories, The University of Texas at Austin, 1989.