Available online at www.sciencedirect.com
ScienceDirect
Mechanical Systems and Signal Processing 21 (2007) 866–884
Mechanical Systems and Signal Processing
www.elsevier.com/locate/jnlabr/ymssp

Symbolic time series analysis of ultrasonic data for early detection of fatigue damage ☆

Shalabh Gupta*, Asok Ray, Eric Keller

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

Received 8 July 2005; received in revised form 20 August 2005; accepted 22 August 2005
Available online 19 October 2005

Abstract

This paper presents a novel analytical tool for early detection of fatigue damage in polycrystalline alloys that are commonly used in mechanical structures. Time series data of ultrasonic sensors have been used for anomaly detection in the statistical behaviour of structural materials, where the analysis is based on the principles of *symbolic dynamics* and *automata theory*. The performance of the proposed method has been evaluated relative to existing *pattern recognition* tools, such as *neural networks* and *principal component analysis*, for detection of small changes in the statistical characteristics of the observed data sequences. This concept is experimentally validated on a special-purpose test apparatus for 7075-T6 aluminium alloy specimens, where the anomalies accrue from small fatigue crack growth.

Keywords: Symbolic time series analysis; Pattern recognition; Neural networks; Anomaly detection

1. Introduction

Gradually evolving changes in the structural parameters of a mechanical system over its service life may generate uncertainties in both dynamic and stationary behaviour. This problem is often addressed by overly conservative estimates of the critical design parameters due to lack of available information. Consequently, the engineering design of mechanical systems suffers from enforcement of large safety factors and results in manufacture of cumbersome and unnecessarily expensive machinery.

Material irregularities, uncertain usage patterns (e.g. random overloads and sudden jerks) and environmental conditions (e.g. fluctuations in temperature and humidity) may adversely affect the service life of mechanical systems to cause performance degradation and unanticipated failures. Damage due to fatigue crack is one of the most commonly encountered sources of structural degradation during both nominal and off-nominal operations in mechanical systems [1]. If not detected at an early stage, the accumulated fatigue damage could potentially cause catastrophic failures in the system, leading to loss of human life and expensive equipment. Therefore, for reliable operation and enhanced availability, it is necessary to develop

☆ This work has been supported in part by Army Research Office (ARO) under Grant No. DAAD19-01-1-0646.
*Corresponding author.
E-mail addresses: szg107@psu.edu (S. Gupta), axr2@psu.edu (A. Ray), eek105@psu.edu (E. Keller).

0888-3270/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ymssp.2005.08.022

capabilities for prognosis of impending failures, such as the onset of wide-spread crack damage in critical structures. In the current state-of-the-art, direct measurements of fatigue damage at an early stage (e.g. crack initiation) are not feasible due to lack of appropriate sensing devices and analytical models. This paper attempts to address this inadequacy by taking advantage of advanced signal processing and pattern recognition tools. Since a vast majority of structural components that are prone to fatigue damage are made of polycrystalline alloys [1], the paper focuses on fatigue damage sensing and prediction for such materials.

Sole reliance on model-based analysis for structural damage monitoring is infeasible because of the difficulty in achieving requisite accuracy in modelling of fatigue damage evolution. For example, no existing model can capture the dynamical behaviour of fatigue damage at the grain level based on the basic fundamentals of molecular physics. In general, these models are critically dependent on the initial defects in the materials, which may randomly form crack nucleation sites. Small deviations in the distribution of initial defects may produce large variations in the evolution of fatigue damage for (apparently) identical specimens under the same loading and environmental conditions [2]. The stochastic nature of fatigue damage stresses the need for online updating of information using various sensing devices which can provide useful estimate of fatigue damage and are capable of issuing early warnings. Consequently, the analysis of time series data [3] from the available sensors mounted on the critical machinery components, is essential for tracking the behaviour pattern of the evolving fatigue damage in real time. Recent literature has demonstrated fatigue crack monitoring using sensor-based information [4,5]. Ultrasonic sensing techniques were applied for fatigue crack detection but they lacked precise signal processing capabilities for small change detection. Information from multiple (e.g. ultrasonic and acoustic emission) sensors was used for fatigue crack monitoring [6] and the results were shown to be in good agreement. However, the issues of early detection and real-time continuous health monitoring were not addressed.

Various signal processing applications deal with the analysis of time series data and attempts have been made to extract maximum useful information from the ensemble of sensor data. The problem of feature extraction from time series data for structural health monitoring has been recently addressed by many researchers [7–9]. The tools of statistical pattern recognition, auto-regressive model analysis, and wavelet analysis were applied to classify faults by different data patterns. However, the critical issue of detecting gradually evolving faults in real time were not addressed. Moreover, no quantifying measure was provided for damage accumulation and growth rate based on statistical information. Recently, techniques of non-linear dynamics have been applied for structural health monitoring [10–12] based on the concepts of attractor-based cross-prediction error between the measured signal and its predicted value. However, since dimension of the phase space may grow unbounded for noisy data, the analysis could be computationally expensive and infeasible for real-time applications. Furthermore, dealing with high dimensions might lead to spurious results and dimension reduction may lead to loss of vital information. To alleviate these difficulties, this paper has adopted a novel method of wavelet-based partitioning [13,14]. Based on this partitioning, the pertinent information is extracted from time series data sets in the form of probability distributions. Slight deviations in these distributions from that under the nominal condition is captured to identify the damage pattern.

Anomaly detection using symbolic time series analysis (STSA) [15] is a pattern recognition method that has been recently developed based upon a fixed-structure, fixed-order Markov chain, called the *D-Markov machine* [13]. This paper presents and experimentally validates this novel concept of fatigue damage prediction for early detection of anomalous patterns in the sensor data. The proposed STSA-based technique is evaluated relative to existing pattern recognition tools, such as *neural networks* and *principal component analysis*, for detection of statistical behaviour changes in the observed data sequence. The experimental platform is a fatigue testing apparatus, where possible anomalies accrue from the growth of small fatigue damage in 7075-T6 aluminium alloy specimens.

The paper is organised in seven sections. Section 2 formulates the problem of anomaly detection and summarises the concept of anomaly detection via symbolic time series analysis [13]. Section 3 describes the details of the experimental apparatus and outlines the basic features of optical microscopy and ultrasonic sensing. Brief descriptions of existing pattern recognition techniques are given in Section 4 and the algorithm of *D*-Markov machine construction is presented in Section 5. The details of data collection and analysis, as needed for experimental validation of the proposed concept, is described in Section 6. The results, derived from the test data, are presented to make a comparative evaluation of the proposed method with other pattern

recognition techniques from the perspectives of early detection of fatigue damage. Section 7 summarises and concludes the paper with recommendations for future work.

2. Problem formulation

Behaviour identification in mechanical systems can be formulated as a two-time-scale problem if the physical process is assumed to have stationary dynamics on the *fast time scale* and any observable non-stationary behaviour is associated with changes occurring on the *slow-time scale*. In other words, the variations in the internal dynamics of the system are assumed to be negligible on the fast time scale, while pattern changes may become significant on the slow-time scale. In general, a long time span in the fast time scale is a tiny (i.e. several order of magnitude smaller) interval in the slow-time scale, over which the system dynamics are assumed to have stationary behaviour. For example, in the context of early detection of fatigue damage in structural materials, small load fluctuations may take place on the fast time scale but the resulting damage evolution and anomaly growth (causing a detectable change in the dynamics of the system) occurs on the slow-time scale; the fatigue damage behaviour is essentially invariant on the fast time scale. Nevertheless, the notion of fast and slow-time scales is dependent on the specific application and operating environment.

From the above perspectives, the problem of anomaly detection is categorised into two subsets: (i) the forward problem and (ii) the inverse problem.

2.1. Forward problem

The primary objective of the forward (or analysis) problem is to characterise the patterns followed by the process dynamics as its behaviour changes on the slow-time scale. The forward problem is executed with an ensemble of collected data and its solution using STSA approach requires the following steps:

- Collection of sensor time series data to construct a pattern set spanning the system behaviour under different operational conditions.
- Generation of symbolic sequences (on the fast time scale) from observed time series data at different epochs of the slow-time scale.
- *D-Markov machine* construction from the generated symbol sequences and computation of the respective state probability vectors.
- Detection of behaviour changes by observing the deviations of the state probability vectors at various slow-time epochs from the one at the nominal condition.

2.2. Inverse problem

The inverse (or synthesis) problem concentrates on inferring the anomalies based on the observed time series data and system response for the purpose of triggering control actions for behaviour control. This problem may not always be well-posed, i.e. there might be no unique solution. Therefore, analysis of time series data sets under various excitations might be necessary for solving the inverse problem.

This paper focuses on analytical formulation and experimental validation of the forward problem, where the source of possible anomalies is fatigue crack initiation and small crack growth.

3. Description of the experimental apparatus

The experimental apparatus, shown in Fig. 1, is a special-purpose uniaxial fatigue testing machine, which is operated under load control or strain control at speeds up to 12.5 Hz; a detailed description of the apparatus and its design specifications are reported in [16]. The apparatus loads the test specimens by a hydraulic cylinder under the regulation of computer-controlled electro-hydraulic servovalves. The feedback signals are generated from either a load cell or an extensometer and are processed by signal conditioners that include standard amplifiers and signal processing units. The damage estimation and life prediction subsystem consists of damage analysis software and the associated computer hardware. The damage analysis software receives

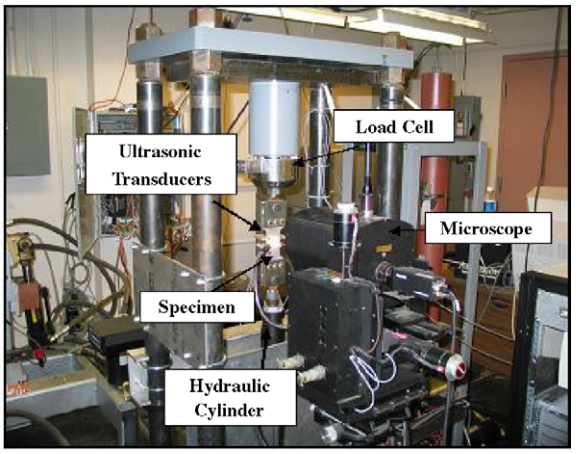

Fig. 1. Special-purpose fatigue test apparatus.

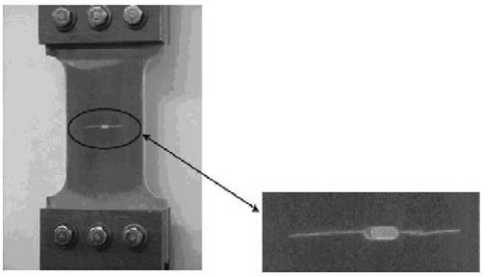

Fig. 2. Cracked centre-notched specimen.

real-time data from the heterogeneous measurement devices and operational data from the process instrumentation and the control module of the fatigue test apparatus, which is briefly described below.

- *Closed loop servo-hydraulic unit*: The instrumentation and control of the computer-controlled uniaxial fatigue test apparatus includes a load cell, actuator, hydraulic system, and controller. The servo-hydraulic unit can provide either random loads or random strains to a specimen for both low- and high-cycle fatigue tests at variable amplitude and multiple frequencies.
- *Computers for data acquisition, signal processing, and engineering analysis*: In addition to the computer controlling the load frame, a second computer is used for data collection from the microscope and image and signal processing. The ultrasonic flaw detection hardware is connected to the third computer. These laboratory computers are interconnected by a local dedicated network for data acquisition, data communications, and control.

3.1. Test specimen

A typical specimen, made of 7075-T6 aluminium alloy, is shown in Fig. 2. The specimen is 3 mm thick and 50 mm wide, and has a slot of 1.58 mm × 4.5 mm at the centre. The central notch is made to increase the stress concentration factor that ensures crack initiation and propagation at the notch ends. The test specimens have been subjected to sinusoidal loading under tension–tension mode (i.e. with a constant positive offset) at a frequency of 12.5 Hz. Since inclusions and flaws are randomly distributed across the material, small cracks appear at these defects and propagate and join at the machined surface of the notch even before microscopically visible cracks appear on the surface.

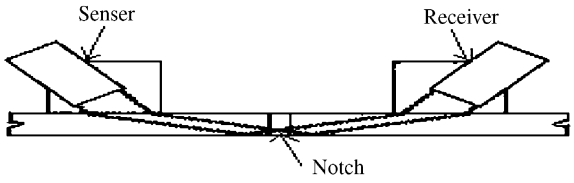

Fig. 3. Ultrasonic flaw detection scheme.

3.2. Sensors for damage detection

The fatigue test apparatus is equipped with a variety of damage sensors. Two types of sensors that have been primarily used for damage detection are: *travelling optical microscope* and *ultrasonic flaw detector*. The advantages of using ultrasonic transducers over microscope are the ease of installation at the desired damage site and detection of early anomalies before the onset of widespread fatigue crack propagation.

(1) *Travelling optical microscope*: The travelling optical microscope, shown as part of the test apparatus in Fig. 1, provides direct measurements of the visible part of a crack. The microscope shifts from left to right side of the central notch and vice versa after every 200 cycles to track crack growth on both sides of the notch. The crack length can be calculated automatically by moving a vertical line in the image to the tip of the crack with the help of the keyboard and the travelled distance is the crack length measured from the notch end. The resolution of the optical microscope is about 2 μm at a working distance of 10–35 cm.

(2) *Ultrasonic flaw detector*: The ultrasonic flaw detector functions by emitting high frequency ultrasonic pulses that travel through the specimen and return back through the receiver transducers. A piezoelectric transducer is used to inject ultrasonic waves in the specimen and an array of receiver transducers is placed on the other side of notch to measure the transmitted signal. In these experiments, an array of 2 receiver transducers was placed below the notch to detect faults on both left and right side of the notch. The ultrasonic waves produced were 5 MHz sine wave signals and they were emitted during a very short portion of every load cycle. In each cycle with sinusoidal loading at 12.5 Hz, ultrasonic waves were triggered at the peak of the load cycle for a very short instant of time. Ultrasonic measurements were taken at stress levels that exceeded the crack opening stress. The crack is open when the specimen is under maximum stress at the peak of a load cycle and this causes maximum attenuation of the ultrasonic waves. Note that if the crack closure occurs at low loads, then an alternative method would be needed to detect anomalies.

The sender and receiver ultrasonic transducers are placed on two positions, above and below central notch, so as to send the signal through the region of crack propagation and receive it on the other side, as seen in Fig. 3. Since material characteristics (e.g. voids, dislocations and short cracks) influence the ultrasonic impedance, a small fault in the specimen is likely to change the signature of the signal at the receiver end. Therefore, the signal can be used to capture some of the minute details and small changes during the early stages of fatigue damage, which may not be possible to detect by an optical microscope [17]. Prior to appearance of a crack on the specimen surface as detected by the optical microscope, deformations (e.g. dislocations and short cracks) inside the specimen may have already caused detectable attenuation and/or distortion of the ultrasonic waves.

It is observed that crack always starts at the stress-concentrated region near the notch but the exact location of the origin of crack can be treated as a random event. Formation of very small cracks is difficult to detect and model due to large material irregularities. Moreover, an optical microscope is only capable of detecting cracks when they appear on the front surface of the specimen. Therefore, this paper focuses on analysing the ultrasonics data for characterisation of fatigue damage in the small crack regime.

4. Pattern recognition for anomaly detection

Technical literature abounds with various techniques of pattern recognition (for example, see citations in [18]). Brief surveys of pattern recognition tools for anomaly detection are reported in [19]. The existing

techniques of pattern recognition, which have been used in this paper for comparison with the proposed STSA approach, are as follows:

- multilayer perceptron neural network (MLPNN);
- radial basis function neural network (RBF);
- principal component analysis (PCA).

These techniques are briefly described for completeness of the paper and in the context of their utilisation for anomaly detection.

4.1. Multilayer perceptron neural network (MLPNN)

The multilayer perceptron (MLP) neural network is a collection of connected processing elements, called nodes or neurons [19–21] whose structure is fixed by choosing the number of layers as well as the (possibly different) number of neurons in each layer. The *training* phase includes modelling of the input–output system architecture and identification of the synapsis weights. A set of inputs is passed forward through the network yielding trial outputs which are then compared to the target outputs to obtain the error. The network parameters (i.e. synapsis weights and biases) are adjusted until the error is within the specified bounds. If the specified bound is exceeded, the error is passed *backwards* through the net and the *training algorithm* adjusts the synapsis weights. The *back-propagation* algorithm has been used in this paper to update the network parameters in the direction in which the performance function decreases most rapidly until the error is within the specified bounds. The *mean square error* criterion is adopted in the recursive algorithm to update the weight vectors $\mathbf{w}_k$ as follows:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \alpha_n \mathbf{g}_n, \tag{1}$$

where $\mathbf{g}_n$ is the gradient $\partial J/\partial \mathbf{w}$ and α_n is the learning rate and J is the training error given by

$$J(\mathbf{w}) = \frac{1}{2}\sum_{j=1}^{q}(\mathbf{o}_j - \mathbf{d}_j)^2, \tag{2}$$

where $\mathbf{o}$ and $\mathbf{d}$ are the actual output vector and the target output vector, respectively, each of dimension q.

Time series data of signals enter into the input layer nodes, progress forward through the hidden layers, and finally emerge from the output layer. Each node i at a given layer k receives a signal from all nodes j in its preceding layer $(k-1)$ through a synapsis of weight w_{ij}^k and the process is carried onto the nodes in the following layer $(k+1)$. The weighted sum of signals x_j^{k-1} from all nodes j of the layer $(k-1)$ together with a bias w_{i0}^k produces the excitation z_i^k that, in turn, is passed through a non-linear *activation function* f to generate the output x_i^k from the node i at the layer k. This is mathematically expressed as

$$z_i^k = \sum_j w_{ij}^k x_j^{k-1} + w_{i0}^k, \tag{3}$$

$$x_i^k = f\left(z_i^k\right). \tag{4}$$

Various choices for the activation function are possible; the hyperbolic tangent function $\tanh(\bullet)$ has been adopted in this paper. For anomaly detection, the MLPNN is trained by setting a set of N input vectors, each of dimension ℓ, and a specified target output vector $\mathbf{d}$ of dimension q. This implies that the input layer has ℓ neurons and the output layer has q neurons. If the time series data are obtained from an ergodic process, then a data set of length $N\,\ell$ can be segmented into N vectors of length ℓ to construct the input and target pattern matrices. The input pattern matrix $\mathscr{P} \in \mathbb{R}^{\ell \times N}$ is obtained from the N input vectors as

$$\mathscr{P} \equiv \left[\mathbf{p}^1\ \mathbf{p}^2\ \cdots\ \mathbf{p}^N\right], \tag{5}$$

where $p^k \equiv \left[y_{(k-1)\ell+1}\ y_{(k-1)\ell+2}\ \cdots\ y_{k\ell}\right]^{\mathrm{T}}$ and each y_k is a sample from the ensemble of the time series data. The corresponding output matrix $\mathscr{O}$ is the output of the trained MLPNN under the input pattern $\mathscr{P}$.

$$\mathscr{O} \equiv \left[\mathbf{o}^1\ \mathbf{o}^2\ \cdots\ \mathbf{o}^N\right], \tag{6}$$

where $\mathbf{o}^i \in \mathbb{R}^q$ is the output of the trained MLPNN under the input $\mathbf{p}^i \in \mathbb{R}^\ell$. The performance vector $\mathbf{u} \in \mathbb{R}^q$ is obtained as the average of the N outputs.

$$\mathbf{u} \equiv \frac{1}{N}\sum_{k=1}^{N} \mathbf{o}^k. \tag{7}$$

The time series data under the nominal condition generates the input pattern matrix $\mathscr{P}_{\mathrm{nom}}$ that, in turn, is used to train the MLPNN with respect to a target output vector $\mathbf{d}$. The resulting output of the trained MLPNN with $\mathscr{P}_{\mathrm{nom}}$ as the input is $\mathscr{O}_{\mathrm{nom}}$ and the performance vector is $\mathbf{u}_{\mathrm{nom}}$. Subsequently, input pattern matrices $\{\mathscr{P}_1, \mathscr{P}_2, \ldots, \mathscr{P}_k, \ldots\}$ are obtained at slow-time epochs $\{t_1, t_2, \ldots, t_k, \ldots\}$ and corresponding output matrices of the trained MLPNN are $\{\mathscr{O}_1, \mathscr{O}_2, \ldots, \mathscr{O}_k \ldots\}$, which yield the respective performance vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \ldots\}$. The behavioural changes are described as anomaly from ideal condition and these anomalies are characterised by a scalar called *anomaly measure*. The anomaly measure at slow-time epoch t_k is obtained as

$$\mathscr{M}_k \equiv d(\mathbf{u}_k, \mathbf{u}_{\mathrm{nom}}), \tag{8}$$

where the $d(\bullet, \bullet)$ is an appropriately defined distance function [22]. In the present analysis, the distance function d is chosen as the Euclidean norm and hence the anomaly measure is given by

$$\mathscr{M}_k \equiv \|\mathbf{u}_k - \mathbf{u}_{\mathrm{nom}}\|_2. \tag{9}$$

4.2. Radial basis function neural network (RBFNN)

In a radial basis function neural network [20], the activation of a hidden unit is determined by the distance between the input vector and the prototype vector. The RBFNN is essentially a nearest neighbour type of classifier. A radial basis function is introduced as

$$f(\alpha, y) = \exp\left(-\frac{\sum_j |y_j - \mu|^\alpha}{N\theta_\alpha}\right), \tag{10}$$

where the exponent parameter $\alpha \in (0, \infty)$; the time series $y \equiv \{y_j\}$ is generated on the fast-time scale; and μ and θ_α are the centre and αth central moment of the data set, respectively. For $\alpha = 2$, $f(\bullet)$ becomes Gaussian, which is the typical radial basis function used in the neural network literature. To perform anomaly detection, the first task is to obtain the sampled time series data when the dynamical system is in the nominal condition and then the mean μ and the central moment θ_α are calculated as

$$\mu = \frac{1}{N}\sum_{j=1}^{N} y_j \quad \text{and} \quad \theta_\alpha = \frac{1}{N}\sum_{j=1}^{N} |y_j - \mu|^\alpha. \tag{11}$$

The distance between any vector y and the centre μ is obtained as $(\sum |y(n) - \mu|^\alpha)^{1/\alpha}$. Following Eq. (10), the radial basis function at the nominal condition is $f_{\mathrm{nom}} = f(\alpha, y)$. Under all conditions including anomalous ones, the parameters μ and θ are kept fixed. However, at slow-time epochs $\{t_1, t_2, \ldots, t_k, \ldots\}$, the radial basis functions $\{f_1, f_2, \ldots, f_k, \ldots\}$ are evaluated from the data sets under possibly anomalous conditions. The anomaly measure at an epoch t_k in the slow-time scale is obtained as a distance from the nominal condition and is given by

$$\mathscr{M}_k = d(f_k, f_{\mathrm{nom}}),$$

where the metric $d(\bullet, \bullet)$ is chosen as the Euclidean distance (see Eq. (9)).

4.3. Principal component analysis (PCA)

The PCA [18,23] is the best known linear feature extractor. The eigenvectors corresponding to the q largest eigenvalues of the $(n \times n)$ (positive semi-definite) covariance matrix generated from the time series data, form the n-dimensional patterns. The linear transformation is then defined as

$$Y = HX, \tag{12}$$

where X is the $(n \times d)$ transposed data matrix, made of n row vectors; H is the $(q \times n)$ linear transformation matrix whose rows represent q feature vectors each of dimension n; and Y is the derived $(q \times d)$ transformed data matrix.

To detect growth in anomaly from the time series data, principal component analysis is performed for dimensionality reduction. Data samples of large enough length ($\ell = n\ d$) can be used to capture the dynamical characteristics of the observed process. The length ℓ of time series data is partitioned into n subsections, each being of length $d = \ell/n$, where $d > n$. The resulting $(d \times n)$ data matrix is processed to generate the $(n \times n)$ covariance matrix that is positive-definite or positive-semidefinite real-symmetric. The next step is to compute the orthonormal eigenvectors $\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^n$ and the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ that are arranged in decreasing orders of magnitude. The eigenvectors associated with the first (i.e. largest) q eigenvalues are chosen as the feature vectors such that

$$\frac{\sum_{i=q+1}^{n} \lambda_i}{\sum_{i=1}^{n} \lambda_i} < \eta, \tag{13}$$

where the threshold $\eta \ll 1$ is a positive real close to 0. The resulting pattern is the matrix, consisting of the feature vectors as columns,

$$\widetilde{M} = \left(\sqrt{\lambda_1}\mathbf{v}_1 \ldots \sqrt{\lambda_q}\mathbf{v}_q\right). \tag{14}$$

The above steps are executed for time series data under the nominal (stationary) condition to obtain $\widetilde{M}_{\text{nom}}$. Then, these steps are repeated at subsequent slow-time epochs, $\{t_1, t_2, \ldots, t_k, \ldots\}$, using the same values of parameters, ℓ, d, n and q, to obtain the respective pattern matrices $\{\widetilde{M}_1, \widetilde{M}_2, \ldots, \widetilde{M}_k \ldots\}$. The anomaly measure at a slow-time epoch t_k is obtained as

$$\mathscr{M}_k \equiv d\left(\widetilde{M}_k, \widetilde{M}_{\text{nom}}\right)$$

where the metric $d(\bullet, \bullet)$ is chosen as the induced Euclidean norm of the matrix difference (see Eq. (9)).

5. Symbolic time series analysis (STSA)

A data sequence (e.g. time series data) can be converted to a symbol sequence by partitioning a compact region Ω of the phase space, over which the trajectory evolves, into finitely many discrete blocks as shown in Fig. 4. Let $\{\Phi_1, \Phi_2, \ldots, \Phi_m\}$ be a partitioning of Ω, such that it is exhaustive and mutually exclusive set, i.e.

$$\bigcup_{j=1}^{m} \Phi_j = \Omega \quad \text{and} \quad \Phi_j \cap \Phi_k = \emptyset \quad \forall j \neq k. \tag{15}$$

Each block Φ_j is labelled as the symbol $\sigma_j\ \varepsilon\ \Sigma$, 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 $\sigma_j\, \varepsilon\, \Sigma$ 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. 4 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. Once the symbol sequence is obtained, the next step is the construction of the finite state machine. These steps are explained in details in the following subsections.

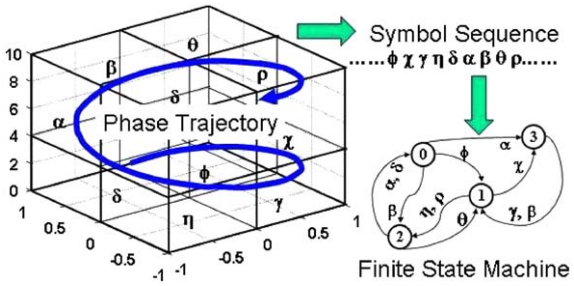

Fig. 4. An example of space partitioning.

5.1. Wavelet space (WS) partitioning

A crucial step in symbolic time series analysis is partitioning of the phase space for symbol sequence generation [15]. Several partitioning techniques have been reported in literature for symbol generation [3,24,25], primarily based on symbolic false neighbours. 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. Moreover, if the time series data is noise-corrupted, then the symbolic false neighbours would rapidly grow in number and require a large symbol alphabet to capture the pertinent information on the system dynamics. Therefore, symbolic sequences as representations of the system dynamics should be generated by alternative methods because phase-space partitioning might prove to be a difficult task in case of high dimensions and in the presence of noise.

This paper has adopted a wavelet-based partitioning approach [14] for construction of symbol sequences from the time series data. The wavelet transform [26] largely alleviates these shortcomings and is particularly effective with noisy data from high-dimensional dynamical systems. In this method, called *wavelet space partitioning* [13], the time series data are first converted by wavelet transform, 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 [27] is defined as

$$S = -\sum_{i=1}^{|\Sigma|} p_i \log(p_i), \tag{16}$$

where p_i is the probability of the ith state and summation is taken over all possible states. Uniform probability distribution of states is a consequence of maximum entropy partitioning that might make the partition coarser in regions of low density of data points and finer in regions of high density of data points. Fig. 5 shows 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.

5.2. State machine construction

The partitioning as described in the previous section is performed at time epoch t_0 of the nominal condition that is chosen to be a healthy 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

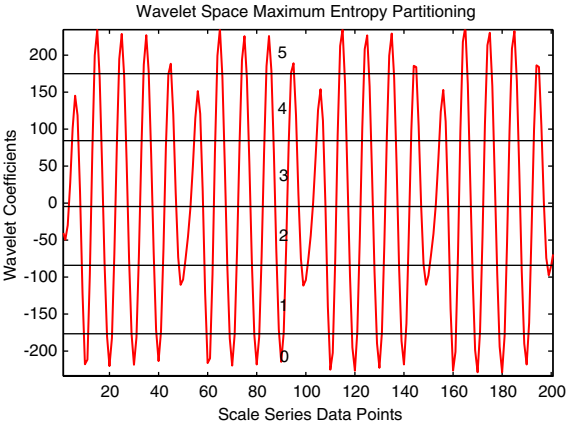

Fig. 5. Maximum entropy partitioning in the wavelet space.

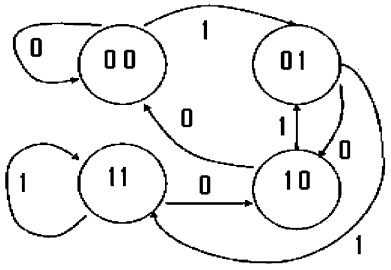

Fig. 6. Finite state automaton with $D = 2$ and $\Sigma = \{0, 1\}$.

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 [13]. 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. $n \leqslant |\Sigma|^D$). 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 a 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 principle of sliding block codes [28] 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 D-Markov machine [13] because of its Markov properties.

Definition 5.1. A symbolic stationary process is called D-Markov if the probability of the next symbol depends only on the previous D symbols, i.e. $P(\sigma_{i_0}/\sigma_{i_{-1}} \ldots \sigma_{i_{-D}} \sigma_{i_{-D-1}} \ldots) = P(\sigma_{i_0}/\sigma_{i_{-1}} \ldots \sigma_{i_{-D}})$.

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. 6 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.

In this analysis, it was found that the combination of $|\Sigma| = 8$ and $D = 1$ was able to capture anomalies earlier than the optical microscope. For $D = 1$, the set of states bears an equivalence relation to the alphabet Σ of symbols [22]. 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_t}. The occurrence of an event at a state may keep the machine in the same state or move it to a new state. The language of the machine is usually incomplete in the sense that all states might not be reachable from a given state.

Definition 5.2. The probability of transitions from state q_j to state q_k belonging to the set Q of states under a transition $\delta : Q \times \Sigma \to Q$ is defined as

$$\pi_{jk} = P(\sigma \in \Sigma \mid \delta(q_j, \sigma) \to q_k); \quad \sum_k \pi_{jk} = 1. \tag{17}$$

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 $|\Sigma|^{D+1}$ non-zero 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 [13]. 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 $\sigma_{i_1}\cdots\sigma_{i_D}\sigma_{i_{D+1}}$ and $\sigma_{i_1}\cdots\sigma_{i_D}$ which are, respectively, denoted by $N(\sigma_{i_1}\cdots\sigma_{i_D}\sigma_{i_{D+1}})$ and $N(\sigma_{i_1}\cdots\sigma_{i_D})$. Note that if $N(\sigma_{i_1}\cdots\sigma_{i_D}) = 0$, then the state $q \equiv \sigma_{i_1}\cdots\sigma_{i_D} \in Q$ has zero probability of occurrence. For $N(\sigma_{i_1}\cdots\sigma_{i_D}) \neq 0$, the transitions probabilities are then obtained by these frequency counts as follows:

$$\pi_{jk} = \frac{P(\sigma_{i_1}\cdots\sigma_{i_D}\sigma_i)}{P(\sigma_{i_1}\cdots\sigma_{i_D})} \approx \frac{N(\sigma_{i_1}\cdots\sigma_{i_D}\sigma_i)}{N(\sigma_{i_1}\cdots\sigma_{i_D})}, \tag{18}$$

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

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. 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.

The *anomaly measure* at slow-time epochs t_k is obtained as

$$\mathscr{M}_k \equiv d(\mathbf{p}_k, \mathbf{p}_{\text{nom}}), \tag{19}$$

where the $d(\bullet, \bullet)$ is chosen as the Euclidean norm of the vector difference (see Eq. (9)).

5.3. Summary of STSA anomaly detection

The following steps summarise the procedure of anomaly detection using symbolic time series analysis (STSA).

- Collection of time series data from appropriate sensor(s) at time epoch t_0 of the nominal condition, where the system is assumed to be in the healthy state (i.e. zero anomaly measure).
- Stacking of the wavelet transform coefficients (obtained with an appropriate choice of mother wavelet and range of scales) that are generated from the time series data at time epoch t_0 to generate the scale series data (see Section 5.1).
- Partitioning of the scale series data into $|\Sigma|$ regions using maximum Entropy partitioning to obtain the symbol sequence.
- Construction of the D-Markov machine states from the chosen alphabet size $|\Sigma|$ and the window length D and generation of the state probability vector $\mathbf{p}_0$ at time epoch t_0 of the nominal condition.

- Generation of time series data sequences at subsequent slow-time epochs, $t_1, t_2, \ldots, t_k, \ldots$, and their conversion to the scale series data in the wavelet domain to generate respective symbolic sequences using the partitioning at time epoch t_0 of the nominal condition.
- Generation of the state probability vectors at slow-time epochs, $t_1, t_2, \ldots, t_k, \ldots$ from the respective symbolic sequences using the finite state machine at time epoch t_0 of the nominal condition.
- Computation of the *anomaly measures* at time epochs, $t_1, t_2, \ldots, t_k, \ldots$ relative to the probability vector at time epoch t_0 of the nominal condition.

5.4. Real time application

Fatigue damage detection using STSA has been implemented in real time. However, the results presented in this paper are generated by off-line analysis. For real time application, the same steps have been followed as described in the previous sections. The nominal condition is chosen after the start of the experiment at a time epoch t_0 where the system attains the steady state and is assumed to be in the healthy condition with zero anomaly measure. The machine states are fixed in advance using a priori determined values of the parameters: alphabet size $|\Sigma|$ and window length D. The partitioning and machine construction were performed based on the time series data at the slow-time epoch t_0 and the resulting information (i.e. partitioning data and the state probability vector at this nominal condition) was stored for future computation of anomaly measures at future slow-time epochs, $t_1, t_2, \ldots, t_k, \ldots$, that were separated by uniform intervals of time in these experiments. The time series data of ultrasonic sensor signals were written as text files so that the STSA algorithm could read the data from the text files to calculate the anomaly measure at those time epochs. The algorithm is computationally very fast 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 health monitoring at any time and is capable of issuing early warnings.

6. Experimental data collection and analysis

The fatigue tests were conducted using specimens, made of the aluminium alloy 7075-T6, at a constant amplitude sinusoidal load for a low-cycle fatigue, where the maximum and minimum loads were kept constant at 87 and 4.85 MPa. (see Section 3 for a description of the test apparatus.) For low-cycle fatigue, the stress amplitude at the crack tip is sufficiently high to observe the elasto-plastic behaviour in the specimens under cyclic loading. A significant amount of internal damage occurs before the crack appears on the surface of the specimen when it is observed by the microscope. This internal damage caused by multiple small cracks, dislocations and microstructural damage affects the ultrasonic waves when they pass through the region where these faults have developed. This phenomenon causes signal distortion and attenuation at the receiver end. The crack propagation stage starts when this internal damage eventually develops into a single large crack. Subsequently, the crack growth rate increases rapidly and when the crack is sufficiently large, complete attenuation of the transmitted ultrasonic signal occurs, as seen at the receiver end.

6.1. Ultrasonic and optical image data collection

The ultrasonic sensing device was triggered at a frequency of 5 MHz at each peak of the ($\sim$12.5 Hz) sinusoidal load to obtain 100 data points in each cycle. Since the ultrasonic frequency is much higher than the load frequency, the data collection was performed for a very short interval in the time scale of load fluctuations. Therefore, it can be implied that the ultrasonic data were collected at the peak of each sinusoidal load cycle, where the stress is maximum and the crack is open causing maximum attenuation of the ultrasonic waves. The slow-time epochs were chosen to be 3000 load cycles (i.e. $\sim$240 s) apart. At the onset of each slow-time epoch, the ultrasonic data points were collected on the fast time scale of 100 cycles (i.e. $\sim$8 s), which produced a string of 10 000 data points. At a frequency of 12.5 Hz, 100 cycles take only 8 s to run and it is assumed that during this fast time scale of 100 cycles, the system remained in a stationary condition and no major changes occurred in the fatigue crack behaviour. This set of time series data collected in the manner

described above at different slow-time epochs was analysed by several techniques to calculate anomaly measures at those slow-time epochs.

The optical images were collected at every 200 cycles until a crack was detected on the specimen surface by the optical microscope. Then onwards the images were taken at user command and the microscope was moved such that it always focused on the crack tip. The vertical line in the images was adjusted by the movement of the microscope until it reached the crack tip. The distance travelled by the microscope determined the crack length.

The nominal condition at the slow-time epoch t_0 was chosen to be 5 kilocycles to ensure that the electro-hydraulic system of the test apparatus had come to a steady state and that no significant damage occurred till that point. This nominal condition was chosen as a benchmark where the specimen was assumed to be in a healthy state, where the anomaly measure was chosen to be zero. The anomalies at subsequent slow-time epochs, $t_1, t_2, \ldots, t_k, \ldots$, were then calculated with respect to the nominal condition at t_0. It is emphasised that the anomaly measure is relative to the nominal condition which is fixed in advance and should not be confused with the actual damage at an absolute level. Any particular value of the anomaly measure greater than zero indicates deviation from the nominal condition and it signifies that some changes have occurred inside the specimen.

6.2. Data analysis and damage evaluation

This section evaluates the collected time series data for early detection of anomalies resulting from fatigue damage. The proposed anomaly detection technique is compared with other pattern recognition tools (see Section 4).

(1) *MLPNN* : Following the MLPNN procedure, described in Section 4.1, the input pattern matrix $\mathscr{P}_{\text{nom}}$ was designed with $N = 200$ columns. Each column, having a length of $\ell = 50$, was generated from the ultrasonic data collected at the nominal condition at time instant t_0 to train the MLPNN. The neural network was designed with an input layer (with 50 neurons), 3 hidden layers (with 40 neurons in layer 1, 30 neurons in layer 2 and 40 neurons in layer 3) and the output layer (with 10 neurons). The target corresponding to each input pattern vector was chosen to be the 10×1 zero vector. The gradient descent back-propagation algorithm was used for network training with an allowable performance mean square error of 1.0×10^{-5}. The input pattern matrices $\mathscr{P}_1, \mathscr{P}_2, \ldots, \mathscr{P}_k, \ldots$, each of dimension (50×200), were then generated from the anomalous data sets at slow-time epochs $\{t_1, t_2, \ldots, t_k, \ldots\}$. Anomaly measures were then calculated with the procedure described in Section 4.1.

(2) *RBF*: Following the RBF procedure for anomaly detection as described in Section 4.2, the standard RBF was chosen with the exponent α to be equal to 2. Lower values of alpha reduced anomaly measure sensitivity and higher values of alpha produced noisy results. An estimate of the parameters, μ and θ_α, were obtained according to Eq. (11) based on the data under the nominal condition at t_0, which produced the requisite radial basis function f_{nom} following Eq. (10). At slow-time epochs $\{t_1, t_2, \ldots, t_k, \ldots\}$ the radial basis functions $\{f_1, f_2, \ldots, f_k, \ldots\}$ were evaluated from the data sets collected at those time epochs. Anomaly measures were then calculated with the procedure described in Section 4.2.

(3) *PCA*: Following the PCA procedure for anomaly detection as described in Section 4.3, a block of sampled ultrasonic data, having length of 10 000, was divided into $n = 8$ segments, each of which was of length $\ell = 1250$; these segments were then arranged to form a 1250×8 data matrix. The resulting 8×8 (symmetric positive-definite) covariance matrix of the data matrix provided a monotonically decreasing set of eigenvalues, $\lambda_1, \ldots, \lambda_8$, and the associated orthonormal eigenvectors $\mathbf{v}^1, \ldots, \mathbf{v}^8$. At time instant t_0 of the nominal condition, the first eigenvalue was found to be the dominant (i.e. $q = 1$) for a threshold of $\eta = 0.05$ such that

$$\frac{\sum_{i=2}^{8} \lambda_i}{\sum_{i=1}^{8} \lambda_i} < \eta.$$

The pattern matrix $\widetilde{M}_{\text{nom}}$ was generated at the nominal condition t_0 and the pattern matrices $\widetilde{M}_1, \widetilde{M}_2, \ldots, \widetilde{M}_k, \ldots$ were generated at time instants $t_1, t_2, \ldots, t_k, \ldots$. Anomaly measures were then calculated following the procedure described in section 4.3.

(4) *STSA*: Following the STSA procedure for anomaly detection as described in Section 5, the alphabet size for partitioning was chosen to be $|\Sigma| = 8$ and window length of $D = 1$, while the mother wavelet chosen to be 'gaus2' [29]. (Absolute values of the wavelet scale series data were used to generate the partition because of the symmetry of the data sets about their mean.) This combination of parameters was capable of capturing the anomalies earlier than the optical microscope. Increasing the value of $|\Sigma|$ further did not improve the results and increasing the value of D created a large number of states of the finite state machine, many of them having very small or zero probabilities, and required larger number of data points at each time epoch to stabilise the state probability vectors. This choice of the parameters $|\Sigma|$ and D was able to capture early anomalies with only 8 states and was computationally very fast in the sense that the code execution time was orders of magnitude smaller than the process response time. The mother wavelet of 'gaus2' provided better results than many other wavelets of the Daubechies family [26] because it closely matched the shape of the sinusoidal signals. State Probability vector $\mathbf{p}_{\text{nom}}$ was obtained at the nominal condition of time epoch t_0 and the state probability vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k, \ldots$ were obtained at other slow-time epochs $t_1, t_2, \ldots, t_k, \ldots$. Anomaly measures were calculated with the procedure described in Section 5.

The six pairs of plates in Fig. 7 show two-dimensional images of a specimen surface and histograms of probability distribution of automaton states at six different time epochs, approximately 5, 30, 40, 45, 60 and 78 kilocycles, exhibiting gradual evolution of fatigue damage. In each pair of plates from (a)–(f) in Fig. 7, the top plate exhibits the surface image of the test specimen as seen by the optical microscope. As exhibited on the top plates in each of the six plate pairs in Fig. 7, the crack originated and developed on the right side of the notch

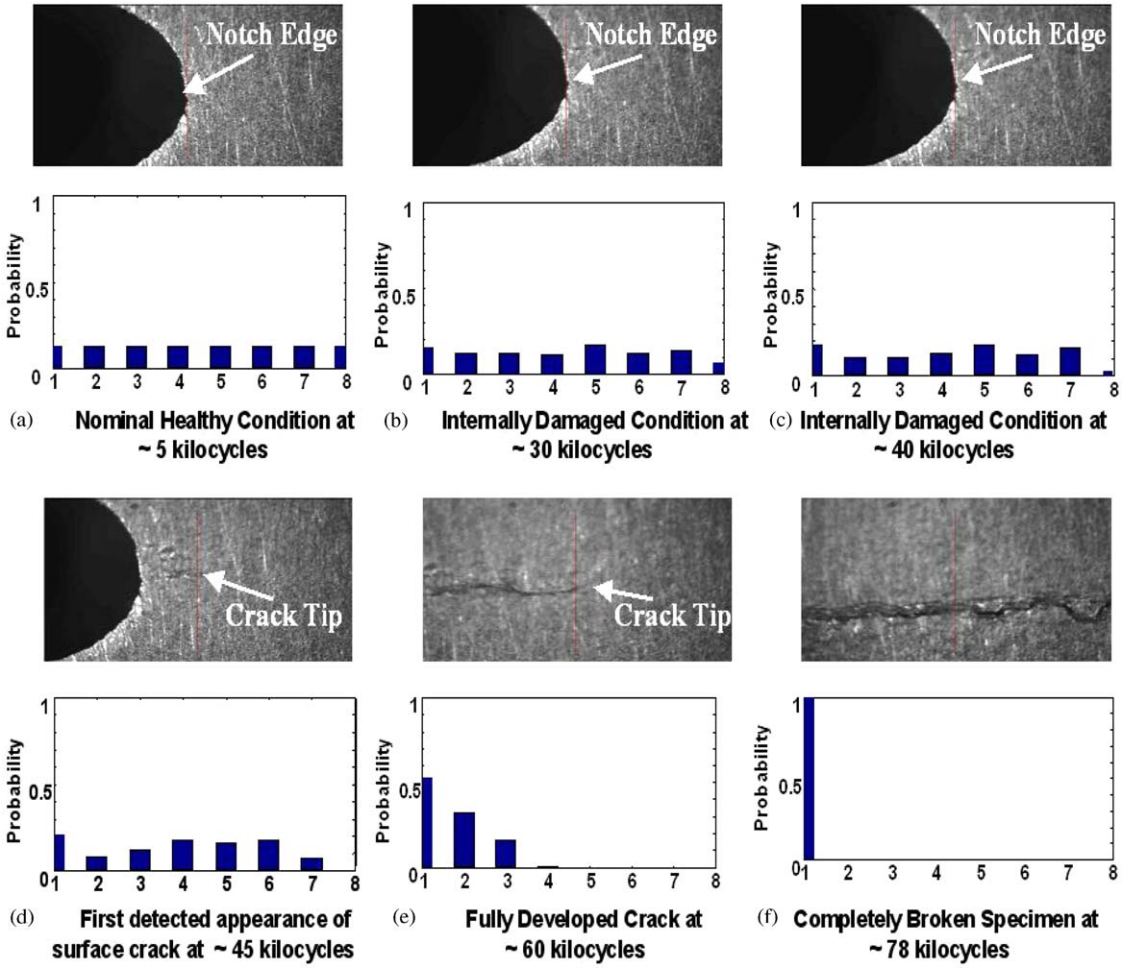

Fig. 7. Pictorial view of crack damage and corresponding probability distribution.

at the centre. Histograms in the bottom plates of six plate pairs in Fig. 7 show the evolution of the state probability vector corresponding to fatigue damage growth on the test specimen at different slow-time epochs, signifying how the probability distribution gradually changes from uniform distribution (i.e. minimal information) to delta distribution (i.e. maximum information).

The top plate in plate pair (a) of Fig. 7 shows the image at the nominal condition (~5 kilocycles) when the anomaly measure is taken to be zero, which is considered as the reference point with the available information on potential damage being minimal. This is reflected in the uniform distribution (i.e. maximum entropy) as seen from the histogram at the bottom plate of plate pair (a).

Both the top plates in plate pairs (b) and (c) at ~30 and ~40 kilocycles, respectively, do not yet have any indication of surface crack although the corresponding bottom plates do exhibit deviations from the uniform probability distribution (see the bottom plate of plate pair (a)). This is an evidence that the analytical measurements, based on ultrasonic sensor data, produce damage information during crack initiation, which is not available from the corresponding optical images.

The top plate in plate pair (d) of Fig. 7 at ~45 kilocycles exhibits the first noticeable appearance of a ~300 μm crack on the specimen surface, which may be considered as the boundary of the crack initiation and propagation phases. This small surface crack indicates that a significant portion of the crack or multiple small cracks might have already developed underneath the surface before they started spreading on the surface. The histogram of probability distribution in the corresponding bottom plate shows further deviation from the uniform distribution at ~5 kilocycles. This is also an indication of having increasingly more information on damage than what was available in the earlier cycles.

The top plate in plate pair (e) of Fig. 7 at ~60 kilocycles exhibits a fully developed crack in its propagation phase. The corresponding bottom plate shows the histogram of the probability distribution that is significantly different from those in earlier cycles in plate pairs (a)–(d), indicating further gain in the information on crack damage. The top plate in plate pair (f) of Fig. 7 at ~78 kilocycles exhibits the image of a completely broken specimen. The corresponding bottom plate shows delta distribution indicating complete information on crack damage.

The observation in Fig. 7 is further clarified by using the notion of entropy (see Eq. (16)). The scale series data at the nominal condition were partitioned using the maximum entropy principle, which led to uniform probability distribution (i.e. maximum entropy) among the states in the bottom plate of plate pair (a) in Fig. 7. In contrast, for the completely broken stage of the specimen, the entire probability distribution is concentrated in only one state of the finite state machine as seen in the bottom plate of plate pair (f) in Fig. 7, which indicates a very large attenuation of the ultrasonic signal. This phenomenon of the sample being completely broken signifies certainty of information and hence zero entropy. Therefore, as the fatigue crack damage evolves, the uniform distribution (i.e. maximum entropy) under nominal condition degenerates toward the delta distribution (i.e. zero entropy) for the broken specimen. In the intermediate stages, gradual degradation can be quantitatively evaluated using this information. Fig. 8 shows the monotonically decreasing profile of entropy versus load cycles, which is a clear evidence of gradual evolution of fatigue crack damage with load cycles. The sharp decrease in entropy from the first appearance of a surface crack at ~45 kilocycles to the complete breakage at ~78 kilocycles can be related to phase transition of first order in the thermodynamic sense [30].

6.3. Comparative evaluation of damage detection techniques

This section makes a comparative assessment of the proposed STSA method with other existing techniques of pattern recognition, described in Section 4 for early detection of fatigue crack damage using the ultrasonic sensor data generated on the test apparatus described in Section 3. The following anomaly detection approaches were investigated.

- multilayer perceptron neural network (MLPNN);
- radial basis function neural network (RBFNN);
- principal component analysis (PCA);
- symbolic time series analysis (STSA).

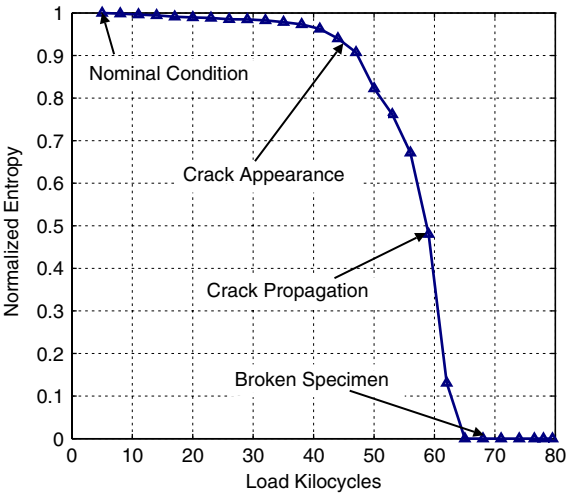

Fig. 8. Evolution of entropy with fatigue damage growth.

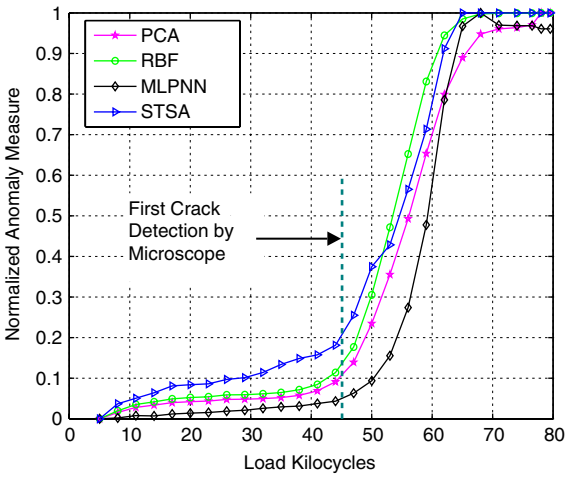

Fig. 9. Performance comparison for fatigue damage detection.

The efficacy of a specific approach is largely determined by its capability for early detection of small anomalies. The four plots in Fig. 9 compare the anomaly measures obtained by using the afore-described four anomaly detection approaches. Each of the curves in Fig. 9 show a possible bifurcation where the slope of the anomaly measure changes dramatically indicating the onset of crack propagation phase. First appearance of fatigue crack was detected by the optical microscope at approximately 45 kilocycles, which is marked by the dashed vertical line in Fig. 9. The crack when it appeared on the surface was ~300 μm long, which implies that a significant amount of fatigue damage including multiple small cracks developed inside the specimen before they finally spread out to the surface. Fig. 10 shows the fatigue crack growth in millimeters vs the load cycles, which was derived from the optical images.

A comparison of the plots in Fig. 9 clearly indicates that the STSA technique for anomaly detection yielded the best performance among all methods. It is observed that small changes can be detected by STSA significantly before the microscope can capture a surface crack. The slope of the anomaly measure represents the anomaly growth rate while the magnitude indicates the changes that have occurred relative to the nominal condition. An abrupt change in the slope (i.e. a sharp change in the curvature) of anomaly measure profile

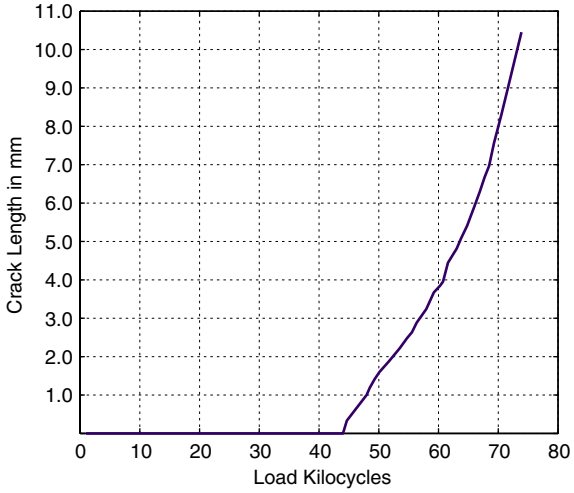

Fig. 10. Crack length plot of microscope data.

provides a clear insight into a forthcoming failure. Each of the four techniques detected a sharp slope change in the anomaly measure when the crack propagation stage started indicating an increased rate of anomaly progression at that stage. STSA, RBFNN and PCA detected this slope change between ~40 and ~44 kilocycles while MLPNN was slightly delayed in capturing this crack propagation phase and detected the slope change after ~44 kilocycles. Upon the onset of crack propagation, all four techniques showed similar patterns as indicated by the region to the right of the vertical dashed line in Fig. 9. However, the critical information lies in the region to the left of the vertical line where no crack was visible on the surface. In this region, STSA clearly demonstrated a significantly superior performance because the other techniques have not shown any significant change in the anomaly measure. MLPNN, RBFNN and PCA were slightly sensitive to the changes before the onset of crack propagation phase and they indicated a very small slope and magnitude in this region. However, the slope of STSA-based anomaly measure curve showed a clear trend of growth of anomaly right after ~15 kilocycles. This was the region where multiple small cracks were possibly formed inside the specimen, which caused small changes in the ultrasonic signal profile. Another increase in anomaly growth rate was observed at ~30 kilocycles. After the multiple small cracks coalesced together to form one single crack and crack propagation stage started, the corresponding changes in the ultrasonic signal were captured by all techniques effectively. A striking feature of the STSA technique with wavelet-based partitioning is that it takes advantage of the vector information, embedded in the histograms of probability distributions, to detect small changes in the signal.

7. Summary and conclusions

This paper presents early detection of fatigue damage in polycrystalline alloys that are commonly used in mechanical structures and its experimental validation on specimens, made of the aluminium alloy 7075-T6. The codes of damage analysis are executable in real time and have been demonstrated in the laboratory environment for on-line detection of fatigue damage, based on the analysis of ultrasonic sensor signals, before any surface cracks are visible through the optical microscope in a special-purpose fatigue testing apparatus. The recently developed tool of STSA [13], which is built upon the principles of *Symbolic Dynamics* and *Automata Theory*, has been adopted. Fatigue damage is detected from small changes in the statistical characteristics of the sensor signals.

The efficacy of the proposed STSA technique has been examined by comparison with other existing techniques of pattern recognition: neural networks (NN) and PCA. The results clearly show that, in comparison with the NN and PCA techniques, the performance of STSA is superior in terms of early detection of precursors that provide early warnings before the onset of widespread fatigue damage. A combination of

maximum-entropy partitioning in the wavelet domain and symbolic dynamics was able to detect fatigue damage growth significantly before the onset of crack propagation phase and yielded better performance than other techniques.

The reported work is a step toward building a reliable instrumentation system for early detection of fatigue damage in polycrystalline alloys; further research is necessary before its usage in industry. The utilisation of this information provided by the 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. While there are many research issues that need to be addressed, the following research topics are being currently pursued.

- Solution of the inverse problem using ultrasonic stochastic data and development of performance bounds for safe reliable operation.
- Statistical analysis of time series data of fatigue damage, collected under identical loading and environmental conditions, to account for manufacturing and material uncertainties.
- Validation of the STSA technique for early detection of fatigue damage under different conditions, such as high-cycle loading, variable-amplitude block loading, and spectral loading [31].
- Interpretation of phase changes in the fatigue damage evolution in terms of those in statistical mechanics.

Acknowledgements

The authors wish to thank their colleague Dr. S. Chin for providing useful information in generating the results.

References

[1] S. Ozekici, Reliability and Maintenance of Complex Systems, vol. 154, NATO Advanced Science Institutes (ASI) Series F: Computer and Systems Sciences, Berlin, Germany, 1996.
[2] K. Sobczyk, B.F. Spencer, Random Fatigue: Data to Theory, Academic Press, Boston, MA, 1992.
[3] H.D.I. Abarbanel, The Analysis of Observed Chaotic Data, Springer, New York, 1996.
[4] D.A. Cook, Y.H. Berthelot, Detection of small surface-breaking fatigue cracks in steel using scattering of Rayleigh waves, NDT&E International 34 (2001) 483–492.
[5] S.I. Rokhlin, J.-Y. Kim, In situ ultrasonic monitoring of surface fatigue crack initiation and growth from surface cavity, International Journal of Fatigue 25 (2003) 41–49.
[6] S. Grondel, C. Delebarre, J. Assaad, J.P. Dupuis, L. Reithler, Fatigue crack monitoring of riveted aluminium strap joints by lamb wave analysis and acoustic emission measurement techniques, NDT&E International 35 (2002) 137–146.
[7] H. Sohn, C.R. Farrar, N.F. Hunter, K. Wordan, Structural health monitoring using statistical pattern recognition techniques, Journal of Dynamic Systems, Measurement and Control 123 (December 2001) 706–711.
[8] Z.K. Peng, F.L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnosis: a review with bibliography, Mechanical Systems and Signal Processing 18 (2004) 199–221.
[9] X. Lou, K.A. Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy interference, Mechanical Systems and Signal Processing 18 (2004) 1077–1095.
[10] J.M. Nichols, M.D. Todd, M. Seaver, L.N. Virgin, Use of chaotic excitation and attractor property analysis in structural health monitoring, Physical Review E 67 (016209) (2003).
[11] W.J. Wang, J. Chen, X.K. Wu, Z.T. Wu, The application of some non-linear methods in rotating machinery fault diagnosis, Mechanical Systems and Signal Processing 15 (4) (2001) 697–705.
[12] L. Moniz, J.M. Nichols, C.J. Nichols, M. Seaver, S.T. Trickey, M.D. Todd, L.M. Pecora, L.N. Virgin, A multivariate, attractor-based approach to structural health monitoring, Journal of Sound and Vibration Processing 283 (2005) 295–310.
[13] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Processing 84 (7) (2004) 1115–1130.
[14] V. Rajagopalan, A. Ray, 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] C.S. Daw, C.E.A. Finney, E.R. Tracy, A review of symbolic analysis of experimental data, Review of Scientific Instruments 74 (2) (2003) 915–930.
[16] E.E. Keller, Real time sensing of fatigue crack damage for information-based decision and control, Ph.D. Thesis, Department of Mechanical Engineering, Pennsylvania State University, State College, PA, 2001.
[17] E.E. Keller, A. Ray, Real time health monitoring of mechanical structures, Structural Health Monitoring 2 (3) (2003) 191–203.
[18] R. Duda, P. Hart, D. Stork, Pattern Classification, Wiley, New York, 2001.

[19] M. Markou, S. Singh, Novelty detection: a review—parts 1 and 2, Signal Processing 83 (2003) 2481–2521.
[20] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press Inc., New York, 1995.
[21] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, Upper Saddle River, NJ, 1999.
[22] A.W. Naylor, G.R. Sell, Linear Operator Theory in Engineering and Science, Springer, New York, 1982.
[23] D.P. Fukunaga, Statistical Pattern Recognition, second ed., Academic Press, Boston, 1990.
[24] R.L. Davidchack, Y.C. Lai, E.M. Bolt, H. Dhamala, Estimating generating partitions of chaotic systems by unstable periodic orbits, Physical Review E 61 (2000) 1353–1356.
[25] M.B. Kennel, M. Buhl, Estimating good discrete partitions form observed data: symbolic false nearest neighbors, Physical Review E 91 (8) (2003) 084102.
[26] S. Mallat, A Wavelet Tour of Signal Processing 2/*e*, Academic Press, New York, 1998.
[27] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.
[28] D. Lind, M. Marcus, An Introduction to Symbolic Dynamics and Coding, Cambridge University Press, Cambridge, UK, 1995.
[29] Wavelet Toolbox, MATLAB, Mathworks Inc.
[30] N. Goldenfeld, Lectures on Phase Transitions and the Renormalization Group, Perseus Books, Reading, MA, 1992.
[31] S. Gupta, A. Ray, E. Keller, Online detection of fatigue failure via symbolic time series analysis, American Control Conference, Portland, OR, June 2005, pp. 3309–3314.