Identification of statistical patterns in complex systems via symbolic time series analysis

Shalabh Gupta,* Amol Khatkhate,[†] Asok Ray,[‡] Eric Keller[§]

The Pennsylvania State University, University Park, PA 16802, USA

Abstract

Identification of statistical patterns from observed time series of spatially distributed sensor data is critical for performance monitoring and decision making in human-engineered complex systems, such as electric power generation, petrochemical, and networked transportation. This paper presents an information-theoretic approach to identification of statistical patterns in such systems, where the main objective is to enhance structural integrity and operation reliability. The core concept of pattern identification is built upon the principles of *Symbolic Dynamics*, *Automata Theory*, and *Information Theory*. To this end, a symbolic time series analysis method has been formulated and experimentally validated on a special-purpose test apparatus that is designed for data acquisition and real-time analysis of fatigue damage in polycrystalline alloys. © 2006 ISA—The Instrumentation, Systems, and Automation Society.

Keywords: Fault/analysis; Forcasting; Goodness monitoring

1. Introduction

Given a subsystem of a human-engineered complex system, the critical issue is whether its dynamics can be adequately described by a mathematically and computationally tractable model. The presence of uncertainties and chaos may often restrict routine applications of the fundamental laws of physics to model such systems [1]. This issue has motivated the study of complex systems during the last few decades, which is continually gaining importance from the perspectives of both fundamental sciences and technological applications. Specifically, sole reliance on model-based analysis for pattern identification in complex systems has been found to be infeasible because of the difficulties to achieve requisite accuracy and precision of the nonlinear spatial-temporal stochastic models. For example, no existing model can capture the dynamical behavior of fatigue damage at the grain level solely based on the fundamental principles of molecular physics. Moreover, small deviations in the initial conditions and critical parameters of the system may eventually produce large bifurcations and chaotic outputs in the expected dynamical behavior [2]. Despite these difficulties, the key problem—identification of statistical patterns—can be formulated in terms of how the inherent dynamics of a complex nonlinear (and possibly nonstationary) process can be inferred from the data generated from sensing devices and ancillary instrumentation. In essence, time series analysis of observed data is needed for tracking the changes in statistical patterns of the evolving system dynamics in real time [3,4].

Human-engineered complex systems, such as

*E-mail address: szg107@psu.edu

[†]E-mail address: amk303@psu.edu

[‡]Corresponding author. E-mail address: axr2@psu.edu

[§]E-mail address: eek105@psu.edu

electric power generation plants, petrochemical plants, and networked transportation systems, must be designed to be information-intensive and have capabilities for early detection and identification of any anomalous behavior to ensure structural integrity and operation reliability. This objective can be achieved via provision of diverse information sources through a network of sensing devices and ancillary instrumentation, installed at selected spatial locations of the complex system. The information provided by the sensor network enhances the capacity of measuring small changes in the dynamical behavior of the system. For example, the information derived from time series data of sensors at different spatial locations provides the capability to capture small parametric and nonparametric perturbations at localized regions and their global effect on the entire system. These sensors must provide online information and quantitative estimates of damage precursors that cause the observed deviations in the statistical pattern.

Various signal processing tools have been employed to extract useful information from the available time-series data. Technical literature abounds with diverse techniques of pattern recognition (for example, see citations in Ref. [5]); a brief survey of pattern recognition tools for anomaly detection is reported in Ref. [6]. Anomaly detection using symbolic time series analysis (STSA) [7] has been recently reported [8] and a comparative evaluation of this novel analytical method shows its superior performance relative to other existing pattern recognition tools in terms of early detection of small changes in dynamical systems [9,10].

This paper presents the theoretical framework of a novel STSA-based method for identification of statistical patterns and its experimental validation for early detection of fatigue damage in structural materials of human-engineered complex systems. The experimental platform consists of a special-purpose electromechanical test apparatus that is designed for data acquisition and analysis of fatigue damage in polycrystalline alloys. The information, needed for anomaly detection and damage analysis, is derived from a network of distributed and heterogeneous fatigue damage sensing devices (e.g., ultrasonic, mechanical displacement, load cell, accelerometer, and optical microscope).

The paper is organized in six sections including

the present one. Section 2 outlines the procedure for identification of statistical patterns in complex systems. Section 3 introduces the underlying concepts of symbolic time series analysis for anomaly detection [8]. Section 4 describes the experimental apparatus on which the anomaly detection method is validated for early detection fatigue damage. Section 5 discusses the results of early detection and identification of fatigue cracks under cyclic loading. The paper is concluded in Section 6 along with recommendations for future research.

2. The procedure for pattern identification

This section outlines the procedure for identification of statistical patterns in complex dynamical systems. The procedure is based on the system response, excited by selected input stimuli (e.g., persistent excitation), or due to self excitation. The pattern identification of the quasistationary process is recognized as a two-time scale (i.e., *fast* and *slow* time scale) problem in the following sense:

- *Fast time scale* refers to the local behavior of the system, where changes in the patterns of the process dynamics are assumed to be insignificant. It is assumed that the statistical distribution of the system dynamics is stationary at the fast time scale [11], i.e., no statistical changes occur during this period.
- *Slow time scale* refers to the long-term behavior of the system, where the patterns of the process dynamics might deviate from those under the nominal conditions. It is assumed that any observable nonstationary behavior pattern is associated with changes occurring on the slow time scale. The pattern changes, if they occur, develop gradually on the slow time scale and may lead to accumulation of anomalies and faults.

The notion of fast and slow time scales is dependent on the specific application and operating environment. In general, the time span in the fast scale, over which data series are collected, is a tiny (i.e., several orders of magnitude small) interval in the slow time scale. While the statistics of the process dynamics are assumed to be locally stationary in these time intervals, it may exhibit nonstationary statistics at different slow-time epochs.

Statistical pattern identification in the above setting is categorized by two inter-related problems:

- the forward (or analysis) problem and
- the inverse (or synthesis) problem.

The primary objective of the forward problem is to identify the patterns in the process dynamics and to track statistical changes occurring over the entire span of slow time. Specifically, the forward problem aims at detecting the deviations in the statistical patterns in the time series data, generated at slow-time epochs, from the nominal behavior pattern. Solutions of the forward problem require the following steps:

- generation of multiple sensor time series data sets under self excitation, or under external stimuli, spanning the system behavior under different operational conditions and
- analysis of the data sets to characterize the system's behavioral pattern based on certain features at different slow-time epochs as the process evolves.

This paper addresses the forward problem and experimentally validates the STSA-based pattern identification method on mechanical structures, where the source of possible anomalies is fatigue crack damage.

The inverse (or synthesis) problem infers the anomalies based on the observed time series data and the information on anomaly characterization, generated in the forward problem. The major role of the inverse problem is to provide information for monitoring and control of the system behavior. The inverse problem is currently under active research and details on the solution method is reported in Ref. [8]; the results on the inverse problem will be reported in a forthcoming publication.

3. STSA

This section briefly describes the technique of STSA [7] for anomaly detection in real time [8]. The STSA method of anomaly detection makes use of the vector information on (finitely many) states of an automaton generated by partitioning the space over which the time-series data evolve. The steps for STSA are as follows:
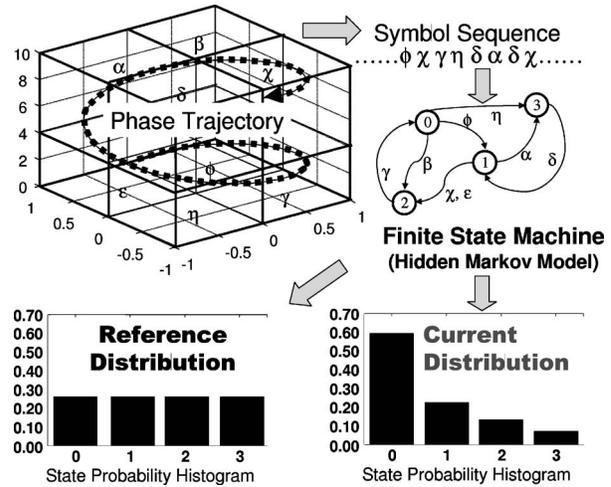
- transformation of time-series data from the

Fig. 1. An example of space partitioning.

continuous domain to the symbolic domain;
- construction of the finite state machine structure based on symbolic sequences and calculation of its state probability vector at various time epochs on the slow time scale; and
- statistical pattern identification based on the deviation of this vector information from the nominal condition.

3.1. Transformation from continuous to symbolic domain

The time-series data are transformed into a symbol sequence by partitioning a compact region Ω in the phase space, over which the data evolves, into finitely many discrete blocks as shown in Fig. 1. Let $\{\Phi_1, \Phi_2, \ldots \Phi_m\}$ be a partitioning of the region Ω, such that it is exhaustive and mutually exclusive set, i.e.,

$$\bigcup_{j=1}^{m} \Phi_j = \Omega \quad \text{and} \quad \Phi_j \cap \Phi_k = \varnothing \quad \forall j \neq k$$

(1)

Each block Φ_j is labelled as the symbol $\sigma_j \in \Sigma$, where the symbol set Σ is called the *alphabet set* consisting of m different symbols. As the system evolves in time, it travels through various blocks in its phase space and the corresponding symbol $\sigma_j \in \Sigma$ is assigned to it, thus converting a data sequence to a symbol sequence $\ldots \sigma_{i_1} \sigma_{i_2} \ldots \sigma_{i_k} \ldots$. Fig. 1 exemplifies the partitioning of the phase space where each block is assigned a particular symbol such that a symbol sequence is generated

from the phase space at a given slow time epoch. Thus, the symbol sequences represent coarse graining of the trajectories' time evolution [11]. Once the symbol sequence is obtained, the next step is the construction of the finite state machine and calculation of the state visiting probabilities to generate state probability vector as shown in Fig. 1. The details of these steps are explained in coming sections.

3.2. Wavelet space partitioning

Several partitioning techniques have been reported in literature for symbol generation [3,12,13], primarily based on symbolic false neighbors. These techniques rely on partitioning the phase space and may become cumbersome and extremely computation intensive if the dimension of the phase space is large. This paper has adopted a wavelet-based partitioning approach [14] because wavelet transform [15] largely alleviates these shortcomings and is particularly effective for noisy data from high-dimensional dynamical systems. In this method, called *wavelet space partitioning* [8], the time series data are first converted to the wavelet transform data, where wavelet coefficients are generated at different scales and time shifts. The graphs of wavelet coefficients versus scale, at selected time shifts, are stacked starting with the smallest value of scale and ending with its largest value and then back from the largest value to the smallest value of the scale at the next instant of time shift. The arrangement of the resulting *scale series* data in the wavelet space is similar to that of the time series data in the phase space. The wavelet space is partitioned with alphabet size $|\Sigma|$ into segments of coefficients on the ordinate separated by horizontal lines such that the regions with more information are partitioned finer and those with sparse information are partitioned coarser. In this approach, the maximum entropy is achieved by the partition that induces uniform probability distribution of the symbols in the symbol alphabet. Shannon entropy [16] is defined as

$$S = -\sum_{i=1}^{|\Sigma|} p_i \log(p_i), \qquad (2)$$

where p_i is the probability of the ith state and summation is taken over all possible states. Uniform probability distribution is a consequence of the maximum entropy partitioning. Fig. 2 shows
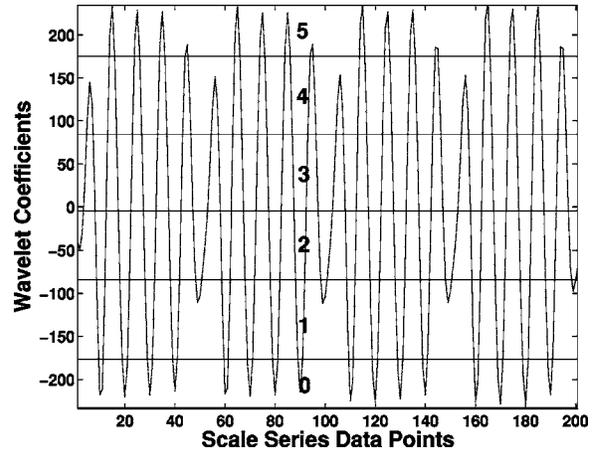

Fig. 2. Maximum entropy partitioning in the wavelet domain.

an example of the maximum entropy partitioning in the wavelet space for alphabet size $|\Sigma|=6$, where the partitioned regions are marked by symbols ranging from 0 to 5.

3.3. State machine construction

The partitioning as described in the previous subsection is performed at (slow-time) epoch t_0 of the nominal condition having zero anomaly measure. A finite state machine is then constructed, where the states of the machine are defined corresponding to a given *alphabet* Σ and window length D. The alphabet size $|\Sigma|$ is the total number of partitions while the window length D is the length of consecutive symbol words forming the states of the machine [8]. The states of the machine are chosen as all possible words of length D from the symbol sequence, thereby making the number n of states to be equal to the total permutations of the alphabet symbols within word of length D, (i.e., $n \leqslant |\Sigma|^D$) where some states may be forbidden and have zero probability of occurrence. The choice of $|\Sigma|$ and D depends on specific experiments, noise level and also the available computation power. A large *alphabet* may be noise sensitive while a small alphabet could miss the details of signal dynamics. Similarly, a high value of D is extremely sensitive to small signal distortions but would lead to larger number of states requiring more computation power. Using the symbol sequence generated from the time-series data, the state machine is constructed on the
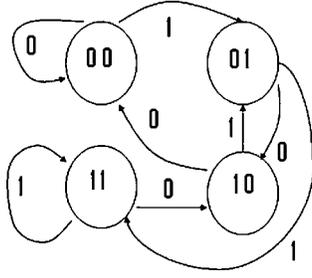
Fig. 3. Finite state automaton with $D=2$ and $\Sigma=\{0,1\}$.

principle of sliding block codes [17] as explained below.

The window of length D on the symbol sequence $\ldots\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_k}\ldots$ is shifted to the right by one symbol, such that it retains the last $(D-1)$ symbols of the previous state and appends it with the new symbol σ_{i_ℓ} at the end. The symbolic permutation in the current window gives rise to a new state. The machine constructed in this fashion is called a D-Markov machine [8] because of its Markov properties.

Definition 3.1: A symbolic stationary process is called D-Markov if the probability of the next symbol depends only on the previous D symbols, i.e., $P(\sigma_{i_0}/\sigma_{i_{-1}}\ldots\sigma_{i_{-D}}\sigma_{i_{-D-1}}\ldots)=P(\sigma_{i_0}/\sigma_{i_{-1}}\ldots \times \sigma_{i_{-D}})$.

The finite state machine constructed above has D-Markov properties because the probability of occurrence of symbol σ_{i_ℓ} on a particular state depends only on the configuration of that state, i.e., previous D symbols. For example, if $\Sigma=\{0,1\}$, i.e., $|\Sigma|=2$ and $D=2$, then the number of states is $n\leqslant|\Sigma|^D=4$; and the possible states are $Q=\{00,01,10,11\}$, some of which may be forbidden. Fig. 3 shows the construction of the finite state machine for the above example where forbidden states, if any, will have zero probability of occurrence.

Once the partitioning alphabet Σ and word length D are determined at the nominal condition (time epoch t_0), they are kept constant for all (slow time) epochs $\{t_1,t_2,\ldots t_k\ldots\}$, i.e., the structure of the machine is fixed at the nominal condition. That is, the partitioning and the state machine structure, generated at the nominal condition serve as the reference frame for data analysis at subsequent time epochs. The states of the machine are marked with the corresponding symbolic word permutation and the edges joining the states indicate the occurrence of an event σ_{i_ℓ}. The occur-

rence of an event at a state may keep the machine in the same state or move it to a new state.

Definition 3.2: The probability of transitions from state q_j to state q_k belonging to the set Q of states under a transition $\delta\colon Q\times\Sigma\to Q$ is defined as

$$\pi_{jk}=P(\sigma\in\Sigma\,|\,\delta(q_j,\sigma)\to q_k);\sum_k\pi_{jk}=1.\quad(3)$$

Thus, for a D-Markov machine, the irreducible stochastic matrix $\Pi\equiv[\pi_{ij}]$ describes all transition probabilities between states such that it has at most $|\Sigma|^{D+1}$ nonzero entries. The left eigenvector $\mathbf{p}$ corresponding to the unit eigenvalue of Π is the state probability vector under the (fast time scale) stationary condition of the dynamical system [8]. On a given symbol sequence $\ldots\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_l}\ldots$ generated from the time series data collected at slow time epoch t_k, a window of length (D) is moved by keeping a count of occurrences of word sequences $\sigma_{i_1}\ldots\sigma_{i_D}\sigma_{i_{D+1}}$ and $\sigma_{i_1}\ldots\sigma_{i_D}$ which are, respectively, denoted by $N(\sigma_{i_1}\ldots\sigma_{i_D}\sigma_{i_{D+1}})$ and $N(\sigma_{i_1}\ldots\sigma_{i_D})$. Note that if $N(\sigma_{i_1}\ldots\sigma_{i_D})=0$, then the state $q\equiv\sigma_{i_1}\ldots\sigma_{i_D}\in Q$ has zero probability of occurrence. For $N(\sigma_{i_1}\ldots\sigma_{i_D})\neq0$, the transitions probabilities are then obtained by these frequency counts as follows:

$$\pi_{jk}\equiv P[q_k|q_j]=\frac{P[q_k,q_j]}{P[q_j]}=\frac{P(\sigma_{i_1}\ldots\sigma_{i_D}\sigma)}{P(\sigma_{i_1}\ldots\sigma_{i_D})}$$

$$\Rightarrow\pi_{jk}\approx\frac{N(\sigma_{i_1}\ldots\sigma_{i_D}\sigma)}{N(\sigma_{i_1}\ldots\sigma_{i_D})}\quad(4)$$

where the corresponding states are denoted by $q_j\equiv\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_D}$ and $q_k\equiv\sigma_{i_2}\ldots\sigma_{i_D}\sigma$.

The time series data under the nominal condition (set as a benchmark) generates the *state transition matrix* Π^{nom} that, in turn, is used to obtain the *state probability vector* $\mathbf{p}^{\text{nom}}$ whose elements are the stationary probabilities of the state vector, where $\mathbf{p}^{\text{nom}}$ is the left eigenvector of Π^{nom} corresponding to the (unique) unit eigenvalue [18]. Subsequently, state probability vectors $\mathbf{p}^1,\mathbf{p}^2,\ldots\mathbf{p}^k\ldots$ are obtained at slow-time epochs $t_1,t_2,\ldots t_k\ldots$ based on the respective time series data. Machine structure and partitioning should be the same at all slow-time epochs. A deviation from the nominal behavior (e.g., the statistical distribution at the nominal condition) is called an anomaly and these anomalies are characterized by a scalar

called *anomaly measure* ($\mathcal{M}$). In the context of fatigue crack phenomena in mechanical structures, the anomaly measure is based on the following assumptions:

- Assumption 1: The evolution of damage $\mathcal{M}(t)$ is an irreversible process, i.e., with zero probability of self healing. This assumption implies the following conditions for all time $t \geq 0$,

  (i) $\mathcal{M} \geq 0$,
  (ii) $d\mathcal{M}/dt \geq 0$.

- Assumption 2: The damage accumulation at a slow time epoch t, when the dynamical system has reached a quasisteady state equilibrium, is a function of the entire path taken to reach that state.

Let us digress in the context of fatigue damage. Although the crack length has been traditionally defined by a straight line joining the starting point to the tip of the crack, the actual crack follows a complicated path, possibly fractal in ductile materials, to reach a particular point. Therefore, the above assumption 2 implies that the anomaly measure should be determined from the actual path traversed and not just the end points. Accordingly, anomaly measure at a slow-time epoch t_k is defined as:

$$\mathcal{M}_k \equiv \sum_{l=1}^{k} d(\mathbf{p}^l, \mathbf{p}^{l-1})^{\alpha};$$

$$d(\mathbf{x}, \mathbf{y}) = \left(\sum_{j=1}^{n} |x_j - y_j|^{\alpha} \right)^{1/\alpha}, \tag{5}$$

where the exponent $\alpha \in [1, \infty)$ depends on the desired sensitivity to small deviations. Large values of α suppress small changes in the signal profile which might be due to noise and spurious oscillations in the signal. Therefore, the choice of alpha is a trade-off between suppression of small fluctuations due to noise and those due to the actual changes in the signal profile resulting from damage growth. In this paper, the exponent is chosen to be $\alpha=2$ implying energy equivalence of the anomaly measure.

4. Experimental validation on a test apparatus

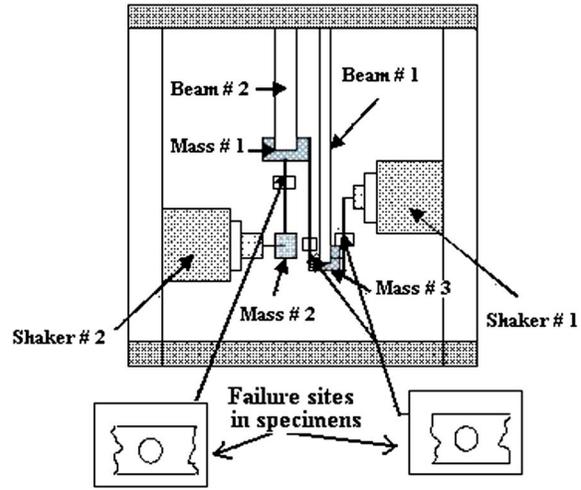This section validates the concept of STSA-based pattern identification for detection of fatigue

Fig. 4. Schematic diagram for the test apparatus.

damage in mechanical structures. With the aim of investigating changes in statistical patterns due to fatigue damage and consequent decision making for damage reduction, a laboratory apparatus has been designed to introduce fatigue damage in critical components [19]. These components are designed to break in a reasonably short period of time to enhance the speed of experiments. From this perspective, the design requirements of the test apparatus include:

(i)   operability under cyclic loading with multiple sources of input excitation;
(ii)   damage accumulation in test specimens (at selected locations) within a reasonable period of time with negligible damage in other components of the test apparatus; and
(iii) accommodation of multiple failure sites for comparative evaluation of structural damage at various spacial locations.

4.1. Description of the test apparatus

The test apparatus is designed and fabricated as a three-degree of freedom (DOF) mass-beam structure excited by oscillatory motion of two shakers. A schematic diagram of the test apparatus and the instrumentation is shown in Fig. 4; dimensions of the pertinent components are listed in Table 1.

The test apparatus is logically partitioned into two subsystems: (i) *the plant subsystem* consisting of the mechanical structure including the test

Table 1
Structural dimensions of the test apparatus.

Component	Material	Length (mm) Mass (kg) (length × width × thickness)
Mass 1	Mild steel	1.0
Mass 2	Aluminium 6061-T6	0.615
Mass 3	Mild steel	2.2
Beam 1	Mild steel	800 × 25.4 × 12.7
Beam 2	Aluminium 6061-T6	711.2 × 22.2 × 11.1
Specimens	Aluminium 6061-T6	203.2 × 22.2 × 11.1

Fig. 5. Side view of failure site on the beam specimen.

specimens to undergo fatigue crack damage, actuators and multiple sensors and (ii) *the instrumentation and control subsystem* consisting of computers, data acquisition and processing, and communications hardware and software. The sensors include: two peizoelectric accelerometers, two linear variable displacement transducers (LVDT), and two load cells for force measurement.

Two of the three major DOFs are directly controlled by the two actuators, shaker No. 1 and shaker No. 2, and the remaining DOF is observable via displacement and acceleration measurements of the three vibrating masses: mass No. 1, mass No. 2, and mass No. 3. The inputs to the multivariable mechanical structure are the forces exerted by the two actuators; and the outputs to be controlled are the displacements of mass No. 2 and mass No. 3.

The three test specimens in Fig. 4 are representatives of plant components, which are subjected to fatigue crack damage. The mechanical structure is excited at one or more of the resonant frequencies so that the critical component(s) can be subjected to different levels of cyclic stresses with no significant change in the external power injection into the actuators. The excitation force vector, generated by the two actuators, serves as the inputs to the multi-DOF mechanical structure to satisfy the requirement No. 1. The failure site in each specimen, attached to the respective mass is a circular hole of diameter 8.5 mm as shown in Fig. 5. The test specimens are thus excited by different
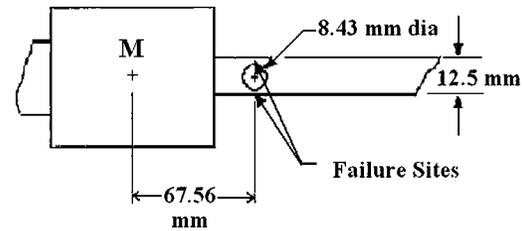
levels of cyclic stresses as two of them are directly affected by the vibratory inputs while the remaining one is subjected to resulting stresses, thus functioning as a coupling between the two vibrating systems. In the present configuration, three test specimens are identically manufactured and their material is 6061-T6 aluminum alloy. In future research, different materials will be selected for individual specimens that may also undergo different manufacturing procedures.

4.2. Hardware implementation and software structure

The electromechanical fatigue damage test apparatus is interfaced with **Keithley** Data Acquisition Boards (DAS 17/18 STDA) having 6 analog/digital (A/D) channels and 4 digital/analog (D/A) channels. Data acquisition is carried out with a sampling rate at 500 Hz for monitoring and control. The time-series data for statistical pattern recognition are decimated as needed. The real-time instrumentation and control subsystem of this test apparatus is implemented on a Pentium personal computer (PC) platform. The software runs on the **Real-Time Linux** Operating System and is provided with A/D and D/A interfaces to the amplifiers serving the sensors and actuators of the test apparatus.

The control software performs real-time communication tasks, in addition to data acquisition and built-in tests (e.g., limit checks and rate checks). The data acquisition is an interrupt service routine (ISR) for the direct memory access (DMA) completion interrupt. The A/D board is initialized to take 12 readings per frame. The DMA controller on the PC motherboard is programmed to read 12 single 16-bit words and store them sequentially in a given memory location for each transfer. When a reading is taken, the result is put into a first in, first out (FIFO) on the A/D board and a DMA request is issued. The DMA
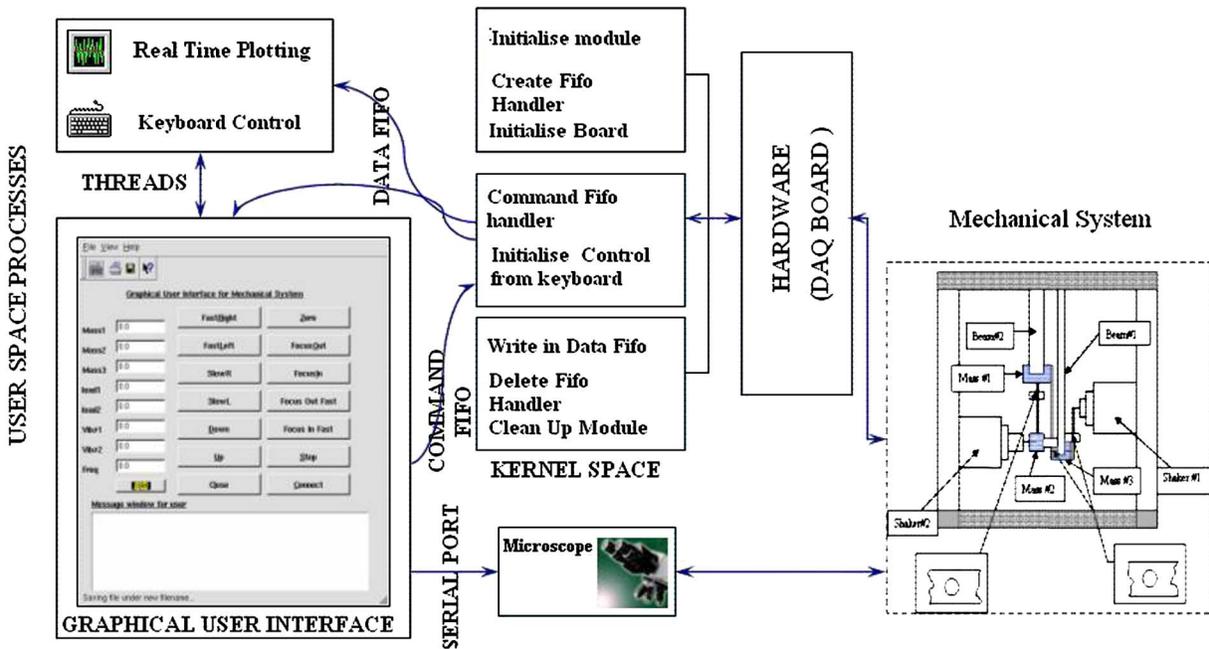
Fig. 6. Graphical user interface of control software.

controller on the motherboard retrieves the data and stores it in the system random access memory in sequence. After every 12th reading is stored, the DMA controller asserts a signal that is looped back to an interrupt line by the A/D board. At this point, control is given to the ISR. As the DMA controller and A/D board are initialized to work together, no periodic programming of either element is necessary for subsequent data acquisition.

Additional hardware devices consist of ultrasonic sensors that provide information about the microstructural changes that take place as the fatigue crack damage evolves. The user space for ultrasonic data acquisition and processing consists of a *c* program that connects through the *Data* FIFO with the data acquisition kernel of the RTLinux operating system and also stores the data to a text file for off-line analysis. There is another program that connects through the *Command* FIFO and instructs the kernel to start or stop the system, increase or decrease the gain, and increase or decrease the excitation frequency. Other additional features of the software are a graphical user interface as shown in Fig. 6 both for real-time data display as well as for connecting to the stage motion of the optical microscope.

4.3. Real time implementation

This subsection presents real-time implementation of STSA-based fatigue damage detection in the laboratory environment. The nominal condition is chosen after the start of the experiment at a time epoch t_0 when the system attains the steady state and is assumed to be in a healthy condition. The states of the automaton (see Figs. 1 and 3) are fixed in advance using *a priori* determined values of the parameters: alphabet size $|\Sigma|$ and window length D. The algorithms for partitioning and machine construction are based on the time-series data at the nominal condition t_0. The resulting information (i.e., the partition and the state probability vector at this nominal condition) is stored for computation of anomaly measures at future slow-time epochs, $t_1, t_2, \ldots, t_k, \ldots$ that are separated by uniform or nonuniform intervals of time. The time series data of multiple sensors are written on text files so that the STSA algorithm can read the data from the text files to calculate the anomaly measure at those time epochs. The algorithm is computationally fast for real-time execution and the results can be plotted on the screen such that the plot updates itself with the most recent anomaly measure at that particular time epoch. This procedure allows on-line condition monitoring at any

time and is capable of issuing early warnings.

4.4. Generation of fatigue crack damage

The mechanical system in Fig. 4 is persistently excited near resonance so as to induce a stress level that causes fatigue damage, ultimately leading to failure. The applied stress is dominantly flexural (i.e., bending) in nature and the amplitude of oscillations is symmetrical about the zero mean level (i.e., a reversed stress cycle [20]). Under cyclic loading, the specimens undergo fatigue cracking where the far-field stress is elastic and plasticity is only localized near the crack site. The fatigue damage occurs at a time scale that is (several order of magnitude) slow relative to the fast time scale dynamics of the vibratory motion and eventually leads to a catastrophic failure. Close observation indicates that fatigue failure develops in the following sequence: (i) the repeated cyclic stress causes incremental crystallographic slip and formation of persistent slip bands; (ii) gradual reduction of ductility in the strain-hardened areas results in the formation of submicroscopic cracks; and (iii) the notch effect of the submicroscopic cracks concentrates stresses to form a large crack which grows till complete fracture occurs. Crack initiation may occur at a microscopic inclusion or at site(s) of local stress concentration. In this experimental apparatus, the sites of stress concentration are localized by creating a hole in each of the three specimens. Since the mechanical structure of the test apparatus consists of beams and masses, the underlying dynamics can be approximated by a finite set of first order coupled differential equations with parameters of damping and stiffness. The damping coefficients are very small and the *stiffness* constants slowly change due to the evolving fatigue crack.

4.5. Sensors for damage detection

The apparatus is equipped with multiple sensors for damage detection including ultrasonic flaw detectors and an optical microscope for localized damage sensing. The advantages of using ultrasonic transducers over optical microscope include the ease of installation at the desired damage site and detection of early anomalies before the onset of widespread fatigue crack propagation. Nevertheless, the optical microscope serves to calibrate the *ultrasonic flaw detectors*.
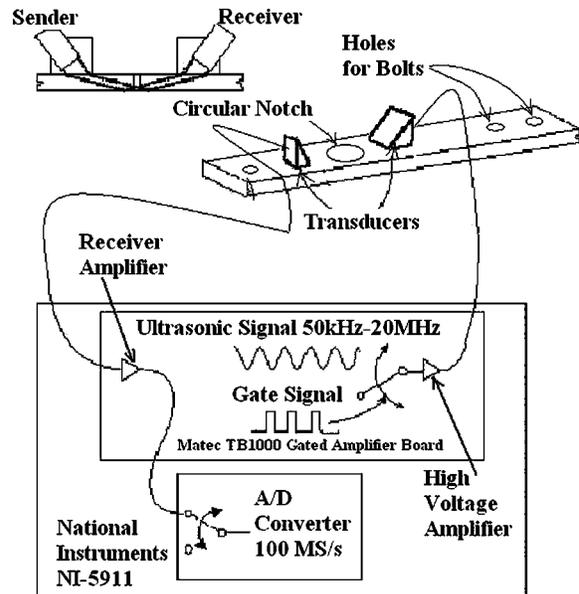

Fig. 7. Ultrasonic flaw detection scheme.

(a) Ultrasonic flaw detector—The ultrasonic flaw detector functions by emitting high frequency ultrasonic pulses that travel through the specimen to the receiver transducers [21]. A piezoelectric transducer is used to inject ultrasonic waves in the specimen and a single receiver transducer is placed on the other side of the circular notch to measure the transmitted signal. The ultrasonic waves produced were 5 MHz sine wave signals and they were emitted during a very short portion of every load cycle.

The sender and receiver ultrasonic transducers are placed on two positions, above and below the central notch, so as to send the signal through the region of crack propagation and receive it on the other side, as shown in Fig. 7. Since material characteristics (e.g., voids, dislocations, and short cracks) influence the ultrasonic impedance, minute damage in the specimen is likely to change the signature of the signal at the receiver end. Therefore, the signal can be used to capture small changes during the early stages of fatigue damage, which may not be possible to detect by an optical microscope. Prior to the appearance of a crack on the specimen surface, deformations (e.g., dislocations and short cracks) inside the specimen may have already caused detectable attenuation and/or distortion of the ultrasonic waves. Therefore, the ultrasonic sensors are utilized for generating the localized information of the growth of fatigue damage.

Table 2
Sensor mountings.

Sensor	Location
LVDT	Mass 1, mass 3
Accelerometer	Mass 1, mass 3
Load cell	Shaker 1, shaker 2
Ultrasonic transducers	Specimen between mass 1 and mass 3

(b) Optical microscope—The traveling optical microscope, shown as part of the schematic in Fig. 6, provides direct measurements of the visible part of a surface crack. The microscope can be shifted from left to right side and vice versa to track the crack tip on the specimen.

(c) Accelerometers, load cells, and displacement transducers — As the body of the accelerometer is subjected to vibrations, the embedded piezoelectric crystal is continually compressed and stretched. The fluctuating force, which is proportional to the instantaneous acceleration, generates time-dependent electric charge that, in turn, is converted to a voltage signal. The load cell functions on a similar principle and is primarily used for measuring the time-dependent forces exerted on the vibrating structure. LVDT's are also mounted on the test apparatus to measure the instantaneous positions of the masses. These sensors are useful for analyzing the effects of local fatigue damage on the global performance of the vibratory system. The locations of various sensors on the test apparatus are listed in Table 2. Raw time series data plots for four sensors (ultrasonic, LVDT, accelerometer, and load cell) are shown in Fig. 8.

It is observed that the crack always starts at the stress-concentrated region on the surface near the center of the circular notch but the exact location of the origin of the crack can be treated as a random event. Formation of very small cracks is difficult to detect and model due to large material irregularities. This paper focuses on integrating the information generated from both the ultrasonic sensors and the sensors measuring the global variables like accelerometers, LVDTs, and load cells for characterization of fatigue damage in the small crack regime.

5. Results and discussion on anomaly detection

STSA method has demonstrated superior performance (in terms of early detection of evolving
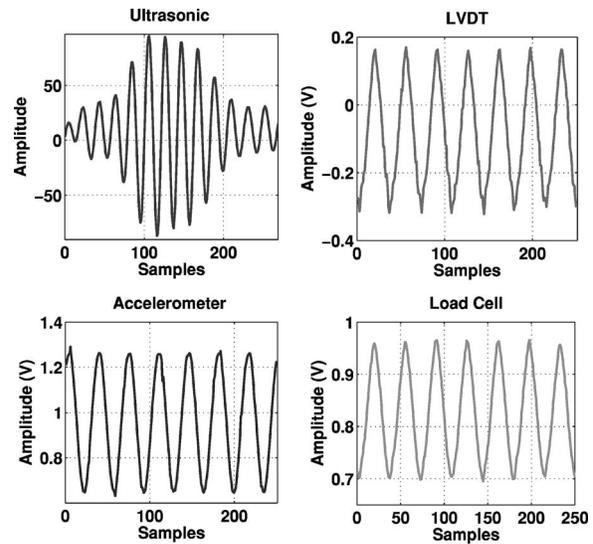

Fig. 8. Raw time series data for four different sensors.

anomalies, and robustness) relative to other existing pattern recognition tools such as principal component analysis and artificial neural networks, in both electronic systems [10] and mechanical systems [9]. This section describes the experimental validation of the symbolic dynamic tools for early detection and quantification of fatigue damage based on time-series data, generated from the sensors mounted on the test apparatus as shown in Fig. 4. In the experiments, both shakers are excited by a sinusoidal input of amplitude 0.85 V and frequency 14.09 Hz (89 rad/s) throughout the run of each experiment. The time-series data from displacement, accelerometer, load cell and ultrasonic sensors, mounted on the experimental apparatus (see Table 2), were collected from the start of the experiments after the system response attained the stationary behavior. The results shown in Fig. 9 are based on the analysis of data collected from sensors mounted on mass No. 1 and shaker No. 1. The time-series data for each sensor were saved for a total period of ~90 kc in 110 files. 1 min of time-series data was stored in each file for all sensors. The total life of the specimen was ~110 min corresponding to ~90 kc. The data set at the beginning stage of experiments served as the reference point representing the nominal behavior of the dynamical system. The anomaly measure at the nominal condition was chosen to be zero and was subsequently updated at approximately 3 min intervals.
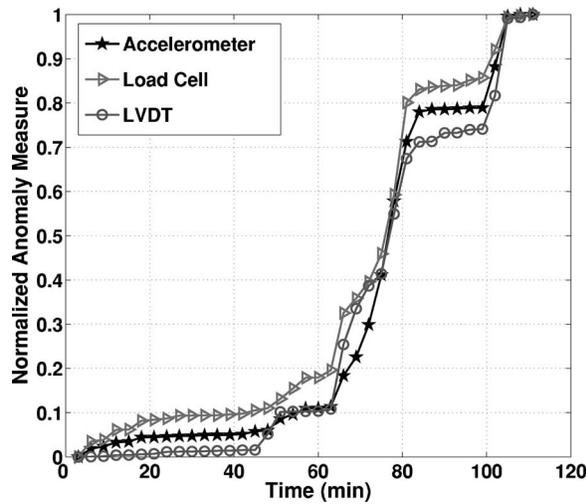
Fig. 9. Normalized anomaly measure plots derived from accelerometer, load cell, and displacement sensors.

Fig. 10. Normalized anomaly measure derived from ultrasonic sensors.

The alphabet size, window length, and the wavelet basis were chosen as $|\Sigma|=8$, $D=1$ (see Sec. 3) and the wavelet basis *"gaus2"* [22], respectively, for time-series data sets of all sensors (displacement, accelerometer, load cell, and ultrasonic). Absolute values of the wavelet scale series data (see Sec. 3.2) were used to generate the partition because of the symmetry of the data sets about their mean. Increasing the value of $|\Sigma|$ further did not improve the results and increasing the value of depth D created a large number of states of the finite state machine, many of them having very small or zero probabilities. The finite state machine constructed with the choice of the parameters $|\Sigma|=8$ and $D=1$ has only eight states and it was able to capture early anomalies. The algorithm is computationally very fast relative to the evolution of fatigue damage. The wavelet basis of gaus2 provided better results than many other wavelets of the Daubechies family [15] because it closely matches the shape of the sinusoidal signals.

The three plots in Fig. 9 represent the evolution of anomaly measure derived from the time-series data of accelerometer, load cell and displacement sensors, respectively. These sensors measure the changes in the global performance of the system due to evolving fatigue damage in a localized region of the system. Fatigue crack growth in the specimen connecting mass No. 1 and mass No. 3 (see Table 2) changes the stiffness of the material. Consequently, the dynamics of the system are al-
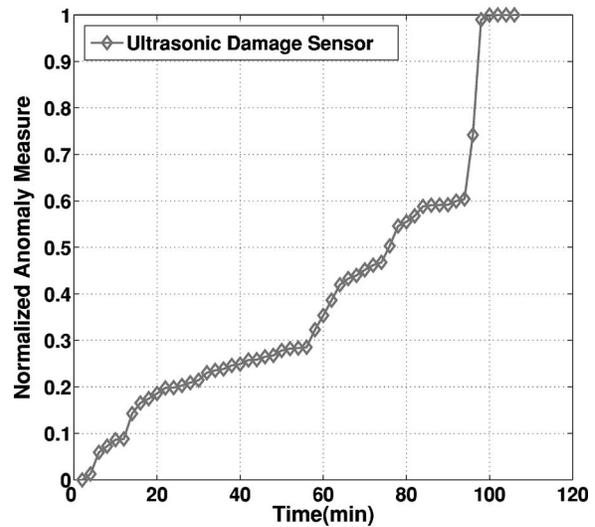
tered because of coupling between different components of the apparatus (see Fig. 4). As a result, the time-series data of these sensors record changes in the statistical patterns as shown in Fig. 9. The three plots in Fig. 9 are normalized by dividing the anomaly measure values with the respective maximum for comparative evaluation.

Fig. 10 exhibits the normalized anomaly measure plot derived from the time-series data of ultrasonic sensors that are mounted on the specimen connecting mass No. 1 and mass No. 3 (see Table 2). This specimen is subjected to a relatively higher load because of resonance and synchronization of two shakers and therefore has higher probability of failure. Ultrasonic sensors capture the effects of small fatigue growth in the localized region of the mechanical structure. The growth of multiple small cracks inside the specimen causes attenuation of the ultrasonic waves. The crack propagation stage starts as multiple small cracks coalesce into a single large crack. At this point, a rapid change is observed in the profile of the ultrasonic signal. Finally, the specimen breakage causes complete attenuation of the received ultrasonic signal because of high impedance of the air gap.

As seen in Fig. 10, ultrasonic sensors capture slow progression of fatigue damage in the localized region of the mechanical structure from the very beginning. The sharp change in the slope at ~58 min indicate the transition from the crack

initiation phase to crack propagation phase. The results from displacement, load cell, and accelerometer sensors, as seen in the three plots of Fig. 9 show similar trends for all three sensors. These plots show little change in the anomaly measure during the crack initiation phase because small fatigue cracks do not significantly affect the global performance; subsequently, the effects of fatigue damage on the overall system behavior become noticeable. The three plots in Fig. 9 exhibit a slope change at ~45 min. At this point, the stiffness change in the specimen caused major change in the dynamics of the system as recorded by these sensors. A change in the stiffness constant of the material alters the natural frequency of vibrations and thereby affecting the global dynamics of the system. These results for localized damage and its effect on the global performance of the system indicate that a control strategy can be built to take predictive actions, based on the derived damage information, for fault mitigation and control. In this experiment, the STSA method captured the gradual progression of faults much earlier than the occurrence of catastrophic failure. This is of paramount practical importance as it can provide ample time for the hierarchical supervisory control system to execute decision and control laws for life extension without significant loss of performance [19]. This is an area of future research.

It is emphasized that the anomaly measure is relative to the nominal condition of that particular data set and may not represent damage in the absolute sense. Any value of anomaly measure greater than zero just indicates some deviation from the nominal condition and it signifies that some small faults might have occurred inside the specimen. However, inferring anomalies and formulation of a decision and control law for life extension is an inverse problem and is a topic of future work.

6. Summary, conclusions, and future work

This paper presents the concept and experimental validation of a novel method for identification of statistical patterns in complex dynamical systems. The underlying principle is based on STSA [7,8] of observed process variable(s), which is built upon the concepts of *symbolic dynamics, automata theory*, and *information theory*. The histograms on state probability distribution are generated from the observed time-series data to serve as

patterns of the evolving behavior change resulting from stiffness reduction due to fatigue damage. The capability of the proposed STSA method for identification of behavior patterns is experimentally validated on a special-purpose laboratory apparatus by a demonstration of how fatigue damage information can be extracted in real time from displacement, accelerometer, load cell, and ultrasonic sensor signals.

The proposed method of statistical pattern identification is useful for performance monitoring and decision making in human-engineered complex systems. The main features are summarized below:

- Information extraction from observed time series data in real time;
- detection and identification of gradually evolving anomalous behavior due to progression of faults; and
- early warnings on incipient faults to avert catastrophic failures.

The reported work is a step towards building a reliable instrumentation system for early detection of faults in human-engineered complex systems, such as electric power generation, petrochemical, and networked transportation. Further research is necessary before its usage in industry. The utilization of the information provided by anomaly measure for appropriate control action for damage mitigation is an area of future work and would require stochastic analysis of multiple data sets generated under identical loading and environmental conditions.

Acknowledgments

References

[1] Badii, R., and Politi, A., *Complexity Hierarchical Structures and Scaling in Physics.* Cambridge University Press, Cambridge, 1997.

[2] Ott, E., *Chaos in Dynamical Systems.* Cambridge University Press, Cambridge, 1993.

[3] Abarbanel, H. D. I., *The Analysis of Observed Chaotic Data.* Springer-Verlag, New York, 1996.

[4] Kantz, H., and Schreiber, T., *Nonlinear Time Series Analysis.* 2nd ed., Cambridge University Press, Cam-

bridge, 2004.

[5] Duda, R., Hart, P., and Stork, D., *Pattern Classification*. Wiley, New York, 2001.

[6] Markou, M., and Singh, S., Novelty detection: a review—Parts 1 and 2, Signal Process. **83**, 2481–2521 (2003).

[7] Daw, C. S., Finney, C. E. A., and Tracy, E.R., A review of symbolic analysis of experimental data, Rev. Sci. Instrum. **74** (2), 915–930 (2003).

[8] Ray, A., Symbolic dynamic analysis of complex systems for anomaly detection, Signal Process. **84** (7), 1115–1130 (2004).

[9] Gupta, S., Ray, A., and Keller, E., Symbolic time series analysis of ultrasonic data for early detection of fatigue damage, Mech. Syst. Signal Process. (in press).

[10] Chin, S., Ray, A., and Rajagopalan, V., Symbolic time series analysis for anomaly detection: A comparative evolution, Signal Process. **85** (9), 1859–1868 (2005).

[11] Beck, C., and Schlogl, F., *Thermodynamics of Chaotic Systems: an Introduction*. Cambridge University Press, Cambridge 1993.

[12] Kennel, M. B., and Buhl, M., Estimating good discrete partitions form observed data: Symbolic false nearest neighbors, Phys. Rev. E **91**(8), 084102 (2003).

[13] Davidchack, R. L., Lai, Y. C., Bolt, E. M., and Dhamala, H., Estimating generating partitions of chaotic systems by unstable periodic orbits. Phys. Rev. E **61**, 1353–1356 (2000).

[14] Rajagopalan, V., and Ray, A., Wavelet-based space partitioning for symbolic time series analysis, *Proceedings of 44th IEEE Conference on Decision and Control and European Control Conference (Seville, Spain)*, December 2005.

[15] Mallat, S., *A Wavelet Tour of Signal Processing* 2/*e*. Academic, New York, 1998.

[16] Cover, T. M., and Thomas, J. A., *Elements of Information Theory*. Wiley, New York, 1991.

[17] Lind, D., and Marcus, M., *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.

[18] Plemmons, R. J., and Berman, A., *Nonnegative Matrices in the Mathematical Science*. Academic, New York, 1979.

[19] Zhang, H., Ray, A., and Phoha, S., Hybrid life extending control of mechanical systems: Experimental validation of the concept. Automatica **36** (1), 23–36 (2000).

[20] Klesni, M., and Lukas, P., Fatigue of Metallic Materials. *Material Science Monographs 71*, Elsevier, New York, 1992.

[21] Keller, E. E., and Ray, A., Real-time health monitoring of mechanical structures, Struct. Health Monit. **2**(3), 191–203 (2003).

[22] Wavelet Toolbox, Matlab, Mathworks Inc.

Shalabh Gupta received a B.Tech (honors) in mechanical engineering from Indian Institute of Technology, Roorkee, India in the year 2001. He has two M.S degrees, one in mechanical engineering and the other in electrical engineering, both from Penn State University. He is currently enrolled in the doctoral program in the Mechanical Engineering Department under the supervision of Dr. Ray. His general research interests include: control theory, signal processing, pattern recognition, fault diagnosis, and statistical mechanics. He is currently working as a graduate research assistant at the Electromechanical Systems Laboratory (EMSL) in the Mechanical Engineering Department of Penn State University.

Amol Khatkhate received a B.E. in mechanical engineering from Veer Mata Jijabai Technological Institute (VJTI), Mumbai, India in the year 2001. He currently has two M.S degrees, one in mechanical engineering and the other in electrical engineering, both from Penn State University. The topic of his dissertation is "Anomaly Detection and Life Extending Control in Electromechanical Systems." His general research interests include: control theory, signal processing, and pattern recognition. His specific areas of interest include diagnostics, prognostics and structural health monitoring. He is currently working as a graduate research assistant at the Electromechanical Systems Laboratory (EMSL), Penn State.

Asok Ray received a Ph.D. degree in mechanical engineering from Northeastern University, Boston, MA, and also graduate degrees in each discipline of electrical engineering, mathematics, and computer science. Dr. Ray joined the Pennsylvania State University in July 1985, and is currently a distinguished professor of mechanical engineering. Dr. Ray had been a senior research fellow at NASA Glenn Research Center under a National Academy of Sciences award. Dr. Ray has authored or coauthored 400 research publications including about 180 scholarly articles in refereed journals such as transactions of ASME, IEEE, AIAA, and research monographs. Dr. Ray is a fellow of IEEE, a fellow of ASME, a fellow of World Innovation foundation (WIF), and an associate fellow of AIAA. Dr. Ray's research experience and interests include: control and optimization of continuously varying and discrete-event dynamical systems, intelligent instrumentation for real-time distributed systems, and modeling and analysis of complex dynamical systems from thermodynamic perspectives in both deterministic and stochastic settings.
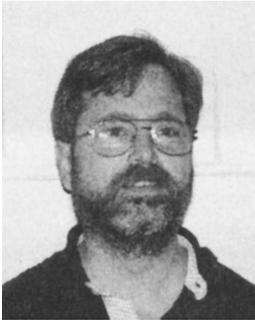
Eric Keller has been a senior research associate at the Applied Research Laboratory at Penn State University since June 2001. His research interests are in the areas of sensor networks, robotics, signal processing, and health monitoring systems. Dr. Keller received his doctoral degree in mechanical engineering from Penn State University in 2001, where his research focused on online detection of fatigue damage. From 1983 to 1994, Dr. Keller was an officer in the U.S. Air Force where he worked in acquisition and logistics engineering. Dr. Keller received a M.S. in aerospace engineering from the Air Force Institute of Technology in 1985 and a B.S. degree in mechanical engineering from Virginia Tech in 1982.